Journal of Technology and Value Addition, Volume 7 (1), January 2025: (30-41)

Research Article

In-Depth Risk Assessment of Covid-19 Vaccination: Unraveling Insights and Strategic Implications Using Data Warehousing Techniques

M.I. Fathima Nihla[1], W.A.D. Akalanka Wicrama Arachchi[2], W.M Dulanji Hansika Wijekoon[2], M.A. Jude Niroshan[2], H.B. Liyanapatabendi[2], and P.P.G. Dinesh Asanka[3]

nihla@seu.ac.lk

[1] South Eastern University of Sri Lanka
[2] Sri Lanka Institute of Information Technology, Sri Lanka
[3] Department of Industrial Management, Faculty of Science, University of Kelaniya, Sri Lanka

Abstract

The global coronavirus challenge has accelerated vaccination implementation. Though these vaccines have strengthened immune defenses, reported side effects have increased public concern. The expeditious dissemination of the coronavirus and its ability to propagate prior to the manifestation of symptoms have rendered immunization against it a worldwide predicament in the twenty-first century. Scientists have developed multiple vaccinations that efficiently enhance the immune system's ability to protect against the virus. Nevertheless, a rise in adverse occurrences linked to the side effects of vaccinations, whether due to chance or casual observation, could potentially jeopardize the public's willingness to embrace immunizations. Therefore, it is imperative to carry out a comprehensive investigation into adverse occurrences associated with immunizations. Analyzing reported data can yield intriguing insights into the occurrences of adverse events related to vaccination. By employing the star model on a selected dataset, we consider several factors such as mortality rates, hospital admissions, gender distribution, immunization types, and fatalities, with particular attention given to vulnerable social populations. This approach provides further insights into adverse events for those who have already been categorized as high-risk. The experiment yields significant findings indicating that Pfizer has greater efficacy in females compared to males, while the Moderna vaccine demonstrates superior effectiveness in males relative to females. Furthermore, the comprehensive dataset demonstrates a higher likelihood of male fatalities compared to female fatalities, accompanied by a significant

recovery rate of 76.4% subsequent to immunization, and no adverse effects were documented. These findings provide crucial data for formulating focused preventative and control strategies aimed at halting the epidemic and mitigating unexpected societal damage.

***Keywords**: Data warehousing, VAERS, Vaccine Adverse Event Reporting System, Covid-19 Vaccination, Adverse, Covid-19, Analytical System*

1. Introduction

The SARS-CoV-2 coronavirus is a respiratory disease that has affected almost every country in the world. It was first reported in Wuhan city of China in late 2019. According to WHO (World Health Organization), this virus has been identified as an air-spreading virus (Richards et al., 2020). The human body's immune system is not capable of defending against the virus. Therefore, scientists have to develop vaccines that can enhance the immune system to prevent major damage to our body from the coronavirus. SARS-CoV-2, discovered in Wuhan, China in late 2019, has had global effects. Due to its great transmissibility and the immune system's inability to fight it alone, vaccine research and deployment became a priority. World Health Organisation emergency use authorisations expedited worldwide immunisation. Vaccines were rushed to market with short trial periods, increasing safety concerns.

These protective antibodies that were developed as vaccines were approved by the WHO for emergency usage. Therefore, multiple vaccines were vaccinated to many people without thorough trial testing for those vaccines. The world was being overrun by the coronavirus, but there was still no preventative management for this serious threat. Lockdowns across the country, case detection and management, contact tracing, health education and awareness campaigns, and other effective government efforts are being used to stop the spread in Sri Lanka (Issadeen et al., 2020). Vaccination programs worldwide generated massive amounts of data on post-vaccination effects. Analyzing this data systematically is essential to understand risks, ensure public safety, and refine vaccine policies. Data warehousing techniques, especially the star schema model, are particularly suited for this analysis as they support multidimensional queries, large-scale integration, and high-performance analytics.

As humans were used as guinea pigs in this process, many people were skeptical about their adverse reactions. Many fatalities were reported with almost all the vaccines, but the world could not afford more time against the viral spread of the

virus. In this paper, we focus on analyzing data based on vaccination: what are the symptoms occurring after vaccination, vaccination age range and gender, how much life-threatening illness and allergies occurring after the vaccination likewise much more information. These vaccination data have been obtained from the VAERS (Vaccine Adverse Event Reporting System) website. They have published adverse data about Pfizer/BioNTech, Moderna, and Johnson & Johnson (Janssen) vaccines.

