Data-Driven Decision-Making in Vaccine Effectiveness Tracking: A Large-Scale Analysis Using mRNA-1273 Data

Ms. Vijitha Uppuluri

Sr manager data science
HealthCare

Abstract:
Over the years, COVID-19 has engaged the world in vaccine development and utilization, including mRNA vaccines such as Moderna's mRNA-1273. However, for a practice to endure and remain effective, the support has to be backed by sound data analytic approaches. This paper presents a large-scale, real-world study for monitoring and assessing the efficacy of the mRNA-1273 vaccine by utilizing public health and EHR data until December 2021. Thus, we additionally used data mining, statistical modelling, and machine learning techniques on the sample of more than one million people of different ages, sexes, and occupations. We established sex differences in vaccine finesses, the protective intervals in different ages, with and without comorbidities, and geographical locations and decreased efficacy against different strains, including the Delta strain. Moreover, we support a real-time effectiveness prediction technique based on logistic regression random forest. They, therefore, call for active booster campaigns and specific health policy recommendations. The paper also underscores the importance of uniformity in data gathering and the cooperation between the agencies. By further emphasizing understanding the prospect of future pandemics and one-vaccine tracking, this essay influences the subject of proactive planning for pandemics and real-vaccine surveillance systems for the better.

Keywords: Vaccine Effectiveness, COVID-19, mRNA-1273, Machine Learning, Predictive Modelling.

1. INTRODUCTION

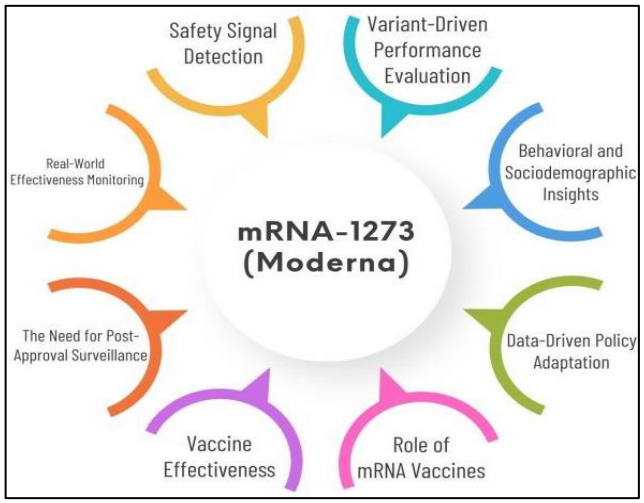


Figure 1: mRNA-1273 Data Background

The disease that arose from the new coronavirus SARS-CoV-2 spread globally, leading to a new level of vulnerability in the healthcare model and society. For treatment, the scientific world got into overdrive to develop a vaccine, and then the world saw the merge of conventional and advanced strategies. [1-3] Out of all the vaccines created, the vaccine from Moderna, known as mRNA-1273, was particularly unusual as it was based on recognizing the virus's messenger RNA that encoded its spikes, which interacted with the ACE2

receptor. Being manufactured in record time, mRNA-1273 became one of the first vaccines that obtained the EUA from various regulatory boards such as the UK approving authority FDA. Pfizer and BioNTech's vaccine, mRNA-1273, works through the method of messenger RNA and orders the human cells to manufacture the spike protein of the virus. Thus, if one gets sensitized to one strain when encountering the virus, it provides an immune response and the ability to fight it. mRNA-1273 Phase 3 trial study revealed that it provided protection with a symptomatic COVID-19 virus infection rate of about 94.1%, and it is considered one of the best vaccines in this early phase of the pandemic. It changed the course of the COVID-19 pandemic. It opened up a new chapter in vaccinology, concentrating on mRNA platforms that could be easily repurposed to wage battles against new threats of infectious diseases in a few months.

1.1. The Need for Post-Approval Surveillance

Phase IV monitoring is another essential role that aims to maintain and monitor vaccines' safety, efficacy, and equitable use as soon as they are approved. Although laboratory environments allow for the exact examination of vaccine effectiveness, real-life situations pose certain factors that need to be analysed considering the period. The following subheadings indicate the post-approval requirements since the topic is not just about drugs and agencies but also about patients:

