Excess Death Analysis

Anandraj Selvam
220164625
a.selvam@se22.qmul.ac.uk
MSc. Big Data Science
Queen Mary, University of London
Professor. Martin Neil

***Abstract*— An in-depth examination of mortality trends above expected levels across numerous populations and nations has become increasingly important in recent years. The goal of this research project is to conduct a thorough investigation of the factors that contribute to this phenomenon, including the global effects of Covid, the long-lasting effects of Covid-19, the impact of vaccination campaigns, the standard of healthcare systems, and the various levels of lockdown measures implemented globally. The underlying dataset will be gathered from official government websites, requiring skillful data gathering, data preprocessing, and wise data refining procedures. The gathered data will be rigorously analyzed using a multidimensional methodology, including statistical approaches like linear regression models. This analysis tries to reveal complex interactions between the components, illuminating both their unique and combined contributions to the development of excessive mortality. The main goal is to identify the basic factors that cause excess mortality, which will help us understand population health dynamics in the context of today's global concerns more thoroughly.**

Keywords*— *Covid-19, Vaccinations, Mortality, Linear Regression, Healthcare, Lockdown.

I. Introduction

This project focuses on conducting a exploratory analysis of excess deaths that surpass expected levels within multiple countries or populations in recent years. The investigation's time frame, which runs from January 2020 to June 2023, covers a vital period that was defined by important world events. Dissecting and understanding the complex web of circumstances that have contributed to the occurrence of these surplus fatalities is the main goal of this endeavor. The analysis aims to investigate various factors contributing to excess deaths: (1) First and foremost, the profound impact of the Covid-19 pandemic takes center stage. The project rigorously scrutinizes the fatalities directly attributed to the virus, aiming to discern the precise influence on the cumulative death toll, (2) The inquiry goes beyond the initial stages of infection and explores long lasting effects of Covid. To understand the complex interplay between mortality rates and persistent health effects, covid contracted in 2020/21/22 that caused death in 2023. (3) It's interesting to note that the study looks beyond the viral world to possible vaccination-related excess fatalities. This entails a thorough analysis of any unanticipated effects linked to vaccination programmes, considering the possibility of unfavorable outcomes and their possible influence on death rates, (4) The impact of different lockdown stringencies implemented worldwide and to understand the complex relationship between mobility limitations and variations in mortality by examining the severity and length of lockdowns enforced globally, (5) Quality of Healthcare across various countries. Understanding how variations in medical services may influence changes in excess deaths requires a comparative analysis of healthcare standards across numerous nations. (Neil, M. and Fenton, N., 2022)

The data necessary for the analysis will be collected from reliable sources, primarily public websites like Our World in Data and the Human Mortality Database. However, the data will be pre-processed by following steps such as data cleaning and manipulation before performing the analysis. A variety of tools are used in the field of data analysis. For calculation and statistical modelling, Microsoft Excel and the adaptable Python programming language emerge as the key components.

The study uncovers correlations and dependencies that lie at the heart of the links between the aforementioned parameters and the frequency of excess deaths through the lens of statistical methods, with a primary focus on linear regression models. It will allow to identify the significant correlations associated with factors and the occurrence of excess deaths. By analyzing the relationships and patterns discovered through statistical analysis, the project seeks to shed light on the primary factors influencing the occurrence of excess deaths in the studied countries or populations.

II. Related Works

Neil, M. and Fenton, N. (2022) has conducted an exploratory analysis to explore the possible causes of excess deaths in 2022. The study aims to consider alternative explanations for the increase in deaths beyond just blaming the Covid-19 vaccination program. The following describes several hypotheses that were tested:(1) Undiagnosed Covid Deaths: The possibility that Covid-19

caused deaths in 2022 that were not officially attributed to the virus. (2) Long-Covid: Deaths in 2022 that could be linked to lingering effects of Covid-19 contracted in 2020/2021. (3) Lockdown Stringency: The level of strictness of lockdown measures in 2020/2021 might have affected excess mortality in 2022. (4) Healthcare Quality: The quality and availability of healthcare services influencing excess deaths. (5) Covid Vaccinations: Whether the Covid-19 vaccinations administered in late 2020 and 2021 are contributing to excess deaths in 2022. Statistical analysis were used to explore these hypotheses, examining data from various countries and time periods. Using linear regression models, scatterplots, and correlation coefficients to evaluate the relationships between Covid-19 cases, vaccination rates, lockdown stringency, healthcare quality, and excess deaths. The conclusions drawn from their analysis are as follows: (1) Undiagnosed Covid Deaths: The study found no statistically significant relationship between Covid-19 cases in 2022 and the excess mortality experienced in the same year. This suggests that Covid-19 might not be a significant driver of the excess deaths. (2) Long-Covid: The data showed a negative correlation between Covid-19 rates in 2020/2021 and excess deaths in 2022, ruling out long-Covid as an explanation for the excess deaths. (3) Lockdown Stringency: The study found no significant statistical correlation between the strictness of lockdown measures and excess mortality. (4) Healthcare Quality: There was no observed relationship between the quality of healthcare and excess mortality, although some interplay between lockdowns and healthcare availability was acknowledged. (5) Covid Vaccinations: The analysis revealed a statistically significant linear relationship between highly vaccinated countries and excess deaths. This correlation was particularly evident after week 12 of 2022, even as Covid-19 cases declined.