Once we deeply analyze, refine, and transform this data, we can get an idea about vaccination doses, hospitalization patterns, recovery status, death count, preference on vaccine manufacturers and much more information. Along the process of analyzing the data, we could design fact tables, dimensions, and attributes according to vaccination data.

It is important to note that this data is bound to a specific time period. Therefore, the results will be slightly different from the most up-to-date and most accurate results. On the other hand, as vaccination programs are ongoing all around the world, this results in high data velocity. Some countries are not open to sharing the vaccination statistics and adverse events publicly. This can lead to high data veracity. Individuals who volunteered to provide all these vaccinations and their reaction data in this data set are from the United States.

2. Research Problem

Many countries around the world are progressing vaccination programs exponentially. This leads to many people getting vaccinated. As there are multiple vaccines manufactured by multiple pharmaceutical companies, it is not visible about the effectiveness and safety of those vaccines for the general public. Identifying the coronavirus vaccine effectiveness, death count and their side effects on different categories like vaccine type, age group, gender, state, manufacturer, etc. will help to make a more realistic prediction on risk analysis of COVID-vaccination.

3. Literature Review

The risk of severe sickness and death from COVID-19 is rising dramatically across the world. As a result, governments of nations affiliated with the WHO are introducing vaccines that may be administered into the public to combat the virus. Although the vaccination process has been welcomed everywhere and is in use

among the public, it is obvious that people are more concerned about its effects. The adverse reactions the vaccination can bring should be reported so that it can be used later. VAERS stands for Vaccine Adverse Event Reporting System, and it is used to track adverse events that occur following immunization. This is used to ensure that vaccines used in the US, particularly COVID-19 vaccines, are continuously monitored for safety (Vaccine Adverse Event Reporting System (VAERS) | CDC, no date). Many nations have licensed COVID-19 vaccines for emergency use. Notably, in low and middle-income countries, there is a need to improve post-marketing surveillance of adverse event data. To avoid and document probable adverse effects, active pharmacovigilance monitoring is required (Shrestha et al., 2021). Trust in the proposed vaccine's safety was the strongest predictor of COVID-19 vaccination intentions.

Those who saw COVID-19 as a serious disease were also more likely to want to get vaccinated against it. In order to obtain substantial vaccination coverage, health officials should focus on disseminating information about the safety of a future COVID-19 vaccine (Karlsson et al., 2021). Also, it is obvious that adverse reaction reporting is required in order to evaluate COVID-19-related hospital visits, ICU enrollment, and deaths avoided under varying considerations about prospective COVID-19 vaccine use, as well as to measure overall population, age-specific potential outcomes of initiating COVID-19 vaccination (Oliver et al., 2020).

Therefore, it is understood that using the available technologies an analytical system must be developed so that the adverse reactions from the vaccine can be determined effectively.

People can both prevent and stay healthy during the pandemic with the use of smart technologies. Long-term technical solutions are developed by engineers and technology experts to support human activities while confined to one's house throughout the epidemic. Some of the concepts that have been created include the Internet of Things, artificial intelligence, wireless communication technologies, and 5G networks. Smart technologies can offer streamlined and safe operations to combat pandemic illnesses like COVID-19. After the Coronavirus pandemic crisis, this study reviewed data from "Smart Technologies" and "COVID-19," and its findings showed that a variety of smart technologies were used in the medical industry to contain the pandemic (Mohamed Nafrees et al., 2022).

In a study conducted by (Schwab et al., 2021), the authors have used two methods for maintaining and reporting sensitive healthcare data, as well as their secure cross-transmission on top of mHealth applications.

4. Methodology

A. Data Set [VAERS (Vaccine Adverse Event Reporting System)]

Covid -19 world vaccine adverse reactions data set (VAERS) is used for this project which was created by the Food and Drug Administration (FDA) and Centers for Disease Control and Prevention (CDC). This data set was updated on 2021.03.31 and it contains around 50,000 records of adverse reactions for covid-19 vaccines. The downloadable VERSE data set contains 3 separate files in CSV format for each year.

B. Data Collection

As the first step, we have configured a Microsoft SQL Server. Then created a database to store our data. VAERS data is accessible by two mechanisms: by downloading raw data in comma separated value (CSV) files for import into a database, spreadsheet, or text editing program. We have considered these 3 csv files as the main data source.

By using this data source, we created staging and then we made a script to upload data to the warehouse in MSSQL.

C. Data Warehouse

The Data warehouse is a process of managing, collecting, and storing data from various sources to enable and support business intelligence activities in analytics. It is a combination of technologies and components which helps with a strategic analysis of Data. Datawarehouse is electronic storage of large amounts of data that is designed for quarrying and analytics. The staging area of the data warehouse is done by cleansing, transforming data that come from the main resource.

