

Risk Factors and Predictive Modeling of Adverse Events Following COVID-19 Booster Vaccination: A Prospective Cohort Study

Shahinur Islam¹, Bipul Nath^{1*}, Subhashis Debnath²

¹Department of Pharmacy, Royal School of Pharmacy, The Assam Royal Global University, Betkuchi, Guwahati, Assam, INDIA.

²Department of Pharmacy, Bharat Pharmaceutical Technology, Amtali, Agartala, Tripura, INDIA.

ABSTRACT

Background: COVID-19 booster vaccination is essential for maintaining protection against severe disease, particularly among older adults; however, substantial interindividual variability exists in post-booster adverse-event responses. **Objectives:** This study aimed to evaluate demographic, clinical, infection-related, and vaccination-associated factors influencing the post-booster adverse-event burden among adults aged ≥ 60 years and develop a predictive score. **Materials and Methods:** A prospective observational cohort study was conducted with 500 recruited adults aged ≥ 60 years, of whom 472 (94.4%) completed follow-up. Adverse events were recorded using a structured follow-up form and analyzed via a composite adverse-event score. Univariable and multivariable linear regression analyses were used to identify independent predictors, and an internally validated, simplified 0–5-point prediction score was constructed. **Results:** The most commonly reported adverse events were injection-site pain (60.6%), fatigue (51.1%), and headache (41.9%). Previous vaccine reaction (standardized $\beta=0.25$), female sex, previous COVID-19 infection, and shorter booster intervals were independently associated with a higher adverse-event burden. The prediction model explained 34% of outcome variability (development RMSE=1.12), and internal validation demonstrated an RMSE of 1.18 with 31% of outcome variability explained, yielding progressive risk stratification across the point-based score. **Conclusion:** The model provides a pragmatic exploratory tool to identify older adults at elevated risk of post-booster adverse events for tailored counseling and monitoring, although external validation in larger, independent cohorts is necessary before clinical implementation.

Keywords: Adverse events, COVID-19 booster, Internal validation, Older adults, Prediction model, Risk factors, Risk stratification.

Correspondence:

Dr. Bipul Nath

Professor, Department of Pharmacy, Royal School of Pharmacy, The Assam Royal Global University, Guwahati-781035, Assam, INDIA.
Email: bipulnath@gmail.com

Received: 21-08-2026;

Revised: 14-09-2026;

Accepted: 05-10-2026.

INTRODUCTION

COVID-19 vaccines have been developed and widely administered worldwide to minimize the risk of severe SARS-CoV-2 infection, hospitalization, and mortality. Nevertheless, the continuous emergence of novel viral variants and the gradual decline in vaccine-induced immunity following completion of the primary vaccination schedule have highlighted the need for booster doses to sustain effective and long-term protection, particularly among older adults and individuals with underlying health conditions (Feikin *et al.*, 2022; Nanishi *et al.*, 2022). Accordingly, booster vaccination has become an important component of the ongoing COVID-19 prevention

strategy. Although booster doses are generally considered safe and well tolerated, Adverse Events Following Immunization (AEFIs), including injection-site pain, fatigue, headache, myalgia, fever, and malaise, remain commonly reported (Sutton *et al.*, 2022; Warkentin *et al.*, 2023).

The incidence and extent of adverse events following vaccination among individuals can be quite different. Variability that may be due, in part, to heterogeneous demographic features; history of prior SARS-CoV-2 infection; history of prior vaccination; comorbidities and intermediate measures of immunity such as T-cell counts and levels between vaccine doses (Sutton *et al.*, 2022; Warkentin *et al.*, 2023; Mori *et al.*, 2023). Differences in the demographic characteristics (e.g., age) of populations can also contribute to disparities in reactogenicity profiles and older adults have previously been found to experience different profile of vaccine-associated AEFIs as compared with younger population (Nanishi *et al.*, 2022; Mori *et al.*, 2023). Also, people with prior SARS-CoV-2 infection or attenuated when vaccinated after receiving earlier mRNA doses may undergo



DOI: 10.5530/ijopp.20270022

Copyright Information :

Copyright Author (s) 2027 Distributed under
Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia. [www.mstechnomedia.com]

immune-modification and reactogenicity (Warkentin *et al.*, 2023; Bardenheier *et al.*, 2021). These interindividual differences in the immune response and vaccine efficacy have implications for more effective vaccine counseling, clinical monitoring, and personalized risk communication.

The majority of vaccine safety studies of COVID-19 have primarily reported on the frequency, incidence or severity of single adverse events individually (Sutton *et al.*, 2022; Warkentin *et al.*, 2023; Ahamad *et al.*, 2023). While these approaches yield valuable insights into vaccine reactogenicity, they may not faithfully capture the spectrum of variability in adverse-event responses within a vaccinated cohort. Statistical variability in an outcome quantified by statistics such as the Standard Deviation of the mean (SD) offers further insights about the extent of heterogeneity in individual responses. Hence, determining which factors are associated with greater variance offers a potential complement to usual analyses based only on means of adverse-event scores or frequencies of adverse events.

The opportunities afforded by predictive modeling enable the formal translation of individual epidemiological and clinical risk factors into an individualized estimate of adverse-event variability. The usefulness of both statistical (Ahamad *et al.*, 2023) and machine-learning techniques for detecting and categorizing adverse effects related to COVID-19 vaccines have been illustrated before (Ahamad *et al.*, 2023; Bardenheier *et al.*, 2021). A valid prediction model might help in identifying individuals who are more likely to show greater heterogeneity of post booster adverse-event responses and would facilitate more tailored counseling and post-vaccination follow up. This approach is especially relevant in older populations, where prior infection, vaccination history, and/or immune senescence may each be involved in large variability between individuals for vaccine reactogenicity (Nanishi *et al.*, 2022; Warkentin *et al.*, 2023; Mori *et al.*, 2023).

While there is ample literature on the safety profile of COVID-19 vaccines, the potential for prediction of variability in adverse-event outcomes following booster vaccination has been examined much less. Also, few studies have aimed at constructing and validating a feasible prediction tool combining several demographic and vaccination-specific variables to estimate the systemic variation of post-booster adverse-event reactions. Previous work has mostly examined the occurrence, frequency and severity of specific adverse events or overall reactogenicity in this population rather than developing tailored models of heterogeneity in composite adverse-event responses (Sutton *et al.*, 2022; Warkentin *et al.*, 2023; Ahamad *et al.*, 2023). Closing this knowledge gap may enhance understanding of factors that drive heterogeneous immunogenicity to vaccination and the provision of a clinically intuitive framework for risk stratification.

Based on this, the current prospective cohort study was conducted to determine demographic, clinical (presence of chronic co-morbidities), infection-related and Vaccination-related predictors for higher variability in adverse-event response following COVID-19 booster vaccination. Design also aimed to derive and internally validate a prediction model for predicting adverse-event variability, with translation of the identifiable predictors into a novel clinical risk-stratification tool. This strategy could potentially offer another evidence-informed framework to help guide personalized patient evaluation and counseling of COVID-19 booster vaccination.

