Real-time analysis of a mass vaccination effort via an Artificial Intelligence platform confirms the safety of FDA-authorized COVID-19 vaccines

Reid McMurry[1+], Patrick Lenehan[1+], Samir Awasthi[1+], Eli Silvert[1], Arjun Puranik[1], Colin Pawlowski[1], AJ Venkatakrishnan[1], Praveen Anand[2], Vineet Agarwal[1], John C. O'Horo[3], Gregory J. Gores[3], Amy W. Williams[3], Andrew D. Badley[3], John Halamka[3], Abinash Virk[3], Melanie D. Swift[3], Katie Carlson[1], Deeksha Doddahonnaiah[1], Anna Metzger[1], Nikhil Kayal[1], Gabi Berner[1], Eshwan Ramudu[1], Corinne Carpenter[1], Tyler Wagner[1], Ajit Rajasekharan[1], Venky Soundararajan[1,2]*

[1] nference, One Main Street, East Arcade, Cambridge, MA 02142, USA
[2] nference Labs, 2nd Floor, 22 3rd Cross Rd, Murgesh Pallya, Bengaluru, Karnataka 560017, India
[3] Mayo Clinic, Rochester, MN 55902, USA
+ Joint first authors
* Address correspondence to Venky Soundararajan (venky@nference.net)

Abstract

As the COVID-19 vaccination campaign unfolds as one of the most rapid and widespread in history, it is important to continuously assess the real world safety of the FDA-authorized vaccines. Curation from large-scale electronic health records (EHRs) allows for near real-time safety evaluations that were not previously possible. Here, we advance context- and sentiment-aware deep neural networks over the multi-state Mayo Clinic enterprise (Minnesota, Arizona, Florida, Wisconsin) for automatically curating the adverse effects mentioned by physicians in over 108,000 EHR clinical notes between December 1st 2020 to February 8th 2021. We retrospectively compared the clinical notes of 31,069 individuals who received at least one dose of the Pfizer/BioNTech or Moderna vaccine to those of 31,069 unvaccinated individuals who were propensity matched by demographics, residential location, and history of prior SARS-CoV-2 testing. We find that vaccinated and unvaccinated individuals were seen in the the clinic at similar rates within 21 days of the first or second actual or assigned vaccination dose (first dose Odds Ratio = 1.13, 95% CI: 1.09-1.16; second dose Odds Ratio = 0.89, 95% CI: 0.84-0.93). Further, the incidence rates of all surveyed adverse effects were similar or lower in vaccinated individuals compared to unvaccinated individuals after either vaccine dose. Finally, the most frequently documented adverse effects within 7 days of each vaccine dose were arthralgia (Dose 1: 0.59%; Dose 2: 0.39%), diarrhea (Dose 1: 0.58%; Dose 2: 0.33%), erythema (Dose 1: 0.51%; Dose 2: 0.31%), myalgia (Dose 1: 0.40%; Dose 2: 0.34%), and fever (Dose 1: 0.27%; Dose 2: 0.31%). These remarkably low frequencies of adverse effects recorded in EHRs versus those derived from active solicitation during clinical trials (arthralgia: 24-46%; erythema: 9.5-14.7%; myalgia: 38-62%; fever: 14.2-15.5%) emphasize the rarity of vaccine-associated adverse effects requiring clinical attention. This rapid and timely analysis of vaccine-related adverse effects from contextually rich EHR notes of 62,138 individuals, which was enabled through a large scale Artificial Intelligence (AI)-powered platform, reaffirms the safety and tolerability of the FDA-authorized COVID-19 vaccines in practice.

Introduction

Following their Emergency Use Authorizations by the Food and Drug Administration (FDA) in December of 2020, more than 52 million doses of BNT162b2 (Pfizer/BioNTech) and mRNA-1273 (Moderna) COVID-19 vaccines have been administered in the United States[1–3]. Phase III trials demonstrate strong efficacy and favorable safety profiles for these vaccines in the cohorts studied. Specifically, the trials showed 95.0% (95% CI, 90.3 to 97.6) and 94.1% (95% CI, 89.3 to 96.8%) efficacy for BNT162b2 and mRNA-1273 respectively. While self-resolving mild to moderate adverse effects were common in vaccinated participants, serious adverse effects occurred rarely and with a frequency comparable to placebo[4,5]. Local adverse effects of any severity reported in these trials included injection site pain (84.1-92%), injection site swelling (10.5-70%), and injection site erythema (9.5-14.6%). Systemic effects of any severity included fatigue (62.9-70%), headache (55.1-64.7%), myalgia (38.3-61.5%), arthralgia (23.6-46.4%), chills (31.9-45.4%), fever (14.2-15.5%), and nausea/vomiting (1.1-23.0%)[6,7].

