

My title*

Mohammed Yusuf Shaikh

March 17, 2024

The survey uncovers how the spread of COVID-19 in England has fluctuated from year to year from 2020 to 2023. We determined a negative binomial regression model and we tested the association between incidence rate (month and year) and the (the overdispersion trait of the data) . Studies have show a marked decrease in the numbers of cases over time this reflects the success of health promotion programs and vaccination campaigns. Aside from that, although there was a slight but noticeable change, there was also a correlate of monthly variations in case rates. This study reinforces that the time changes in COVID-19 are an obstacle, and the response has to be subject to constant public health monitoring and flexibility. After designing a surveillance system, policymakers and public health authorities can consequently leverage the lessons learnt first hand in informed planning for future outbreaks. With in-depth analysis, we focus on the significance of the time related to the transmission of infectious diseases and productivity of the field of research by which time can play a role.

1 Introduction

COVID-19 pandemic is brought about by SARS-CoV-2 virus and first appeared in the end of 2019, growing to be the world 's deadly virus in a very short time. Existing worldwide healthcare systems, countries' economies, and everyday life have experienced frequent disruptions by this outbreak on the single hand, it drove the alarming need for better infectious disease surveillance and management strategies. Of the lot of lessons that have been attained, the know-how of the temporal dimension in the tranmission of diseases is currently the most important aspect of the successful public health reaction.

Temporal analysis that captures disease indicators changes can't be left out when there is a need to appraise outbreak trends, forecast epidemics and assess the effectiveness of the interventions. One example is, during COVID-19, these studies offered the introspect to the

*Code and data are available at: [LINK](#).

extent public health actions, seasonal changes in transmission, and the implication of the immunization campaigns do impact them. Notwithstanding the right amount of resources invested in the research, the very complex temporal factors and the COVID-19 cases continue to be studied because most interventions are made in the region where most of the options for intervention and behavior, and environmental conditions are incomprehensible.

There is the case of England which has a diverse population and decentralized public health service as well that makes it a unique objective to study these dynamics. During this pandemic the county has seen the rise and fall of infection rates which were caused by national lockdown, local tier system and vaccination roll out. These factors, along with the populations' behavior and the change of seasons, have resulted in the intricate networks of disease spread. It is essential to notice such patterns to develop strategies for future public health and also mitigate the consequences of the resurgence or new viral threats.

This study shall seek to elucidate the spatiotemporal trends of COVID-19 transmission in England, by revealing how the case rates per 100,000 population have transmuted from 2020 to 2023. Through the use of a negative binomial regression model, we account for the overdispersion that usually happens in infectious disease count data. With this, we present a more sophisticated analysis on the relationship of time (month and year) with the COVID-19 case rates. Implementation of this approach therefore makes it possible to track with precision the seasonal trends and the efficacy of contextual public health interventions over time.

By doing this, we have an ambition of bringing more understanding about the chronological characteristics of Covid-19 and helping to make informative decisions that can guide policy making, public health planning and future research. This study, however, shows the importance of temporal analysis in the ongoing struggle to manage the incidence of COVID-19 worldwide. It provides a foundation for combating infectious diseases with more effectiveness and efficiency in the future.

2 Data

2.1 Source and Collection

The dataset this study is derived from are open data records from Leicester City Council containing test data to the day, with 7-day rates per 100,000 for all England, covering the time period from 2020 until now. These figures are based on the most recent population datasets that have been published by the Office for National Statistics (ONS) for the middle of the year of 2019, thus making possible to perform comparative analysis across various periods of time. This unprecedented dataset allows for the identification of COVID-19's pattern of transmission throughout England over time, which can then reveal what impact certain factors have had on the spread of the virus at various stages.