Study Design and Setting

An ongoing open-label prospective observational cohort study was performed on individuals aged ≥ 60 years of age who received a COVID-19 booster vaccine. Participants were recruited before or at the time of booster vaccination and were prospectively followed over a predefined post-vaccination observation period (Collins *et al.*, 2024). The study aimed to evaluate demographic, clinical, infection-related, and vaccination-related factors associated with variation in post-booster adverse-event responses and to develop an exploratory prediction model for identifying individuals likely at greater predicted burden of adverse events. The prediction model was designed, developed, validated, and reported according to contemporary recommendations for clinical prediction models (Heus *et al.*, 2020; Najafabadi *et al.*, 2020; Collins *et al.*, 2024).

Study Population

The study included adults aged ≥ 60 years who were eligible for and received COVID-19 booster dose. Inclusion criteria for participants occurred if they had:

- Received the recommended primary series of COVID-19 vaccines

A booster dose was received during the study period.

- Provided written informed consent.
- Agreed to get the follow-up assessments done according to the schedule.

However, individuals with incomplete vaccination records; an acute illness at enrolment that could interfere with assessment of post-vaccination symptoms (e.g., symptomatic SARS-CoV-2 infection); an inability to complete the required follow-up period or voluntarily withdrew consent were not included in the analysis.

Various Sampling Size and Follow Up of Participants

We did a prospective cohort study with the aim of evaluating post-booster adverse events and possible demographic, clinical, and vaccination-related factors linked with the study outcome among older adults who received a COVID-19 booster dose. Screener was initially applied to 500 elderly with inclusion and

exclusion criteria. Sample sizes were planned according to the anticipated prevalence of main risk factors, expected effect size for each predictor, number of candidate predictors and development/internal validation requirements for the multivariable prediction model. Methodological recommendations guided sample-size considerations for the development and validation of prediction models (Riley *et al.*, 2021; Riley *et al.*, 2020).

Written informed consent was provided after all participants were informed about the objectives and procedures of the study, prior to enrolment. Study design: Participants were followed up prospectively after administering the COVID-19 booster dose.

A structured case-record form was used to gather demographic, clinical, infection-related and vaccination-related information at baseline. Variables were registered: age, sex, previous COVID-19 infection history, diabetes mellitus, hypertension history, previous significant vaccine-related reaction and receiving at least two doses of the previous COVID-19 vaccination. Data on previous COVID-19 infection and reaction to any vaccinations were recorded as participant-reported history complemented by either available medical or vaccination records, where possible.

The COVID-19 booster dose was administered to all enrolled participants in accordance with the relevant vaccination schedule. Enrollment of participants was conducted prospectively over the pre-defined post-booster observation period. Participants were evaluated for the incidence and severity of a series of local and systemic adverse events at follow-up, in accordance with standard outcomes for reactogenicity following vaccination against COVID-19 (Sutton *et al.*, 2022; Warkentin *et al.*, 2023; Mori *et al.*, 2023; Sutton *et al.*, 2024). Only subjects who completed the planned follow-up assessments were included in final analyses.

Data Collection

A structured case-record form was used to collect baseline information. Data were collected from participant interviews, clinical records and documentation of vaccinations (where relevant).

The following variables were recorded:

Demographic characteristics

- Age.
- Sex.

Clinical characteristics

- History of previous COVID-19 infection.
- Diabetes mellitus.
- Hypertension.
- Other relevant comorbidities, where available.

Vaccination-related characteristics

- Previous COVID-19 vaccine doses received.
- History of adverse reaction to previous COVID-19 vaccine.
- Interval from the last previous dose and booster dose.

After booster vaccination, participants were evaluated by a median of days at predefined follow-up intervals for the occurrence and severity of solicited adverse events. The local and systemic adverse events reported in the study were consistent with those commonly described in reactogenicity studies of COVID-19 vaccines (Sutton *et al.*, 2022; Warkentin *et al.*, 2023; Mori *et al.*, 2023; Sutton *et al.*, 2024).

Assessment and Frequency of Adverse Events Post-Booster Dose

Adverse events following booster vaccination reported through a structured follow-up form were assessed among the 472 participants who completed the scheduled post-booster follow-up. Assessment of local and systemic adverse events occurring during the post-vaccination observation period.

Local adverse events were defined as those that occurred at the injection site, including injection site pain, injection-site swelling, and injection-site redness; systemic adverse events included fatigue, headache, myalgia, fever/chills/arthritis/nausea. Multiple adverse events could be reported by a participant and were recorded individually. These adverse-event domains are similar to those assessed in COVID-19 vaccine reactogenicity assessments and systematic reviews (Sutton *et al.*, 2022; Warkentin *et al.*, 2023; Mori *et al.*, 2023; Sutton *et al.*, 2024).

Each event was documented according to the predefined study assessment criteria for presence and severity of each adverse event. Finally, these individual adverse-event responses were aggregated to form a composite post-booster adverse-event score, which served as the primary outcome measure for the subsequent statistical and predictive analyses.

The frequency of each AE was reported as the number and percentage of subjects with an event [n (%)]. Denominator was the number of participants who completed follow-up ($n=472$): The percentage.

$$\text{Adverse - event frequency (\%)} = 100 \times \frac{\text{Number of participants with the event}}{n = 472}$$

We used descriptive statistics to summarize the distribution of individual adverse events. As participants could experience more than one adverse event, the frequencies of individual events were calculated independently and thus are not anticipated to result in summation to 100%. The results were reported as n (%) in the table and appeared graphically on the x-axis plane (horizontal bar chart), sorted in descending order. We then conducted univariable and multivariable analyses of the summary composite

adverse-event score to determine factors associated with variation in post-booster adverse-event responses.

Study Outcome and Outcome Variability

The main endpoint was the composite adverse-event score following booster vaccination, calculated based on predefined local and systemic adverse events reported during follow-up. It was summarized with mean $\pm$ SD corresponding to interindividual variability, while the observed adverse-event burden.

Outcome score was compared by sex, previous COVID-19 infection (yes/no), previous vaccine reaction (yes/no), age group (60-65 years, 66-70 years and ≥ 70 years) and booster-dose interval. For two-group and three-group comparisons, independent-samples *t*-test or one-way ANOVA was conducted respectively along with appropriate non-parametric tests if parametric assumptions were not met. Statistical significance was defined as a two-sided *p*-value < 0.05 . We performed these exploratory assessments to provide some preliminary information about assets before regression model fitting.

Univariable Risk-Factor Analysis

We conducted univariable linear regression to individually examine the association of composite adverse-event score with age, sex, history of SARS-CoV-2 infection prior to vaccination, past reaction after previous vaccine doses, short interval between booster and second COVID-19 vaccines, diabetes and hypertension. This resulted in estimates for each predictor, the corresponding β coefficient, 95% Confidence Interval (CI), and *p*-value. All variables with *p* < 0.05 or the clinical relevance were included in multivariable analysis. The methods used for prediction modeling and transparent writing were in accordance with current recommendations regarding clinical prediction-modeling research (Collins *et al.*, 2024; Heus *et al.*, 2020; Najafabadi *et al.*, 2020; Collins *et al.*, 2024).