Consistent with CDC recommendations, early vaccination efforts (Phase 1a) in the United States have targeted healthcare workers, residents and staff at long term care facilities, who are at elevated risk for COVID-19 relative to the general public[8–11]. As the vaccines continue to be administered more broadly, it will be critical to continuously evaluate real world safety and efficacy data from all those who have received these vaccines. This approach may validate existing findings or highlight differences in the larger population compared to the clinical experience with respect to vaccine efficacy and adverse effects. Of note, the populations undergoing vaccination may differ meaningfully from the trial populations. For example, healthcare workers and long-term care residents receiving the first wave of vaccines in Phase 1a likely comprised a small fraction of the Phase 3 trial populations. Additionally, some pregnant women are receiving COVID-19 vaccines in the real world[12], while they were excluded from both trials[4,5]. Finally, continuous monitoring will help to identify and better quantify the frequency of rare severe adverse effects such as anaphylaxis, which was widely reported but only observed in a small number of individuals after approval of both vaccines[13–17].

One approach to post- authorization surveillance of vaccine efficacy and safety is via the real-time analysis of patient data stored in electronic health record (EHR) systems. We have previously developed and described augmented curation methods to rapidly create and compare cohorts of COVID-19 patients within a large EHR system, and have recently applied these methods to assess the real world efficacy of both BNT162b2 and mRNA-1273 in over 31,000 individuals receiving these vaccines at the Mayo Clinic and associated health system[18–22]. Here we expand on these efforts to study the adverse effects experienced by individuals after COVID-19 vaccination in the clinical environment.

It should be noted that monitoring vaccine-associated adverse effects in a clinical trial setting and outside of trial environments are quite different. In clinical trials, participants are aware that they are receiving an experimental product, and adverse effect reporting is encouraged or solicited. On the other hand, individuals receiving a COVID-19 vaccine during the mass vaccination campaign are informed of adverse effects that are likely to be experienced and can even be discouraged from seeking medical attention unless the symptom is particularly severe. Thus, adverse effects are likely to be captured only in the EHR of individuals who experience

symptoms which are severe or persistent enough to warrant a return to clinic, or who happen to have a previously scheduled routine clinical visit in the post vaccination time period. Accordingly, the intention of our analysis is not to determine whether real world data recapitulates the adverse effect frequencies reported in prior trials. Instead, it is to (1) establish the rates at which individuals actually report potential vaccine-associated adverse effects to health care practitioners (HCPs) in several defined time intervals after vaccination, and (2) determine whether these rates of adverse effect reporting rates are unexpectedly high.

To address the latter point, it is critical to establish the baseline frequency at which each adverse effect is expected to be documented in the clinical notes of our vaccinated cohort. We thus employed a 1-to-1 propensity matching procedure to derive a cohort of unvaccinated individuals who are balanced for demographic factors, residential location, and history of SARS-CoV-2 PCR testing (see **Methods**)[23], and we similarly curated their clinical notes over the same time period to quantify the frequency of the defined symptoms of interest. This propensity matched group serves a purpose similar to that of the placebo arm in clinical trial safety assessments, allowing us to contextualize and better interpret the absolute rates of adverse effects documented in the notes of vaccinated individuals.

We find that individuals undergoing COVID-19 vaccination do not return to the clinic at unexpectedly high rates in the 7, 14, or 21 days following either vaccine dose. Further, the rates of adverse effects (headache, myalgia, arthralgia, fatigue, fever, diarrhea, nausea/vomiting, anaphylaxis, facial paralysis, lymphadenopathy, and erythema) documented in EHR notes over these time intervals are not higher in vaccinated individuals compared to the unvaccinated control cohort. This study supports the safety and tolerability of BNT162b2 and mRNA-1273 in practice, further strengthening the case for the rapid and broad distribution of vaccines to the public.