Chudov, I. (2022). The research claims to have found a statistically significant correlation between the uptake of COVID-19 booster shots and excess mortality (higher deaths than expected) in the year 2022.The extract offered underlines a thorough examination of the probable link between the use of COVID-19 booster shots and the occurrence of increased mortality in 2022. To establish a strong statistical relationship, the inquiry carefully applies statistical procedures, using linear regression analysis. Since booster uptake accounts for a significant fraction of the observed variance in excess mortality, the computed R-squared value of 40% indicates this. An exceptionally low P-value of 0.0002 emphasises this statistical importance, supporting the validity of the results. The Short-Term Mortality Database and the Our World in Data repository were used as the primary data sources for the synthesis that forms the study's basis. The analytical approach is dependent on estimating excess mortality, which is done by comparing "deaths per week" for the year 2022 with comparable periods from 2017 to 2019. Notably, the author cautiously tackles restrictions, such as data exclusions resulting from a lack of data from particular countries. While the study's primary objective centres on establishing a correlation, the author judiciously refrains from drawing causal inferences. This sensible strategy recognises the complexity of demographic changes and the need for in-depth research. Notably, the data raises important questions about the possible causes of the link that was identified, such as post-vaccination mortality, changed booster efficacy, increased rates of reinfection, long-lasting effects of repeated immunisations, and prolonged spike protein production. This exploration raises important issues in the broader academic conversation and emphasises the need for more research. The complex interaction between booster uptake and excess mortality emphasises the significance of addressing the causal relationships in a thorough and nuanced manner while considering the complexity of demographic dynamics.

Chudov, I. (2022a). A significant investigation into the connection between COVID-19 vaccination rates and excess mortality rates across various countries. The investigation reveals a strong and obvious correlation between vaccination rates, including booster shots, and excess death rates in a few different countries, greater vaccination rates may be linked to greater mortality rates, according to this found association. The detailed discussion of the research methods used to examine the excess mortality data for time periods. The author emphasises the importance of low p-values, which support the notion that the relationships found are not coincidental occurrences. Over time, vaccination rate’s ability to explain excess mortality has progressed incrementally. It is noteworthy that this explanatory power increased from 27% in the initial phase (weeks 10-35) to 49% in the latter phase (weeks 20-44), demonstrating a strengthening relationship between immunisation and mortality. The study reveals a counterintuitive pattern in which the association between immunisation and increased mortality gets stronger with time. Contrary to what is often expected, correlations between variables don't always weaken over time. First off, a delayed effect of vaccination that eventually results in higher mortality may exist. Alternately, the diminishing protective effects of immunisation may over time reveal harmful effects, magnifying the relationship with death. This comparison highlights the unusual relationship between vaccination and mortality, which strengthens with time in the opposite direction of anticipated trends. The study refrains from offering conclusive explanations for the growing correlation between vaccination and excess mortality because it recognises the inherent uncertainty in

the situation. It raises intriguing questions and proposes possible directions for additional study and investigation.

III. Methodology

A. Dataset

The dataset for this exploratory analysis of excess deaths is collected from reliable sources such as "Our World in Data" (OWID) and the "Short Term Human Mortality Database." The Covid dataset has 68 features and 316579 instances, it holds the covid related entries on daily basis and the Human mortality database has 19 features and 6064 instances, it holds the mortality entries on weekly basis. (Jdanov, D.A., Galarza, A.A., et al., 2021)

The countries data used in the analysis is displayed on the charts using ISO-3166 country code system. These sources provide comprehensive and validated mortality data for multiple countries and populations. The dataset includes information on deaths, time periods, geographical locations, Covid-19 cases, vaccinations, healthcare indicators, lockdown stringencies, and other relevant variables. (Wikipedia.,2022)

B. Proposed Technique:

The analysis uses a multidimensional methodology that includes machine learning, statistical analysis, and data preprocessing. To guarantee data quality and consistency, the process starts with extensive data cleansing and modification. For an analysis to be accurate and trustworthy, this phase is crucial. Excess Mortality is calculated by :

Excess Mortality in 2023 = (Mortality Rate in 2023 / Average Mortality rate between 2017-2019) - 1

The data is initially preprocessed used Excel since the data entry period on both datasets are not same. Initially, removed the unnecessary columns and then aggregate the values on Weekly basis for both the dataset to be merged. Preprocessing is performed to address missing values (replacing them with zeros) and to ensure specific columns are of desired data types. It also involves grouping, aggregation, and calculation of various statistics based on different criteria, such as year and week numbers.

C. Machine Learning Model (Linear Regression):

Linear regression is a foundational machine learning technique used to explore relationships between variables and make predictions based on those relationships. It aims to find the best-fit straight line that represents the relationship between the variables in a way that minimizes the difference between the actual values of the dependent variable and the predicted values generated by the line. The equation for a simple linear regression model is often represented as:

$$y = mx+b$$

where,

- y is the dependent variable
- x is the independent variable
- m is the slope of the line
- b is the y-intercept

In the context of prediction, you can use a trained linear regression model to predict the values of the dependent variable for new or unseen values of the independent variable. In the context of understanding relationships, linear regression provides insights into how changes in the independent variable(s) are associated with changes in the dependent variable. In this project, linear regression is used to analyze connections between numerous variables and the occurrence of excess deaths, including Covid-19 cases, extended Covid deaths, vaccinations, healthcare quality, and lockdown requirements. The excess deaths are the dependent variable in a linear regression, and the factors are the independent variables. The study uses linear regression to quantify the impact of several factors on excess fatalities and to emphasize the extent to which each factor contributes to differences in mortality rates between populations and countries.