Multivariable Risk-Factor Analysis

A multivariable linear regression model identified independent predictors of the composite adverse-event score after adjustment for potential confounders. The model included age, sex, previous history of COVID-19 infection, previous adverse vaccine reaction, interval to booster-dose acquisition, diabetes and hypertension. β , standardised β (95% CI and *p* values) were estimated where *p* < 0.05 was considered statistically significant. Hence, a regression-based methodology was chosen because the primary outcome was a continuous composite adverse-event score. Current TRIPOD+AI reporting recommendations cover regressionbased clinical prediction models (Collins *et al.*, 2024; Collins *et al.*, 2024). Subsequently, significant independent predictors were used for the development of the simplified clinical prediction score.

Assessment of Regression Model Assumptions

All the assumptions of considered multivariable linear regression model were checked before interpreting the results. We assessed linearity, independence of observations, normality and homoscedasticity of residuals, and multicollinearity among predictors.

The Variance Inflation Factor (VIF) and tolerance values were used to assess multicollinearity. Diagnostics plots (residual and diagnostic) were checked for linearity, normality, homoscedasticity and model adequacy. Objective: In prediction-model development, it is important to avoid the risk of model instability and overfitting by assessing model assumptions and predictive performance (Martin *et al.*, 2021).

Construction of Prediction Score and Relative Contribution of Predictors

This simplified clinical prediction score was derived from independent multivariable regression model predictors. Regression coefficients of clinical relevance were allocated weighted points for predictor variables. Items were then summed to yield an overall prediction score, where higher scores indicated a predicted post-booster adverse-event burden.

Participants were then stratified into low-, intermediate-, and high-risk cohorts based on predetermined score ranges. These categories must be interpreted as predicted adverse-event burden categories (instead of probabilities of a binary clinical event), because the modeled outcome was a continuous adverse-event burden score.

Standardized β coefficients from the final multivariable regression model were used to rank independent predictors. Since standardized β coefficients are unitless, they were compared in absolute magnitude to estimate relative standardized association with the outcome. We ordered the predictors on the basis of their absolute β standardized value, with higher values indicating greater relative association. The results were summarized in a

Table 1: Baseline demographic and clinical characteristics.

Variable	Participants, <i>n</i> (%)
Total recruited	500
Completed follow-up	472 (94.4)
Mean age, years	>60
Female	270 (54.0)
Male	230 (46.0)
Previous COVID-19 infection	152 (30.4)
Diabetes mellitus	54 (10.8)
Hypertension	86 (17.2)
Previous significant vaccine reaction	71 (14.2)
≥ 2 previous COVID-19 vaccine doses	500 (100)

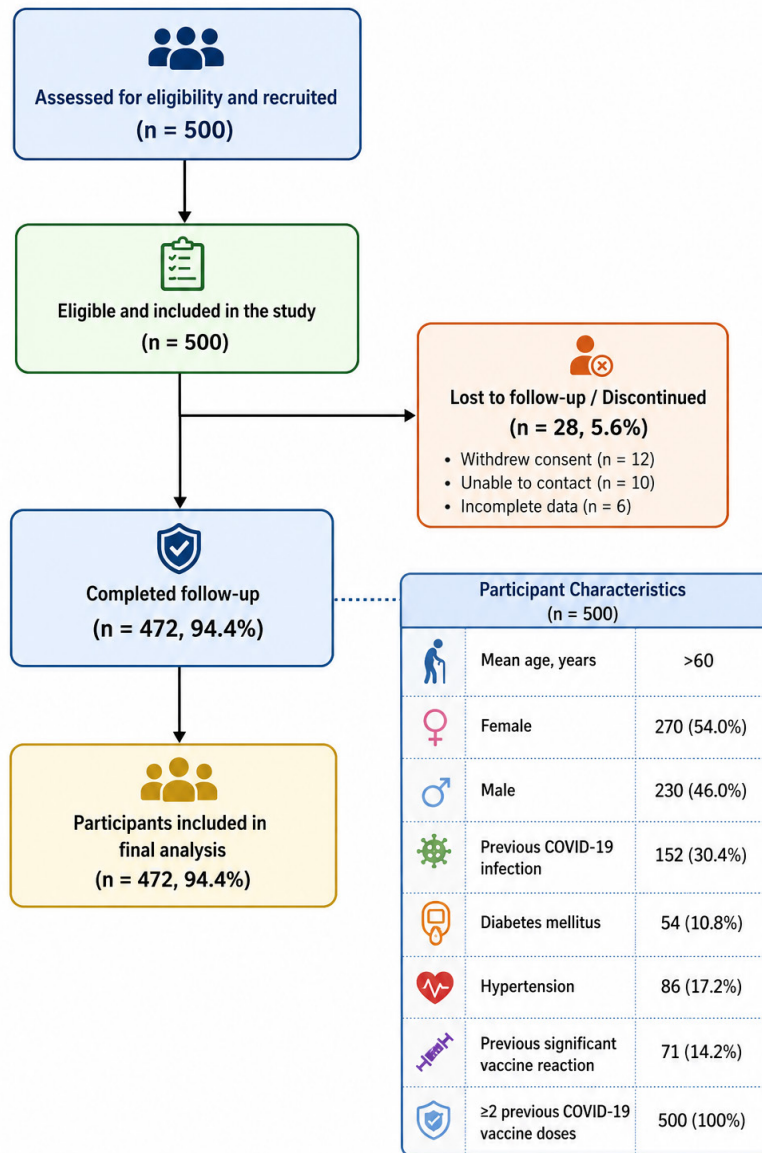


Figure 1: Participant recruitment and follow-up flow diagram.

table and illustrated with horizontal bar diagrams (Figure 5) from highest to lowest magnitude.

Performance of prediction models and internal validation

The metrics used to assess the efficacy of prediction model were R^2 , adjusted R^2 , RMSE, and MAE. R^2 was the proportion of variance in the outcome explained by the model and adjusted R^2 reflected update number of predictors included. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) measured prediction accuracy where lower the value the better. The explained variability is presented as a percentage (R^2). Model performance assessment and internal validation was based on current guidelines for prediction-model development and

validation (Collins *et al.*, 2024; Riley *et al.*, 2021; Riley *et al.*, 2020; Martin *et al.*, 2021).

Model calibration was evaluated by comparing observed and predicted outcome scores across prediction ranges. Model validity, stability and risk of overfitting were assessed using internal validation methods with bootstrap resampling (10,000 samples) or k-fold cross-validation. Estimating optimism and variability in predictive performance estimates is probably the most common area where we rely on bootstrap-based assessment methods, as discussed above (Martin *et al.*, 2021).

Performance metrics evaluated in the development and internal validation datasets were compared to assess temporal stability and possible performance degradation of models. Development versus validation performance was assessed as a potential measure of overfitting.

The performance estimates were given in Table 6. To estimate the reliability and stability of prediction model, we examined to what extent development performance mattered from validation performance (Collins *et al.*, 2024; Martin *et al.*, 2021).

Observed-versus-Predicted Outcome Scores

Graphical assessment of agreement between observed outcome scores and those predicted by the clinical prediction model. The observed composite adverse-event score was compared with the predicted score based on the proposed scoring system for each participant.

The observed score over predicted score was plotted on the X axis and Y axis with 45° reference line showing perfect agreement (observed = predicted). To examine how many of the observations were above and below the reference line, and where on that continuum they fell.