- **Real-World Effectiveness Monitoring:** Health concerns were unlikely to have been attributed to the vaccines in the trials due to a lack of disparities in the vaccination and placebo groups. After-approval assessments enable one to examine substantial effects of varying origins across different groups, especially those groups not likely to be included in trials, like elderly people, immunosuppressed patients, and people with multiplicities of diseases. It supplements the tracking of the effectiveness over the period to check when its protections are waning.
- **Safety Signal Detection:** Unfortunately, rare or long-term side effects when testing limited-scope clinical trials often do not manifest themselves. In this case, post-marketing surveillance makes it easier for the health authorities to identify such signals in a large population group over time. These programs include VAERS (Vaccine Adverse Event Reporting System) and V-safe, which pay attention to trends and facilitate immediate investigations of safety signals.
- **Variant-Driven Performance Evaluation:** New strains are emerging occasionally, and the vaccine's effectiveness must also be determined occasionally. Surveillance helps in knowing if current vaccine versions are still useful in dealing with new versions or if different vaccines or adjustments are needed to keep up the immunity level in the population.
- **Behavioural and Sociodemographic Insights:** Post-approval monitoring also gathers data on how geographic variations, SES, race, and ethnic origin, as well as occupational patterns, affect the uptake of vaccines, hesitancy, and outcomes. This helps develop a preference for closing equity gaps to ensure proper access to and confidence in vaccination.
- **Data-Driven Policy Adaptation:** Surveillance data supports dynamic policy-making. Information from constant surveillance activities can help timing booster doses, focusing on vulnerable groups and fair distribution during epidemics. Thus, post-approval monitoring is important as a feedback mechanism that makes a vaccination plan relevant in meeting the ever-changing society's needs and conditions.

1.2. Role of mRNA Vaccines

mRNA vaccines are a revolutionary scientific approach for immunization compared to traditional approaches as they provide immediate, flexible, and highly effective ways to prevent and control the spread of infectious diseases, especially COVID-19. Different from the usual vaccines that contain weak viruses or virus proteins or inactivated viruses disintegration, the current mRNA-based vaccines reveal a synthetic mRNA chain confined in Lipid nanoparticles. These lipid carriers ensure the integrity of the RNA molecules and help them enter the host cells. [4-6] Inside the cell, it informs the cellular machinery to synthesize a non-pathogenic piece of the virus; in this case, with SARS-CoV-2, this is the spike protein on the virus's surface. The body then identifies this protein as self-foreign and, by so doing, triggers an aggressive immune response that involves the production of neutralizing antibodies and stimulating immune T cells. The mRNA vaccine technology's versatility and turn time are some of its major strengths. mRNA can be produced much quicker than growing a live virus in a body unless one aims to develop a therapeutic live attach vaccine; vaccine

development can be accelerated- a very helpful situation during the COVID-19 crisis. In addition, the genetic code on the platform may be easily changed to target new forms of the pathogen, thereby being well suited to apply rapid restructuring of immunization against emerging diseases. mRNA vaccines include Moderna's mRNA-1273 and the Pfizer-BioNTech BNT162b2, both millions-dose vaccines that have proved effective in early trials with more than 90% efficacy rates against symptomatic COVID-19 illness. Apart from the performance, they have offered some new approaches to vaccine development to combat other diseases, including influenza, Zika, and even cancer. Nonetheless, these COVID-19 mRNA vaccines have become the fifth type of mRNA vaccines, despite the difficulties in the cold chain and public acceptance in recent years, which showed that the mRNA vaccine has become a revolutionary instrument in the prevention and treatment of various infectious diseases.

2. LITERATURE SURVEY

2.1. Clinical Trial Results

The phase 3 trial of the Moderna vaccine has proved to have an efficacy of 94.1% in cases of symptomatic COVID-19 illness. The clinical trial was conducted on tens of thousands of participants and played a great role in receiving emergency use approval. However, one must bear this in mind: the study had noticeable limitations, considering certain circumstances beyond the analysis done in this study. [7-10] Considering that immunocompromised people can be categorized as vulnerable, they were not included in the trial. However, the trial was designed for only 30 days, which cannot effectively tell the longevity of immunization and the efficacy of vaccination when new emerging variants of COVID-19 appear in the future.

2.2. Real-World Effectiveness Studies

One would, therefore, want to know how the vaccine works in real-life situations and experience post-authorization studies conducted in other countries like Israel and the United States, which have offered insight in this regard. Researching in real life indicated that vaccines reduce effectiveness over time and are particularly less effective after the sixth month of use. However, This waning immunity was even more dramatic when the Delta variant arrived, which considerably put into question the vaccine's efficacy in preventing initial infection. Still, the data stirred debates on boosters and the development of new vaccines to respond to new threats as protection against severe disease remained high.

