Dean&Francis
ISSN 2959-409X

MS003762

The Cardioprotective Effects of Vaccinations: An Integrative Analysis

Hanwei Cao[1,&,*], **Yiqiu Pan**[2,&], **Guoliang Yin**[3,&]

[1] *Department of Statistics, Columbia University, New York, NY, United States, lukatzao123@gmail.com*
[2] *Department of Mathematics, University of Edinburgh, Edinburgh, Scotland, United Kingdom, s2167460@ed.ac.uk*
[3] *Department of Medicine, Soochow University, Suzhou, Jiangsu, China, 2130416020@stu.suda.edu.cn*
**corresponding author*
[&] *Hanwei Cao, Yiqiu Pan, and Guoliang Yin contributed equally to this work and should be considered co-first authors.*

Abstract:

This study examines the relationship between vaccinations (influenza, pneumonia, tetanus, and COVID-19) and heart disease prevalence using data from the Behavioral Risk Factor Surveillance System (BRFSS). Using logistic regression and Random Forest Classification (RFC) models, the study identified key predictors of cardiovascular disease (CVD). The RFC model demonstrated high accuracy (93.89%) and suggested that vaccinations may reduce CVD risk, though pre-existing health conditions could influence results. The study concludes that vaccinations may be valuable in reducing CVD risk and calls for further research to confirm these findings and understand the mechanisms involved.

Keywords: cardiovascular disease, vaccinations, influenza, pneumonia, tetanus, COVID-19, logistic regression, random forest, public health.

1. Introduction

Cardiovascular disease (CVD) remains the leading cause of mortality worldwide, responsible for approximately 20.5 million deaths in 2021 alone, comprising about one-third of all global deaths [1]. This significant public health challenge necessitates a comprehensive understanding of various influencing factors to develop effective prevention strategies. Preventive actions, such as lifestyle modifications and medical interventions, have been implemented to reduce CVD incidence. Among these interventions, vaccinations have emerged as a potential strategy to mitigate CVD risk, although their impact remains underexplored and, at times, controversial.

Recent studies suggest that certain vaccinations, typically aimed at preventing infectious diseases, may also offer cardiovascular benefits. Respiratory infections like influenza and pneumonia can exacerbate cardiovascular conditions, leading to severe complications such as myocardial infarction and heart failure. Vaccinations against these infections have demonstrated efficacy in reducing their incidence and severity, potentially offering protective cardio-

ISSN 2959-409X

vascular effects [2,3]. Additionally, vaccines like tetanus may improve cardiovascular health through mechanisms involving reduced systemic inflammation and improved immune function [4,5]. The American Heart Association (AHA) underscores the importance of understanding the broader health benefits of vaccinations [6]. Studies have shown that influenza vaccination is associated with a reduced risk of major adverse cardiovascular events, while pneumonia vaccination has been linked to lower rates of cardiovascular complications [7,8]. These findings suggest that vaccinations could serve dual preventive roles, addressing both infectious and cardiovascular diseases.

However, comprehensive research on the cardiovascular benefits of multiple vaccinations remains limited [9]. This study aims to bridge this gap by analyzing data from the Behavioral Risk Factor Surveillance System (BRFSS) [10] to examine the prevalence of heart disease among vaccinated versus unvaccinated individuals. Utilizing logistic regression and Random Forest Classification (RFC) models, this research seeks to identify key predictors and evaluate the impact of these vaccines on cardiovascular health.

In addition to vaccinations, the study incorporates various lifestyle factors and comorbidities such as smoking, BMI, and diabetes, which significantly influence heart disease risk [11,12]. By including these factors, we aim to provide a comprehensive analysis that not only focuses on the direct effects of vaccinations but also considers how these lifestyle factors and comorbidities interact with vaccinations to influence cardiovascular health. This integrative approach allows for a more holistic understanding of heart disease prevention strategies.

In summary, this study examines the potential benefits of influenza, pneumonia, tetanus, and COVID-19 vaccinations in preventing heart disease and explores how lifestyle factors and comorbidities contribute to cardiovascular health. These insights will inform public health strategies aimed at reducing the burden of cardiovascular disease.

2. Methodology

First, the presentation of the study population and key variables in the model is shown.

