Deciphering the Temporal Patterns of COVID-19: A Comprehensive Analysis of Case Rates in England*

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.

---

*Code and data are available at: https://github.com/Yusuf365/Data-Analysis-Regression.git

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 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.

Data used in this paper was downloaded from '(Data 2023) , cleaned and analyzed with the programming language R (R Core Team 2022). Also with support of additional packages in R: `tidyverse` (Wickham et al. 2019), `janitor` (Firke 2021), `dplyr` (Wickham et al. 2022), `readr` (Wickham, Hester, and Bryan 2024), (Ripley et al. 2023), (Gabry et al. 2023), (Robinson and Hayes 2022), (Arel-Bundock 2023) .

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
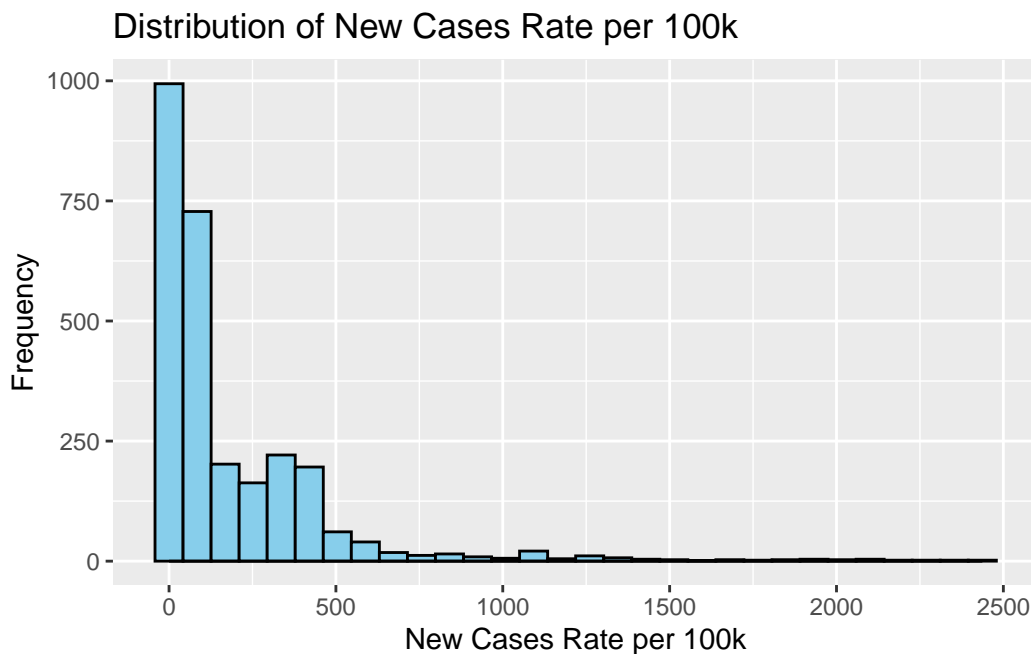
Distribution of New Cases Rate per 100k

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

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 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 NewCases-RatePer100k 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 COIVD—19 cases per 100,000 population, which is a count of occurrences within a specified population size and time period, the generalized linear model (GML) 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 of 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 infections rates. Further on, the model lacks an explanation of the spatial variability or non-time covariates which could inspire future studies

4 Results

Table 1: Model Summary of Negative Binomial

Term	Estimate	Std..Error	z.value	Pr…z..
(Intercept)	1620.0000	42.970000	37.71	<2e-16
month	-0.1059	0.006973	-15.19	<2e-16
year	-0.7986	0.021250	-37.58	<2e-16

The Table 1 was applied in order to investigate the association between time (measured in months and years) and the New Cases Rate per 100k with a Negative Binomial regression model. As mentioned, the summary model will give the regression estimates their standard errors, z-values, and the p-values.

The model features an intercept and coefficients for month and year. The estimate for the intercept was above zero, suggesting the baseline new case rate per 100k was highest at the reference time. The month as well as the year indicators were negative and highly significant ($p < 0.001$), which implies that the new case rate statistically significantly falls down as time changes forward. A negative sign for the month suggests that new case rates decrease slightly with each passing month. On the other hand, the year sign is clearly more significant and refers

to the drastic impact of various means and the way pandemic situations evolve. The Negative Binomial model evaluated with an estimated value for its dispersion parameter θ equal to 0.6998 and standard error of 0.0163 demonstrated overdispersion in the count data correlated to what could have been expected under a Poisson model. This illuminates the reason of choosing this Negative Binomial model. Concerning model fit, the residual deviance (3287.1) on 2741 degrees of freedom indicates that the model is adequate where it has been specified appropriately. The AIC (Akaike Information Criteria) value of the model is 33351. Having smaller values of AIC than the Null model which only consists of the intercept terms (no predictors) proposed that incorporating month and year as predictors, provide a meaningfully better fit to the data.

