PROJECT 3

Predictive Analytics for COVID-19 Infection and Mortality Rates Among Healthcare Personnel

Bharosa Gauchan, Mardochee Duffaut, Alexia Vasquez

Bellevue University- May 26, 2024

TABLE OF CONTENTS

INTRODUCTION

The COVID-19 pandemic has underscored the vulnerabilities of healthcare systems worldwide, with healthcare personnel (HCP) facing unprecedented risks due to their continuous exposure to infected patients. Understanding the infection and mortality rates among HCP is crucial for developing effective strategies to mitigate these risks. Predictive analytics offers powerful tools for forecasting trends and identifying high-risk scenarios, enabling proactive measures to safeguard healthcare workers. Healthcare workers are essential in managing pandemics, but they also represent a high-risk group for infection. Early in the COVID-19 pandemic, numerous reports highlighted significant infection rates among HCP, leading to severe workforce shortages and increased mortality. The integration of predictive analytics in healthcare can provide valuable insights into infection patterns and mortality risk factors, aiding in the development of protective measures.

Our project involves the use of statistical techniques, machine learning, and data mining to analyze current and historical data to make predictions about future events. In the context of COVID-19, predictive models will analyze various factors such as demographics (age, gender, underlying health conditions), workplace exposure (intensity of patient contact, protective measures), and environmental factors (hospital infrastructure, regional infection rates).

To accomplish this project, we will use a comprehensive dataset from the Centers for Disease Control and Prevention's COVID-19 Response, which includes various metrics related to weekly cases among healthcare personnel (version date: September 14, 2023).

MMWR week :represents the specific week of the year as defined by the CDC's MMWR calendar

Case count: refers to the total number of recorded instances (cases) of a particular disease or condition within a specified period and geographic area

Death count: refers to the total number of deaths recorded due to a specific disease, condition, or event within a specified period and geographic area

BUSINESS PROBLEM/HYPOTHESIS

Business Problem:

The COVID-19 pandemic has placed extraordinary stress on healthcare systems worldwide, with healthcare personnel (HCP) being disproportionately affected due to their direct exposure to the virus. High infection and mortality rates among HCP not only compromise the health and safety of these essential workers but also exacerbate staffing shortages, impacting the overall capacity of healthcare facilities to manage patient care. Effective management and mitigation strategies are crucial to protect HCP and maintain the integrity of healthcare services.

Hypothesis:

By leveraging predictive analytics, it is possible to accurately forecast COVID-19 infection and mortality rates among healthcare personnel. This forecasting capability can identify high-risk scenarios and underlying risk factors, enabling targeted interventions that reduce infection rates and improve safety outcomes for healthcare workers. Specifically, predictive models can help:

1. Anticipate infection hotspots within healthcare settings.
2. Optimize the allocation of personal protective equipment (PPE) and other critical resources.
3. Implement timely interventions based on predicted trends.
4. Identify demographic and workplace factors contributing to higher infection and mortality rates among HCP.

Testing this hypothesis involves developing and validating predictive models using comprehensive datasets, such as the weekly COVID-19 cases and deaths among healthcare personnel provided by the Centers for Disease Control and Prevention (CDC). The success of these models in accurately forecasting infection and mortality trends will demonstrate their potential as tools for enhancing pandemic response strategies in healthcare environments.

METHODS/ANALYSIS

Data Collection:

The primary dataset for this project is the "COVID-19 Cases and Deaths Among Healthcare Personnel, by week" from the Centers for Disease Control and Prevention (CDC), with the version dated September 14, 2023. This dataset includes various metrics such as the number of confirmed COVID-19 cases, deaths, and demographic information among healthcare personnel on a weekly basis.

Additional Data Sources:

- **Health records** from hospitals and clinics, providing detailed patient interactions and health conditions.
- **Demographic information** of healthcare personnel, including age, gender, and pre-existing health conditions.
- **Workplace exposure data**, encompassing PPE usage, patient contact intensity, and hospital infrastructure.
- **Regional COVID-19 infection rates** and **hospital capacity** data to understand the broader context of the healthcare environment.