IV. Results and Discussion

(1) Covid Cases in 2023 causing Excess Deaths

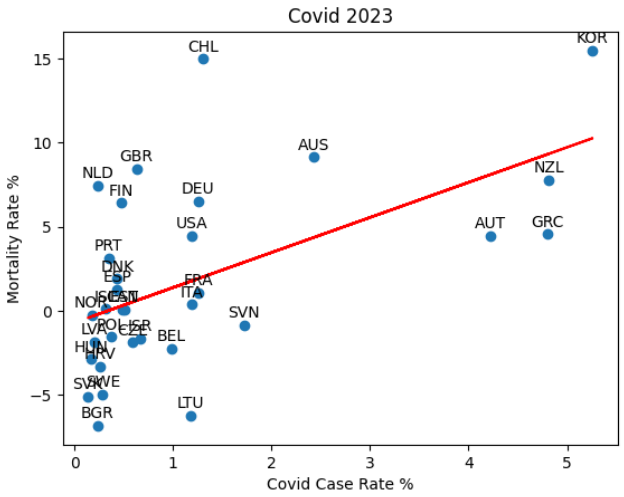

Fig. 1. Covid case rate vs Mortality rate

The scatter plot (Fig.1) shows the distribution of countries based on their Mortality Rate and Case Rate in

2023. Countries with higher case rates tend to have somewhat higher mortality rates, including Chile and South Korea, are outliers that depart from the general pattern. Although the death rates are relatively low, these nations have higher case rates. Most nations have lower rates of covid but higher and lower than average mortality rates. Regression line suggest a positive correlation between case rate and mortality rate. However, the spread of points and the presence of outliers indicate that case rate alone is not the sole determinant of mortality rate. The relationship is likely influenced by various other factors.

TABLE I.
Calculation of R-Square and P-value

Scenario	**R-square value**	**P-value**
Covid	0.302	0.001
Historic Covid (2020-2022)	0.036	0.295
Lockdown Stringency	0.167	0.020
Covid Vaccinations	0.402	0.009

(2) Covid Cases in 2020-2022 causing Excess Deaths in 2023

Another case is that the deaths in 2023 that could be linked to lingering effects of Covid-19 contracted in 2020/2021/2022. Like the previous analysis, there are outliers that deviate from the general trend, CHL (Chile) still stand out, indicating that it has lower case rate, but higher Mortality rate as shown in Fig.2. It is noticeable that countries like Slovenia, France and Denmark have higher covid rate but very low mortality rate. Bulgaria has the lowest mortality of all.

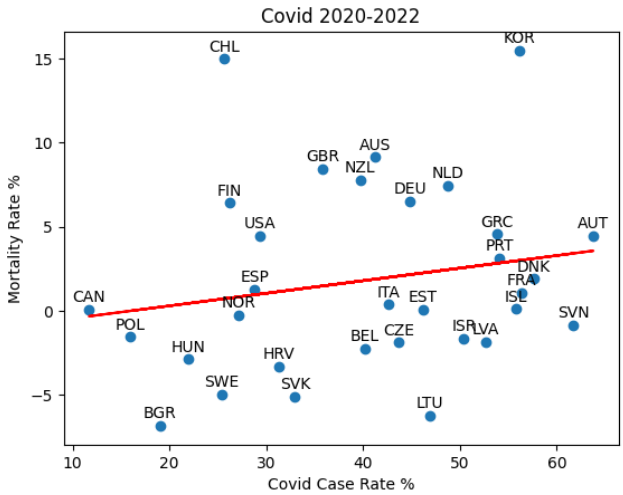

Fig. 2. Covid cases from 2020-2022 vs Mortality

From the Table I, the p-value is relatively high, it indicates weaker evidence against the null hypothesis. The R-square value is very close to zero, it shows less variation in the data. There is no significant correlation between the Historic Covid and Mortality Rates. Hence, the covid cases during 2020-2022 is not causing the excess deaths in 2023.

(3) Lockdown Stringency in 2020-2022 causing excess deaths in 2023

The level of strictness of lockdown measures implemented worldwide might have affected excess mortality in 2023. Unfortunately, no data has available for 2023 since most of the countries removed the lockdown. But for the analysis, the data which has been used is recorded between 2020-2022. The Table I shows that p-value is lower than the threshold 5%, hence it shows that excess deaths had significant relationship with lockdown stringency, and the R-square value of 0.03 is very low and the scatterplot (Fig.3) shows that many points deviate from the line.

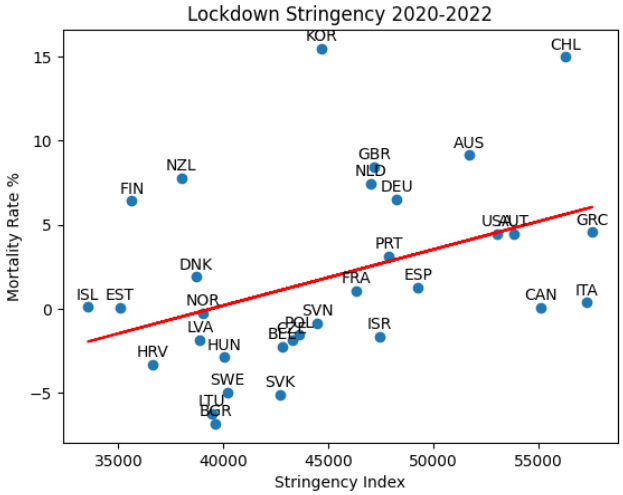

Fig. 3. Lockdown Stringency Index vs Mortality

(4) Covid Vaccinations causing excess deaths in 2023

This scenario implies whether the Covid-19 vaccinations administered from late 2020 are causing excess deaths in 2023. From the scatterplot (Fig.4), it is evident that countries like Chile and South Korea are highly vaccinated and have high mortality rate. It shows that vaccinations not effective in that region, it is same as last year analysis. Nearly 93% of the countries in this analysis have more that 50% of fully vaccinated rate. While comparing with 2022 analysis, nearly half of the countries in the Fig.4 have negative mortality rate, means there is no excess.