Hence, the findings indicate a clear and consistent temporal trend in the rate of new COVID-19 cases in England. This aligns with what would be expected given the introduction of public health measures, the seasonal impact on virus transmission, and the progression towards widespread vaccination over the years studied. The significant p-values for both time predictors reiterate their importance in explaining the variability in new case rates.
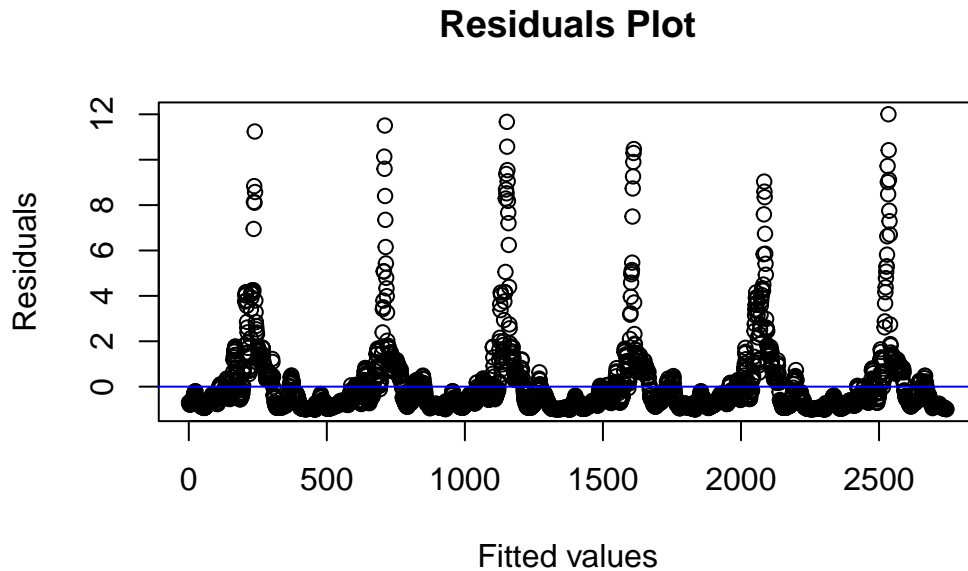
Residuals Plot

Figure 2: Residuals Plot for the Negative Binomial Regression Model

Residuals plot is a very powerful tool for validating the model of regression with respect to the fitting. In Figure 2: shows the negative binomial model residuals plotted versus fitted values. In an ideal case, the residuals should randomly deviate around the horizontal line at zero, which shows that the model unbiasedly predicts the predicted values regardless of their size.

Consistently, Figure 2, we find that the residuals are clustered close to the central vertical line, which is regarded as a good representation of the model behavior. Yet there are definitive

patterns which take the shape of different vertical lines. These are probably due to the noise in discretize data as well as the deficit of the case between what the model predicted and the existed days.
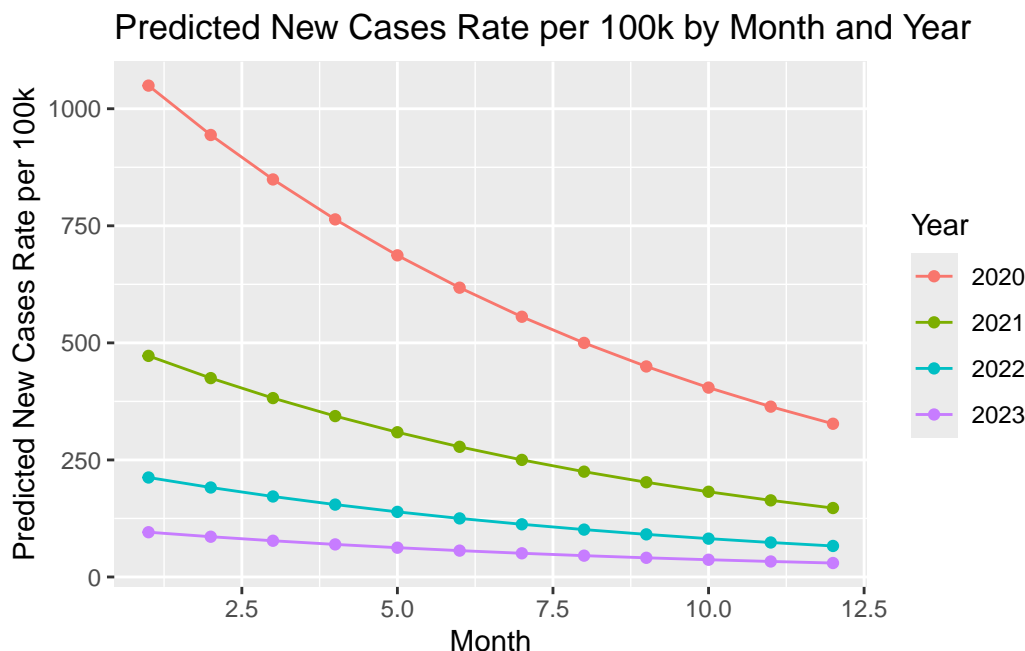