Phase 1: Data Preprocessing:

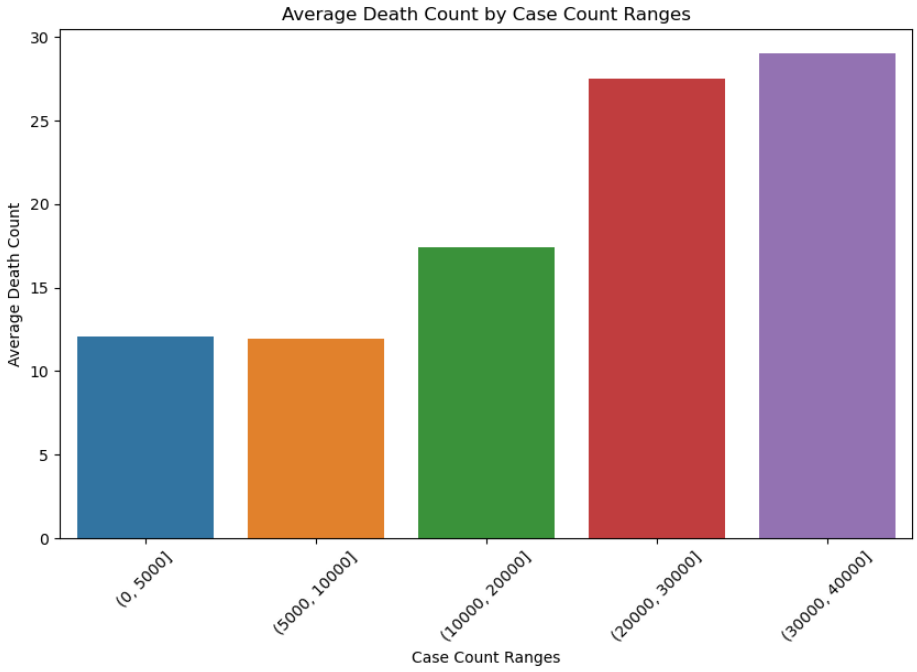
1. **Data Cleaning:** Removing duplicates, handling missing values, and correcting any inconsistencies.
2. **Data Normalization:** Standardizing data formats and scales to ensure compatibility across different datasets.

3. **Feature Engineering:** Creating new features that could be relevant for the predictive models

Phase 2: Model Development:

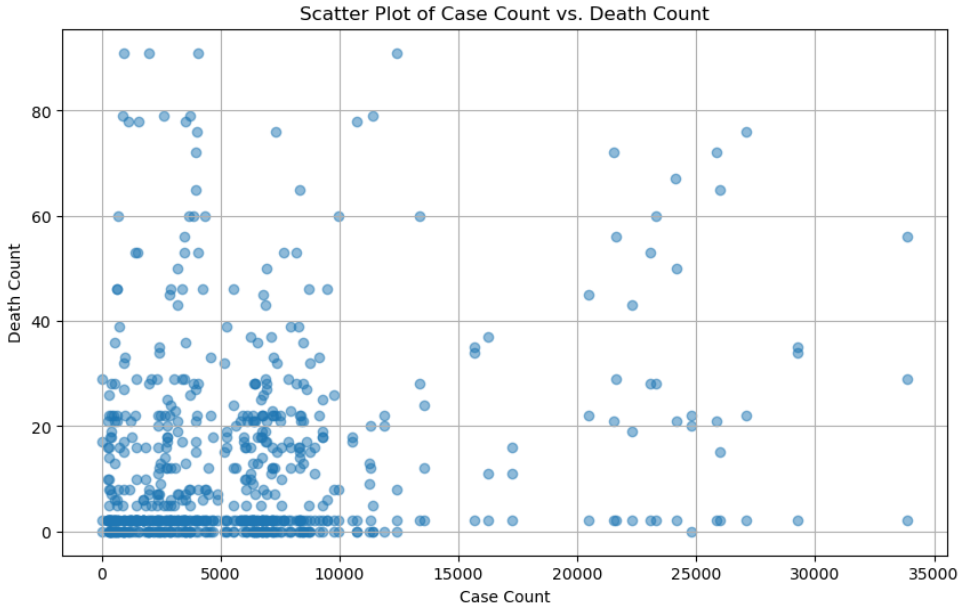
1. Exploratory Data Analysis (EDA):

- Visualizing trends and patterns in the data using histograms, scatterplot, and line chart.