The distribution of observations being more clustered around the 45° reference line indicated better agreement in predicted and observed outcomes. This graphical assessment supplemented the quantitative measures of R², RMSE, MAE and calibration that

were used to evaluate overall model performance. Clear reporting of observed and predicted outcomes is in line with good reporting practice for prediction models (Collins *et al.*, 2024; Heus *et al.*, 2020; Najafabadi *et al.*, 2020).

Ethical Considerations

The study was performed according to the principles outlined in the Declaration of Helsinki, and with prior approval from the Institutional human ethics committee of Assam Royal Global University. Informed written consent was obtained from all participants prior to their inclusion in the study.

Confidentiality of study participants was maintained. A unique study identification code was assigned to each participant, and all personal identifying information remained secure from unauthorized disclosure. The data collected were solely used for research purposes only and analyzed in an anonymous or coded form.

RESULTS AND DISCUSSION

Methods Participant recruitment, follow-up and baseline characterization

500 subjects aged ≥60 years were recruited for the prospective cohort study. Of these participants, 28 were lost to follow-up (i.e. did not complete the predefined follow-up) leaving a total of 472 participants (94.4%) with data for analysis (Figure 1). The high follow-up completion rate suggests excellent retention of participants and provided sufficient numbers for further evaluation of booster associated adverse events and their risk factors.

Among these included 270 (54.0%) females and 230 (46.0%) males, which showed a balanced sex distribution pattern with a slight female predominance. As far as clinical history is concerned, 152 participants (30.4%) reported previous COVID-19 infection. Diabetes mellitus was reported in 54

Table 2: Frequency of post-booster adverse events among participants completing follow-up.

Adverse event	n	Percentage (%)
Injection-site pain	286	60.6
Fatigue	241	51.1
Headache	198	41.9
Myalgia	176	37.3
Fever	128	27.1
Chills	96	20.3
Arthralgia	84	17.8
Nausea	62	13.1
Injection-site swelling	58	12.3
Injection-site redness	43	9.1

Table 3: Comparison of outcome scores according to selected risk factors.

Risk factor	Group	Mean outcome score±SD	p-value
Sex	Male	3.42±1.01	0.002
	Female	4.01±1.18	
Previous COVID-19 infection	No	3.51±1.04	0.008
	Yes	4.02±1.21	
Previous vaccine reaction	No	3.49±1.03	<0.001
	Yes	4.35±1.28	
Age	60-65 years	3.62±1.08	<0.001
	66-70 years	3.88±1.12	
	≥70 years	4.16±1.19	
Booster-dose interval	<6 months	4.02±1.20	0.015
	≥6 months	3.58±1.05	

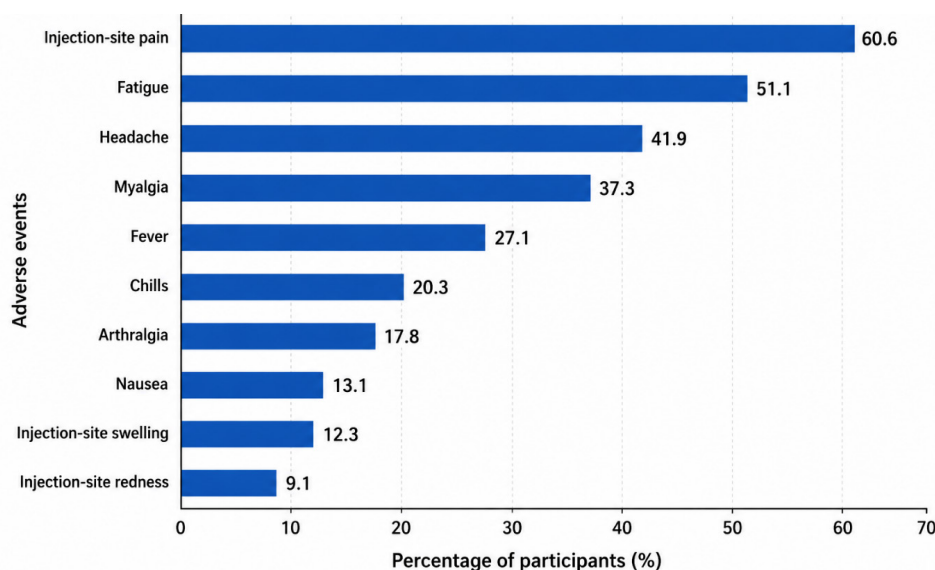


Figure 2: Frequency of post-booster adverse events.

(10.8%) subjects, hypertension in 86 patients (17.2%). Previous notable vaccine reactions were reported by 71 participants (14.2%). Including the vaccinations received in the past 12 months prior to randomization and post-randomization up to study day 14, all 500 participants (100%) had previously received at least two doses of COVID-19 vaccine, confirming an adequate vaccination history for outcome evaluation post-booster. Summary of Baseline Demographic and Clinical Characteristics are displayed in Table 1. The variation seen in sex, previous COVID-19 infection, comorbidities and prior vaccine reactions thus offered a potential utility for determining their association with the burden of post-booster adverse-events. Univariable and multivariable regression analyses were then performed to assess these variables along with other a priori identified covariates.

The proportion of participants with a history of strong vaccine reactions was especially interesting at 14.2% of our sample, since previous responses to vaccines may identify individuals most susceptible to adverse events as the result of subsequent booster vaccination. Previous COVID-19 infection was another clinically important candidate predictor. After multivariable analyses to adjust for other potential predictors, these factors were found to have independent contributions.

In summary, the high completion rate for 12-month follow-up and comprehensive demographic, clinical, and vaccination-history data available provided a solid basis from which risk-factor analysis, multivariable regression, prediction-model generation and risk classification could be undertaken. An overview of the participant recruitment and follow-up process is summarized as: 500 assessed for eligibility → 28 lost to follow-up → 472 participants completed follow up and analysed (94.4%).

Frequency of Post-Booster Adverse Events

Table 2 shows the frequency of post-booster adverse events among the 472 participants who completed follow-up. Participants were permitted to report multiple adverse events and as such, percentages represent the proportion of patients reporting each event rather than being expected to sum to 100%.

The most common adverse event was pain at the site of injection, which occurred in 286 participants (60.6%). After that, fatigue (241; 51.1%) and headache (198; 41.9%); Myalgia was the most frequently reported systemic adverse event ($n=176$; 37.3%), with other frequent events being fever ($n=128$; 27.1%), chills ($n=96$; 20.3%), and arthralgia ($n=84$; 17.8%).

Nausea (62; 13.1%), injection site swelling (58; 12.3%), and injection site redness (43; 9.1%) were less-reported events. In terms of local reactions, pain at the injection site was significantly more common than swelling or redness.

Overall, local pain and common systemic symptoms including fatigue and headache were the majority of post-booster adverse events observed in our cohort. An emerging pattern indicates that most reactions reported were frequent local or systemic events after booster vaccination. These observations set out the descriptive groundwork for using demographic information, prior COVID-19 infection history, prior vaccine reactions, booster-dose interval and clinical comorbidities in future analyses to determine whether they are associated with differences in post-booster adverse-event burden.

Figure 2 uses a horizontal bar chart to show the frequency distribution of adverse events, which include injection-site discomfort, fatigue, headache, myalgia, fever, chills, arthralgia, nausea, injection-site edema, and injection-site redness. The events are displayed in descending order of frequency.