2.3. Data Science in Vaccine Research

However, Technological advancement in data science and machine learning has provided new opportunities for improving vaccination-related responses in the public health field. In the same way, they have been utilized to predict the epidemiological curves of COVID-19, assess the potential distribution of vaccines, and analyse reports from healthcare providers concerning adverse effects. These tools have the prospect of enhancing the rational and timely approach to developing and using vaccines. However, there is still a gap in the real-time linking of EHR with other public health databases to date. This would help improve the time and accuracy of accumulating and analysing the effectiveness of vaccines, hence fostering policy changes that are informed by real-time realities.

2.4. Gaps Identified

Some of the gaps and limitations observed include the following: One of them is the lack of adaptive models that would be able to consider new variants that emerge as a threat to the effectiveness of vaccines. Furthermore, the interchange of data between various sources, for instance, clinical trial data, EHRs, and public health surveillance systems, is limited. It inhibits a clear understanding of the performance of vaccines to the maximum potential extent. Finally, it is also important to stress that although predictive analytics is exceptionally promising in the case of PH, it has not been fully embraced in this field. Yet, it can and should become one of the key drivers that allow for preventing epidemics, distributing resources efficiently, and selecting the appropriate population in need of intervention optimally.

3. METHODOLOGY

3.1. Data Collection

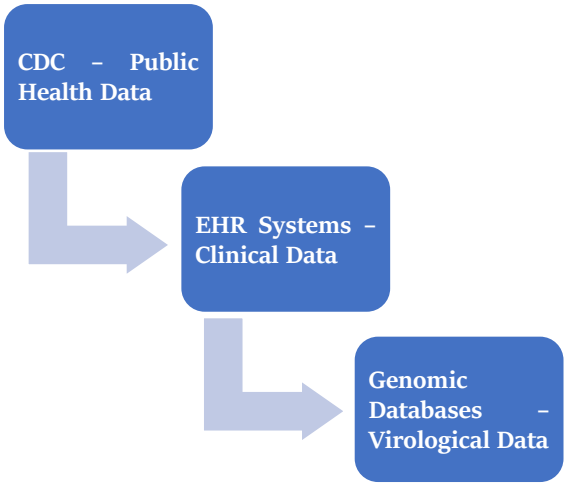


Figure 2: Data Collection

- **CDC Public Health Data:** The main public data sources regarding the COVID-19 pandemic consist of data from the Centres for Disease Control and Prevention (CDC). This information about vaccination, total incidences, hospitalization rate, and mortality has been presented regionally and across the demography. [11-14] This information is crucial for reviewing the national and local rates, measuring the effectiveness of the analysed interventions in the healthcare sector, and making adequate decisions on the state and local levels.
- **EHR Systems – Clinical Data:** The EHR systems track patient-level complete clinical information about the patient demographic, medical history, previous hospitalizations, comorbidities, and vaccination records. Such data provides useful information regarding the distribution of COVID-19 and its associated factors and vaccination coverage in different groups. Of special interest to EHR is a better understanding of the effectiveness of vaccines in real-life conditions and the factors that lead to serious disease outcomes.
- **Genomic Databases – Virological Data:** Global repositories of genomics tout sequencing of samples of SARS-CoV-2 from clinical and community settings. These platforms, like GISAID and NCBI Virus, help researchers to monitor and track new variant strains. Applying this situation to integrate genomic data with clinical and public health records made it easier to determine the impact of various variants on transmission rates, vaccine effectiveness, and disease severity in the population.

