

Article

Identifying Modifiable Predictors of COVID-19 Vaccine Side Effects: A Machine Learning Approach

Sara Abbaspour^{1,2,*}, Gregory K. Robbins^{3,4}, Kimberly G. Blumenthal^{4,5,6,7}, Dean Hashimoto^{4,8,9}, Karen Hopcia⁹, Shibani S. Mukerji^{4,10}, Erica S. Shenoy^{3,11}, Wei Wang^{2,12} and Elizabeth B. Klerman^{1,2,12}

- ¹ Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, USA
- ² Division of Sleep Medicine, Harvard Medical School, Boston, MA 02114, USA
- ³ Department of Medicine, Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA 02114, USA
- ⁴ Harvard Medical School, Boston, MA 02114, USA
- ⁵ The Mongan Institute, Massachusetts General Hospital, Boston, MA 02114, USA
- ⁶ Division of Rheumatology, Allergy, and Immunology, Massachusetts General Hospital, Boston, MA 02114, USA
- ⁷ Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
- ⁸ Department of Medicine, Brigham and Women's Hospital, Boston, MA 02114, USA
- ⁹ Occupational Health Services, MassGeneralBrigham, Boston, MA 02114, USA
- ¹⁰ Department of Neurology, Division of Neuroimmunology and Neuro-Infectious Diseases, Massachusetts General Hospital, Boston, MA 02114, USA
- ¹¹ Infection Control Unit, Massachusetts General Hospital, Boston, MA 02114, USA
- ¹² Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA 02114, USA
- * Correspondence: sabbaspour@mgh.harvard.edu



Citation: Abbaspour, S.; Robbins, G.K.; Blumenthal, K.G.; Hashimoto, D.; Hopcia, K.; Mukerji, S.S.; Shenoy, E.S.; Wang, W.; Klerman, E.B. Identifying Modifiable Predictors of COVID-19 Vaccine Side Effects: A Machine Learning Approach. *Vaccines* **2022**, *10*, 1747. <https://doi.org/10.3390/vaccines10101747>

Academic Editor: Giuseppe La Torre

Received: 20 September 2022

Accepted: 15 October 2022

Published: 19 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Side effects of COVID-19 or other vaccinations may affect an individual's safety, ability to work or care for self or others, and/or willingness to be vaccinated. Identifying modifiable factors that influence these side effects may increase the number of people vaccinated. In this observational study, data were from individuals who received an mRNA COVID-19 vaccine between December 2020 and April 2021 and responded to at least one post-vaccination symptoms survey that was sent daily for three days after each vaccination. We excluded those with a COVID-19 diagnosis or positive SARS-CoV2 test within one week after their vaccination because of the overlap of symptoms. We used machine learning techniques to analyze the data after the first vaccination. Data from 50,484 individuals (73% female, 18 to 95 years old) were included in the primary analysis. Demographics, history of an epinephrine autoinjector prescription, allergy history category (e.g., food, vaccine, medication, insect sting, seasonal), prior COVID-19 diagnosis or positive test, and vaccine manufacturer were identified as factors associated with allergic and non-allergic side effects; vaccination time 6:00–10:59 was associated with more non-allergic side effects. Randomized controlled trials should be conducted to quantify the relative effect of modifiable factors, such as time of vaccination.

Keywords: vaccination; COVID-19; side effects; allergy; time-of-day-effects; machine learning; model explanation

1. Introduction

COVID-19 vaccines have been distributed to billions of individuals worldwide and have reduced serious illness, hospitalizations, and death [1]. As of July 2022, only 61% of the world's population has been fully vaccinated against COVID-19 [2]. An important factor for vaccine hesitancy is concern about vaccine safety, efficacy, and side effects [3,4]. Understanding risk factors for vaccine-related side effects—especially ones that may be modifiable—is important for clinicians, for patient safety, for patient expectations and planning, and possibly for reducing hesitancy to be vaccinated.

At MassGeneralBrigham (MGB), a large integrated healthcare system, surveys were used to obtain information about post COVID-19 vaccination symptoms through email, text message, phone, and smartphone application links as part of employee health monitoring. Several publications using epidemiological analytic techniques have reported important clinical information using this dataset, including the incidence of allergic and non-allergic side effects after mRNA COVID-19 vaccination and their association with variables such as demographic characteristics, allergy history, and prior known infection with SARS-CoV-2 [5–12]. This report explores new potential covariates, including time-of-day for COVID-19 vaccinations (which is an easily modified factor), and utilized a different analysis approach—Machine Learning (ML).

Age, sex, race, hormones, and body mass index (BMI) have previously been associated with vaccination-related adverse effects for different vaccines [13–15]. We studied time-of-day of vaccination as a potential predictor because of the well-known impact of circadian rhythms on physiology [16] including immune responses [16,17] and because it is relatively easily modified. Effects of time-of-day of COVID-19 vaccination on anti-Spike antibody responses have been documented: late afternoon vaccination, being female and being younger were associated with higher response [18]. Different side effects related to time-of-day of other vaccines have also been reported [19,20].

ML is a powerful technique for diagnosis, detection, prediction, and prognosis in medicine. Studies have used ML-based approaches to explain the contribution of different variables (e.g., age, tumor size, and number of removed lymph nodes) in prediction of 10-year overall survival of breast cancer [21], identify the most critical factors in predicting the prevalence of stroke [22], and predict the risk of hypoxemia during general anesthesia and provide explanations of the risk factors (e.g., age, sex, BMI, blood pressure, temperature, and medication) [23].

In this study, our aim was to identify predictors of post COVID-19 vaccine-related side effects using ML methods on self-reported side effects for 3 days after the first vaccination of an mRNA-based vaccine.

2. Materials and Methods

2.1. Data

The dataset was created from MGB electronic health record's COVID-19 Datamart and a REDCap (Research Electronic Data Capture [24]) survey that collected self-reported symptoms for 1–3 days after each vaccination. Detailed information about the REDCap side effect survey have been previously reported [5]. For this report, we only used data from the first vaccination.

Data from individuals were included if they (i) received an mRNA COVID-19 vaccine at a MGB site between December 2020 and April 2021, (ii) responded to at least one day of the REDCap survey, (iii) had documented time-of-day of administration of COVID-19 vaccine and (iv) did not have a COVID-19 diagnosis or positive polymerase chain reaction test within one week after the vaccination (because of overlap of disease and side-effect symptoms [5]). Individuals who received the Janssen adenovirus-based vaccine were excluded from this analysis for two reasons: only 1486 (<3%) individuals who met our inclusion criteria received a Janssen vaccine, and people with high risk of allergic reaction were recommended for the Janssen vaccine and this could introduce a bias in our dataset [25].