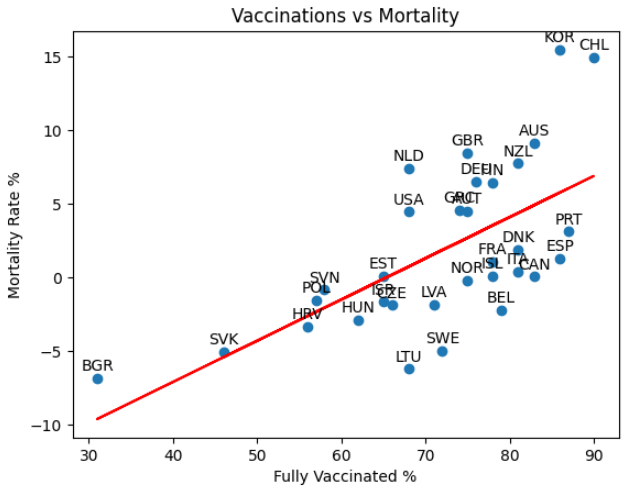

Fig. 4. Vaccinations vs Mortality

Table I shows that p-value is 0.009 less than the threshold 0.05 and making it statistically significant, and the R-square value is relatively high and has positive correlation which suggests that Vaccinations has a meaningful impact on Mortality rate. (Neil, M. and Fenton, N., 2022)

(5) Quality of Healthcare causing Excess Deaths in 2023

Unfortunately, there is no data on quality of health care for the period 2020-2023. It is not possible to detect any connection between excess mortality and quality of healthcare. The analysis has performed with 2015 healthcare data, healthcare quality has no relationship with excess mortality. (Neil, M. and Fenton, N., 2022)

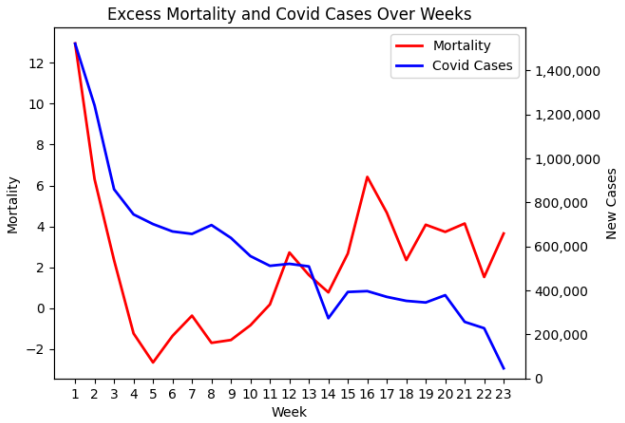

Fig. 5. Aggregated Covid cases and Mortality in 2023

While aggregating the mortality of all the countries in the analysis, it is evident that excess mortality matches with the covid trend (Fig.5), and it also shows that overall covid cases and excess mortality rate has been falling in 2023. From week-1 to week-8 there's huge drop in the mortality rate and Covid cases, and around week-9 excess mortality rate starts increasing gradually, but the covid cases rate still dropping. The excess mortality signal does not separate from the covid signal until around week 13, after which the excess stays high while covid falls.

From the last year analysis (2022), the scatterplot (Fig.6) shows that excess mortality follows the covid trend closely: when covid increases, excess increases shortly after. The excess mortality signal does not separate from the covid signal until around week 12, after which the excess stays high while covid falls. The overall excess mortality rate and number of covid cases is low in 2023, while comparing with year 2022. (Neil, M. and Fenton, N., 2022)

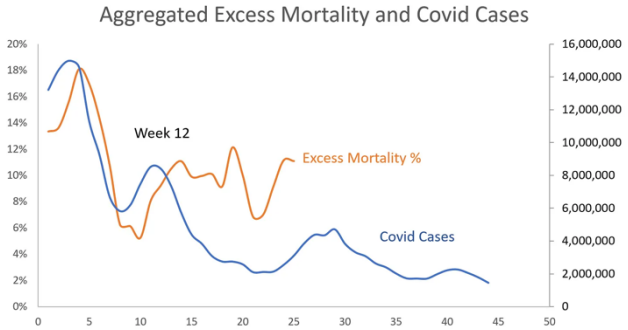

Fig. 6. Aggregated Covid cases and Mortality in 2022

A. Temporal Analysis

To examine if covid-19 or vaccines causes excess deaths at various times of the year (weeks), the temporal analysis has been performed.

B. Covid Temporal Analysis

The scatterplot shows a general trend where, as the case rate increases, the average mortality rate tends to increase as well. There are a few points where the general trend is notably different. Examples include nations with high death rates but relatively low case rates (outliers above the trend line) and nations with low case rates but high mortality rates (outliers below the trend line). as shown in Fig.6(a).

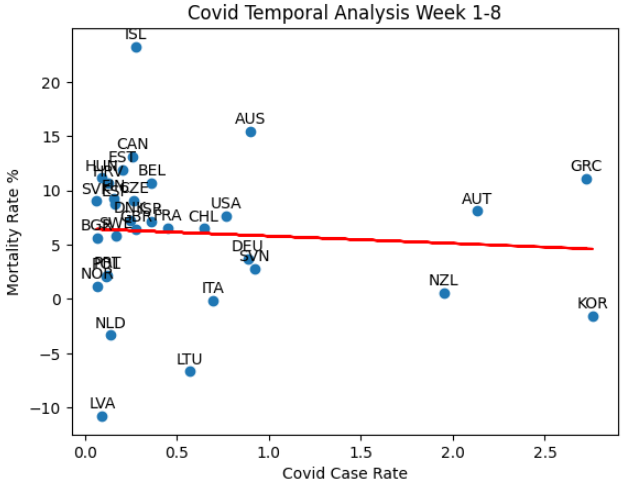

Fig. 6(a). Covid Temporal analysis (Week 1-8)

The slope of the line is negative, indicating a negative correlation between case rate and mortality rate. From the Table II, If the null hypothesis were true, there is a 67.4% chance of seeing this result (or something even more extreme), according to the p-value of 0.674. Statistical significance is typically defined as a p-value less than a significance level, which is typically 0.05. Because the p-value in this instance is substantially higher than 0.05, we are unable to rule out the null hypothesis, indicating that there is insufficient data to draw a conclusion about the statistical significance of the association between "Case Rate" and "Mortality Rate." The degree to which the regression line matches the data points is indicated by the R-squared value.