The graph shows an increasing trend in the average death count as the case count range increases. This suggests that higher case counts are associated with higher average death counts.

- Identifying correlations between different variables cases count and deaths count



General Trend:

There doesn't appear to be a simple, linear relationship between case count and death count. The scatter plot shows a lot of variation, especially in the lower-case count ranges.

High death counts can occur even with relatively low case counts, possibly indicating the presence of outbreaks or particularly severe cases in those data points.

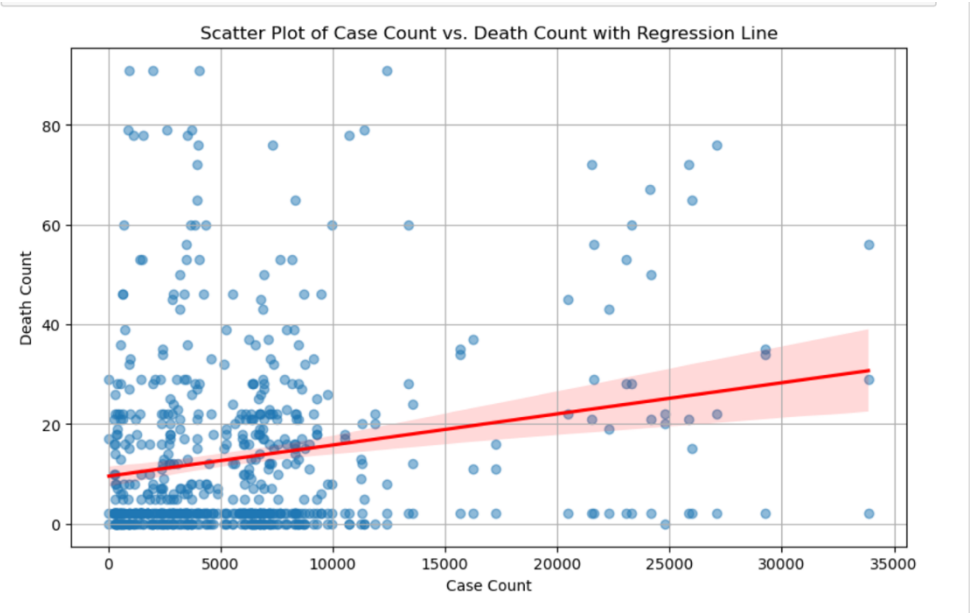
Clusters:

Low Case Counts, High Death Counts: There is a notable cluster of points with low case counts but high death counts, suggesting that in some instances, a small number of cases can result in a high number of deaths.

High Case Counts, Moderate Death Counts: For higher case counts, the death counts seem to be more moderate, indicating that larger case numbers do not necessarily result in proportionally higher death counts.

2. Model Selection:

- **Statistical Models:** Logistic regression to estimate probabilities of infection and mortality.



The regression line slopes upwards, indicating a positive relationship between case count and death count. As the number of cases increases, the number of deaths tends to increase as well.

The slope of the line is relatively shallow, suggesting that while there is a positive correlation, it is not very strong.

For lower case counts (0 to 5000), there is a high density of points with a wide range of death counts (0-80), showing a lot of variability.

For higher case counts (5000 to 35000), the points are more spread out, and the death counts generally increase, but the variability in death counts is also present.

a) **Model Training and Validation:**

- Splitting the data into training and testing sets to evaluate model performance.
- Using cross-validation techniques to ensure robustness and generalizability.
- Tuning hyperparameters to optimize model accuracy.

b) **Model Evaluation:**

- **Accuracy:** The proportion of correctly predicted cases out of the total cases.
- **Precision and Recall:** To measure the performance of the classification models in identifying true positives and minimizing false negatives.
- **ROC-AUC Curve:** To evaluate the trade-off between true positive and false positive rates.
- **Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE):** For assessing the performance of time series models.

c) **Implementation of Predictive Models:**

- Integrating the predictive models into a real-time dashboard for healthcare administrators.
- Providing actionable insights and recommendations based on model predictions, such as:
    - Allocating additional PPE to high-risk areas.
    - Adjusting staffing levels in anticipation of infection surges.
    - Implementing targeted training and awareness programs for at-risk groups.