Variables considered as potentially associated factors for side effects were: demographics (age, sex (Female or Male), race (Asian, White, Black, or Other Race (i.e., Unknown/Missing, Other, Two or More, Declined, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander)), and ethnicity (Hispanic, Non-Hispanic, or Other Ethnicity (i.e., Unknown/Missing, or Declined))); any note of Epinephrine Autoinjector Prescription (e.g., EpiPen); any history of allergy (divided into Food, Vaccine, Medication, Insect sting, Seasonal, Latex and Other categories using coded and free-text data within the electronic health record allergy list); any COVID-19 diagnosis/positive test prior to

vaccination; vaccine manufacturer (Pfizer: New York, NY, USA or Moderna: Cambridge, MA, USA); and clock time of vaccine administration and/or appointment. The number of different allergy history categories for an individual was summed. The variables considered for analysis were selected based on expert knowledge, feature selection approaches, and their availability in the database.

Outcomes of interest were side effects queried at days 1, 2 and 3 post-vaccination by a REDCap survey. The questions and response options were:

- Allergic symptoms: (Yes/No) (i) Rash or itching; (ii) Hives; (iii) Swollen lips, tongue, eyes, or face; (iv) Respiratory symptoms (wheezing, chest tightness, or shortness of breath).
- Non-allergic symptoms: (None/lower severity/higher severity) (i) New headache; (ii) New fatigue; (iii) Joint pain; (iv) Muscle pain; (v) Fever.

This study was approved by the MGB human research committee and MGB Occupational Health Services.

2.2. Pre-Processing

Race and ethnicity were grouped into 4 categories to reduce the number of categories tested: White/Non-Hispanic, Non-White/Non-Hispanic, Any-Race/Hispanic, and Any-Race/Other-Ethnicity. For time-of-day of vaccination, there were two data entries available: immunization time and appointment time. Immunization time was used preferentially; appointment time was used only if immunization time was not available. Time-of-day groupings for these analyses were 6:00–10:59, 11:00–15:59, and 16:00–21:59. Individuals receiving vaccines before 6:00 and after 22:00 were excluded because of low numbers (10 participants total).

To prepare the input features for analysis, categorical variables were converted into dummy variables using OneHotEncoder, a scikit-learn (version 1.0.1) preprocessing package in Python. To avoid collinearity effect between the input variables, a Variance Inflation Factor analysis was conducted (threshold = 5) [26].

To increase the classification performance, dimensionality reduction in the feature set is often necessary. In this study, two feature selection methods were applied: Shapley Additive exPlanations (SHAP) feature importance values (details below), and forward feature selection and backward feature elimination. Forward feature selection and backward feature elimination consist of adding features one by one to the feature set. If an added feature produced higher accuracy rate, it would stay in the feature set; otherwise, it would be removed. Once all features were evaluated, features in the obtained feature set were removed in inverted order if their subtraction did not negatively affect accuracy. After obtaining SHAP feature importance values from the ML Extreme Gradient Boosting (XGB) model, we chose the 8 best ranked features, since using more features did not improve our model's performance (specifics below).

For allergic symptoms, if the response to an allergic symptom for any of the 3 days was "yes", it was grouped as class 1 (=yes). If the response to an allergic symptom for all of the 3 days was "no", then it was grouped as class 0 (=no). If a participant completed the survey for only one or two days and the responses on those days were both "no", that entry was removed, because we do not know if the response on the missing day would have been "yes". For non-allergic symptoms, similar logic was used to group as class 1 (=higher severity) or class 0 (=none/lower severity) or removal of entry.

2.3. Machine Learning Model

XGB, a tree-based ML model, was selected because of its execution speed and performance [21], high interpretability, and the possibility of identifying the strongest predictors by applying a model explanation such as TreeExplainer [26]. XGB (with max_depth = 3, number of estimators = 50, and learning rate = 0.1) was applied to predict any allergic (yes vs. no) and any non-allergic (higher severity vs. none and lower severity) side effects reported for 3 days after the vaccination. The model was parametrized using a randomized

search of different parameter settings with a 5-fold cross validation. Since the dataset is not balanced in respect to the dependent variables, up sampling (RandomOverSampler an imbalanced-learn (version 0.7.0) over_sampling package in Python) was used to increase the number of samples for the minority classes (i.e., the yes responses to allergic symptoms and the higher severity responses to non-allergic symptoms) in the training sets.

2.4. Evaluation

A stratified k-fold ($k = 5$) cross-validation was used to validate the performance of the ML model. This method uses a large part of the data (80% of the data) to train the model, and a small part of the data (20% of the data) to test the model. The stratified cross-validation was repeated 10 times and the average and the standard deviation (SD) of F-score (Equation (1)) was calculated [27]. This evaluation metrics has a range of 0 to 1, a higher value shows a better performance.

$$F - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (1)$$

2.5. Explainability

One concern about ML is that the results are “black box” and not interpretable. To address this, we chose TreeExplainer that uses SHAP values, a game theory method for assigning an importance value to variables based on their contribution to the model [26], and to explain the magnitude and direction of the contribution of each of the variables to the model prediction [21]. This property therefore allows providing both new insights into the model’s variables and the relations between them. SHAP values were generated using the SHAP package (version 0.39.0 in Python). These values were used to obtain a visualization of the overall feature importance for the model. Then, to show how each variable contributed to the model’s output, we generated SHAP boxplots (by applying a seaborn (version 0.11.2) boxplot package in Python) for categorical variables (e.g., sex, race/ethnicity) and SHAP scatter plots for continuous variables (e.g., age). We also used SHAP’s local explainability feature (SHAP Waterfall plot) to display the effect of each of the variables for individual predictions.

The processing stages designed for this study are illustrated in Figure 1A,B. All analyses were performed using open-source libraries in Python 3.7.