TABLE II.
Covid and Vaccines Temporal Analysis

Temporal Analysis	Covid-19		Vaccines	
Weeks	R-square value	p-value	R-square value	p-value
1 to 8	0.005	0.674	0.004	0.773
9 to 16	0.202	0.009	0.407	0.001
17 to 23	0.234	0.005	0.368	0.007

With an R-squared value of 0.00598 the "Case Rate" can only account for around 0.598% of the variation in the "Mortality Rate." This suggests that the linear relationship between these two variables is incredibly weak.

Note that during week 1-8, South Korea (KOR) has negative mortality rate. But note that during week 9-16, South Korea has the second highest mortality rate as shown in Fig.6(b). The outliers might be affecting the regression line's overall slope.

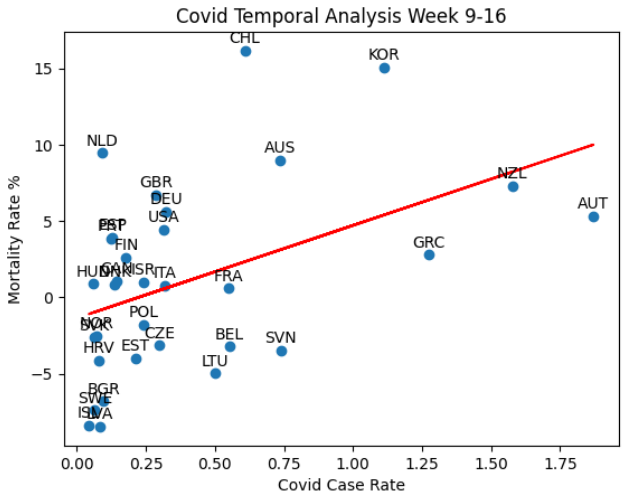

Fig. 6(b). Covid Temporal analysis (Week 9-16)

There are instances where the case rate is low, but the average mortality rate is relatively high (e.g., Netherlands and Spain). This highlights that while the overall trend is positive, the relationship is not entirely consistent for all countries. Some countries have zero new cases, which leads to zero case rates. This suggests that in the context of this dataset, there is a significant association between these two variables. According to the moderate R-squared value of 0.20297, variations in case rates between the nations account for about 20.30% of the variation in mortality rates. This shows that in this dataset, "Case Rate" is a significant predictor of changes in "Mortality Rate." Despite the statistical significance of the association, it's vital to keep in mind that correlation does not imply causality. There may be additional variables impacting this association that the analysis has not considered. These data points could have an impact on the slope and interpretation of the regression line (Fig.6(b)).

During the week 17-23, 87% of the countries have very less covid case rate and not that South Korea reaches the highest mortality rate and covid rate. Given that the p-value is less than the significance level of 0.05, there is strong evidence refuting the null hypothesis that there is no significant correlation between "Case Rate" and "Mortality Rate." According to the R-squared value of 0.23429, the "Case Rate" may account for about 23.43% of the variation in the "Mortality Rate." This suggests that there is a moderate correlation between the two variables.

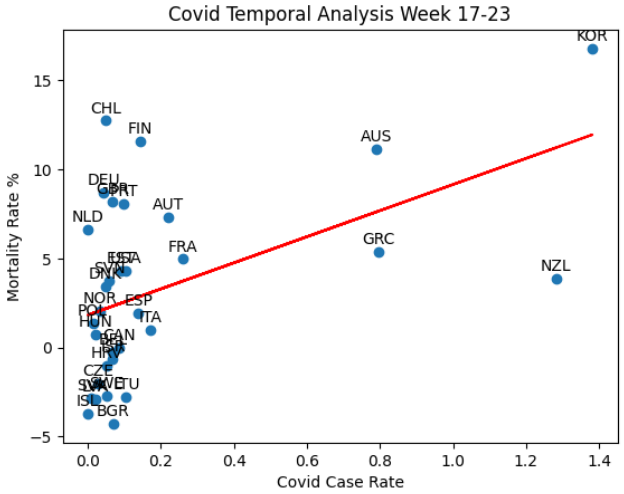

Fig. 6(c). Covid Temporal analysis (Week 17-23)

The R-squared values for covid temporal analysis are typically low across all time intervals as shown in Table II, indicating that a significant amount of the variability in the dependent variable (perhaps the average fatality rate) throughout these weeks is not clearly explained by the linear model. The observed connections over the first period are not statistically significant, as shown by the comparatively high p-values for those times. The second- and third-time interval (weeks 17 to 23) has a lower p-value than the first, which may indicate a stronger hint of a

significant association during this period, though it may not be very strong.

C. *Vaccine Temporal Analysis*

To examine whether vaccines cause excess mortality at various time intervals in 2023, the data recorded over the weeks has been split into three periods. Note, Countries like United Kingdom, Hungary, Iceland, Latvia, Netherlands, Norway, Slovakia, Slovenia are not reporting the vaccination data. Few countries like Australia, New Zealand, Chile, Canada have reported vaccinations data only for few weeks. Hence, the temporal analysis has been done only with available data. From the Fig. 7(a), it is evident that CHL(Chile) is the most vaccinated and have mortality rate over the average and stopped reporting the data after week 8. The R-squared value for the time period Weeks 1 to 8 is very low (0.004), indicating that the vaccination variable explains very little variance in the data. The strong p-value (0.773) indicates that there is no statistically significant correlation between the observed data during this period and vaccinations.

