Unraveling COVID-19 Dynamics in the UK: Insights from Multivariate Analysis

Sreelekshmi Girijadevi Purushothaman(22200034), Vinothini Balasubramani (22202390)

MSc - Data and Computational Science, University College Dublin.

Introduction

In the wake of the COVID-19 pandemic, nations worldwide grappled with unprecedented public health challenges. With the UK being among the hardest-hit countries and promptly initiating vaccination campaigns, it raised crucial inquiries: What has been the tangible impact of vaccinations on infections, hospitalizations, and fatalities? Can we discern the efficacy of the vaccine campaign against the backdrop of other influential factors?

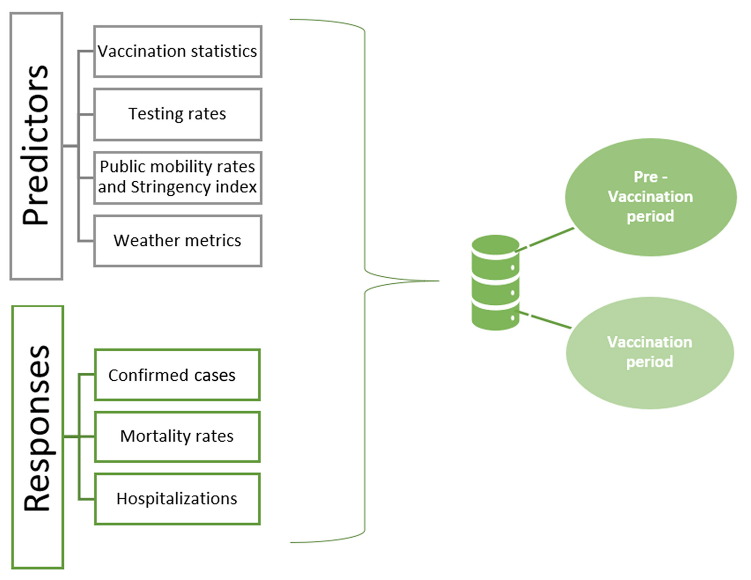
Research Question: How do the outcomes of COVID-19 (infections, hospitalisations, and mortality rates) differ between the pre-vaccination and vaccination periods, and to what extent do vaccines and other external factors contribute to these differences?

Significance: Analysing COVID-19 data from pre-vaccination and vaccination periods through multivariate linear regression offers real-world insights into the vaccine's effectiveness. This research underscores the tangible benefits of vaccination amidst other influencing factors, guiding public health strategies and reinforcing the imperative of widespread vaccination to combat the pandemic.

Objective: Analyse the correlation between vaccination rates and COVID-19 outcomes in the UK. Explore the impact of human mobility on the virus's transmission during pre-vaccination and vaccination periods. Examine potential associations between weather patterns and COVID-19 metrics to elucidate any indirect influences.

Data

The COVID-19 impact analysis on the UK is being conducted using Google's open dataset. This dataset covers a range of variables, including confirmed cases, deaths, testing rates, hospitalizations, vaccination statistics, mobility trends, weather metrics, and government stringency index. The dataset's comprehensive nature allows for a comprehensive exploration of COVID-19 dynamics, enhancing the depth of analysis for this project.

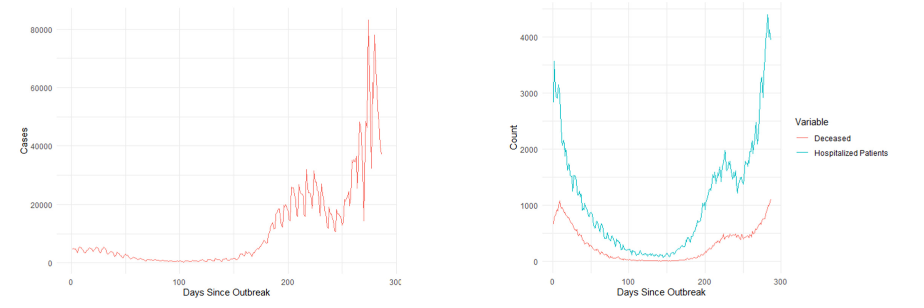


Dataset Period Division and Analysis Approach:

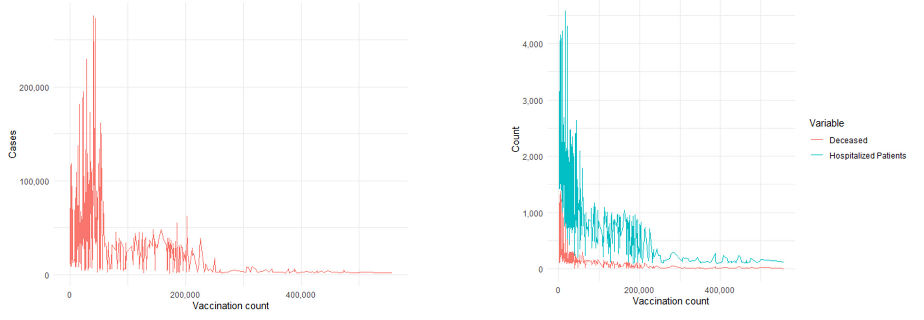
The dataset is systematically divided into two distinct periods: pre-vaccination and vaccination phases. This division allows us to deeply analyze the effects of various factors, such as testing rates, mobility patterns, stringency index, and weather metrics, both before and after the commencement of vaccination efforts. During the vaccination period, our analysis extends beyond these predictors to encompass the impact of vaccination metrics on the dynamics of cases deaths, and hospitalizations. This comprehensive approach enables a comprehensive assessment of the factors influencing the COVID-19 landscape.

To effectively communicate the trends, we employ a visual approach. For the pre-vaccination period, we present the count of cases over time in one plot and the count of hospitalizations and deaths in another plot. In the context of the vaccination period, we adopt a unique approach. Here, the count of cases, hospitalizations, and deaths is juxtaposed against the rate of vaccination.

Pre-Vaccination Period - Covid Cases, Deaths & Hospitilisations



Vaccination Vs Covid Cases, Deaths & Hospitilisations



Methods

Our research project aimed to comprehensively analyze the impact of COVID-19 vaccinations on infection rates, fatalities, and hospitalizations in the UK. Leveraging the power of Multiple Linear Regression (MLR), a suitable approach for exploring complex relationships among multiple variables, we systematically investigated the changing dynamics before and after the vaccination rollout.

Harnessing the power of Multivariate Linear Regression (MLR), a sophisticated statistical technique, we delved into the intricacies of the COVID-19 pandemic. Splitting our dataset into pre-vaccination and vaccination periods, we tailored MLR models for each epoch. This strategic division enabled us to uncover evolving relationships between predictors and responses, offering invaluable insights into the pandemic's dynamics. MLR's significance lies in its ability to decipher complex interdependencies among variables. By amalgamating diverse predictors, MLR paints a comprehensive picture of their combined influence on cases, deaths, and hospitalizations.

As we strive to comprehend the pandemic's nuances, the MLR framework emerges as a robust analytical tool, shedding light on the intricate factors shaping COVID-19 outcomes. Our formulated equation below captures the essence of MLR, providing a comprehensive analysis of the multifaceted dimensions of the pandemic.

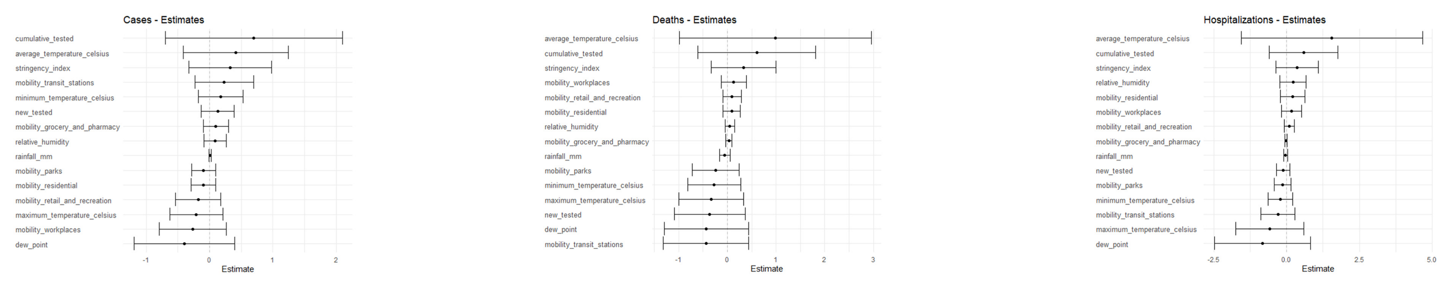
$$\begin{bmatrix} \text{cases}_1 & \text{deaths}_1 & \text{hospitalizations}_1 \\ \text{cases}_2 & \text{deaths}_2 & \text{hospitalizations}_2 \\ \vdots & \vdots & \vdots \\ \text{cases}_n & \text{deaths}_n & \text{hospitalizations}_n \end{bmatrix} = \begin{bmatrix} 1 & \text{vaccination rates}_1 & \text{test rates}_1 & \text{mobility rates}_1 & \text{weather metrics}_1 & \text{stringency index}_1 \\ 1 & \text{vaccination rates}_2 & \text{test rates}_2 & \text{mobility rates}_2 & \text{weather metrics}_2 & \text{stringency index}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \text{vaccination rates}_n & \text{test rates}_n & \text{mobility rates}_n & \text{weather metrics}_n & \text{stringency index}_n \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_{\text{vaccination rates}} \\ \beta_{\text{test rates}} \\ \beta_{\text{mobility rates}} \\ \beta_{\text{weather metrics}} \\ \beta_{\text{stringency index}} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- Data Preprocessing: Manage missing data, clean datasets, and ensure data consistency.
- EDA: Explore variable distributions, relationships, and potential outliers for data understanding.
- Model Building: Leveraging the Multivariate Linear Regression technique, we adeptly constructed models
- using datasets spanning both pre-vaccination and vaccination periods.
- Model Evaluation: Assess model performance using R-squared and cross-validation techniques.
- Comparison Analysis: Compare MLR model results to unveil vaccination impact.