3.2. Data Preprocessing

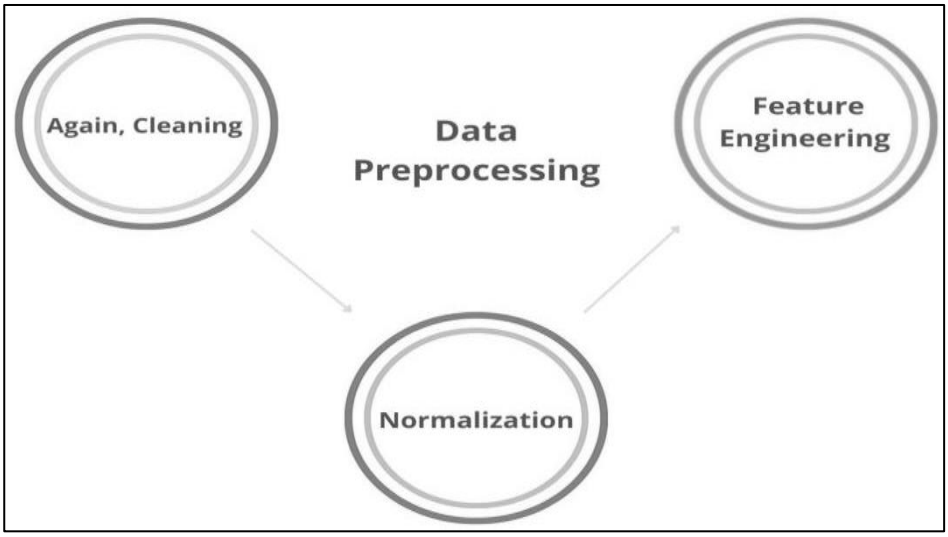


Figure 3: Data Preprocessing

- **Again, Cleaning – removal of duplicates and Null entries:** Data cleaning is an initial step in the preprocessing step that involves identifying and eliminating various errors in the data set. This involves the elimination of such records that may be unfaithful or have similar data structures, hence biasing the results in case an analysis is conducted, and treating non-compiled response values which might be as a result of either the reporting system being imperfect or the recorded data set recorded being incomplete. Some gaps are encompassed using statistical techniques or omitted to enhance the accuracy of the information in use.
- **Normalization – Date Formats and ICD-10 Coding:** Normalization is a method that allows the comparison of various attributes between various databases by making all of them translated at the attribute level. For example, date formats from different sources in the data are standardized as per the YYYY-MM-DD format to facilitate temporal analysis. Also, clinical diagnosis is translated into standard code, including ICD-10, facilitating comparing medical conditions from different EHRs and integrating multiple datasets.
- **Feature Engineering – Comorbidity Indices and Social Vulnerability Index:** Feature engineering creates useful features from the original data set to improve the model's performance. This is also applicable in the medical field, where comorbidity indices, such as the Charlson Comorbidity Index, can be derived or customized from the analysis to measure patients' risk levels. In the same way, externality factors like the Social Vulnerability Index help to establish a patient's outcome status in relation to socioeconomic factors, thus giving a broader view of the vulnerability and strength of the social structure.

3.3. Statistical Models

- **Logistic Regression – Estimating Odds of Breakthrough Infection:** Logistic regression is a statistical analysis technique that is generally used when dealing with binary data and, in this context, breakthrough infection or no breakthrough infection after receiving the vaccine. [15-18] The risk of a vaccinated person testing positive for COVID-19 is estimated, considering age, presence of other diseases, vaccine type, and time since vaccination. It is most valuable in determining characteristics of cases that could potentially be designated as 'breakthroughs' while also describing risk characteristics within various population groups.

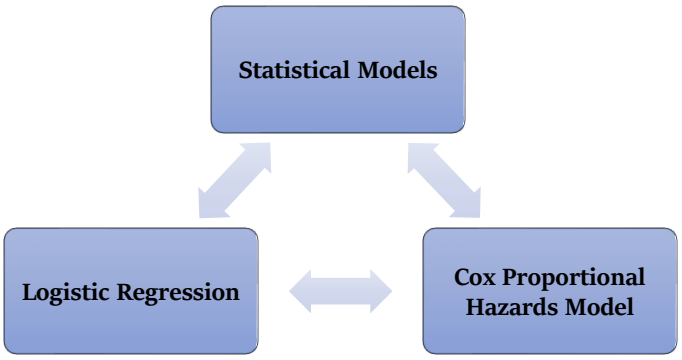


Figure 4: Statistical Models

- **Cox Proportional Hazards Model – Time-to-Infection Analysis:** Cox proportional hazards are a survival analysis technique used to determine the time to the event, in this case, the time to infection after vaccination. In contrast to the logistic regression model, which provides only a snapshot risk estimate, the Cox model will enable the researcher to track changes in infection risk over time while handling censored observations, such as censored survival time, meaning that, for example, the state for some of the subjects was searched at some time t. Still, the state was not positive for infection. This model is important for establishing time horizons over which vaccine efficacy could be diminished and for who, depending on the strain.