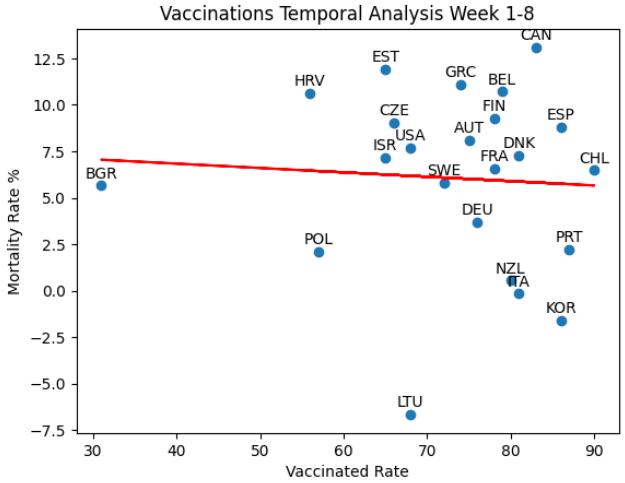

Fig. 7(a). Vaccine Temporal analysis (Week 1-8)

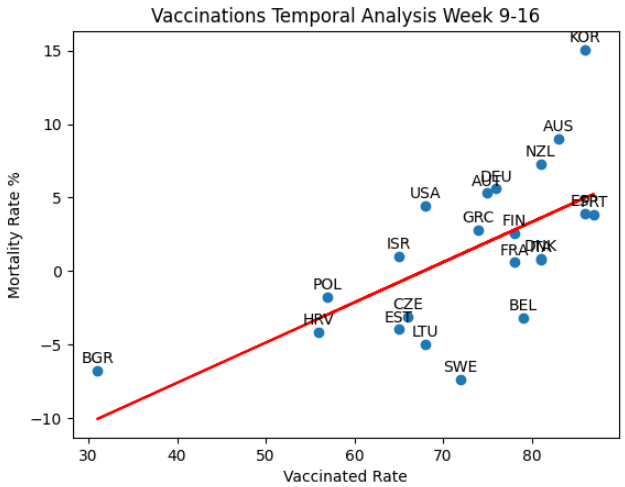

Fig. 7(b). Vaccine Temporal analysis (Week 9-16)

During week 9 to 16, the overall mortality rate has been rising gradually across the countries as shown in Fig.5 and there's a huge rise of mortality rate in South Korea from week 9 as shown in Fig. 7(b). Only few European countries like Sweden, Belgium, Croatia, Poland falls below the excess mortality rate. Even though countries like Portugal, Spain, Finland, Australia, and New Zealand are highly vaccinated these countries have high mortality rate, which shows that vaccines are not effective as it was expected.

The R-squared value is greater (0.407) for the time period Weeks 9 to 16, pointing to a stronger correlation between the observed data and the vaccinations. The little p-value (0.001) indicates that there is statistical significance in this association.

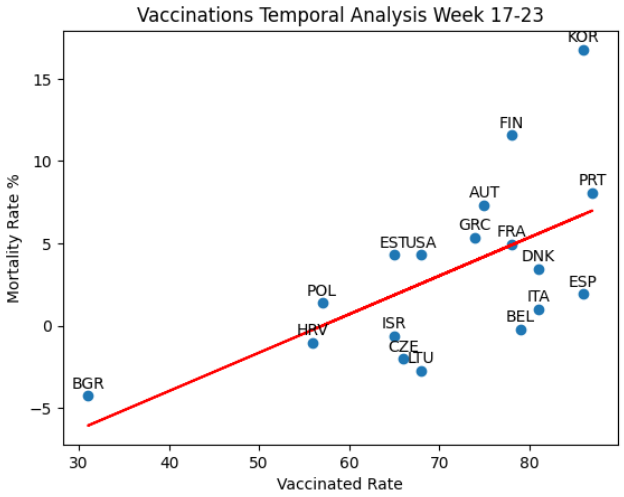

Fig. 7(c). Vaccine Temporal analysis (Week 17-23)

During week 17-23, the overall covid case rate falls exponentially, and the mortality rate has risen up for all countries as shown in Fig.7(c). It shows that Covid cases in 2023 are not responsible for excess mortality. Hence, indicates that the vaccine-related variable is linked to the result, but there is still a sizable degree of variability that cannot be accounted for by this variable alone. The R-squared value is similarly rather high (0.368) for the time period "Weeks 17 to 23," and the p-value (0.007) shows that the association between vaccinations and the observed data during this time period is also statistically significant. (Neil, M. and Fenton, N., 2022)

V. Conclusion

First and foremost, it's important to acknowledge that the study's use of somewhat coarse temporal period leaves room for more precise ones. It is important to consider the influence of reporting lags and delays, which is a key element. Clearly, the increase in Covid-19 cases and its impact on excess mortality highlight the limitations of vaccine effectiveness.

The temporal analysis supports that the vaccines are not effective. This observation does not offer any new insights and is consistent with existing knowledge. Although an assessment of healthcare quality was made, concerns about the suitability of the criteria in this area still exist. Many countries have not maintained data properly or stopped reporting like number of fully vaccinated people, number of covid cases and mortality rates. Likewise, if the clear data has been provided, it is possible to find the root cause with strong evidence to support it.

There is a strong indication that at least some of the increased death rate is being caused, if not all, by the vaccination programme. The vaccines don't seem to be safe based on this evidence.

A. Future Works

Due to inconsistencies in the available data, reporting delays, and other limitations in the analysis, arriving at a conclusive understanding is challenging. A solid foundation of evidence is necessary to draw firm conclusions. With access to accurate and comprehensive data, we can conduct a thorough analysis to uncover the underlying reasons behind the elevated mortality rates. This analysis will play a crucial role in guiding further research and facilitating the implementation of new policies.