In the data warehouse, the schema defines the way to organize the system with all the database entities (fact tables, dimension tables) and the logical association. In this project, we use the star schema which contains one fact table in the middle and the 9 associated dimension tables. Facts and dimensions in the data warehouse provide space to store data in a denormalized environment. The cube will provide better performance for all analytical reports.

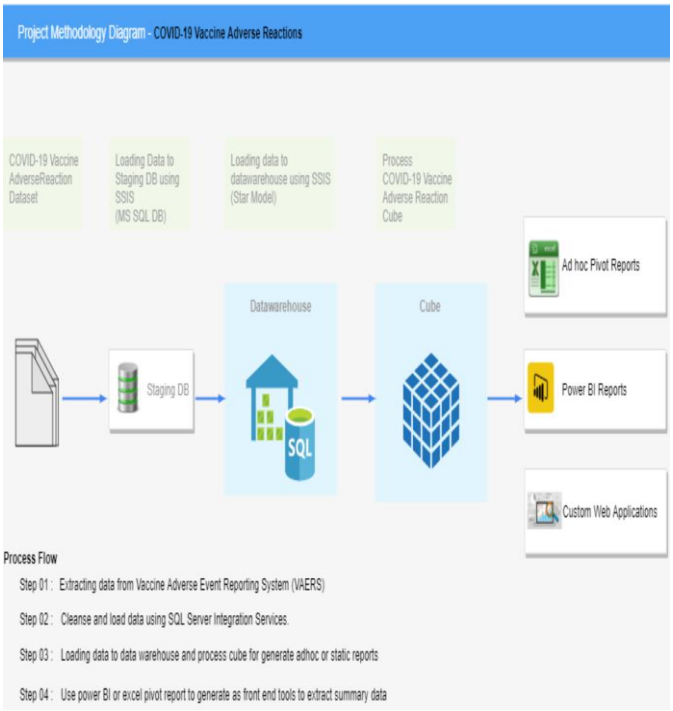
D. ETL Process

ETL is a process of data integration from the various sources into a data warehouse which encompasses three steps extraction, transformation, and loading. Collecting the data from multiple sources, converting data into a standard format (filtering, cleansing, and joining) and loading data into the data warehouse are the main stages of the ETL process. Basically, extracting the data from the CSV file, cleansing data, and loading it into the data warehouse are basic steps of the ETL Process of this project, using Microsoft SQL Server Integration Services (SSIS).

E. Cube Designing

The data cube in the data warehouse is used to store multidimensional structured data. At the OLAP cube, it can be straightforward to analyze data in the data model. In this project, the SQL Server Analysis Services (SSAS) was used to create a cube in the data warehouse. In order to perform the analysis, we have used it as the pivot table. The Cube is designed by including a fact and dimension table.

Figure 01: Flow of Creating Data Warehouse and Cube

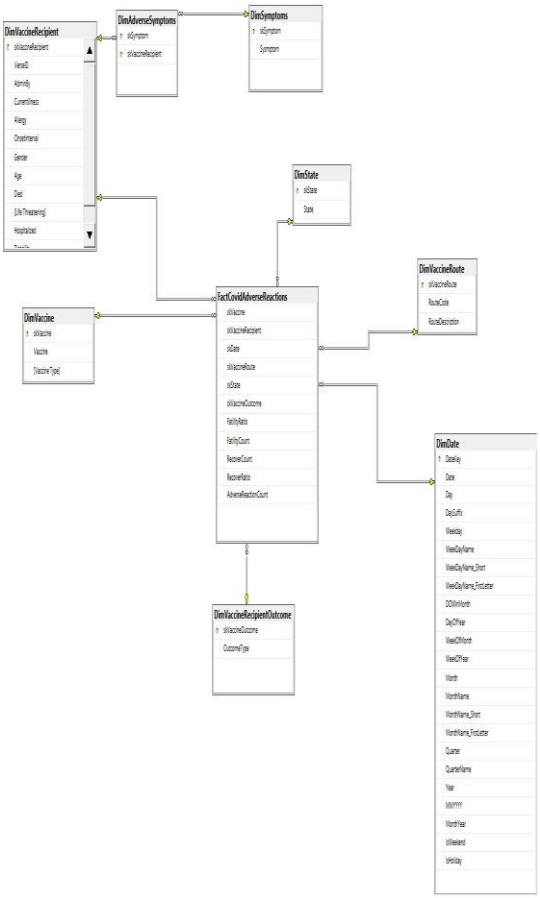


5. Implementation

The data warehouse is created based on fact and dimension structures as shown in figure 2. All dimension tables are connected with one fact table by using its surrogate key (SK). This structure has been created according to the star schema format. We have set up a data

warehouse primarily in order to create a cube. Cube is a data structural model that can be used to visualize multidimensional data in an elegant way. That cube can be used to generate a pivot table. This pivot table will act as the base for our data analysis process.