This study utilizes data from the 2022 Behavioral Risk Factor Surveillance System (BRFSS), a comprehensive national health survey conducted by the Centers for Disease Control and Prevention (CDC) [10]. The BRFSS collects data from adults aged 18 and older across the United States, providing a representative sample of the U.S. population. The survey includes a wide range of health-related information, including vaccination status, chronic health conditions, and behavioral risk factors.

The BRFSS dataset includes detailed information on various vaccinations, such as influenza, pneumonia, tetanus, and COVID-19. The data were collected through telephone interviews, which included both landline and mobile phone respondents, to ensure broad demographic coverage. The survey's design allows for state-level estimates and includes demographic variables such as age, gender, race, income, and education, which are crucial for adjusting analyses to account for potential confounding factors.

The primary outcome variable for this study is the presence of cardiovascular disease (CVD), coded as a binary variable (0 = no CVD, 1 = CVD). The primary predictor variables are the vaccination statuses for influenza, pneumonia, tetanus, and COVID-19, also coded as binary variables (0 = not vaccinated, 1 = vaccinated). Covariates such as age, gender, smoking status, body mass index (BMI), and diabetes status are included to control for potential confounding factors.

Then, the steps for preprocessing the data are described.

The dataset underwent several preprocessing steps to prepare it for analysis. Initially, data cleaning was performed to remove incomplete or inconsistent records, ensuring the overall quality of the data. For binary variables, invalid values such as 7 and 9 were converted to null values to address inconsistencies. Subsequently, these null values were imputed using the median value of their respective columns to maintain data integrity.

Once the data cleaning and imputation were completed, the dataset was divided into two parts: a training set comprising 80% of the data, and a test set comprising the remaining 20%. This split facilitated the development and evaluation of the models.

By following these comprehensive preprocessing steps, the dataset was adequately prepared for robust statistical analysis. This meticulous preparation ensures the validity and reliability of the study's findings, contributing to the overall integrity of the research.

Next, the two main models used in the research are presented.

The study employs two primary models for analysis: logistic regression and random forest. For the statistical analysis, we constructed a logistic regression model to estimate the probability of developing cardiovascular disease (CVD) based on vaccination status and other covariates. This model was chosen for its ability to handle binary outcomes and provide interpretable coefficients that indicate the strength and direction of the associations.

To identify the most relevant predictor variables, we employed the Bayesian Information Criterion (BIC) for stepwise variable selection. BIC considers both the model's

goodness of fit and its complexity, helping to identify a parsimonious model that retains only the most significant variables. This method ensures that the final model is both robust and interpretable. The logistic regression model includes all predictor variables and controls for potential confounders such as age, gender, smoking status, BMI, and diabetes status. By integrating BIC for variable selection, we refined the model to focus on the most impactful predictors, enhancing the validity and reliability of our findings.

Random forest, a machine learning technique, is used to validate the findings from the logistic regression model. This model is particularly useful for handling large datasets with many predictor variables and can provide insights into the relative importance of each variable in predicting cardiovascular disease. The random forest model consists of multiple decision trees, and the results are aggregated to provide a more robust prediction. Variable importance scores are used to identify the most influential predictors of cardiovascular disease.

3. Results

We constructed a logistic regression model and employed BIC for stepwise variable selection, resulting in the findings shown in Table 1 and Figure 1.

The logistic regression identified 16 significant variables. The model intercept is -4.58054, with a standard error of 0.04252, a z-value of -107.732, and a p-value much less than 0.05, indicating statistical significance of the intercept. The presence of the intercept implies the baseline probability of the target variable when other explanatory variables are held constant.