Outcomes stratified by each major risk factor

Regression analysis was then conducted with the selected demographic and clinical risk factors. A composite score of adverse events attributed to each vaccine candidate post-booster was computed per study participant, and scores were compared across demographic and clinical stratifications. Table 3 shows the results and Figure 3 presents it graphically. Differences in the outcome were summarised across risk-factor groups according to mean score $\pm$ SD.

The mean outcome score was 4.01 ± 1.18 in females and 3.42 ± 1.01 in males ($p=0.002$). Likewise, participants with previous COVID-19 infection had a higher mean score compared to individuals without previous infection (4.02 ± 1.21 vs. 3.51 ± 1.04 ; $p=0.008$). The greatest difference was seen by prior vaccine reaction. The mean outcome score was significantly higher among participants with a history of previous vaccine reaction than among those without (4.35 ± 1.28 vs. 3.49 ± 1.03 ; $p<0.001$).

An increasing trend in mean outcome score was also observed across age groups. Among participants aged 60-65 years, the mean score was 3.62 ± 1.08 ; this increased to 3.88 ± 1.12 for subjects aged 66-70 years and further increased to 4.16 ± 1.19 for those aged ≥ 70 ($p<0.001$). The mean outcome score was greater in participants who received the booster within <6 months than for those whose interval was ≥ 6 months (4.02 ± 1.20 vs 3.58 ± 1.05 ; $p=0.015$). Overall, all these findings suggest that composite outcome score is significantly different depending on sex and previous COVID-19 infection of participants, previous vaccine

reaction expertise of the participants, their age group size as well as booster-dose interval. Previous vaccine reaction showed the greatest difference in mean score (0.70 points) and was selected as an important candidate predictor for subsequent regression analysis among these factors.

Univariable Risk-Factor Analysis

A univariable linear regression was performed to examine the demographic and clinical characteristics associated with the composite post-booster adverse-event score. The summary of the results is displayed in Table 4. Positive independent predictors of a higher outcome score included female sex ($\beta=0.24$, 95% CI: 0.09-0.39; $p=0.002$). A positive correlation was also found for previous COVID-19 infection ($\beta=0.19$; 95% CI: 0.05-0.33; $p=0.008$). The highest association was for previous vaccine reaction ($\beta=0.31$, 95% CI: 0.17-0.45; $p=0.05$); there was no significant relationship with the outcome. Likewise, changes in diabetes mellitus ($\beta=0.07$, 95% CI: -0.04 to 0.18; $p=0.214$) or hypertension ($\beta=0.05$, 95% CI: -0.04 to 0.14; $p=0.301$) were not statistically significant either with the differences in health utilities over time.

As a result, female sex, prior COVID-19 infection, previous vaccine reaction and shorter booster-dose interval were recognized as candidate predictors from the univariable analysis. These variables were then tested in the multivariable model to see whether their associations remained significant after adjusting for potential confounders.

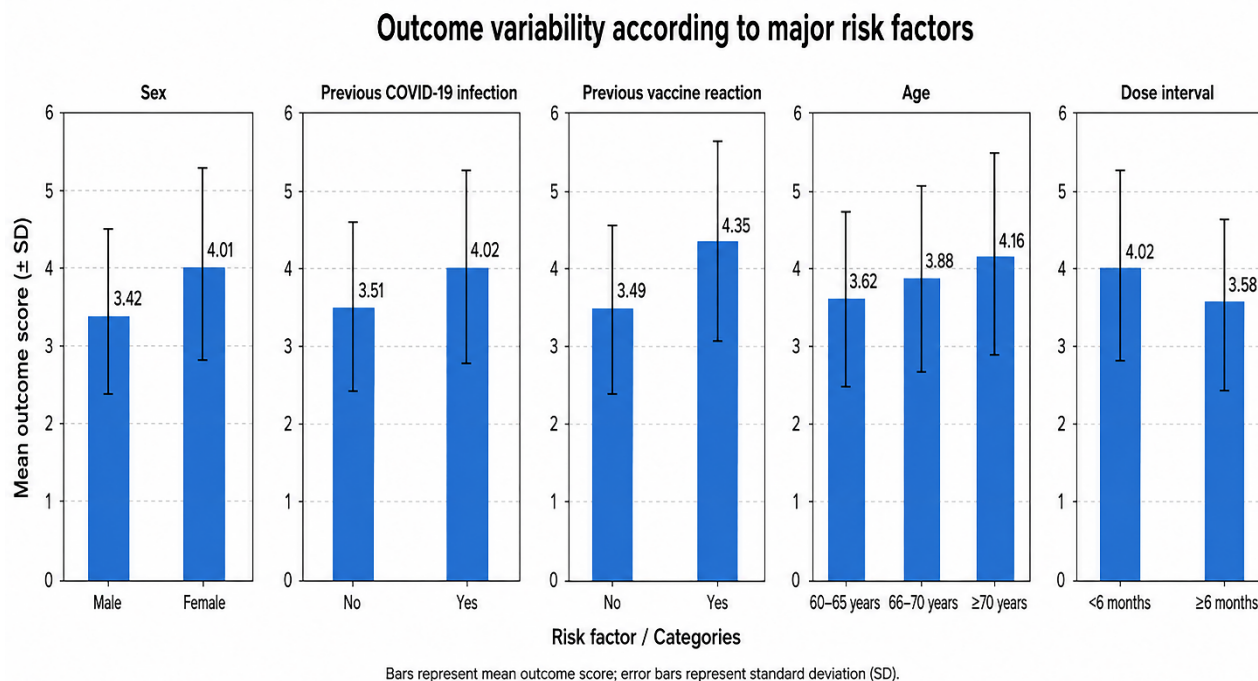


Figure 3: Outcome variability according to major risk factors. Bars represent mean outcome scores and error bars represent Standard Deviation (SD).

Table 4: Univariable analysis of potential risk factors.

Variable	β	95% CI	p-value	Interpretation
Age	0.00	-0.00 to 0.00	>0.05	Not significant
Female sex	0.24	0.09-0.39	0.002	Significant
Previous COVID-19 infection	0.19	0.05-0.33	0.008	Significant
Previous vaccine reaction	0.31	0.17-0.45	<0.001	Strongest predictor
Short booster-dose interval	0.16	0.03-0.29	0.015	Significant
Diabetes mellitus	0.07	-0.07 to 0.21	0.214	Not significant
Hypertension	0.05	-0.05 to 0.15	0.301	Not significant

Table 5: Multivariable regression analysis.

Predictor	β	SE	Standardized β	p-value
Intercept	2.31	0.28	-	<0.001
Age	0.002	0.018	0.01	0.912
Female sex	0.42	0.14	0.18	0.003
Previous COVID-19 infection	0.35	0.13	0.15	0.008
Previous vaccine reaction	0.68	0.16	0.25	<0.001
Shorter booster-dose interval	0.29	0.12	0.12	0.016
Diabetes mellitus	0.11	0.09	0.06	0.221
Hypertension	0.08	0.08	0.05	0.318

β , regression coefficient; SE, standard error.