Results

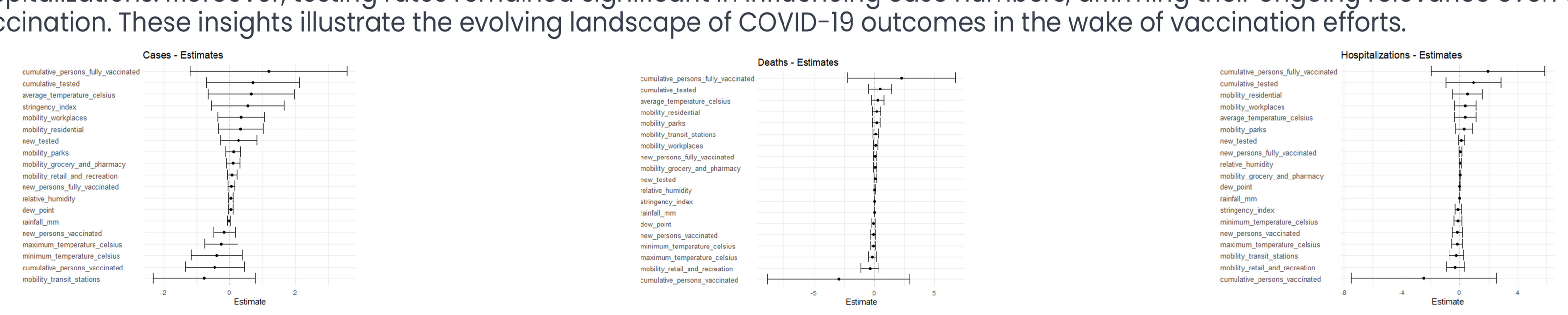
Pre-Vaccination Analysis Interpretation:

For the pre-vaccination period, we conducted Multivariate Linear Regression to unravel the relationships between COVID-19 responses — confirmed cases, deceased cases, and hospitalized patients — and an array of predictors including testing rates, mobility patterns, weather metrics, and the stringency index. Following the evaluation using the test data, the model has exhibited a satisfactory Mean Squared Error of 0.102. Notably, the analysis has provided compelling insights into these interrelationships, with each response attaining an R-squared value of roughly 0.80 or higher. Notably, the cumulative number of tests administered exhibited a strong positive association with confirmed cases, deceased cases, and hospitalized patients, indicating the pivotal role of testing in identifying cases. Additionally, higher stringency index values were linked to increased cases, implying the effectiveness of stringent measures in controlling the spread of the virus. Mobility patterns, particularly reduced activity in retail, recreation, and transit stations, were negatively correlated with cases, hinting at the impact of reduced movement on case numbers. These findings demonstrate the complex interplay of diverse factors influencing COVID-19 outcomes during the pre-vaccination phase.



Vaccination Analysis Interpretation:

In the context of the vaccination period, we extended our multivariate regression analysis to comprehend the shifting dynamics after starting vaccination. Upon model fitting, evaluation using the test data has demonstrated a notably improved Mean Squared Error compared to the Multivariate Linear Regression (MLR) model using pre-vaccination data. Specifically, the achieved Mean Squared Error is 0.0702. Furthermore, the responses have exhibited a strong fit, with R-squared values comfortably exceeding 0.80. Our models revealed intriguing patterns between vaccination metrics and COVID-19 responses. Notably, the cumulative number of persons vaccinated individuals exhibited a significant negative relationship with infection cases, deceased cases, and hospitalized patients, underscoring the protective effect of vaccination against severe outcomes. Mobility changes played a crucial role as well, with decreased activity in transit stations associated with reduced cases and hospitalizations. Moreover, testing rates remained significant in influencing case numbers, affirming their ongoing relevance even during vaccination. These insights illustrate the evolving landscape of COVID-19 outcomes in the wake of vaccination efforts.