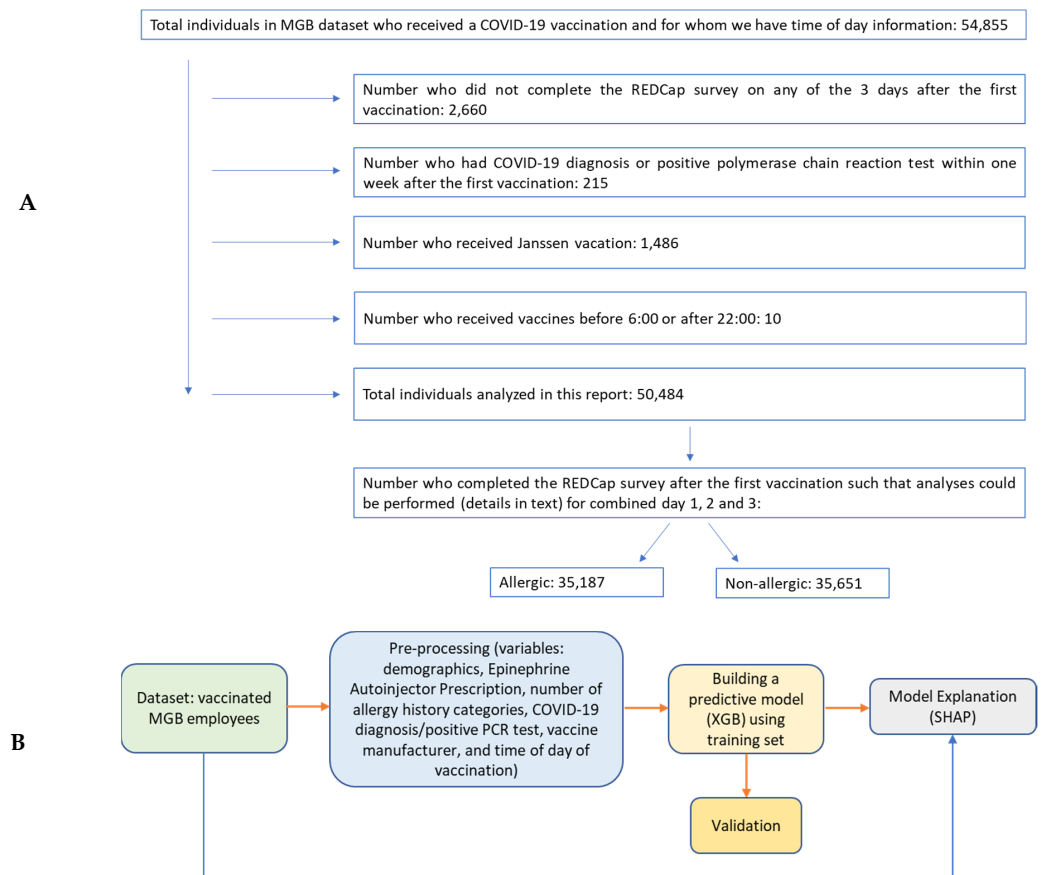


Figure 1. (A): Creation of dataset; (B): Block diagram illustrating the processing stages used in this study to identify predictors of post COVID-19 vaccine-related side effects.

3. Results

Data analyzed were from 50,484 individuals (Table 1, Figure 1A). Of these individuals, 60% received the Moderna vaccine; 73% were female; ages ranged from 18 to 95 years old; 8% had prior COVID-19 diagnosis or positive test. A total of 2% had Epinephrine Autoinjector Prescription; 28% had any history of allergy documented; and 34% received their vaccine from 6:00 to 10:59, 44% from 11:00 to 15:59 and 22% from 16:00 to 21:59. When both immunization and appointment time were available, the number of times for whom there was difference in immunization vs. appointment time group (i.e., 6:00–10:59, 11:00–15:59, 16:00–21:59) was 2.6% (156 of 6011 appointments).

Table 1. N: number of individuals and (%) percent of total participants.

Variables	N	(%)
Age in Years		
Age Group 1 (18–40)	25,213	50
Age Group 2 (41–60)	18,529	37
Age Group 3 (61–95)	6742	13
Total	50,484	100
Sex		
Female	36,801	73
Male	13,683	27
Total	50,484	100
Race/Ethnicity		
White/Non-Hispanic	28,408	56
Non-White/Non-Hispanic	8066	16
Any Race/Hispanic	2662	5
Any Race/Other Ethnicity	11,348	23
Total	50,484	100
Prescription History		
Epinephrine Autoinjector Prescription	1246	2
Allergy History		
Any History of Allergy	14,197	28
COVID-19 Diagnosis/Positive PCR Test		
Any Before Vaccination 1	3797	8
Vaccine Manufacturer		
Pfizer	20,324	40
Moderna	30,160	60
Total	50,484	100
Clock Time of Vaccine Administration/Appointment		
Time 1 (6:00–10:59)	17,254	34
Time 2 (11:00–15:59)	22,367	44
Time 3 (16:00–21:59)	10,863	22
Total	50,484	100

Medical conditions including any history of thrombosis, myocardial infarction, or stroke were variables that did not show any impact on the model's accuracy, and therefore, were not included in the final feature set. Eight variables/features (age, sex, race/ethnicity, Epinephrine Autoinjector Prescription, number of allergy history categories, any prior COVID-19 diagnosis or positive test, vaccine manufacturer, and time-of-day of vaccination) were used to build two predictive ML models; one for predicting any allergic symptoms (yes vs. no) and one for predicting any non-allergic symptoms (higher severity vs. none/lower severity) reported for 3 days after vaccination. The models showed predictive F-score values of 84% (SD = ± 0.01) for allergic symptoms and 81% (SD = ± 0.01) for non-allergic symptoms.

A SHAP feature importance plot was created using mean absolute SHAP values of the ML model for predicting allergic symptoms (Figure 2A): this plot orders the input variables (top to bottom along the y-axis) according to their importance to the ML model. The most important predictors to predict any allergic symptoms were, in descending order: number of allergy history categories, sex, race/ethnicity, age, Epinephrine Autoinjector Prescription, any prior COVID-19 diagnosis or positive test, vaccine manufacturer, and time-of-day of vaccination (Figure 2A).