Multivariable Risk-Factor Analysis

Multivariable linear regression analysis was used to assess the independent contribution of demographic and clinical factors on the set of composite post-booster adverse-event score after adjusting for other variables. The results are shown in Table 5. Chronic Female sex was the only independent predictor of outcome- ($\beta=0.42$, SE=0.14, standardized $\beta=0.18$; $p=0.003$). In the same way, prior COVID-19 infection was significantly and independently associated with a higher outcome score ($\beta=0.35$, SE=0.13, standardized $\beta=0.15$; $p=0.008$).

Previous vaccine reaction was the strongest independent predictor ($\beta=0.68$, SE=0.16, standardized $\beta=0.25$; $p<0.001$). The standardized β coefficient was the highest among all tested predictors (multivariable model), indicating that the relative contribution of this predictor to be maximal. The shorter interval for booster-dose remained independently associated with the outcome ($\beta=0.29$, SE=0.12, standardized $\beta=0.12$; $p=0.016$), but its relative contribution in COVID-19-no-results was smaller than that of previous vaccine reaction, female sex and previous COVID-19 infection. In contrast, age had no significant independent effect on the outcome ($\beta=0.002$, SE=0.018, standardized $\beta=0.01$; $p=0.912$). Similarly, diabetes mellitus ($\beta=0.11$, SE=0.09, standardized $\beta=0.06$; $p=0.221$) and hypertension ($\beta=0.08$, SE=0.08, standardized $\beta=0.05$; $p=0.318$) were not statistically significant either after adjustment as well.

In summary, multivariable analysis revealed that history of past vaccine reaction, female sex, history of SARS-CoV-2 infection prior to vaccination as well as shorter interval from the booster-dose were independent predictors for increased composite post-booster adverse-event score. Of these factors, prior vaccine reaction had the largest relative contribution as indicated by the highest standardized β coefficient.

The independent predictors were used to then construct a simple prediction score, making these findings actionable. The independent association of age, diabetes and hypertension status in this study suggests that these variables played a relatively minor role in the final adjusted model for the overall cohort.

In the sequential analyses, previous vaccine reaction, female sex, previous COVID-19 infection and shorter booster-dose interval were all found to be positively correlated with higher post-booster adverse-event scores. Whilst the age-group analysis described resulted in weirdly consistently better outcome scores with the increasing age, and certainly seemed to indicate that older patients fared better (an observation supported by other reports), it transpired that modelling age as a continuous predictor was not appropriate since neither univariable or adjusted regression analyses found any significant association. This is a reminder to perform multivariable modeling to determine independent predictors.

On univariable analysis, history of vaccine reaction was the strongest predictor (highest β) and among all evaluated factors it

had the highest standardized β when considered in a multivariable context. This evidence statistics support the inclusion of past vaccine reaction, individual sex, prior infection status and interval for doses in this next clinical score/risk classification tool.

Relative Importance of Independent Predictors

The standardized β coefficients from the final multivariable regression model were used to establish the relative contribution of each independent predictor. Results are displayed in Table 6 and Figure 4.

The greatest relative contribution to the composite post-booster adverse-event score was observed for previous vaccine reaction, with the highest standardized β coefficient ($\beta=0.25$). After this, female sex ($\beta=0.18$), preceding COVID-19 (e.g., infection) ($\beta=0.15$), and shorter booster-dose interval between two doses compared with the second dose of vaccine more often to induce CD4+ T cell responses ($\beta=0.12$). Age had a very small independent effect (standardized $\beta=0.01$) and did not add substantively to the model.

It was found that the levels of predictors, rating were as follows:

Predictors of the intensity of local pain (0.25): Prior vaccine reaction > Female sex (0.18) > Previous COVID-19 infection (0.15) > Shorter dose interval (0.12) > Age (0.01). Highly previous vaccine reaction is advocated as a potential predictor of individuals who would better identify those participants with greater post-booster adverse-event burden. Female sex was the next strongest predictor, and this complements its statistically significant outcome in multivariable models. A recent prior infection with COVID-19 also showed a moderate independent effect, indicating that previous SARS-CoV-2 exposure may be important in evaluating post-booster responses.

The contribution of a shorter boost-dose interval, while smaller in magnitude than previous vaccine reaction, female sex, and previous infection, was found to be statistically significantly associated with seropositivity suggesting the vaccination interval may provide greater predictive utility as a risk factor when considered binned with others.

On the other hand, age was relatively of no importance (standardized $\beta=0.01$; $p=0.912$). Despite the differences between age categories identified by descriptive analysis, chronological age alone did not independently enhance prediction ability after adjusting for other covariates. These findings imply that, even in this older cohort, age is a less informative predictor than vaccination history and selected demographic characteristics.

In summary, the standardized coefficients show that the most important predictor in the final model was previous vaccine reaction and age played little role. Such results offer a quantitative rationale to emphasize prior vaccine reaction, sex, history of a past COVID-19 infection, and booster-dose interval in the next established clinical prediction score. Standardized β coefficients should not to be interpreted as risk percentage or other probabilities, but rather represent the relative contributions of predictors in the regression model.

Table 6: Standardized β coefficients from the multivariable regression model.

Predictor	Standardized β
Previous vaccine reaction	0.25
Female sex	0.18
Previous COVID-19 infection	0.15
Shorter dose interval	0.12
Age	0.01

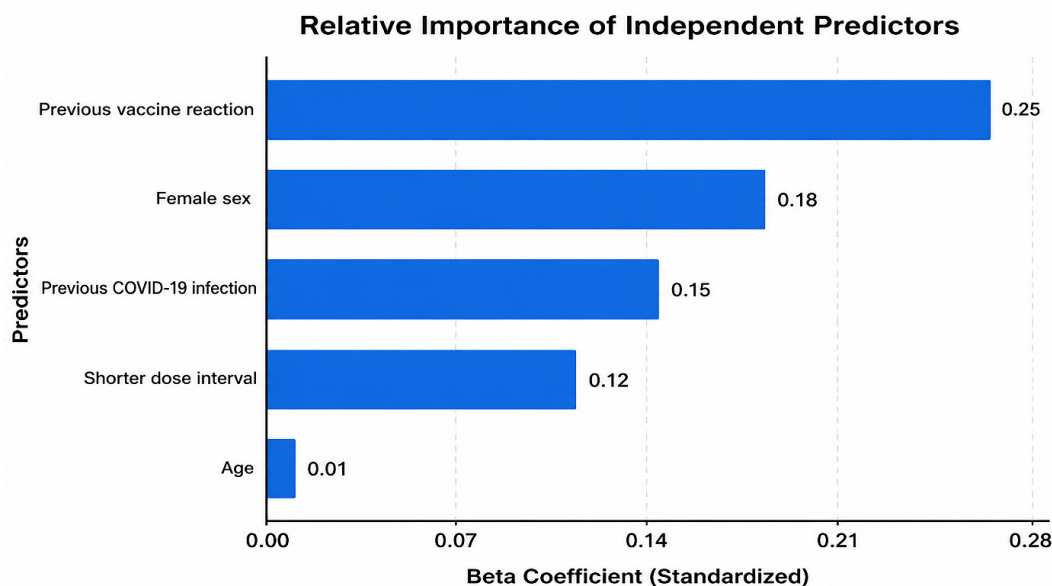


Figure 4: Relative importance of independent predictors.