Methods

Study design, setting and population

This is a retrospective study of individuals who underwent polymerase chain reaction (PCR) testing for suspected SARS-CoV-2 infection at the Mayo Clinic and hospitals affiliated with the Mayo Clinic Health System. This study was reviewed by the Mayo Clinic Institutional Review Board (IRB) and determined to be exempt from the requirement for IRB approval (45 CFR 46.104d, category 4). Subjects were excluded if they did not have a research authorization on file.

The cohorts of vaccinated and unvaccinated individuals considered for this study are identical to the cohorts considered in a previous analysis: “FDA-authorized COVID-19 vaccines are effective per real-world evidence synthesized across a multi-state health system”[23]. In total, there were 507,525 individuals in the Mayo electronic health record (EHR) database who received a PCR test between February 15, 2020 and February 8, 2021. To obtain the study population, we defined the following inclusion criteria: (1) at least 18 years old; (2) no positive SARS-CoV-2 PCR test before December 1, 2020; (3) resides in a locale (based on Zip code) with at least 25 individuals who have received BNT162b2 or mRNA-1273. This population included 249,708 individuals, of whom 31,623 have received BNT162b2 or mRNA-1273 and 218,085 have no

record of COVID-19 vaccination. Vaccinated individuals who had tested positive for SARS-CoV-2 by PCR between December 1, 2020 and the date of their first vaccine dose were excluded, resulting in 31,299 individuals. Individuals with zero follow-up days after vaccination (i.e., those who received the first vaccine dose on the date of data collection) were also excluded, leaving 31,069 individuals in the final vaccinated cohort.

The propensity matched unvaccinated cohort was selected from the previously derived set of 218,085 unvaccinated individuals. The purpose of this cohort was to establish the baseline frequency of EHR documentation for each adverse effect of interest in a cohort which is clinically similar to our vaccinated cohort. These baselines, or expected, frequencies can then be compared to the observed frequencies to determine whether or not these adverse effects are reported at unexpectedly high rates among patients receiving a COVID-19 vaccine.

The selection algorithm and its associated counts are summarized in **Figure 1**. More details on the propensity score matching procedure are provided in the prior manuscript[23]. This includes a table showing that the cohorts were indeed balanced for all covariates included in propensity score matching (standardized mean difference < 0.1 for each one) and figures illustrating the distributions of age and total follow-up time between the first vaccine dose and the end of the data collection period.

Definition of time intervals for safety analyses

For each vaccinated individual, we defined the date of their first vaccine dose as Day V_1 and the date of their second vaccine dose as Day V_2. In the Results section, these are referred to as "actual" dates of vaccination. For each unvaccinated individual, Day V_1 and Day V_2 were designated as identical to their matched vaccinated individual. In the Results section, these are referred to as "assigned" dates of vaccination.

Definition of adverse effects of interest

The adverse effects considered were primarily derived from those assessed in phase 3 trials of BNT162b2 and mRNA-1273[4,5], including fatigue, fever, myalgia, arthralgia, headache, lymphadenopathy, erythema, diarrhea, and vomiting. We also included anaphylaxis and facial paralysis (Bell's palsy), as these rare events have been reported in individuals receiving COVID-19 vaccines as well. Each adverse effect was mapped to a set of synonyms intended to capture the most common ways that a given phenotype would be referenced in the context of a clinical note.

Curation of adverse effects from clinical notes

To curate the adverse effects experienced by each patient from the electronic health record, we used a BERT-based neural network model[24] to classify the sentiment for the phenotypes (described above) mentioned in the clinical notes. Specifically, this classification model categorizes phenotype-containing sentences into one of four categories: (1) confirmed diagnosis (2) ruled-out diagnosis, (3) possibility of disease, and (4) alternate context (e.g., family history). This classification model was trained on 18,500 sentences and has shown an out-of-sample accuracy of 93.6% with precision and recall scores above 95%[25]. For each individual, we

applied the sentiment model to the clinical notes in the Mayo Clinic electronic health record during our defined intervals of interest for each individual: (1) Day V_1 to 7, 14, or 21 days after Day V_1, and (3) Day V_2 to 7, 14, or 21 days after Day V_2. For each phenotype, we identified the first date on which the given individual had at least one sentence in which the phenotype was categorized as "confirmed diagnosis" with a confidence score of at least 90%. For the phenotypes anaphylaxis and facial paralysis, each such sentence was manually reviewed to verify the positive sentiment (i.e. confirmed diagnosis) and to assess the tense of this sentiment (i.e. past vs. present). Only sentences which confirm a present diagnosis were used to count adverse effects in this study.