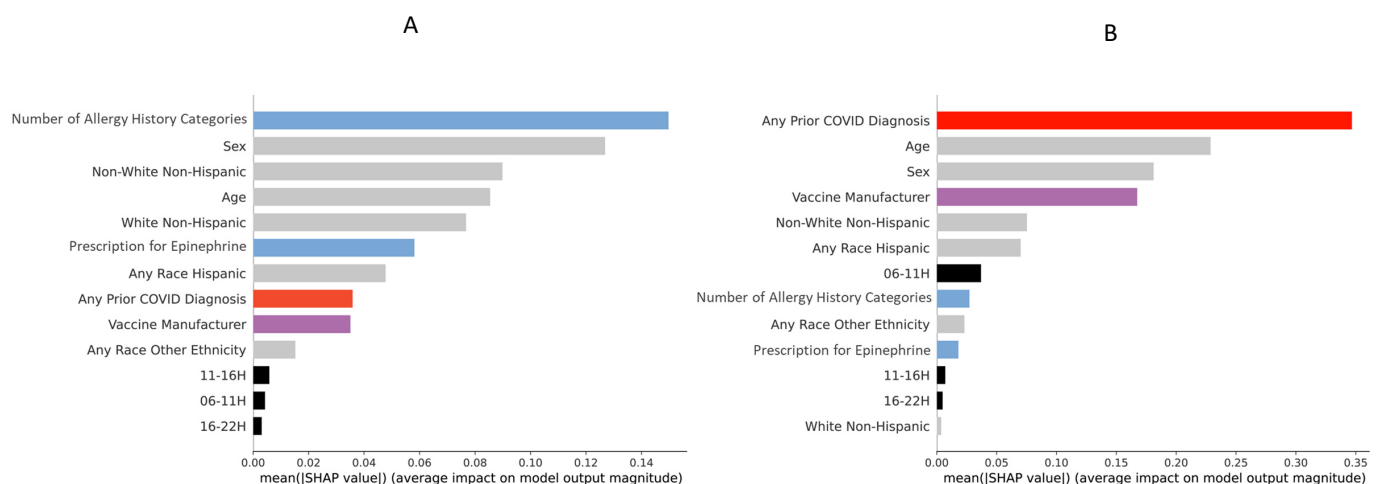


Figure 2. Feature importance plot using the mean absolute SHAP values for (A): allergic side effects and (B): non-allergic side effects. Colors: grey—demographics, blue—allergy history category/prescription for Epinephrine, red—any prior COVID diagnosis or positive test, purple—vaccine manufacturer, black—time-of-day of vaccination.

The SHAP boxplot (Figure 3A) shows the direction of impact of the categorical variables on the model output in predicting report of allergic symptoms for 3 days after vaccination. Positive SHAP values are associated with a higher likelihood of reporting symptoms and negative SHAP values are associated with a lower likelihood of reporting symptoms. Females, Non-White/Non-Hispanic, Any-Race/Hispanic, Epinephrine Autoinjector Prescription, those who had any prior COVID-19 diagnosis or positive test, and people who received the Moderna vaccine were more likely to report allergic symptoms. No significant (i.e., the SHAP value of 0 was within 5–95% distribution of values) effect of time-of-day was found for allergic symptoms. For the continuous variable of age, younger adults of both sexes were more likely to report allergic side effects (Figure 3B), with magnitude approximately constant for ages 30–60 and then a decline in likelihood starting at ~60 years of age. There was a monotonically increasing effect of number of allergy history categories on the likelihood to report allergic side effects (Figure 3C).

For any non-allergic symptoms, the most important predictors in descending order of magnitude were any prior COVID-19 diagnosis or positive test, age, sex, vaccine manufacturer, race/ethnicity, time-of-day of vaccination, number of allergy history categories, and Epinephrine Autoinjector Prescription (Figure 2B). For non-allergic symptoms, in addition to the factors significant for allergic symptoms, significant effects of time-of-day (morning vaccinations (6:00–10:59) were also associated with more non-allergic side effects (Figure 3D); there was also a more linear effect of age on likelihood of non-allergic symptoms (Figure 3E). A lower magnitude of the monotonically increasing effect of the number of allergy history categories was also seen (Figure 3F).

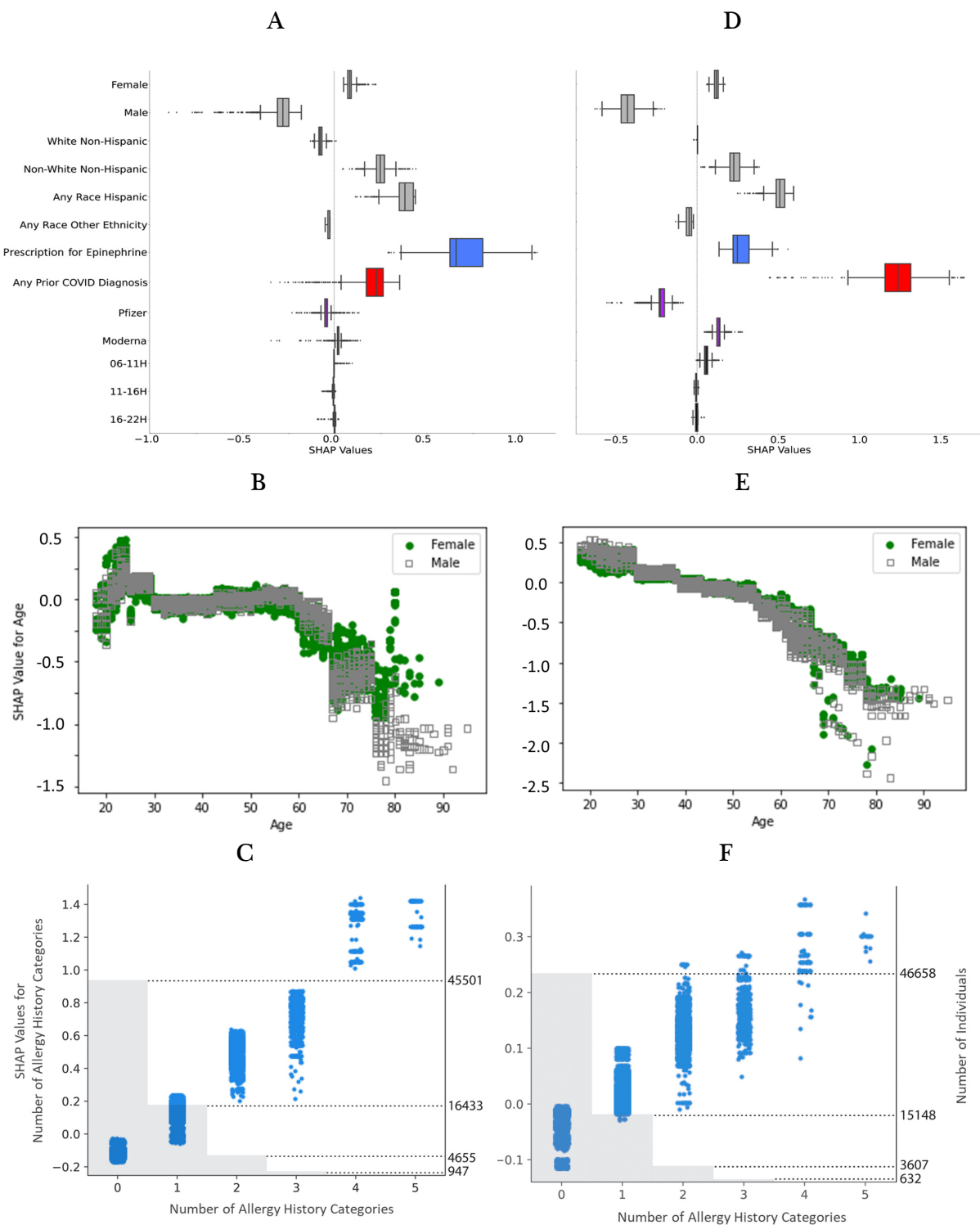
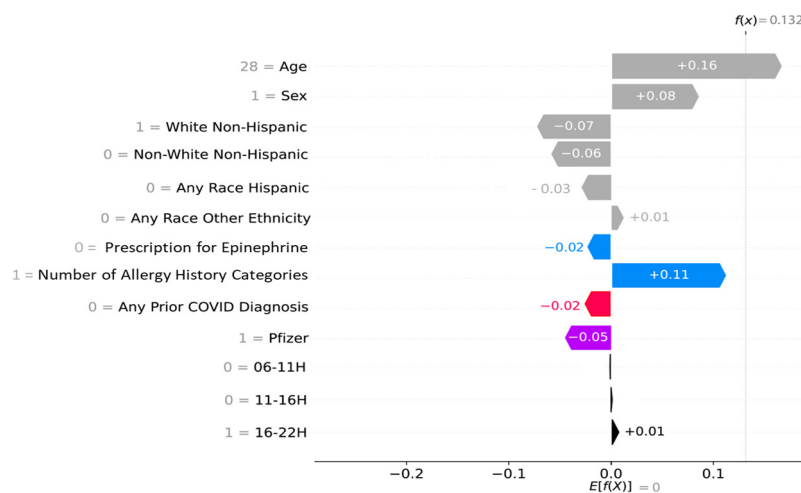