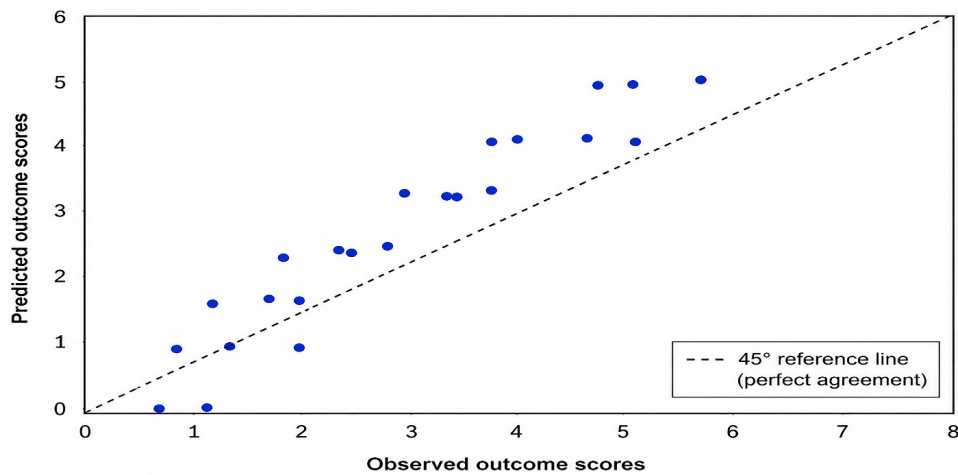


Figure 5: Observed-versus-predicted composite post-booster adverse-event scores. The scatter plot shows the relationship between observed and predicted scores; the 45° reference line represents perfect agreement.

Table 7: Prediction-model performance.

Parameter	Development cohort	Internal validation
R ²	0.34	0.31
Adjusted R ²	0.32	-
RMSE	1.12	1.18
MAE	0.86	0.91
Explained variability	34%	31%
Calibration	Acceptable	Acceptable

Internal Validation and Performance of Prediction Model

The prediction model performed through explained variability, prediction error and calibration in the development cohort was trained and validated internally. Results can be found in Table 7. In the development cohort, an R² of 0.34 (adjusted R² 0.32) was seen after as well as indicating that around the composite post-booster adverse-event score was explained by only those predictors included in the model (around 34% of variability). The marginal difference between R² and adjusted R² values indicates that including the number of predictors has not significantly inflated the explanatory power of the model.

In the development cohort, the model showed RMSE 1.12 and MAE 0.86. The MAE suggests that the average absolute error of prediction was approximately 0.86 score units converting predicted outcome scores into observed score differences. The results showed a moderate predictive accuracy, but also some variability that remained unexplained.

On the other hand, the model ended up having an R² of 0.31 during internal validation which means it was able to explain ~31% variability. This led to only minor reductions (34%-31%) in explained variability. The corresponding RMSE increased from

1.12 to 1.18 (similar for MAE increasing from 0.86 to 0.91). The small variations in these performance indicators imply that the predictive performance of the model remained rather steady within internal validation.

Calibration was satisfactory in both the development and internal validation cohorts indicating fair performance of predicted vs observed outcome scores across the prediction spectrum. In summary, despite a small deterioration in R² and measures of prediction error only during validation, the findings indicated a reasonable level of stability in the models (evidence against overfitting) within each internal validation context.

It thus showed only moderate explanatory and predictability power, reflecting 34% explained variability of the development cohort as well as 31% in internal validation. The results show that the chosen predictors capture a substantial amount of variation in post-booster adverse-event responses. About two-thirds of the outcome variability remained unexplained, suggesting that unmeasured interindividual demographic, biological, behavioral and vaccination-related factors may play a role in determining post-booster response.

While the model was promising with reasonable internal performance, these results must be considered exploratory. Internal validation, which tests model stability in the available study population but does not prove generalizability to other populations/other settings. External validation using an independent cohort is needed before this prediction model can be made available for routine clinical or public-health purposes.

Observed-versus-Predicted Outcome Scores

We then assessed the agreement between (i) the clinical composite post-booster adverse-event score and its corresponding predicted clinical score in a scatter-plot format (Figure 5). The correlation between observed and predicted scores was found to be generally positive, with higher observed scores tending to correspond to

Table 8: Proposed clinical prediction scoring system.

Predictor	Criteria	Points
Age	≥60 years	0
Sex	Female	1
	Male	0
Previous COVID-19 infection	Yes	1
	No	0
Previous significant vaccine reaction	Yes	2
	No	0
Booster-dose interval	<6 months	1
	≥6 months	0
Maximum possible score		5

Table 9: Stratification of risk by prediction score.

Total score	Risk category	Mean predicted outcome score
0-1	Low	1.2
2-3	Moderate	2.8
4-5	High	4.3

higher predicted scores. The distribution of observations along the 45° reference line representing perfect agreement was reasonable for showing direct concordance between observed and predicted values, although some deviations were noted.

The observed-versus-predicted findings were aligned with the comprehensive performance of the model using R^2 values for all disease severities were 0.34 in the development cohort and 0.31 during internal validation. The RMSE was 1.12 and 1.18, and MAE was 0.86 and 0.91, respectively. This relatively small decrease in performance on validation indicates a reasonable amount of stability within the model, but not perfect individual-level prediction.

Overall, the findings suggest that the proposed model could provide moderate predictive performance and accounts for an important percentage of variation in post-booster adverse-event scores. On the other hand, other predictors not included in this current model may contribute to differences among individuals in their responses following adverse-event.

Development of the Clinical Prediction Score

This was undertaken to derive a clinical prediction score by using the independent predictors from the final multivariable regression model. A score of 2 points was given for a previous vaccine reaction, each point is describing as the sex female (1 point), previous COVID-19 infection (1 point), and booster-dose interval <6 months (1 point), figures out a total score between 0 and 14. Age received 0 points for the independent contribution to the model was negligible.

Total scores from 0 to 5: Higher scores predicted a greater subsequent adverse-event burden after the booster (Table 8). This scoring system is a simplified summary of the regression results and allows for initial risk stratification.

Consistent with the higher contribution from previous vaccine reaction in the multivariable model, a larger weight has also been assigned to measure 1. So, the score is a simple point-based system that incorporates principal predictors into its calculation based on their importance.

The greater weighting assigned to previous vaccine reaction reflects its comparatively higher contribution in the multivariable model. Thus, the score incorporates the relative importance of the principal predictors into a simple point-based system.

Risk-Based Classification with respect to the Prediction Score

Participants were classified as low-risk (0-1), moderate-risk (2-3), and high-risk (4-5) based on their total prediction score as shown in Table 9. Mean predicted outcome scores were also averaged 1.2, 2.8 and 4.3 respectively compared to the equivalent achievement group those outlined above.

Across the three risk categories, there was an incremental increase in the mean predicted outcome score from 1.2 (low) to 2.8 (moderate) and 4.3 (high). The finding that risk scores appearing in a stepwise pattern bolsters the internal consistency of our proposed scoring system and potential implications for use as an initial risk-stratification tool. Nevertheless, the suggested cut-offs need to be regarded as exploratory since this study population is both derived and evaluated within the same population. Validation of the scoring system in an independent cohort is needed to assess its clinical applicability.