REFERENCES

Chudov, I. (2022). *PROVEN RELATIONSHIP: COVID Boosters and Excess Mortality in 2022*. [online] www.igor-chudov.com. Available at: https://www.igor-chudov.com/p/proven-relationship-covid-boosters?utm_source=substack&utm_campaign=post_embed&utm_medium=web

Chudov, I. (2022a). Association Between Vaccines and EXCESS MORTALITY Getting Stronger -- and is Discussed in UK Parliament. [online] www.igor-chudov.com. Available at: https://www.igor-chudov.com/p/association-between-vaccines-and?utm_source=substack&utm_campaign=post_embed&utm_medium=web

Craig, C., Neil, M., Fenton, N., McLachlan, S., Smalley, J., Guetzkow, J., Engler, J., Russell, D. and Rose, J., 2022. Official mortality data for England reveal systematic undercounting of deaths occurring within first two weeks of Covid-19 vaccination.

Dee, J. (2023). COVID vs. COVID (part 1). [online] John Dee's Almanac. Available at: https://jdee.substack.com/p/do-covid-vaccines-work-part-13.

Doshi, P. (2020). Will covid-19 vaccines save lives? Current trials aren't designed to tell us. BMJ, [online] 371. doi:https://doi.org/10.1136/bmj.m4037.

Fenton, N.E., Neil, M. and McLachlan, S., 2021. Paradoxes in the reporting of Covid19 vaccine effectiveness: Why current studies (for or against vaccination) cannot be trusted and what we can do about it.

Haklai, Z., Aburbeh, M., Goldberger, N. and Gordon, E.-S. (2021). Excess mortality during the COVID-19 pandemic in Israel, March–November 2020: when, where, and for whom? Israel Journal of Health Policy Research, 10(1). doi:https://doi.org/10.1186/s13584-021-00450-4.

Islam, N., Shkolnikov, V.M., Acosta, R.J., Klimkin, I., Kawachi, I., Irizarry, R.A., Alicandro, G., Khunti, K., Yates, T., Jdanov, D.A., White, M., Lewington, S. and Lacey, B. (2021). Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries. BMJ, p.n1137. doi:https://doi.org/10.1136/bmj.n1137.

Jdanov, D.A., Galarza, A.A., Shkolnikov, V.M., Jasilionis, D., Németh, L., Leon, D.A., Boe, C. and Barbieri, M. (2021). The short-term mortality fluctuation data series, monitoring mortality shocks across time and space. Scientific Data, [online] 8(1), p.235. doi:https://doi.org/10.1038/s41597-021-01019-1.

Kepp, K.P., Björk, J., Kontis, V., Parks, R.M., Bæk, K.T., Emilsson, L. and Lallukka, T. (2022). Estimates of excess mortality for the five Nordic countries during the COVID-19 pandemic 2020−2021. International Journal of Epidemiology, 51(6), pp.1722–1732. doi:https://doi.org/10.1093/ije/dyac204.

Levitt, M., Zonta, F. and John (2023). Excess death estimates from multiverse analysis in 2009–2021. European Journal of Epidemiology. doi:https://doi.org/10.1007/s10654-023-00998-2.

Msemburi, W., Karlinsky, A., Knutson, V., Aleshin-Guendel, S., Chatterji, S. and Wakefield, J. (2022). The WHO estimates of excess mortality associated with the COVID-19 pandemic. Nature. doi:https://doi.org/10.1038/s41586-022-05522-2.

Neil, M. and Fenton, N. (2022). *The Devil's Advocate: An Exploratory Analysis of 2022 Excess Mortality*. [online] Where are the numbers? by Norman Fenton and Martin Neil. Available at:

https://wherearethenumbers.substack.com/p/the-devils-advocate-an-exploratory#footnote-1-90143623.

Neil, M., Fenton, N. and McLachlan, S., 2021. Discrepancies and inconsistencies in UK Government datasets compromise accuracy of mortality rate comparisons between vaccinated and unvaccinated.

Neil, M., Fenton, N., Smalley, J., Craig, C., Guetzkow, J., McLachlan, S., Engler, J., Russell, D. and Rose, J., 2022. Official mortality data for England suggest systematic miscategorisation of vaccine status and uncertain effectiveness of Covid-19 vaccination. *Published online January*, *12*.

Neil, M., Fenton, N.E., Smalley, J., Craig, C., Guetzkow, J., Mclachlan, S., Engler, J. and Rose, J., 2021. Latest statistics on England mortality data suggest systematic mis-categorisation of vaccine status and uncertain effectiveness of Covid-19 vaccination. *ResearchGate, December*.

Rancourt, D., Baudin, M. and Mercier, J., 2021. Nature of the COVID-era Public Health Disaster in the USA, from All-cause Mortality and Socio-geo-economic and Climatic Data.". *Denis Rancourt*, *21*.

Rizzi, S. and Vaupel, J.W., 2021. Short-term forecasts of expected deaths. *Proceedings of the National Academy of Sciences*, *118*(15), p.e2025324118.