Evaluating rates of return to clinic after vaccination

To evaluate the likelihood of returning to the clinic after vaccination, we counted the number of individuals who had at least one clinical note in the 7, 14, and 21 days after Day V_1 and Day V_2. The fraction of individuals with clinical follow-up was calculated as the number of individuals with at least clinical note in the time window divided by the total number of individuals in each group (n = 31,069 for first vaccine dose; n = 17,067 for second vaccine dose). The difference in clinical follow-up rates was assessed by calculating the odds ratio (OR) along with its corresponding 95% confidence interval (CI) using the fisher.test function in R version 1.3.959[26]. The null hypothesis was that the OR falls between 0.91 and 1.1 (i.e., the larger rate is at most 10% larger than the smaller rate); thus, an OR was considered significant if the upper bound of the 95% CI was less than 0.91 or the lower bound of the 95% CI was greater than 1.1.

Evaluating vaccine adverse effects among the total cohorts

To evaluate adverse effects associated with receiving a COVID-19 vaccine in the clinical setting, we compared the two populations described above and summarized in **Figure 1**: (1) 31,069 individuals with follow-up who received BNT162b2 or mRNA-1273 and did not have a prior positive SARS-CoV-2 PCR test ("vaccinated"), and (2) 31,069 propensity matched individuals who have never received a COVID-19 vaccine and did not have a positive SARS-CoV-2 PCR test before the first vaccination date (dose 1) of their matched individual ("unvaccinated").

The incidence of a given adverse effect after each vaccine dose was assessed by computing the incidence rate ratio (IRR) of the vaccinated and unvaccinated cohorts. Specifically, we evaluated adverse effects which were documented in clinical notes within 7 days of receiving the first vaccine dose (Day V_1 to 7 days after Day V_1) or the second vaccine dose (Day V_2 to 7 days after Day V_2). For each cohort in a defined time period, incidence rates were calculated as the number of individuals experiencing the given adverse effect in that time period divided by the total number of at-risk person-days contributed in that time period. For each individual, at-risk person-days are defined as the number of days from the start of the time period in which the day on which the individual experienced the adverse effect or died, or four days prior to testing positive for SARS-CoV-2. The IRR was calculated as the incidence rate of the vaccinated cohort divided by the incidence rate of the unvaccinated cohort, and its 95% CI was computed using an exact approach described previously[27]. The IRR was considered to be statistically significant if the 95% CI did not include 1.

Evaluating vaccine adverse effects in individuals who returned to clinic

Because only a fraction of individuals in the vaccinated and unvaccinated cohorts contributed clinical notes during the defined intervals after their actual or assigned vaccination dates, we also assessed the rates of adverse effects documented in the EHR specifically among the individuals who had returned to clinic. To select the cohorts for this analysis, we considered any pair of matched vaccinated and unvaccinated individuals who both contributed at least one clinical note which contained at least one phenotype term in the 7, 14, or 21 days after the actual or assigned date of first or second vaccination. The number of such matched pairs in each time interval is provided in **Table S1**. For each time interval We then computed incidence rates for each adverse effect in the vaccinated and unvaccinated cohorts, along with the incidence rate ratios and 95% CI as described above. The IRR was considered to be statistically significant if the 95% CI did not include 1.