2.2 Variables

The dataset comprises several key variables critical for the analysis: The dataset comprises several key variables critical for the analysis:

Date: The study was conducted from Jan 1 2020 to Dec 31 2023 with the data of interest. This time of variable is of great importance for the consonance of temporal analysis and is additional developed into year, month and day components in order to be clearer and more detailed. **AreaName:** Finds the place of this study like England only. The geographical situation will be the constant assessment theme. **NewCasesBySpecimenDate:** The amount of new infection cases appear in the report of specimen date. This gradable idea acting as a measure of virus's advance. **NewCasesRatePer100k:** The rate of new cases (per 100,000 population) and standardized using ONS 2019 intermediate as a measure. This pace in like manner creates the basis for making comparisons over different time periods and populations. **NewCasesRollingSum:** A running total representing the new cases which offers a trending indicator which presents a smooth metric that can be used to show trends over a specific time period.

2.3 Data Preprocessing

To prepare the dataset for analysis, several preprocessing steps were undertaken: To prepare the dataset for analysis, several preprocessing steps were undertaken:

Cleaning: In the dataset, the values that were wrong were, such as missing items or inaccurate entries, were checked for any inconsistencies. Because the outcome of the analysis can have profound implications, accurate data was absolutely necessary. **Transformation:** Date variable was disaggregated to year, month, and day components so that a cross-sectional time analysis could be done. This process lets switch from seasonal observation to daily activity recording directly seeing the impact of an event or activity on a given day. **Normalization:** Rates were calculated per 100,000 people into population to provide comparisons over time and across different population sizes. That data was standardized. **Aggregation:** Different data types were sometimes added to monthly or yearly scales to discover wide-ranging trends and sort out any noise found in the daily records.

2.4 Descriptive Analysis

The histogram plotted in Figure 1 is the distribution of New Cases Rate per 100k in the England region, from the year 2020 to 2023. The x axis symbolizes the new case rate per 100,000 population and y-axis signifies the frequency of these rates within the dataset. Graphically shown, the data is right-skewed implying most data is at the lower end of the spectrum. Many of the peaks are high and separated by long tails to the right showing high rates of transmission. A significant component of the pandemic waves, including the sporadic outbreaks and possible