3. **Ethical Considerations:**

- Ensuring data privacy and security by anonymizing healthcare personnel data.
- Addressing potential biases in predictive models to avoid discrimination and ensure fairness.
- Implementing transparent and explainable AI practices to build trust among stakeholders.

RESULTS

For our analysis, we examined the correlation between COVID-19 infection cases and deaths among healthcare workers. The correlation coefficient was found to be 0.204, which indicates a weak positive linear relationship. This relationship suggests that while increases in infection case counts do lead to increases in death counts, the relationship is very weak.

Interpretation of Findings:

- Weak Correlation: The weak correlation (0.204) implies that the number of infection cases among healthcare workers is not strongly predictive of the number of deaths. In other words, although there is a slight tendency for death counts to increase as case counts increase, this tendency is not strong.

- Public Health Implications: This finding is crucial for public health strategies. It implies that direct interventions solely aimed at reducing the number of infection cases might not be sufficient to significantly reduce the number of deaths among healthcare workers.

RECOMMENDATIONS/ETHICAL CONSIDERATIONS

Based on our findings from the final analysis, we would recommend the following strategies:

1. Enhance Treatment Protocols: Improving the quality of medical care and treatment protocols for healthcare workers who contract COVID-19 can help reduce mortality rates.

2. Provide Adequate PPE: Ensuring that healthcare workers have continuous access to high-quality personal protective equipment (PPE) is essential to prevent infections.

3. Overall, Health Support: Offering comprehensive health support, including mental health services, vaccination programs, and regular health check-ups, can help maintain the overall well-being of healthcare workers.

By implementing these recommendations, healthcare systems can better protect their workers and potentially reduce both infection and mortality rates among this critical workforce.

CONCLUSION

The application of predictive analytics in managing COVID-19 infection and mortality rates among healthcare personnel has the potential to significantly enhance pandemic response strategies. By accurately forecasting trends and identifying high-risk scenarios, healthcare systems can implement targeted interventions to protect their workforce and maintain operational capacity. This project aims to demonstrate the feasibility and effectiveness of such predictive models using comprehensive datasets from the CDC and other sources.

10 QUESTIONS

1. What is the focus of the project?

The focus of the project is to use predictive analytics to forecast COVID-19 infection and mortality rates among healthcare personnel, thereby helping healthcare institutions to strategize and allocate resources proactively.

2. What types of data are being analyzed in this project?

The project analyzes weekly data on COVID-19 cases and deaths among healthcare personnel, including demographic information, healthcare settings affected, and geographic distribution by state or region.

3. Why is predictive analytics important for managing COVID-19 cases and deaths among healthcare personnel?

Predictive analytics is important because it helps anticipate future trends, enabling healthcare institutions to implement early interventions, allocate resources effectively, and ensure the well-being of healthcare workers.

4. What are some of the potential challenges mentioned in the project related to data analysis?

Potential challenges include inconsistencies or missing data, the influence of external factors, the evolving nature of the pandemic, ensuring data privacy, and the complexity of developing accurate and transparent predictive models.

5. How does the project plan to address the issue of evolving COVID-19 variants and public health interventions?

The project plans to address this issue by frequently updating predictive models to reflect new data and changes in public health interventions, ensuring the models remain accurate and relevant.

6. What are the benefits of accurately forecasting COVID-19 trends among healthcare personnel?

Accurate forecasting allows healthcare institutions to anticipate surges in cases, allocate resources efficiently, implement targeted measures, and ultimately protect frontline healthcare workers.