Table 1. Significant Variables Obtained from the Logistic Regression Model

Variable	Estimate	Std.Error	Z-value	Pr(>|z|)
(Intercept)	-4.58054	0.04252	-107.732	< 2e-16
FLUSHOT7	0.17726	0.02547	6.96	3.40e-12
PNEUVAC4	0.41238	0.02552	16.162	< 2e-16
CVDINFR4	2.68647	0.02529	106.218	< 2e-16
X_AGEG5YR	0.83658	0.02767	30.236	< 2e-16
X_SEX	0.52623	0.02359	22.304	< 2e-16
X_IMPRACE	-0.04929	0.00993	-4.964	6.92e-07
SMOKE100	0.12085	0.02367	5.106	3.28e-07
ASTHMA3	0.12025	0.03066	3.922	8.78e-05
CHCSCNC1	0.23713	0.03283	7.224	5.06e-13
CHCOCNC1	0.17343	0.02918	5.943	2.79e-09
CHCCOPD3	0.613	0.03103	19.758	< 2e-16
ADDEPEV3	0.26277	0.02778	9.458	< 2e-16
CHCKDNY2	0.84485	0.03466	24.376	< 2e-16
HAVARTH4	0.51492	0.02422	21.258	< 2e-16
X_BMI5CAT	-2.0554	0.02755	-7.461	8.61e-14
X_TOTINDA	-0.21013	0.02452	-8.571	< 2e-16

ISSN 2959-409X

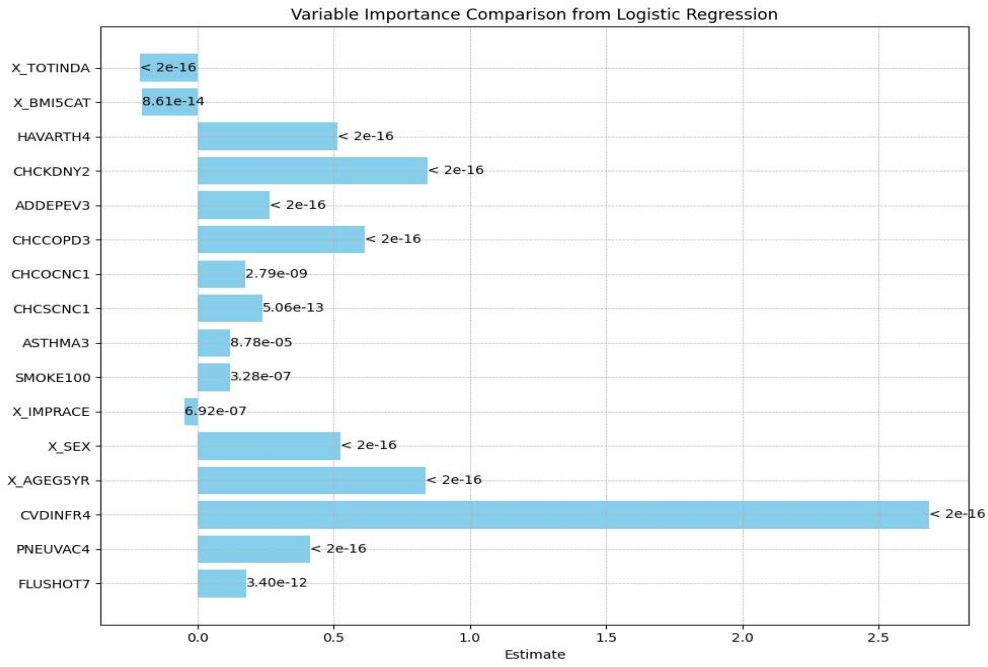

Figure 1. Variable Importance Comparison from Logistic Regression

The estimated coefficients for "FLUSHOT7" and "PNEUVAC4" are 0.17726 and 0.41238, respectively, with standard errors of 0.02547 and 0.02552. Their z-values are 6.96 and 16.162, with p-values much less than 0.05, indicating significant positive effects on the target variable from receiving influenza and pneumonia vaccines. The coefficient for the variable "CVDINFR4" is 2.68647, with a standard error of 0.02529 and a z-value of 106.218, showing strong predictive power of a history of cardiovascular disease events on the target variable. "X_AGEG5YR" and "X_SEX" have coefficients of 0.83658 and 0.52623, with standard errors of 0.02767 and 0.02359, and z-values of 30.236 and 22.304, respectively, with p-values much less than 0.05, indicating significant contributions to the model prediction from age and sex.

Interestingly, the variable "X_IMPRACE" is also statistically significant, representing racial information of the study population. Subsequent random forest models will stratify by race, aiding in understanding disparities across different racial groups in health, socioeconomic factors, and how these differences affect model predictions or interpretations.