Figure 3. (A): SHAP value boxplot that shows the direction of impact of each variable on model's output for allergic side effects. Positive SHAP values are indicative of having side effects, while negative SHAP values are indicative of not having side effects. Box limits indicate 25th and 75th percentile, vertical line within the box indicates 50th percentile, and other vertical lines indicate 5th and 95th percentiles; (B): SHAP feature independent plot for age vs. sex showing the impact of age on model output and (C): SHAP value scatter plot for number of allergy history categories showing the impact of this variable on model output. (D–F): as in (A–C) except for non-allergic side effects. Colors in (A,D): grey—demographics, blue—prescription for epinephrine, red—any prior COVID diagnosis or positive test, purple—vaccine manufacturer, black—time-of-day of vaccination.

SHAP values can also be used to create a “local” explanation for every observation/individual in the dataset (in addition to the global effects detailed above). To illustrate a local explanation for specific individual predictions, Waterfall plots were used: Figure 4 presents four examples. Each row in SHAP Waterfall plot shows the positive or negative contribution (x -axis) of each input variable (y -axis) to the overall likelihood of having allergic (Figure 4A,B) or non-allergic side effects (Figure 4C,D). For example: (i) characteristics such as being young (Age = 28), female, Hispanic, with a history of allergy, receiving Moderna and being vaccinated 6:00 to 10:59 and 16:00 to 21:59 (06-11H = 1, 16-22H = 1) increase the chance of having side effects after vaccination (ii) characteristics such as being male, White Non-Hispanic, any race other ethnicity, no history of Autoinjector Epinephrine Prescription, no history of allergy, no prior COVID diagnosis or positive test, receiving Pfizer, and being vaccinated between 11 to 15:59 (11-16H = 1) decrease the chance of having side effects after vaccination.

A



B

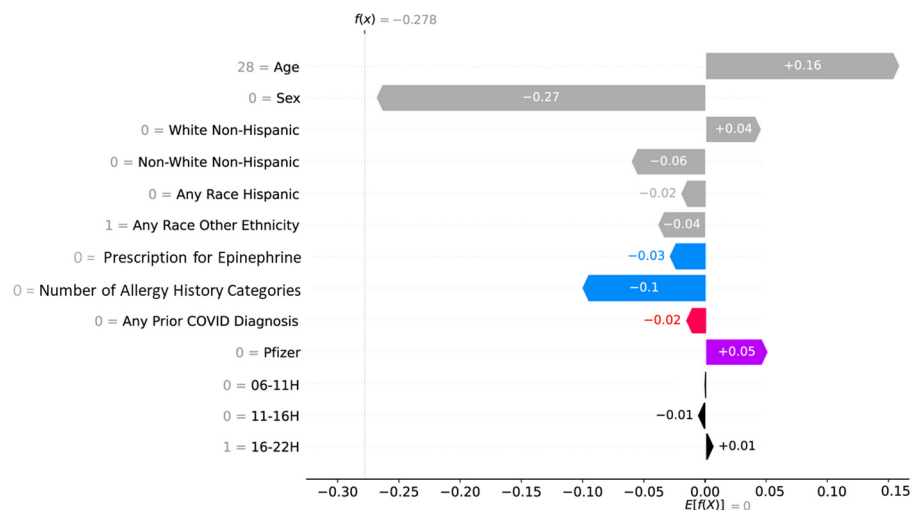


Figure 4. Cont.