7. What methods are used in the project to develop predictive models?

The methods include data collection, data preprocessing, exploratory data analysis (EDA), feature engineering, model development, model evaluation, interpretation of results, and documentation and reporting.

8. How do public health policies and interventions impact the trends in COVID-19 cases and deaths among healthcare personnel?

Variations in public health policies and interventions can significantly impact weekly trends, either by reducing or increasing the spread of the virus among healthcare workers, thereby affecting case and death rates.

9. What role do demographic factors play in the analysis of COVID-19 cases and deaths among healthcare personnel?

Demographic factors such as age and sex can influence the susceptibility to infection and mortality rates, helping to identify which groups of healthcare personnel are most at risk and require targeted interventions. This Project did not touch on those features, but in the dataset, they prove to show important patterns.

10. What are the key sources of data and references used in this project?

Key sources include the CDC COVID Data Tracker, NYC Health COVID-19 Vaccination Data, World Health Organization reports, Pfizer-BioNTech COVID-19 Vaccine Study, The Lancet, Journal of Public Health Policy, National Institutes of Health, and various economic and psychological studies related to vaccination programs.

REFERENCES

1. CDC COVID Data Tracker – From the CDC website this is the primary source for COVID-19 vaccination data in the United States.
2. NYC Health COVID-19 Vaccination Data - The info from the NYC campaign and includes demographics and geographical distribution.
3. World Health Organization (WHO) - Shows report and guidelines on vaccination programs, and their global public health impact.
4. Pfizer-BioNTech COVID-19 Vaccine Study - Insights into vaccine efficacy and safety profiles through clinical trial data.
5. The Lancet - Vaccine Efficacy Analysis - Articles that have been peer-reviewed on their efficacy on various COVID – 19 vaccines in different populations.
6. Journal of Public Health Policy -Impacts of public health policies on vaccination rates and outcomes.
7. National Institutes of Health (NIH) - Long-term health outcomes following vaccination.
8. Economic Impacts of Vaccination Programs - An analysis on the economic benefits of widespread vaccination efforts.
9. Psychological Factors Influencing Vaccine Uptake - Studies on factors that affected public willingness to get vaccinated.
10. New England Journal of Medicine - Different reports on vaccine breakthrough cases and their implications for public health strategies.
11. Hogan, C. M., Parzuchowski, A. S., Lyu, X., Goldstick, J., & Resnicow, K. (2022). Characterization of US State COVID-19 Vaccine Incentive Programs. *JAMA Network Open*, *5*(10). https://doi.org/10.1001/jamanetworkopen.2022.35328
12. COVID-19 Deaths Among Healthcare Personnel, by week - ARCHIVED | Data | Centers for Disease Control and Prevention (cdc.gov)
13. COVID-19 Cases Among Healthcare Personnel, by week - ARCHIVED | Data | Centers for Disease Control and Prevention (cdc.gov)
14. Hogan, C., Parzuchowski, A., Lyu, X., Goldstick, J. E., & Resnicow, K. (2022). Characterization of US state COVID-19 vaccine incentive programs. JAMA Network Open, 5(10), e2235328. https://doi.org/10.1001/jamanetworkopen.2022.35328
15. Hacisuleyman, E., Hale, C., Saito, Y., Blachère, N. E., Bergh, M., Conlon, E. G., Schaefer-Babajew, D., DaSilva, J., Muecksch, F., Gaebler, C., Lifton, R. P., Nussenzweig, M. C., Hatziioannou, T., Bieniasz, P. D., & Darnell, R. B. (2021). Vaccine Breakthrough Infections with SARS-CoV-2 Variants. New England Journal of Medicine/the New England Journal of Medicine, 384(23), 2212–2218. https://doi.org/10.1056/nejmoa2105000
11. J.R Zahar, D. Seytre, P. Moenne-Locoz, A. Lomont, & Y. Tandjaoui-Lambiotte. (2022). *Spread of viruses, Which measures are the most apt to control COVID-19?* National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9746078/

16. 12 COVID-19 vaccination strategies for your community. (2022, November 29). Centers for Disease Control and Prevention. https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence/community.html

APPENDIX