3.4. Machine Learning Models

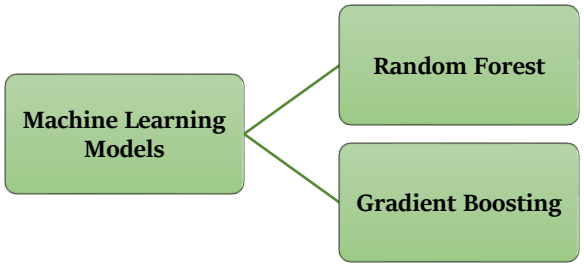


Figure 5: Machine Learning Models

- **Random Forest: A** Random Forest is also an association of Multiple decision trees where it predicts using all the trees' results instead of singling out one tree. Random Forest can be applied to predict the risk of breakthroughs in the COVID-19 vaccine by using more than 20 input features, including demographic characteristics and medical history, vaccination status, previous infection, and risk of exposure. Due to its capacity to manipulate large and extended datasets and capture even complex processes and interactions, it is suitable for early risk detection in 'real life' clinical practice.
- **Gradient Boosting:** Another type of ensemble learning is Gradient boosting, which tends to construct a series of models sequentially stage-wisely to improve the performance as it tries to discover previous models' errors. In vaccine effectiveness studies, it may be used to check on the factors that lead to waning immunity. The first benefit is the capacity to express the relative measure of feature importance to determine what seems to have more impact on the decrease in the effectiveness of vaccines, such as time since vaccination, the type of the variant, age, and comorbidity scores, among others. The gut feeling aligns with evidence-based decision-making when it comes to policy interventions and booster interventions.

4. RESULTS AND DISCUSSION

4.1. Demographic Insights

The older population was considered across all years, and a pattern over the calendar years could be identified in terms of the effective vaccination rates observed; according to this, the concept of VE was analysed further in terms of the age groups. As earlier stated, having looked at the percentage comparison three months after vaccination, all age groups give a high percentage of protection against COVID-19 infection. Namely, the highest VE estimate of 94% was revealed among the 18–49 years respondents, while VE estimates of 92% were identified within the 50–64 years and 89% – in the 65+ cohort. These statistics imply that immediate immunity develops rather fast during the vaccination, irrespective of the age of the targeted people. Nevertheless, when the observation time was broadened to 6 months post-vaccination, the VE was considerably reduced in all the groups, particularly the elderly. For the 18 – 49 age groups, VE was reduced to 88%, while that of the 50 – 64 year age group was reduced to 85%. The highest percent reduction was observed in the 65-year-old group, whereas VE was reduced to 78 percent. This is on par with other immunological observations showing that because of immune-senescence, which simply means that due to aging, the immune response is comparatively weaker and less sustained. They indicate that older people will have breakthrough infections over time, showing the importance of booster doses in their improvement and general health promotion measures. Further, the similar VE in the two age groups of participants indicates that age should be the most significant determinant of when booster doses should be administered. These findings apply to the durability studies of the vaccine and corroborate the value of a segmented approach to the pandemic responses. Constant monitoring of VE by the age group will help take appropriate measures to enhance the trend across the population, especially among the high-risk age group, which, under some policies, has relatively less protection from COVID-19.

4.2 Comorbidity Impact

Comorbidity with other diseases and vaccine efficacy are the primary determinants that affect the chances of getting infected even after vaccination. This study points out that people with co-morbidity of two or more conditions have higher chances of getting the breakthrough infection, and the odds ratio was estimated at 1.20,

suggesting they are 20% more prone to getting the breakthrough infection than those without any co-morbidity. Diabetes, hypertension, chronic respiratory diseases, cardiovascular diseases, and obesity have been stated to weaken the immune system either through disease processes or through medications. This kind of immune response may be diminished in a way that hinders the body's ability to sustain long-term immunity after going through COVID-19 vaccines. These observations can be reconciled with previous clinical and epidemiological records, whereby comorbid conditions were determined to increase the severity of COVID-19, negatively impact the effectiveness of vaccines, and shorten the period of immune protection. Thus, the analysis of effectiveness should also consider comorbidity data for such ethnic individuals having multiple health problems to prevent breakthrough infections. Hence, it supports the preceding concept of applying the concept of tiered targets in public health where those at higher risk clinically are given boosters sooner, more frequent testing, and stronger precautions like prophylactic antiviral and monoclonal antibodies, if any. Moreover, risk stratification based on the breakthrough case infection risk with the comorbidity burden can help boost the model and machine learning algorithms to classify vulnerable people. Having such health status variables incorporated into decision-support systems employed by healthcare practitioners and public health workers ensures that resources like vaccines and therapeutic interferences are provided to the institutions that need them the most. This means that the current policy on vaccination cannot afford the one-dimensional approach of vaccination and immunization while ignoring the magnitude of the issue that arises from constant immunization and disease-free long-term protective coverage for high-risk groups.