Results

Vaccinated individuals do not return to clinic at unexpectedly high rates after either vaccine dose

To assess rates of clinical follow-up after study enrollment, we compared the number of vaccinated and unvaccinated individuals with EHR notes within 7, 14, or 21 days of each actual or assigned vaccination date. In the 7 days after the first vaccine dose, 6,121 of the 31,069 (19.7%) vaccinated individuals had at least one EHR note, compared to 5,789 of 31,069 (18.6%) unvaccinated individuals (Odds Ratio = 1.07; 95% CI: [1.03-1.12]) (**Table 1**). This difference was not considered significant as the lower bound of the 95% CI was less than 1.1 (see **Methods**). The rates of return to clinic were also similar between these cohorts within 14 and 21 days of the first dose (Odds $\text{Ratio}_{14\ \text{Days}}$ = 1.11 [1.07-1.15]; Odds $\text{Ratio}_{21\ \text{Days}}$ = 1.13 [1.09-1.16]) (**Table 1**)

Within 7 days of the second dose, 2,513 of 17,067 (14.7%) unvaccinated individuals contributed EHR notes, compared to 2,778 (16.3%) of 17,067 unvaccinated individuals (Odds Ratio = 0.89 [0.84-0.94]) (**Table 2**). This difference was not considered significant as the upper bound of the 95% CI was greater than 0.91 (see **Methods**). Similarly, the rates of return to clinic were similar between these cohorts within 14 and 21 days of the second dose (Odds $\text{Ratio}_{14\ \text{Days}}$ = 0.88 [0.83-0.93]; Odds $\text{Ratio}_{21\ \text{Days}}$ = 0.89 [0.84-0.93]) (**Table 2**).

In summary, these findings show that individuals receiving COVID-19 vaccines do not tend to return to the clinic in the subsequent weeks at higher-than-expected rates, which suggests favorable tolerability of these vaccines.

Documented vaccine associated adverse effects in EHR notes are rare compared to the frequencies reported in clinical trials

Among the 31,069 individuals in each cohort, the most commonly documented symptoms in the 7 days after the first vaccine dose included arthralgia (183 [0.59%] vaccinated individuals; 171 [0.55%] unvaccinated individuals), diarrhea (179 [0.58%]; 266 [0.86%]), erythema (159 [0.51%]; 184 [0.59%]), myalgia (125 [0.40%]; 135 [0.43%]), and fever (85 [0.27%]; 198 [0.64%])

(**Table 3**). The same events were also the most frequently documented in the 7 days following the second vaccine dose (n = 17,067 individuals per group): arthralgia (67 [0.39%] vaccinated individuals; 81 [0.47%] unvaccinated individuals), diarrhea (56 [0.33%]; 120 [0.70%]), erythema (53 [0.31%]; 93 [0.54%]), myalgia (58 [0.34%]; 62 [0.36%]), and fever (31 [0.31%]; 109 [0.64%]) (**Table 4**).

Notably, these rates of adverse effects documented in EHR notes were markedly lower than the rates of adverse effects observed in clinical trials (e.g. fatigue: 63-70%; myalgia: 38-62%; arthralgia: 24-46%; fever: 14-16%; erythema: 10-15%)[6,7]. This is to be expected, as individuals vaccinated in the real world are advised that it is normal to experience these adverse effects, and so they are less likely to report them to a healthcare provider. In line with this, the vaccine associated adverse effects which are captured in EHR notes are likely to be those that are severe or persistent enough to cause an individual to return to clinic or otherwise notify their HCP.

Adverse effects are not documented more frequently in the EHR notes of vaccinated individuals compared to those of propensity-matched unvaccinated individuals

To test whether documentation of each adverse effect was actually associated with receiving a COVID-19 vaccine, we computed the incidence rate ratio (IRR) for each adverse effect between the vaccinated and propensity matched unvaccinated cohorts. Most symptoms, including myalgia, arthralgia, fatigue, and headache, had similar incidence rates during the 7 days after the first or second actual or assigned vaccine dose, with the confidence intervals including 1 (**Tables 3-4**). This was also true of the more serious adverse effects, anaphylaxis and facial paralysis (**Tables 3-4**). Several symptoms actually had a significantly lower incidence rate in the 7 days after both doses in the vaccinated cohort, including fever ($IRR_{dose\ 1}$ = 0.43 [0.33-0.55]; $IRR_{dose\ 2}$ = 0.28 [0.18-0.43]), diarrhea ($IRR_{dose\ 1}$ = 0.67 [0.55-0.81]; $IRR_{dose\ 2}$ = 0.46 [0.33-0.64]), and lymphadenopathy ($IRR_{dose\ 1}$ = 0.55 [0.4-0.75]; $IRR_{dose\ 2}$ = 0.41 [0.24-0.68]).