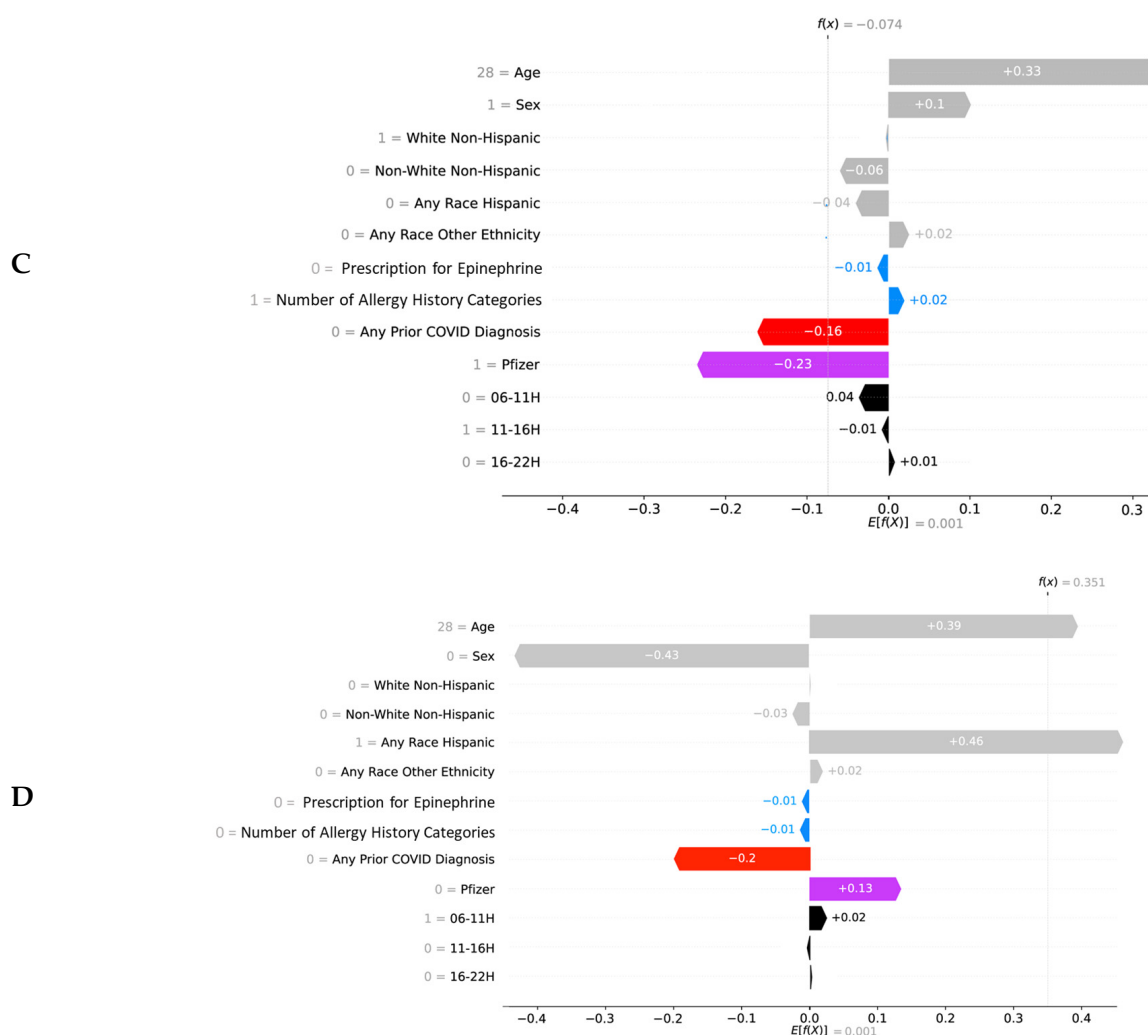


Figure 4. SHAP Waterfall plots exemplifying local/individual predictions for 4 individuals showing the contribution of each variable to the prediction. The gray text in front of each variable name is the value of the particular variable. The baseline value ($E[f(X)]$) is displayed below the x -axis, indicating the expected value of the model. The model output for each individual ($f(x)$) is shown on top of each panel; it is the sum of SHAP values calculated for all variables. Positive SHAP values push the model to predict having side effects, while negative SHAP values push the model to predict no side effects. Allergic side effects: (A): a 28 year old white Non-Hispanic male who received Moderna between 16 and 21:59. (B): a 28 year old male with any race Other Ethnicity who received Moderna between 16 and 21:59; Non-allergic side effects: (C): a 28 year old white Non-Hispanic female who received Pfizer between 11 and 15:59, and (D): a 28 year old male with any race Hispanic who received Moderna between 06 and 10:59. Absolute SHAP values < 0.01 were not presented on the figures. Colors: grey—demographics, blue—allergy history category/prescription for epinephrine, red—any prior COVID diagnosis or positive test, purple—vaccine manufacturer, black—time-of-day of vaccination.

4. Discussion

We used an explainable ML method to identify predictors of post-COVID-19 vaccine side effects in a large dataset. Our results are consistent with several recent publications [5,6,8,9] that documented both non-allergic and allergic type side effects after COVID-19 vaccination and identified the effect of different factors influencing the severity of reported side effects after COVID-19 vaccinations. These findings include: (i) association of reported non-allergic symptoms after vaccination with demographic characteristics and prior COVID-19 diagnosis or positive test [5,8]. We also documented that female, younger individuals, Non-White race, Hispanic ethnicity, and those with prior COVID-19 infection

were more likely to report non-allergic side effects after vaccination. (ii) association of reported allergic reactions to mRNA COVID-19 vaccines with a history of allergic reaction [6] (iii) the Moderna vaccine is associated with more allergic reactions than the Pfizer vaccine.

We also identified previously unreported factors affecting side effects reported after the first dose of an mRNA COVID-19 vaccine: number of allergy history categories, a history of Autoinjector Epinephrine Prescription and the modifiable factor, time-of-day of vaccination. Circadian rhythms are physical, mental, and behavioral changes that display a period of approximately 24 h [28]. These rhythms influence almost all areas of physiology, including the sleep-wake cycle, body temperature, blood pressure, and heart rate [16]. Chronomedicine aims to incorporate knowledge of biological rhythms to increase treatment effectiveness, including reduction in side effects. Timing the administration of a drug to coincide with peak levels of its physiologic target has shown clinical benefits in hypertension, hypercholesterolemia, cancer, and other areas [29]. Currently, health care professionals rarely consider time-of-day in their diagnosis and treatment administration [30] and often vaccination times are chosen by convenience. The information about time-of-day may be used to better define relevant physiology (i.e., the multiple components of a response to vaccination, some of which may differ by time-of-day) and improve clinical care. For example, altering the time of COVID-19 vaccination to lower unwanted side effects would be a relatively low-cost and scalable change in practice.

Electronic health data provide the opportunity to improve healthcare. Handling these large and complex datasets requires special computational techniques that can deal with these datasets. ML techniques have broad applications in healthcare and are helpful in identifying patterns in large datasets [31]. Developments in the area of ML and model explanation, and strong methods to compute and visualize the magnitude and direction of impact of input variables on model's outputs, can help translate knowledge from science to practice [26,32]. Given our multidimensional datasets, the application of ML can be useful since its strength includes dealing with many input variables.

Limitations of this work are that the data are from an observational study. Randomized clinical trials should be performed to further test our hypotheses of time-of-day effects. Collecting time-of-day of vaccination data and data about sleep obtained before or after the vaccination [33–36] should be included in future studies. Future work should also (i) explore side effects occurring during the 3 days after the second dose using these techniques: (ii) the impact of night shift work before and/or after COVID-19 vaccination on self-reported side effects, and (iii) target underlying physiological reasons.

5. Conclusions

In this study, we used XGB, a ML model to predict the occurrence of self-reported COVID-19 vaccination side effects using a range of variables (e.g., demographics, history of allergy, vaccine manufacturer and time-of-day of vaccination). We then used a model explanation technique (SHAP) to identify the important predictors of COVID-19 vaccine-related side effects and explain the effect of the input variables on model's output. Our results demonstrate that demographics, any history of allergy, any prior COVID diagnosis or positive test, vaccine manufacturer, and time-of-day-of-vaccination (6:00–10:59 associated with significantly more non-allergic side effects) effects on side effects reported for three days after the first dose of a COVID-19 vaccination. This information can be used to understand the risk factors of adverse events and for planning of possible time-out of work for healthcare workers and patients (e.g., reduce risk for needing to miss work after vaccination).