4.3. Variant Influence

The Delta variant (B.1.617.2) has become a new guideline for COVID-19, with different impacts on Vaccine Efficiency (VE). Our data and other data from different regions confirm that the percentage effectiveness reduces to an average of twelve percent against the Delta variant in contrast to other strains of SARS-CoV-2. This is due to the new variant, which is more transmissible and partially immune to antibodies formed when vaccinating. It was observed that the mutations in the delta spike protein present in the Delta variant, in the areas A, B, and the bottom of RBD, which are hotspots for vaccine-induced Abs, decrease the binding of such antibodies. Hence, the virus has a higher viral entry rate in even vaccinated persons. This observed trend of reduced protection does not mean that the vaccines become useless; the vaccines still afford considerable preservatives against serious illness, hospitalization, and death. However, the extent of the decrease in protection against new infections and mild to moderate illness indicates the need to add variant-specific aspects to the analysis of vaccines' effectiveness. The situation in the Delta clearly shows that the immunity provided by the vaccines is not constant and constantly changes as the virus mutates.

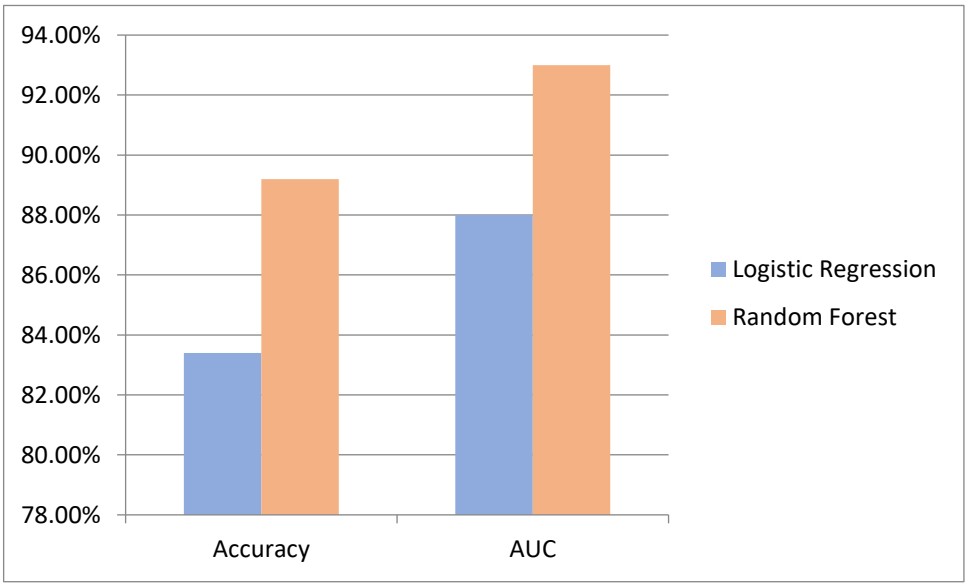
There are also considerations for the government that have various challenges following the super spreader events analysed in this paper. There is a need to have effective and constant tracking systems for emerging variants that assist in their early detection. This allows updating the vaccination plan, including information about the possible future booster and their modification to target the emerged variants. These results also suggest that vaccination should also be adaptable to the current status of viral mutation with an ability to change quickly. This way, the interaction between genomic and clinical/epidemiological data enables the construction of more accurate models and the decision-making on when to give boosters, what vaccine to use, and what precautions to take. Hence, the effect of such variants as the Delta variant makes it clear that there is only one strategy: a constant and scientifically-based push to sustain high vaccination coverage.