WHO (2020). WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020. [online] World Health Organization. Available at: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

Wikipedia. (2022). List of ISO 3166 country codes. [online] Available at: https://en.wikipedia.org/wiki/List_of_ISO_3166_country_codes

MSc Project - Reflective Essay

Project Title:	Excess Death Analysis
Student Name:	**Anandraj Selvam**
Student Number:	**220164625**
Supervisor Name:	Professor. Martin Neil
Programme of Study:	MSc Big Data Science FT

I. Introduction

This project delves into an exploratory analysis of the phenomenon of excess deaths surpassing anticipated levels within various countries or populations in recent years. This research project aims to conduct a thorough investigation of the factors that contribute to this phenomenon, such as the global effects of Covid, the long-lasting effects of Covid-19, the impact of vaccination campaigns, the quality of healthcare systems, and the various levels of lockdown measures implemented globally. The underlying dataset will come from official government websites, necessitating careful data collection, pre-processing, and refining techniques. A multifaceted methodology will be used to thoroughly analyse the collected data, using statistical techniques like linear regression models. This research seeks to illuminate the intricate relationships between the factors, highlighting both each factor's individual and collective contributions to the emergence of excessive mortality. To better understand population health dynamics in the context of current global issues, it is important to identify the fundamental variables that contribute to excess mortality.

II. Research Process

The data used for this exploratory research of excess deaths was carefully gathered from reputable sources, including "Our World in Data" (OWID) and the "Short Term Human Mortality Database." The ISO-3166 country code system is used to represent nation-specific data in the study, enabling crystal-clear chart visualisation. These sources are well known for providing reliable and verified mortality data covering a wide range of nations and demographics. This dataset contains a wide range of useful information, including data on mortality, time intervals, geographic locations, Covid-19 occurrences, immunisation rates, healthcare measures, degrees of lockdown limitations, and other relevant factors.

The method of analysis is fundamentally multidimensional, involving statistical inspection, careful data pre-processing, and machine learning. The foundation of this process is the careful vetting and enhancing of the data to guarantee its consistency and integrity. This first stage is crucial for ensuring the accuracy and reliability of the subsequent analysis. Excel is used to perform preliminary pre-processing on the data, taking into account the differences in timing between the two datasets data entry. The earliest steps entail the elimination of unnecessary columns, followed by the weekly aggregate of values, which harmonises the data in preparation for eventual amalgamation. Data gaps are addressed (and filled with zeroes in subsequent pre-processing steps), and selected columns are aligned with the desired data types.

In order to uncover the complex interactions between multiple variables and the occurrence of excess deaths, linear regression emerges as the analytical technique of choice. The following factors are being examined: vaccination rates, the standard of healthcare, Covid-19 cases, protracted Covid-related mortality, and the rigour of lockdown procedures. The concept of excess deaths, which serves as the dependent variable in the context of linear regression, is the foundation of this research. On the other hand, the many aspects that have been clarified serve as the independent variables. This study uses linear regression to measure the impact that various factors have on the occurrence of excess mortality. This methodology is used in the investigation in an effort to show how much each individual element affects death rates differently in different populations and nations.

III. Challenges and Learnings

There are several difficulties and chances for learning when doing an exploratory investigation of excess fatalities, such as the one described in this project. Below are the challenges. Making sure the dataset is reliable and consistent is one of the biggest hurdles in data analysis. Dealing with data gathered from many sources might result in inconsistencies, missing values, and mistakes. To get reliable findings, you must learn how to pre-process and clean data correctly. Pre-processing various datasets can be complicated and time-consuming, especially when they have different entry periods and structures. By overcoming this obstacle, we can learn useful techniques for manipulating data and wrangling it, which will help us comprehend data integrity better. It can be difficult to choose which variables to include in the study and comprehend the relationships between them. Although linear regression offers insightful information, accurate interpretation of the data requires a firm understanding of statistical ideas. It is essential to learn how to draw insightful conclusions while avoiding misunderstandings. Working with delicate data, such as mortality rates and healthcare quality, necessitates navigating ethical issues with data protection, appropriate analysis, and avoiding biased results.

IV. Limitations

The division of data into distinct time periods follows a coarse approach, potentially overlooking more refined temporal patterns that could yield deeper insights. The quality of the data and the definitions used are absolutely necessary for the analysis to be reliable and accurate. Any biases or inconsistencies in the input data could be carried over into the results of the study. The existence of reporting lags and delays raises serious concerns. Particularly in situations when events are changing quickly, like the context of a pandemic, the timeliness of data reporting can greatly affect how events are portrayed. As an exploratory project, the main objective is to discover new ideas rather than definitive conclusions. Given the inherent limitations of the study's scope and objectives, this preliminary approach means that the findings might not offer a completely comprehensive or conclusive perspective.

V. Future Works

Since there are many limitations with the analysis due to data inconsistency, reporting lags and delays, we can’t come to a conclusion without a strong evidence. With the proper data, we can able to analyse and find the root cause of excess mortality and

helps to continue further research and taking measures by implementing new policies. Although the study's research indicates that Covid-19 vaccines significantly lower death rates, it would be beneficial to look into the immunisations' long-term consequences. A longer time of data collection and analysis could reveal information on the longevity of vaccine protection and the possible appearance of new variants. Despite the fact that the report found no association between healthcare quality and excess mortality throughout the study period, a more thorough investigation of the capability and quality of healthcare systems may provide valuable information. It might be beneficial to look into how healthcare infrastructure, medical resource accessibility, and healthcare access affect excess mortality rates. Data from 2023 and later would let researchers to evaluate the long-term impacts of Covid-19, vaccines, and other variables because the research paper's data only goes up to 2022. This might offer insightful information on tendencies that go beyond the time period under study.

VI. Conclusion

This research report provided a thorough examination of the intricate connections between the prevalence of Covid-19, mortality rates, lockdown procedures, vaccines, and excess deaths. The results demonstrated the complex nature of these interactions, with diverse trends across various nations and epochs. While some trends did show up, outliers and data fluctuations highlighted the need for careful interpretation. According to the data, Covid-19 vaccines significantly lowered mortality rates whereas other factors, such as the severity of lockdowns, did not appear to be significantly associated with an increase in deaths. Overall, the report presented suggestions for further research and policy concerns and offered insightful information about the complex dynamics of the Covid-19 pandemic's effect on mortality.