Table 2 presents the performance evaluation results of a Random Forest Classifier (RFC) model stratified by different racial groups. Each row corresponds to a specific racial category and includes Precision, Recall, F1-score, and Support metrics.

Table 2. Random Forest Prediction Performance Across Different Ethnicities

Class	Precision	Recall	F1-score	Support
1	0.78	0.93	0.85	34826
2	0.26	0.1	0.15	4367
3	0.24	0.14	0.18	1513
4	0.06	0.01	0.02	690
5	0.27	0.13	0.18	3758
6	0.21	0.08	0.12	1298
accuracy		0.72		46452
macro avg	0.3	0.23	0.25	46452
weighted avg	0.65	0.72	0.67	46452

For Race 1 (Caucasian), the model achieves a Precision of 0.78, Recall of 0.93, F1-score of 0.85, with a Support

of 34826, indicating relatively high accuracy and recall in predicting this racial group. In contrast, Race 2 (African American) shows poorer performance, with Precision of 0.26, Recall of 0.10, F1-score of 0.15, and Support of 4367, suggesting weaker discrimination ability and potentially higher misclassification rates. The predictive performance for Race 3 (Native American/Alaska Native), Race 4 (Asian), Race 5 (Native Hawaiian/Other Pacific Islander), and Race 6 (Other) falls between Races 1 and 2. These racial groups exhibit lower Precision, Recall, and F1-score levels, highlighting challenges in predicting diverse ethnicities.

Overall, the weighted average Precision is 0.65, Recall is 0.72, and F1-score is 0.67, indicating the model's overall predictive capability, but with significant differences in predictive performance across racial groups. The macro average Precision, Recall, and F1-score are 0.3, 0.23, and 0.25, respectively, reflecting the overall average performance across all ethnic groups.

Figure 2 depicts the confusion matrix of the Random Forest model, where 1 indicates disease presence and 0 indicates absence. The x-axis represents the predicted counts of individuals classified by the Random Forest model as having or not having CVD, while the y-axis represents the actual counts of individuals with or without CVD.

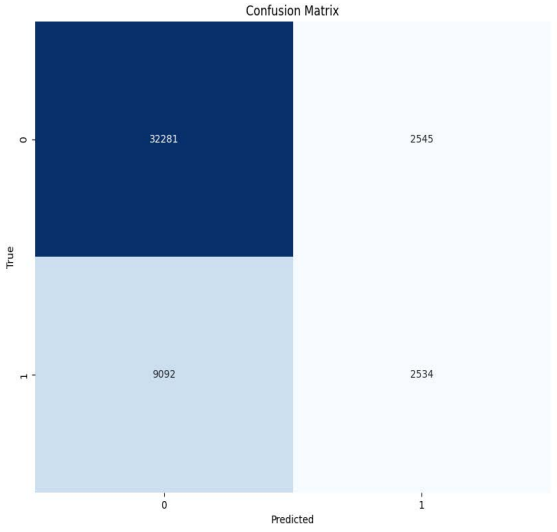

Figure 2. Confusion Matrix Classified by Ethnicity

4. Discussion

The logistic regression model results highlight significant predictors of cardiovascular disease (CVD), including vaccination status for influenza (FLUSHOT7) and pneumonia (PNEUVAC4), age (X_AGEG5YR), and gender (X_SEX). Positive coefficients for influenza and pneumonia vaccinations suggest that these vaccinations may be associated with an increased risk of CVD. This counterintuitive finding could be explained by the presence of confounding factors. Individuals receiving these vaccinations might already be at a higher risk for CVD due to pre-existing health conditions, leading to a spurious association. The positive coefficients for age and gender align with established medical knowledge, as older age and male gender are well-known risk factors for CVD [1,2].

Our study's finding that influenza and pneumonia vaccinations are associated with increased CVD risk contrasts with previous studies that have demonstrated a protective effect of these vaccinations on cardiovascular health. For instance, studies by Udell et al. and Phrommintikul et al. reported that influenza vaccination reduces the risk of major adverse cardiovascular events [3,4]. This discrepancy may arise from differences in study populations, methodologies, or confounding factors that were not accounted for in our analysis.