Figure 02: Used Fact and Dimension Table to Create Data Warehouse



6. **Results**

- **Gender-Based Fatality Analysis:**

Findings indicate higher fatality rates among males post-vaccination. This trend suggests the need for gender-specific public health strategies when assessing vaccine risks.

- **Vaccine Performance by Gender:**

Moderna demonstrates higher effectiveness in males, whereas Pfizer shows better results in females. These distinctions could inform tailored vaccine deployment to minimize adverse outcomes.

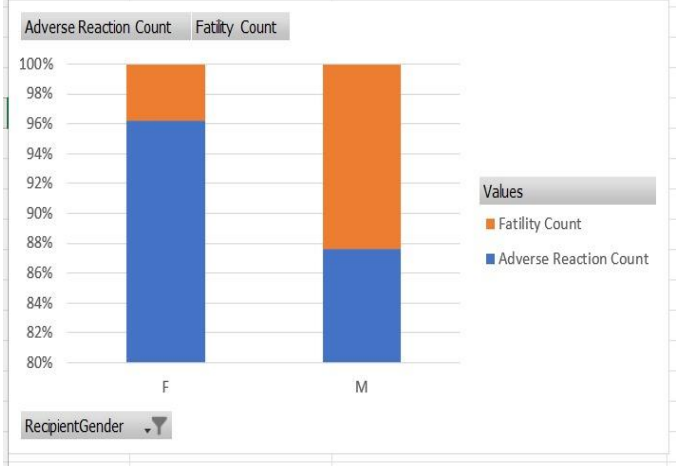
- **Overall Outcomes:**

Recovery rate: 76.94%
Death rate: 6%
Life-threatening reactions: 4%
Hospitalizations: 10%
Disabilities: 1.7%

While a 6% fatality rate post-vaccination appears concerning, it must be contextualized against COVID-19's significantly higher fatality rate, particularly in unvaccinated populations. Therefore, the benefits of vaccination continue to outweigh the risks. The effectiveness of the vaccination is observed by examining the gender.

Figure 03: Gender wise Fatality Count

Gender	Adverse Reaction Count	Fatility Count
F	16400	643
M	5157	728
Grand Total	**21557**	**1371**



Considering the whole sample set getting vaccinated, the male fatality count is higher than the female fatality count. According to this analysis, vaccination has more risks for males that can lead to side effects. The male fatality ratio is higher than the female fatality ratio after vaccination.

Figure 04: Gender based fatality count analysis for vaccine

Row Labels	Fatility Count
MODERNA	
F	328
M	436
U	2
PFIZER\BIONTECH	
F	301
M	285
U	5
UNKNOWN MANUFACTURER	
F	14
M	7
U	0
Grand Total	1378

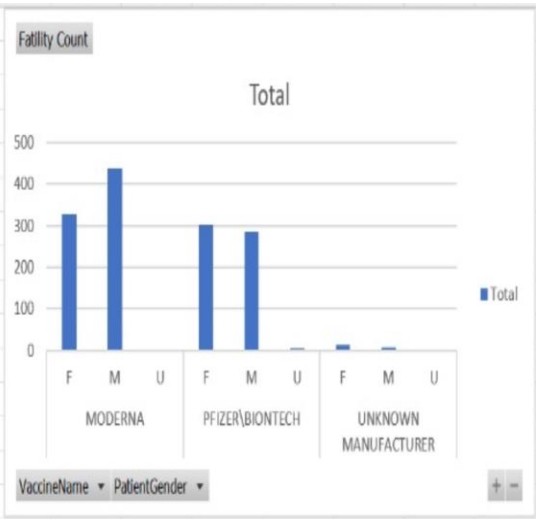


Figure 4 data shows the Moderna vaccine effectively works for males rather than females. On the other hand, the Pfizer vaccine effectively works for females rather than males. Considering the fatality values, Moderna fatality count is low for females, but the Pfizer fatality count is low for males.

If some countries use both vaccines, they can recommend Pfizer for the male population and Moderna for the female population. This helps to control their fatality count ratio in the country.