The observed-versus-predicted analysis and risk-classification results offer early evidence of the validity of our clinical prediction model. The model showed moderate predictive performance, and the consistent increase in predicted outcome across risk categories indicates a potentially useful stratification of participants according to their post-booster adverse-event burden.

However, the model only accounted for approximately 31-34% of the variance observed, suggesting that a considerable amount of variation remained unexplained. Thus, the proposed score should be considered as a hypothesis-generating predictive tool that is in need of external validation in a larger independent cohort before clinical implementation.

CONCLUSION

This study is a cohort study that aimed to establish exploratory prediction model for risk factors associated with post-booster adverse-event burden variability among older adults. Of 500 participants recruited, follow-up was completed by 472 (94.4%) individuals. Commonly reported adverse events were

injection-site pain, fatigue, and headache. Multivariable analysis revealed significant independent predictors including previous vaccine reaction, sex, previous COVID-19 infection and shorter booster-dose interval; the strongest relative contribution being from previous vaccine reaction. A 0-5 point prediction score showing progressively higher levels of predicted outcome across low-, moderate- and high-risk groups. In the development cohort, the model explained 34% of outcome variance with accepted prediction error and calibration (31% during internal validation). A practical solution for preliminary risk stratification of post-booster adverse-event burden could thus lie in the proposed scoring system. This model, however, is only exploratory and needs validation in larger external cohorts to guide clinical use.

ABBREVIATIONS

AEFI: Adverse Events Following Immunization; **CI:** Confidence Interval; **MAE:** Mean Absolute Error; **RMSE:** Root Mean Square Error; **R²:** Coefficient of Determination; **SD:** Standard Deviation; **SE:** Standard Error; **TRIPOD+AI:** Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis + Artificial Intelligence; **VIF:** Variance Inflation Factor.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCE

- Ahamad, M. M., Aktar, S., Uddin, M. J., Rahman, T., Alyami, S. A., Sarker, I. H., et al. (2023). Adverse effects of COVID-19 vaccination: Machine learning and statistical approach to identify and classify incidences of morbidity and postvaccination reactogenicity. *Healthcare*, 11(1), 31. <https://doi.org/10.3390/healthcare11010031>
- Archer, L., Snell, K. I. E., Ensor, J., Hudda, M. T., Collins, G. S., & Riley, R. D. (2021). Minimum sample size for external validation of a clinical prediction model with a continuous outcome. *Statistics in Medicine*, 40(1), 133-146. <https://doi.org/10.1002/sim.8766>
- Bardenheier, B. H., Gravenstein, S., Blackman, C., Gutman, R., Sarkar, I. N., Marcum, Z. A., et al. (2021). Adverse events following one dose of mRNA COVID-19 vaccination among U.S. nursing home residents with and without a previous SARS-CoV-2 infection. *Journal of the American Medical Directors Association*, 22(11), 2228-2232. <https://doi.org/10.1016/j.jamda.2021.08.024>
- Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., et al. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385, Article e078378. <https://doi.org/10.1136/bmj-2023-078378>
- Feikin, D. R., Higdon, M. M., Abu-Raddad, L. J., Andrews, N., Araos, R., Goldberg, Y., et al. (2022). Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: Results of a systematic review and meta-regression. *The Lancet*, 399(10328), 924-944. [https://doi.org/10.1016/S0140-6736\(22\)00152-0](https://doi.org/10.1016/S0140-6736(22)00152-0)
- Heus, P., Reitsma, J. B., Collins, G. S., Damen, J. A. A., Scholten, R. J. P. M., Altman, D. G., et al. (2020). Transparent reporting of multivariable prediction models in journal and conference abstracts: TRIPOD for abstracts. *Annals of Internal Medicine*, 173(5), 403-408. <https://doi.org/10.7326/M20-0193>
- Martin, G. P., Riley, R. D., Collins, G. S., & Sperrin, M. (2021). Developing clinical prediction models when adhering to minimum sample size recommendations: The importance of quantifying bootstrap variability in tuning parameters and predictive performance. *Statistical Methods in Medical Research*, 30(12), 2545-2561. <https://doi.org/10.1177/09622802211046388>
- Mori, M., Takahashi, K., Iijima, H., Kanou, K., Sawamura, K., Hiramatsu, Y., et al. (2023). Impact of sex and age on vaccine-related side effects and their progression after booster mRNA COVID-19 vaccine. *Scientific Reports*, 13(1), 19285. <https://doi.org/10.1038/s41598-023-46823-4>
- Nanishi, E., Levy, O., & Ozonoff, A. (2022). Waning effectiveness of SARS-CoV-2 mRNA vaccines in older adults: A rapid review. *Human Vaccines & Immunotherapeutics*, 18(5), 2045857. <https://doi.org/10.1080/21645515.2022.2045857>
- Riley, R. D., Debray, T. P. A., Collins, G. S., Archer, L., Ensor, J., van Smeden, M., et al. (2021b). Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Statistics in Medicine*, 40(19), 4230-4251. <https://doi.org/10.1002/sim.9025>
- Riley, R. D., Ensor, J., Snell, K. I. E., Harrell, F. E., Jr., Martin, G. P., Reitsma, J. B., et al. (2020). Calculating the sample size required for developing a clinical prediction model. *BMJ*, 368, Article m441. <https://doi.org/10.1136/bmj.m441>
- Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E., Jr., et al. (2021a). Minimum sample size for external validation of a clinical prediction model with a continuous outcome. *Statistics in Medicine*, 40(1), 133-146. <https://doi.org/10.1002/sim.8766>
- Sutton, N., San Francisco Ramos, A., Beales, E., Purohit, A., Bate, J., Baillie, J. K., et al. (2022). Comparing reactogenicity of COVID-19 vaccines: A systematic review and meta-analysis. *Expert Review of Vaccines*, 21(9), 1301-1318. <https://doi.org/10.1080/14760584.2022.2098719>
- Sutton, N., San Francisco Ramos, A., Beales, E., Purohit, A., Bate, J., Baillie, J. K., et al. (2024). Comparing reactogenicity of COVID-19 vaccine boosters: A systematic review and meta-analysis. *Expert Review of Vaccines*, 23(1), 273-286. <https://doi.org/10.1080/14760584.2024.2315089>
- Warkentin, L., Werner, F., Zeschick, N., Stangl, L., Schober, K., Lenski, D., Grill, E., Diefenbach, M., et al. (2023). Reactogenicity and safety of COVID-19 primary immunisation and booster vaccination regimens: A comparative observational cohort study. *BMC Medicine*, 21(1), 218. <https://doi.org/10.1186/s12916-023-02924-5>
- Zamanipoor Najafabadi, A. H., Ramspek, C. L., Dekker, F. W., Heus, P., Hooft, L., Moons, K. G. M., et al. (2020). TRIPOD statement: A preliminary pre-post analysis of reporting and methods of prediction models. *BMJ Open*, 10(9), e041537. <https://doi.org/10.1136/bmjopen-2020-041537>

Cite this article: Islam S, Nath B, Debnath S. Risk Factors and Predictive Modeling of Adverse Events Following Covid-19 Booster Vaccination: A Prospective Cohort Study. *Indian J Pharmacy Practice*. 2027;20(1):32-44.