Figure 3: Predicted New Cases Rate per 100k by Month and Year

This line chart from Figure 3 represents the expected number of new COVID-19 cases per 100,000 population, which is a product of not only the months but also the years, according to our negative binomial regression model. According to this paragraph, every year is a separate line.

It is evident that new cases rate per 100k persons are decreasing as we get months that are far from January. The persisting downward trend might mean that seasonality, public health interventions, and/or natural course of the pandemic, including the effects of vaccination, are to blame for the falling rates of current cases. Steepest decline is clearly seen in year 2020, displayed by the line. This could be attributed to the initial high incidence followed by a sharp decrease most probably reflecting the implementation of early and harsh lockdown restrictions purposed to handle the progress of the disease. The next years—2021 in (green line), 2022 in (blue line), and 2023 in (purple line) do not show the drastic decline in the estimated percentages. Such flattening may stand for the good progress towards virus bio neutralization accomplished by the ongoing public health measures, the development of herd immunity, or the effect of vaccination programs. It is worth pointing out that the forecast rates in 2023 are considerably lower than the actual rates in 2020 with a substantial decrease in the number of new cases over the years implying a good vaccine and therapy for the COVID-19 treatment that is getting more and more effective. The model seems to depict the deceleration of the

COVID-19 cases, being so as to identify the strategies used. Nevertheless, further studies are required in order to check if the applied interventions and other conditions could impact on the effectiveness eventually.
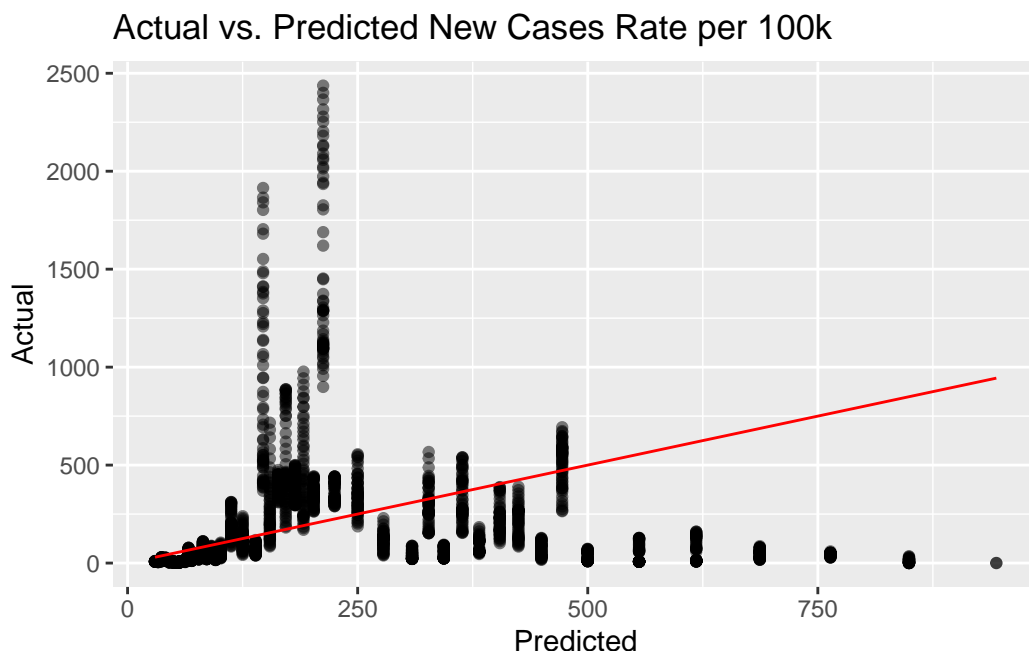

Figure 4: Actual vs. Predicted New Cases Rate per 100k

The Figure 4 graph shows a comparison between the actual and predicted the first new COVID-19 cases rates per 100,000 population. Predictions were checked by a negative binomial regression model. As the data points move nearer to the red line. A deviation appears at cases predicted rate as the amplification of predicted case rate indicates that the model underestimated in higher transmission times. The graph shows clearly that the model is in good shape for minor case surges but more factors still have to be taken into considerations in order to improve the predictions concerning case rates at the higher end of the spectrum.