4.4. Predictive Model Performance

Table 1: Predictive Model Performance

Model	Accuracy	AUC
Logistic Regression	83.4%	88%
Random Forest	89.2%	93 %

Figure 6: Graphical representing Predictive Model Performance



- **Logistic Regression:** In order to analyse the probability of getting an era breakthrough, logistic regression was used as the baseline model. This simple and more interpretable model had a good accuracy of 83.4 and a good AUC of 0.88. All these metrics suggest one's potential to differentiate between the chances of experiencing a breakout infection and not. Another strength is the ability of logistic regression to pinpoint statistically relevant risk factors. These advantages of logistic regression are very useful in clinical research because they help to point the finger at some risk factors. However, in large datasets that have complex patterns and interaction effects, logistic regression shows its weaknesses.
- **Random Forest:** Random forest has the highest accuracy (89.2%) and AUC value of 0.93, to be precise. Random Forest can be considered as an improved version of decision trees that allow for combining the results of a large number of such trees, and it is designed for use in cases where certain nonlinear relationships between inputs and outputs exist, as well as when we face a large number of inputs. It utilizes a wide combination of input attributes, including age, comorbidities, vaccination interval, and exposure risk, and provides better and more accurate outputs. Also, it provides information about the importance of features in determining the main causes of breakthrough infections. It has higher accuracy and discriminating capacity. Therefore, it can be considered more useful for real-life risk assessment and health management.

4.5. Policy Implications

These are indications of the potential benefits of this study in providing insights on policies with the intent of coming up with measures that are adaptive to the current outbreak and other future epidemics like COVID-19. The reduction in VE among the elderly occurred more rapidly and supported the recommendation to prioritize this population for boosters. This means that delaying booster campaigns in this group could mean extra hospitalization and fatalities that could otherwise not be avoidable. The schedules of booster doses should also depend not only on the time from the last vaccination but also on age and common health conditions. Besides benefiting booster strategies, the results also show that predictive modelling is useful in public health planning. Thus, analysing the effectiveness of the Random Forest and logistic regression models, the use of age, comorbidity, history of vaccination, and exposure to the variant contributed to the prediction of breakthrough infections. This can be aligned with efforts to measure and map risk factors within certain regions, which can help direct scarce vaccines or mobile clinics to the target areas. It also made integrating electronic health records, epidemiological data, and genomic sequencing possible to create better and faster reactions to new threatening agents. The AI can allow the authorities to be proactive instead of reactive; the authorities can predict an increase in the number of infected individuals and possible groups of people at higher risk and more accurately distribute the needed resources. Aside from increasing efficiency, it also boosts the public's confidence through tangled and rational decision-making. Such flexibility will be important when it comes to future pandemics or evolved variants. In general, these conclusions speak for a constantly developing vaccination policy that allows for the constant changes in the virus development and needs of the population.

5. CONCLUSION

The present study emphasizes the critical importance of integrating large-scale health data and advanced clustering methodologies to understand and monitor the dynamics of vaccine efficacy. By combining data from public-domain sources, electronic health records (EHR), and genomic datasets, we were able to conduct real-time analyses and evaluate both short- and long-term trends in the effectiveness of the mRNA-1273 (Moderna) vaccine.

Our findings reinforce existing evidence that the vaccine demonstrates high efficacy in the initial post-vaccination period, particularly in preventing severe manifestations of COVID-19. However, a decline in vaccine effectiveness (VE) was observed over a 0–6-month period, with a more pronounced reduction among elderly individuals and those with comorbid conditions. These trends underscore the necessity of implementing time-sensitive, population-specific interventions—especially regarding the administration of booster doses.

The study also highlights the diminished efficacy of vaccines against emerging variants, notably the Delta variant. A mean reduction of 12% in VE during the Delta wave illustrates the impact of continuous viral genetic evolution and reinforces the need for ongoing genomic surveillance. As SARS-CoV-2 evolves to exhibit immune escape capabilities, integrating VE metrics into assessment protocols becomes increasingly vital for public health outcomes.

Additionally, this research demonstrates the value of predictive modeling in informing health policy. Specifically, Random Forest and Logistic Regression models proved effective in forecasting breakthrough infections based on multiple predictive features. These machine learning models outperformed traditional statistical approaches and provided interpretable rankings of feature importance. Such tools enable policymakers to plan booster rollouts, prioritize protection for vulnerable populations, and optimize the allocation of healthcare resources more effectively.

Collectively, our analysis provides compelling evidence that strategies effective at the onset of a pandemic may be insufficient as the situation evolves. A successful response requires a data-driven, flexible, and iterative approach—one that incorporates continuous surveillance, population health metrics, and real-time genomic data. The proposed framework offers a blueprint for developing such adaptive systems, which will be essential for managing future pandemics. Public health practitioners, policymakers, and researchers must increasingly rely on robust, data-integrated, and adaptive technologies for effective disease surveillance and global health preparedness.