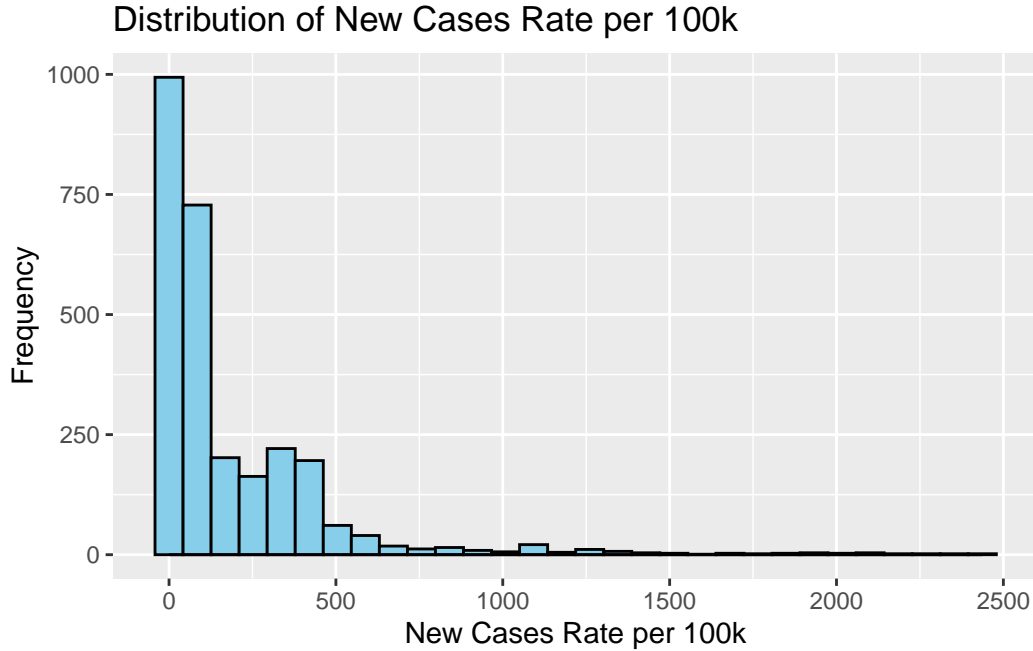


Figure 1: Distribution of New COVID-19 Cases Rate per 100k in England

superspreader events, which underlie the highly episodic nature of the infectious disease, can be seen here. The shape of the distribution is emblematic of the need for a statistical model that incorporates such overdispersion — where the variation is larger than the mean — because such modeling has innumerable implications on the kind of model to be chosen.

The total number of records in the dataset and the range of dates covered indicated the comprehensive nature of the data, capturing the entirety of the pandemic’s progression in England up to 2023. Summary statistics for `NewCasesBySpecimenDate`, `NewCasesRatePer100k`, and `NewCasesRollingSum` highlighted the variability in COVID-19 transmission rates over time, reflecting waves of infection and periods of relative calm. As the distribution of `NewCasesRatePer100k` revealed a right-skewed pattern, indicative of periods with exceptionally high transmission rates contrasted against more typical levels of new cases. Temporal trends were visualized through time series plots, illustrating the impact of national lockdowns, vaccination rollouts, and other public health measures on COVID-19 transmission rates.

3 Model

3.1 Model Setup

One main purpose of our model is to review the periods of COVID-19 cases per 100,000 population in England, considering how the data changed from 2020 to 2023 and whether it

is linked to the seasons in particular. Through analysis of time (conveyed through months and years) and COVID-19 transmission we are endeavoring to expose the networks of how the pandemic gets changed when different public health measures and societal situations have an impact. Taking into account the observation being independent variable namely COVID—19 cases per 100,000 population, which is a count of occurrences within a specified population size and time period, the generalized linear model (GLM) framework can be applied. Variation becomes the main issue when it comes to the choice between Poisson and negative binomial GLMs. The regulation of count data depends on the distribution and variance of the count data. Originally, the Poisson regression applied were the uncomplicated one which assumed that the mean and variance of the distribution were the same. Nevertheless, this assumption fails to hold in the analysis of the real data especially in large counts like the COVID-19 case rates, where the process is often overdispersed (variance exceeds the mean). To deal with the phenomenon of overdispersion, we identified negative binomial regression model as the most appropriate. This model echoes the Poisson model but has an inherent feature that increases the likelihood of extra variation occurring. This model should become a basis for the data we have as the variance exceeds the mean by a large margin. This negative binomial model, therefore, is more flexible and precise, having been able to accommodate the observed high level of variability observed on various time periods.

3.2 Model Specification

The model is specified as follows:

$$\text{NewCasesRatePer100k} \sim \text{NegativeBinomial}(\mu, \theta)$$

The parameters μ is the mean number of cases per 100,000 population, modeled as a function of time (month and year). θ is the dispersion parameter, accounting for overdispersion in the data.

The relationship between the mean of the dependent variable and the independent variables (month and year) is expressed through a log link function:

We use a log-linear model for the mean number of cases μ :

$$\log(\mu) = \beta_0 + \beta_1(\text{month}) + \beta_2(\text{year})$$

Where:

- β_0 is the intercept, representing the log of expected case rate when month and year are zero (baseline level).
- β_1 and β_2 are coefficients representing the effect of each month and year on the log of the case rate, respectively.

The negative binomial regression model is justified based on empirical evidence. At first, the overdispersion was present in the case rates data which can be explained by Poisson model inadequacy. The simple Poisson model does not account for all variable factors in infectious disease transmission.

3.3 Assumptions and Limitations

The negative binomial model is more suitable for our data than other count models, with some assumptions and restrictions however. These factors encompass the assumption of mutual independence of observations and omitted heterogeneity affecting infection rates. Further on, the model lacks an explanation of the spatial variability or non-time covariates which could inspire future studies.

4 Results

5 References