Author Contributions: Conceptualization, S.A., G.K.R., K.G.B. and E.B.K.; methodology, S.A., G.K.R., K.G.B., E.S.S. and E.B.K.; software, S.A.; validation, S.A. and E.B.K.; formal analysis, S.A.; investigation, S.A., G.K.R., K.G.B., D.H., K.H., S.S.M., E.S.S. and E.B.K.; resources E.B.K.; data curation, S.A., W.W. and E.B.K.; writing—original draft preparation, S.A.; writing—review and editing, all authors; visualization, S.A.; supervision, E.B.K.; project administration, S.A. and E.B.K.; funding acquisition, E.B.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by S.A.; D.H.; K.H.: None. G.K.R.: SEED, Gilead, Citius Pharm, Emergent Biosolutions, Leonard Meron Bioscience, Pfizer. K.G.B.: NIH K01-AI125631 and Massachusetts General Hospital Department of Medicine Transformative Scholar Award. S.S.M.: NIH K23-MH115812, James S. McDonnell Foundation, and Rappaport Fellowship. E.S.S.: CDC, ASPR (HHS), and MIT/Quanta Computers. W.W.: Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health Award UL1 TR002541). E.B.K.: NIH R01-NS099055, U01NS114001, U54-AG062322, R21DA052861, R01NS114526-02S1, R01-107064; DoD W81XWH201076; Leducq Foundation for Cardiovascular Research.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of MassGeneral Brigham IRB #2021P001080. Approved 4/16/21.

Informed Consent Statement: Not required by the MGB human research committee.

Data Availability Statement: The data are not available to external parties.

Acknowledgments: The authors thank many colleagues in the MassGeneralBrigham health system for the design and implementation of the COVID-19 vaccination program, including Paul D. Bidinger, Thomas D. Sequist, Upeka Samarakoon, Rajesh Patel, Leeann Ouimet, Allen Judd, Anna R. Wolfson, Rebecca Saff, Aidan A. Long, Lily Li, Tanya M. Laidlaw, David I. Hong, Anna M. Feldweg, Katrin Stinson, Amanda J. Centi, Lynn Simpson, Nahal Beik, BCPS, Christian M. Mancini, Amelia S. Cogan, and Aubree E. McMahon. No compensation was received.

Conflicts of Interest: S.A.; D.H.; K.H.; S.S.M.: None. G.K.R.: Trial support from Leonard Meron Bioscience, Consulting for Teradyne Inc; outside the submitted work. Member of the DHHS OI guidelines. K.G.B.: Grant support from Phadia Ab (Thermo Fisher Scientific); personal fees for legal case review from Weekley Shulte Valdes Murman Tonelli, Piedmont Liability Trust, Vasios Kelly and Strollo PA, and Publix Supermarkets; and royalties from UpToDate, all outside the submitted work. E.S.S.: a writer for Up To Date; outside the submitted work. W.W.: Consultancy for National Sleep Foundation, outside the submitted work. E.B.K.: Consulting for American Academy of Sleep Medicine Foundation, Circadian Therapeutics National Sleep Foundation, Sleep Research Society Foundation, Yale University Press; travel support from European Biological Rhythms Society; partner owns Chronsulting. All outside the submitted work.

References

1. CDC COVID-19 Vaccines Work. 2021. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/effectiveness/work.html> (accessed on 1 August 2022).
2. Coronavirus (COVID-19) Vaccinations. 2022. Available online: <https://ourworldindata.org/covid-vaccinations> (accessed on 1 August 2022).
3. Azarpanah, H.; Farhadloo, M.; Vahidov, R.; Pilote, L. Vaccine hesitancy: Evidence from an adverse events following immunization database, and the role of cognitive biases. *BMC Public Health* **2021**, *21*, 1686. [CrossRef]
4. Biswas, N.; Mustapha, T.; Khubchandani, J.; Price, J.H. The nature and extent of COVID-19 vaccination hesitancy in healthcare workers. *J. Community Health* **2021**, *46*, 1244–1251. [CrossRef] [PubMed]
5. Shenoy, E.S.; Wickner, P.G.; West, L.R.; Banerji, A.; Blumenthal, K.G.; Centi, A.J.; Gottlieb, A.; Hashimoto, D.M.; Kim, E.; Kim, M.; et al. Symptom monitoring after coronavirus disease 2019 (COVID-19) vaccination in a large integrated healthcare system: Separating symptoms from severe acute respiratory coronavirus virus 2 (SARS-CoV-2) infection. *Infect. Control Hosp. Epidemiol.* **2021**, 1–8. [CrossRef]
6. Blumenthal, K.G.; Robinson, L.B.; Camargo, C.A.; Shenoy, E.S.; Banerji, A.; Landman, A.B.; Wickner, P. Acute allergic reactions to mRNA COVID-19 vaccines. *JAMA* **2021**, *325*, 1562–1565. [CrossRef]
7. Blumenthal, K.G.; Freeman, E.E.; Saff, R.R.; Robinson, L.B.; Wolfson, A.R.; Foreman, R.K.; Hashimoto, D.; Banerji, A.; Li, L.; Anvari, S.; et al. Delayed large local reactions to mRNA-1273 vaccine against SARS-CoV-2. *N. Engl. J. Med.* **2021**, *384*, 1273–1277. [CrossRef] [PubMed]
8. Robinson, L.B.; Fu, X.; Hashimoto, D.; Wickner, P.; Shenoy, E.S.; Landman, A.B.; Blumenthal, K.G. Incidence of cutaneous reactions after messenger RNA COVID-19 vaccines. *JAMA Dermatol.* **2021**, *157*, 1000–1002. [CrossRef] [PubMed]
9. Robinson, L.B.; Landman, A.B.; Shenoy, E.S.; Hashimoto, D.; Fu, X.; Camargo, C.A.; Wickner, P.; Blumenthal, K.G. Allergic symptoms after mRNA COVID-19 vaccination and risk of incomplete vaccination. *J. Allergy Clin. Immunol. Pract.* **2021**, *9*, 3200–3202.