Discussion

The central endeavour aimed to discern the impact of COVID-19 vaccination and other influential factors on infections, mortality, and hospitalisations through multivariate linear regression. Specifically, we contrasted data from the pre-vaccination and vaccination periods to grasp the pandemic's evolving dynamics.

Theoretical and Practical Implications:

The study's results emphasize the nuanced relationship between mobility, environmental factors, and COVID-19 outcomes. While vaccinations undeniably aid in pandemic control, a holistic approach, factoring in mobility patterns and climatic influences, becomes essential in epidemiological modelling. Practically, governments can use such insights to formulate targeted mobility restrictions or adjust vaccination campaigns.

Future Directions:

Further studies could probe the impact of vaccination on emerging COVID-19 variants, the role of vaccine boosters, and the synergistic effects of public health measures combined with vaccination.

Key Takeaways:

i) For both pre-vaccination and vaccination phases, increased testing correlated positively with confirmed cases, yet vaccination rates, particularly complete vaccinations, generally showed negative associations with infections and mortalities.

ii) Mobility patterns presented mixed results; for instance, mobility in parks, grocery, and pharmacy settings indicated potential hotspots for virus transmission.

iii) Intriguingly, climatic variables, such as higher temperatures and dew points, showed unexpected correlations with outcomes, pointing to possible external or unaccounted influences.

The findings underscore the paramount importance of vaccinations in controlling the pandemic. The negative correlation between complete vaccinations and mortality rates or infections solidifies the vaccine's protective role. Surprising positive correlations, like that between complete vaccinations and new deaths or between higher temperatures and infections, suggest the multifaceted nature of the pandemic's progression, demanding a more holistic perspective when strategizing health responses.

Conclusion

In our extensive analysis of COVID-19 dynamics and vaccination impact, several pivotal insights surfaced:

1. Testing remains paramount in understanding the pandemic's trajectory, with a direct correlation seen across essential metrics
2. Vaccination emerged as a crucial mitigator, with data overwhelmingly supporting its effectiveness in curbing the virus's devastation
3. Mobility, particularly in public spaces, significantly influences the spread, suggesting the importance of monitoring and managing public movements
4. Stringency measures by governments, though effective to an extent, yield optimal results when synergized with a robust vaccination drive
5. Weather, intriguingly, has its indirect yet noteworthy influence, with warmer climates showing a promising correlation with diminished infections

In essence, the investigation underscores the importance of a multi-pronged approach to Covid-19 control, where testing, vaccination, controlled mobility, stringent government measures, and an understanding of weather interplays are vital in steering us towards a safer tomorrow.

References

1. Richard A. Johnson and Dean W. Wichern. Multivariate Linear Regression Models in Applied Multivariate Statistical Analysis. (6th ed.) 360-418 Pearson Prentice Hall,2007.
https://ostad.hormozgan.ac.ir/ostad/UploadedFiles/863845/97050509-3761826667770356.pdf
2. R. Suganya, R.Arunadevi, & Seyed M.Buhari. COVID-19 Forecasting using Multivariate Linear Regression. Research Square.
https://www.researchsquare.com/article/rs-71963/v1
3. Lewis, F.I., Ward, M.P. Improving epidemiologic data analyses through multivariate regression modelling. Emerg Themes Epidemiol 10, 4 (2013).
https://doi.org/10.1186/1742-7622-10-4
4. Keeling MJ, Dyson L, Guyver-Fletcher G, et al. Fitting to the UK COVID-19 outbreak, short-term forecasts and estimating the reproductive number. Statistical Methods in Medical Research. 2022;31(9):1716-1737. doi:10.1177/09622802211070257
https://journals.sagepub.com/doi/full/10.1177/09622802211070257
5. Rencher, Alvin C. Methods of Multivariate Analysis. 2nd ed., John Wiley & Sons, Inc., 2002, pp. 337-339.
https://www.ipen.br/biblioteca/slr/cel/0241

Dataset source: Google API - Covid19 Open Data - https://storage.googleapis.com/covid19-open-data/v3/location/GB.csv