5 Discussion

This study presents an article that comprises of the seasonal pattern, strategic plans, and the trajectory of the COVID-19 case rate decline in England as reflected in the substantial drop of the number of cases over time. The negative binomial regression model, which we selected because it can address over dispersion, exemplified a drop every month and a year in the new cases number per 100,000 of population. The lower transmission rate is predicted in the Model which is seen in Figure 3, however peak periods are forecasts less accurately in the model. This can as well illustrate the effects of other unexpected things happened, say a more contagious

subtype appeared or the shifting behaviors of people, that are not linked with the general flow of time only.

The model model's strong point is its ability to forecast with a high degree of accuracy cases that are not as highly transmitable as the initial ones. Following the provision of this purpose is to help healthcare systems and policymakers in the task of allocating resources to operations in normal, routine times. Besides the close correlation with the downward trend in case incidence, the above pattern appears as a big picture of short- and long-term implementation of the public health interventions and vaccination drive. An overlook of case rates at high prevalence points remains among the things that the researchers should not disregard. Their argument is subject to criticism. The missing factors in variables such as disruption of public health policy changes, population following the designs to the letter may also be an explain the gap. Besides, the presumption that the given month and year are standard units of time without any considerations for their interdependence may draw some simplistic real-world complexity of pandemic progress. This article is the beginning of many others which hopefully will encourage further exploration of the causal link between poverty and underachievement. Building upon the existing variables with additional ones such as mobility data or public health policies or even vastly different virus strains might be useful as long as it increases the model's predictive power. The intra-monthly patterns or the impact of social get-togethers on the level of happiness could also throw more detailed information about the global scenario.

6 Conclusion

Our study has gone into a trend analysis of COVID-19 cases in England, using a negative binomial regression model to determine at what rate the cases have changed monthly and yearly over the period 2020 to 2023. The trend-line showed the changes in the number of cases increasing gradually which may be a result of the changes in the season, public health intervention methods and the development of the community immunity level over a period of time.

When the model was tested, it proved its strength for predicting flat transmission, which is the foundation of the process of planning and managing of healthcare. Yet, it showed less accuracy in anticipating the high transmission rate since there was lack of interactivity between some of the factors that can lead to pandemic. This underlines the need to consider other modelling variables in the future besides public health policy, cultural behaviors and viral properties to factor in their multilayered impact on the pandemic spread.

The model has its limitations, but it is often helpful for about what will happen during the epidemic and in the context of planning to prevent disease outbreaks. The paper stresses the relevance of new discoveries and regular data assessment. This will enable the development of practical guidelines, especially where new variations emerge or the behaviors of other people change.

In general, when the world is striving to conquer the contagiousness of the coronavirus, it is obvious that the evidence-based, meticulous and comprehensive modelling techniques are imperative. These methods both give us the information of previous situations in history and prepare us for future epidemics also. The studies that we carry out are part of a team effort that strives to find answers to the current Covid 19 problem. In turn, these facts reveal the need for a modified and comprehensive reaction to this extraordinary global health crisis.

References

Arel-Bundock, Vincent. 2023. *Modelsummary: Summary Tables and Plots for Statistical Models.* https://vincentarelbundock.github.io/modelsummary/.

Data, UK Government. 2023. *Covid-19 - Positive Tests and Rate Per 100,000 Population for Leicester and England.* https://www.data.gov.uk/dataset/2348636d-3e8c-406b-a331-f536939f5c5c/covid-19-positive-tests-and-rate-per-100-000-population-for-leicester-and-england.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

Gabry, Jonah, Ben Goodrich, Jonah Gabry, Paul-Christian Bürkner, and Imad Ali. 2023. *Rstanarm: Bayesian Applied Regression Modeling via Stan.* https://mc-stan.org/rstanarm.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ripley, Brian, Bill Venables, Douglas M. Bates, Kurt Hornik, Albrecht Gebhardt, and David Firth. 2023. *MASS: Support Functions and Datasets for Venables and Ripley's MASS.* https://cran.r-project.org/package=MASS.

Robinson, David, and Alex Hayes. 2022. *Broom: Convert Statistical Analysis Objects into Tidy Data Frames.* https://CRAN.R-project.org/package=broom.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data.* https://readr.tidyverse.org.