Figure 05: Adverse Reaction Count

Row Labels	Adverse Reaction Count
DIED	1378
Disabled	391
Hospitalized	2280
Life Threatening	969
Recoverd	16746
Grand Total	**21764**

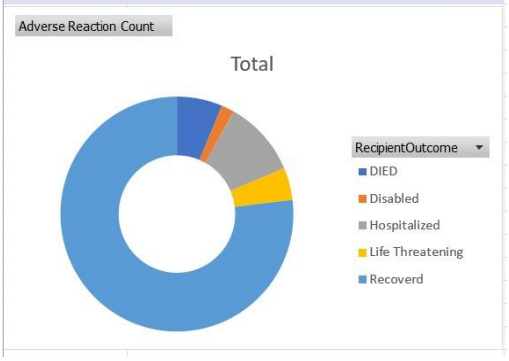


Figure 5 report depicts adverse events like dead count, disable count, hospitalized count, life threatening count against recovery count for people who got caught by the coronavirus.

The Figure 5 also shows that the recovery count is higher after getting the vaccine without any side effects. As a percentage, it is 76.94%. However, it shows:

- Death count as 6%
- Life-threatening count as 4%
- Disabled count as 1.7%
- Hospitality counts as 10%

According to this analysis, vaccination is a fairly good approach against the COVID-19. But it is worth mentioning that vaccination is 6% deadly regardless of the vaccine type.

7. Conclusion

Many government healthcare officials are not capable of deeply analyzing the vaccination data. Setting up a data warehouse, aggregating data from multiple data sources and pushing data into a cube is not a simple task. After analyzing a selected sample data set, it is possible to observe information that is not clearly visible from an operational database. To obtain more accurate and precise data, sample data set size can be increased with the monthly updates that are being published to VAERS from the operational database. This can be automated to view more accumulated results which provide more reliable information for the decision makers.

This study shows how data warehousing can reveal key insights from large vaccination datasets. Such technologies help improve response plans by monitoring unfavorable events in real time for governments and health organizations. For accuracy, future study should automate data updates and expand datasets across more locations. This study also emphasizes the importance of combining public health data with analytical methods for pandemic management. This method can be applied to chronic disease surveillance and pharmacovigilance.

References

Issadeen, S. et al. (2020) Assessing COVID-19 related knowledge and perceptions of preventive health care employees working in health region Kalmunai-Sri Lanka, European Journal of Molecular and Clinical Medicine, 7(6), pp. 418–426.

Karlsson, L. C. et al. (2021) 'Fearing the disease or the vaccine: The case of COVID-19', Personality and Individual Differences, 172, p. 110590. doi: 10.1016/J.PAID.2020.110590.

Mohamed Nafrees, A. C. et al. (2022) 'Smart Technologies to Reduce the Spreading of COVID-19: A Survey Study', in Sharma, H. et al. (eds) Proceedings of the International Conference on Intelligent Vision and Computing (ICIVC 2021). Cham: Springer International Publishing, pp. 250–265.

Oliver, S. E. et al. (2020) 'The Advisory Committee on Immunization Practices' Interim Recommendation for Use of Pfizer-BioNTech COVID-19 Vaccine — United States, December 2020', MMWR. Morbidity and Mortality Weekly Report, 69(50), pp. 1922–1924. doi: 10.15585/MMWR.MM6950E2.

Richards, T. J. et al. (2020) 'Cities policy responses', Tackling coronavirus (COVID-19) Contributing to a global effort, 8(1), pp. 2–25. Available at: https://repositorio.cepal.org/bitstream/handle/11362/45445/1/S2000286_es.pdf%0Ahttps://www.stlouisfed.org/on-the%0Ahttp://arxiv.org/abs/2003.12347%0Ahttps://ourworldindata.org/coronavirus-source-data,%0Ahttps://doi.org/10.1016/j.prostaglandins.2020.10640.

Schwab, J. D. et al. (2021) 'Perspective on mHealth Concepts to Ensure Users' Empowerment-From Adverse Event Tracking for COVID-19 Vaccinations to Oncological Treatment', IEEE Access, 9, pp. 83863–83875. doi: 10.1109/ACCESS.2021.3087315.

Shrestha, S. et al. (2021) 'Adverse events related to COVID-19 vaccines: the need to strengthen pharmacovigilance monitoring systems', Drugs & Therapy Perspectives, 37(8), p. 1. doi: 10.1007/S40267-021-00852-Z.

Vaccine Adverse Event Reporting System (VAERS) | CDC (no date). Available at: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/vaers.html (Accessed: 20 November 2021).