10. Li, L.; Robinson, L.B.; Patel, R.; Landman, A.B.; Fu, X.; Shenoy, E.S.; Hashimoto, D.M.; Banerji, A.; Wickner, P.G.; Samarakoon, U.; et al. Association of Self-reported High-Risk Allergy History with Allergy Symptoms After COVID-19 Vaccination. *JAMA Netw. Open* **2021**, *4*, e2131034.
11. Shavit, R.; Maoz-Segal, R.; Iancovici-Kidon, M.; Offengenden, I.; Yahia, S.H.; Maayan, D.M.; Lifshitz-Tunitsky, Y.; Niznik, S.; Frizinsky, S.; Deutch, M.; et al. Prevalence of allergic reactions after Pfizer-BioNTech COVID-19 vaccination among adults with high allergy risk. *JAMA Netw. Open* **2021**, *4*, e2122255.
12. Kadali, R.A.K.; Janagama, R.; Peruru, S.; Gajula, V.; Madathala, R.R.; Chennaiahgari, N.; Malayala, S.V. Non-life-threatening adverse effects with COVID-19 mRNA-1273 vaccine: A randomized, cross-sectional study on healthcare workers with detailed self-reported symptoms. *J. Med. Virol.* **2021**, *93*, 4420–4429.
13. Gonzalez-Dias, P.; Lee, E.K.; Sorgi, S.; de Lima, D.S.; Urbanski, A.H.; Silveira, E.L.; Nakaya, H.I. Methods for predicting vaccine immunogenicity and reactogenicity. *Hum. Vaccines Immunother.* **2020**, *16*, 269–276. [\[CrossRef\]](#)
14. Pondo, T.; Rose, C.E.; Martin, S.W.; Keitel, W.A.; Keyserling, H.L.; Babcock, J.; Parker, S.; Jacobson, R.M.; Poland, G.A.; McNeil, M.M. Evaluation of sex, race, body mass index and pre-vaccination serum progesterone levels and post-vaccination serum anti-anthrax protective immunoglobulin G on injection site adverse events following anthrax vaccine adsorbed (AVA) in the CDC AVA human clinical trial. *Vaccine* **2014**, *32*, 3548–3554. [\[PubMed\]](#)
15. Klein, S.L.; Jedlicka, A.; Pekosz, A. The Xs and Y of immune responses to viral vaccines. *Lancet Infect. Dis.* **2010**, *10*, 338–349. [\[CrossRef\]](#)
16. Allada, R.; Bass, J. Circadian mechanisms in medicine. *N. Engl. J. Med.* **2021**, *384*, 550–561. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Long, J.E.; Drayson, M.T.; Taylor, A.E.; Toellner, K.M.; Lord, J.M.; Phillips, A.C. Morning vaccination enhances antibody response over afternoon vaccination: A cluster-randomised trial. *Vaccine* **2016**, *34*, 2679–2685. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Wang, W.; Balfe, P.; Eyre, D.W.; Lumley, S.F.; O'Donnell, D.; Warren, F.; Crook, D.W.; Jeffery, K.; Matthews, P.C.; Klerman, E.B.; et al. Time of day of vaccination affects SARS-CoV-2 antibody responses in an observational study of health care workers. *J. Biol. Rhythms* **2022**, *37*, 124–129. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Langlois, P.; White, R.; Glezen, W. Diurnal variation in human response to influenza vaccination? A pilot study of 125 volunteers. *Ann. Rev. Chronopharmacol.* **1986**, *3*, 123.
20. Langlois, P.H.; Smolensky, M.H.; Glezen, W.P.; Keitel, W.A. Diurnal variation in responses to influenza vaccine. *Chronobiol. Int.* **1995**, *12*, 28–36. [\[CrossRef\]](#)
21. Moncada-Torres, A.; van Maaren, M.C.; Hendriks, M.P.; Siesling, S.; Geleijnse, G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Sci Rep.* **2021**, *11*, 6968. [\[CrossRef\]](#)
22. Hu, L.; Liu, B.; Ji, J.; Li, Y. Tree-Based Machine Learning to Identify and Understand Major Determinants for Stroke at the Neighborhood Level. *J. Am. Heart Assoc.* **2020**, *9*, e016745. [\[CrossRef\]](#)
23. Lundberg, S.M.; Nair, B.; Vavilala, M.S.; Horibe, M.; Eisses, M.J.; Adams, T.; Liston, D.E.; Low, D.K.-W.; Newman, S.-F.; Kim, J.; et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2018**, *2*, 749–760. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Harris, P.A.; Taylor, R.; Thielke, R.; Payne, J.; Gonzalez, N.; Conde, J.G. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. methodology and workflow process for providing translational research informatics support. *J. Biomed. Informat.* **2009**, *42*, 377–381. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Banerji, A.; Wickner, P.G.; Saff, R.; Stone, C.A., Jr.; Robinson, L.B.; Long, A.A.; Wolfson, A.R.; Williams, P.; Khan, D.A.; Phillips, E.; et al. mRNA vaccines to prevent COVID-19 disease and reported allergic reactions: Current evidence and suggested approach. *J. Allergy Clin. Immunol. Pract.* **2021**, *9*, 1423–1437. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Geurkink, Y.; Boone, J.; Verstockt, S.; Bourgois, J.G. Machine Learning-Based Identification of the Strongest Predictive Variables of Winning and Losing in Belgian Professional Soccer. *Appl. Sci.* **2021**, *11*, 2378. [\[CrossRef\]](#)
27. Chan, T.M.; Li, Y.; Chiau, C.C.; Zhu, J.; Jiang, J.; Huo, Y. Imbalanced target prediction with pattern discovery on clinical data repositories. *BMC Med. Inform. Decis. Mak.* **2017**, *17*, 47. [\[CrossRef\]](#)
28. Circadian Rhythms. Available online: <https://nigms.nih.gov/education/fact-sheets/Pages/circadian-rhythms.aspx> (accessed on 1 July 2022).
29. Ruben, M.D.; Hogenesch, J.B.; Smith, D.F. Sleep and circadian medicine: Time of day in the neurologic clinic. *Neurol. Clin.* **2019**, *37*, 615–629. [\[CrossRef\]](#)
30. Ruben, M.D.; Smith, D.F.; FitzGerald, G.A.; Hogenesch, J.B. Dosing time matters. *Science* **2019**, *365*, 547–549. [\[CrossRef\]](#)
31. Caruana, R.; Lou, Y.; Gehrke, J.; Koch, P.; Sturm, M.; Elhadad, N. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; pp. 1721–1730.
32. Wiens, J.; Shenoy, E.S. Machine learning for healthcare: On the verge of a major shift in healthcare epidemiology. *Clin. Infect. Dis.* **2018**, *66*, 149–153. [\[CrossRef\]](#)
33. Spiegel, K.; Sheridan, J.F.; Van Cauter, E. Effect of sleep deprivation on response to immunization. *JAMA* **2002**, *288*, 1471–1472. [\[CrossRef\]](#)
34. Lange, T.; Dimitrov, S.; Bollinger, T.; Dieckmann, S.; Born, J. Sleep after vaccination boosts immunological memory. *J. Immunol.* **2011**, *187*, 283–290. [\[CrossRef\]](#)

-
35. Prather, A.A.; Hall, M.; Fury, J.M.; Ross, D.C.; Muldoon, M.F.; Cohen, S.; Marsland, A.L. Sleep and antibody response to hepatitis B vaccination. *Sleep* **2012**, *35*, 1063–1069. [[CrossRef](#)] [[PubMed](#)]
 36. Lange, T.; Perras, B.; Fehm, H.L.; Born, J. Sleep enhances the human antibody response to hepatitis A vaccination. *Psychosom. Med.* **2003**, *65*, 831–835. [[CrossRef](#)] [[PubMed](#)]