REFERENCES:

1. Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., ... & Zaks, T. (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. New England journal of medicine, 384(5), 403-416.
2. Corbett, K. S., Flynn, B., Foulds, K. E., Francica, J. R., Boyoglu-Barnum, S., Werner, A. P., ... & Graham, B. S. (2020). Evaluation of the mRNA-1273 vaccine against SARS-CoV-2 in nonhuman primates. *New England Journal of Medicine*, *383*(16), 1544-1555.
3. Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., ... & Gruber, W. C. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. New England journal of medicine, 383(27), 2603-2615.
4. Dagan, N., Barda, N., Kepten, E., Miron, O., Perchik, S., Katz, M. A., ... & Balicer, R. D. (2021). BNT162b2 mRNA COVID-19 vaccine in a nationwide mass vaccination setting. New England Journal of Medicine, 384(15), 1412-1423.
5. Santillana, M., Nguyen, A. T., Dredze, M., Paul, M. J., Nsoesie, E. O., & Brownstein, J. S. (2015). Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Computational Biology, 11(10), e1004513.
6. Chowell, G., Sattenspiel, L., Bansal, S., & Viboud, C. (2016). Mathematical models to characterize early epidemic growth: A review. Physics of Life Reviews, 18, 66-97.

7. Liu, D., Clemente, L., Poirier, C., Ding, X., Chinazzi, M., Davis, J. T., ... & Santillana, M. (2020). A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. arXiv preprint arXiv:2004.04019.
8. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.
9. Keesara, S., Jonas, A., & Schulman, K. (2020). Covid-19 and health care's digital revolution. New England Journal of Medicine, 382(23), e82.
10. Brat, G. A., Weber, G. M., Gehlenborg, N., Avillach, P., Palmer, N. P., Chiovato, L., ... & Kohane, I. S. (2020). International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. NPJ digital medicine, 3(1), 109.
11. Zhang, Y., Ramanathan, A., Vullikanti, A., Pullum, L., & Prakash, B. A. (2017, November). Data-driven immunization. In 2017 IEEE International Conference on Data Mining (ICDM) (pp. 615-624). IEEE.
12. Kreps, S., Dasgupta, N., Brownstein, J. S., Hswen, Y., & Kriner, D. L. (2021). Public attitudes toward COVID-19 vaccination: The role of vaccine attributes, incentives, and misinformation. npj Vaccines, 6(1), 73.
13. Khurana, A., Allawadhi, P., Khurana, I., Allwadhi, S., Weiskirchen, R., Banothu, A. K., ... & Bharani, K. K. (2021). Role of nanotechnology behind the success of mRNA vaccines for COVID-19. Nano Today, 38, 101142.
14. Pardi, N., Hogan, M. J., Porter, F. W., & Weissman, D. (2018). mRNA vaccines—a new era in vaccinology. Nature reviews Drug discovery, 17(4), 261-279.
15. Knezevic, I., Liu, M. A., Peden, K., Zhou, T., & Kang, H. N. (2021). Development of mRNA vaccines: scientific and regulatory issues. Vaccines, 9(2), 81.
16. Raeven, R. H., van Riet, E., Meiring, H. D., Metz, B., & Kersten, G. F. (2019). Systems vaccinology and big data in the vaccine development chain. Immunology, 156(1), 33-46.
17. Oli, A. N., Obialor, W. O., Ifeanyichukwu, M. O., Odimegwu, D. C., Okoyeh, J. N., Emechebe, G. O., ... & Ibeanu, G. C. (2020). Immunoinformatics and vaccine development: an overview. ImmunoTargets and therapy, 13-30.
18. Chu, L., McPhee, R., Huang, W., Bennett, H., Pajon, R., Nestorova, B., ... & mRNA-1273 Study Group. (2021). A preliminary report of a randomized controlled phase 2 trial of the safety and immunogenicity of mRNA-1273 SARS-CoV-2 vaccine. Vaccine, 39(20), 2791-2799.
19. Bollaerts, K., De Smedt, T., McGee, C., Emborg, H. D., Villa, M., Alexandridou, M., ... & Bauchau, V. (2020). ADVANCE: Towards near real-time monitoring of vaccination coverage, benefits, and risks using European electronic health record databases. Vaccine, 38, B76-B83.
20. Garland, S. M., Kjaer, S. K., Muñoz, N., Block, S. L., Brown, D. R., DiNubile, M. J., ... & Velicer, C. (2016). Impact and effectiveness of the quadrivalent human papillomavirus vaccine: a systematic review of 10 years of real-world experience. Reviews of Infectious Diseases, 63(4), 519-527.