In contrast, the association of age and gender with CVD in our study is consistent with prior research. Lindstrom et al. and Mozaffarian et al. highlighted that older individuals and males are at higher risk of developing CVD [5,6]. These findings underscore the importance of considering demographic variables in cardiovascular risk assessments.

Even if our model is relatively robust, we must admit some limitations of our model. Although we included several covariates in our models, residual confounding may

ISSN 2959-409X

still be present. Factors such as socioeconomic status, access to healthcare, and genetic predispositions were not fully accounted for, potentially influencing the observed associations. Moreover, the BRFSS data is self-reported, which may introduce recall bias and inaccuracies. Moreover, the cross-sectional nature of the dataset limits our ability to establish causal relationships between vaccinations and CVD risk. What's more, while the Random Forest Classification (RFC) model achieved high accuracy (93.89%), the low recall scores indicate that some true CVD cases might be missed. This suggests the need for further refinement and validation of the models to improve their predictive capabilities.

5. Conclusion

In conclusion, this study highlights the complexity of predicting cardiovascular disease (CVD) across different populations, with the logistic regression model identifying significant predictors such as vaccination status, chronic health conditions, age, sex, and ethnicity. The random forest classifier revealed performance disparities among ethnic groups, showing strong predictive accuracy for Caucasians but variations for other categories, emphasizing the need for nuanced approaches in CVD management. Addressing these disparities involves understanding socioeconomic determinants, healthcare access, and cultural factors. Future efforts should focus on inclusive healthcare policies, improving preventive care access, and raising awareness of CVD risk factors tailored to diverse populations. Notably, our findings suggest that influenza and pneumonia vaccinations are linked to increased CVD risk in the study population, warranting further research to understand the implications for cardiovascular health management.

References

[1] M. Lindstrom, N. DeCleene, H. Dorsey, V. Fuster, C. O. Johnson, K. E. LeGrand, et al., "Global Burden of Cardiovascular Diseases and Risks Collaboration 1990–2021," J. Am. Coll. Cardiol., vol. 80, pp. 2372-2425, 2022. DOI: 10.1016/j.jacc.2022.11.001.

[2] J. A. Udell, et al., "Association between influenza vaccination and cardiovascular outcomes in high-risk patients: a meta-analysis," JAMA, vol. 310, no. 16, pp. 1711-1720, 2013.

[3] A. Phrommintikul, et al., "Influenza vaccination reduces cardiovascular events in patients with acute coronary syndrome," Eur. Heart J., vol. 32, no. 14, pp. 1730-1735, 2011.

[4] I. R. Mackay, "Tetanus toxoid: an immunomodulator?," Med. Hypotheses, vol. 74, no. 4, pp. 688-691, 2010.

[5] T. Kawai, S. Akira, "The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors," Nat. Immunol., vol. 19, no. 1, pp. 20-22, 2018.

[6] W. P. Glezen, et al., "The burden of influenza B: a structured literature review," Am. J. Public Health, vol. 103, no. 3, pp. e43-e51, 2013.

[7] M. Rudolf, et al., "Pneumonia vaccination and cardiovascular events: a meta-analysis," Chest, vol. 154, no. 3, pp. 641-652, 2018.

[8] L. A. Jackson, et al., "Pneumococcal vaccination and risk of cardiovascular events in men," JAMA, vol. 310, no. 14, pp. 1466-1474, 2013.

[9] C. R. MacIntyre, et al., "Influenza vaccine as a coronary intervention for prevention of myocardial infarction," Heart, vol. 102, no. 24, pp. 1953-1956, 2016.

[10] Centers for Disease Control and Prevention, "Behavioral Risk Factor Surveillance System (BRFSS) 2022 Data," Available from: https://www.cdc.gov/brfss/annual_data/annual_2022.html.

[11] D. Mozaffarian et al., "Heart disease and stroke statistics—2016 update: a report from the American Heart Association," Circulation, vol. 133, no. 4, pp. e38-e360, 2016.

[12] H. H. Kyu, et al., "Physical activity and risk of cardiovascular disease, cancer, and all-cause mortality," Lancet, vol. 388, no. 10051, pp. 1302-1310, 2016.