These trends of vaccinated individuals showing similar or lower incidence rates for each considered adverse effect were also true in the 14 and 21 days following actual or assigned vaccination dates (**Tables S2-S5**). In summary, this data supports the tolerability of BNT162b2 and mRNA-1273, as individuals receiving COVID-19 vaccines do not return to the clinic to report adverse effects at higher rates than propensity matched unvaccinated individuals.

Among matched pairs who return to clinic, adverse effects are reported at similar rates in the vaccinated and unvaccinated cohorts

To complement our previous analysis, we also considered the rates of adverse effect documentation in only the subset of matched pairs who each contributed at least one EHR note in the 7, 14, or 21 days following the first or second actual or assigned vaccine dose. In the 7 days after the first actual or assigned vaccination date (n = 1,432 pairs), the IRR 95% CIs for every adverse effect except diarrhea included 1, and the incidence rate of diarrhea was higher in the unvaccinated cohort (IRR = 0.47 [0.29-0.73]) (**Table 5**). After the second vaccination date (n = 944 pairs), the IRR 95% CIs all included 1 except for myalgia, which showed a slightly higher incidence rate in vaccinated individuals (IRR = 2.7 [1.1-7.6]) (**Table 6**).

Again, vaccinated individuals who returned to the clinic also showed similar or lower incidence rates for each considered adverse effect in the 14- and 21-days post vaccination compared to their matched unvaccinated individuals who also returned to the clinic (**Tables S6-S9**). Taken together, this analysis further corroborates the conclusion that adverse effects are not documented in the EHR of vaccinated individuals at unexpectedly high rates.

Discussion

This study demonstrates that the two currently FDA-authorized COVID-19 vaccines, BNT162b2 and mRNA-1273, are safe and tolerated beyond the confines of a clinical trial setting. This conclusion is consistent with the extensive safety and tolerability assessments conducted in Phase I/II and Phase III trials over the past nine months[4,5,28,29]. Here we assessed real world safety and tolerability by longitudinally curating the EHR documentation of adverse effects in 31,069 individuals receiving at least one dose of a COVID-19 vaccine compared to a propensity matched unvaccinated cohort of the same size. Compared to this control cohort, vaccinated individuals were not more likely to return to the clinic or to report any of the surveyed adverse effects within 7, 14, or 21 days after the first or second vaccine dose. When considering only the pairs of matched individuals contributing at least one phenotype-containing clinical note in these follow-up windows, only myalgia was documented more frequently in EHR notes of vaccinated patients within 7 days of the second vaccine dose (IRR = 2.7; 95% CI: 1.1-7.6), while the other adverse effects were documented at similar rate or lower rates in the vaccinated cohort.

Our finding that EHR notes from vaccinated and unvaccinated individuals record similar rates of potential vaccine associated adverse effects differs from the data obtained in Phase III trials, wherein vaccinated participants experienced higher rates of symptoms. Further, the absolute rates of adverse effects documented in these EHR notes are well below the rates reported in clinical trials. These discrepancies are likely attributable to differences in methodology of reporting and recording symptoms and events. Both trials included a 7-day post-vaccination period in which symptoms were actively solicited from some or all individuals as well as longer periods in which unsolicited adverse effects and serious adverse reactions were recorded from all individuals. In contrast, our methods rely exclusively on the recording of unsolicited symptoms or events in the EHR. Given that individuals are warned of the likely vaccine associated adverse effects at the time of vaccination in the real world setting, it is likely that most mild or moderate symptoms are never actually reported and thus are not documented in an EHR note. It is also possible that HCPs, who comprise a significant proportion of the vaccinated population thus far, are less likely to report adverse effects given their clinical expertise and ability to self-assess the severity of their illness.

That said, serious safety concerns by definition require medical care and thus are likely to be documented in the EHR. For example, should an individual experience anaphylaxis, this individual will likely require emergent care, during which one or more clinical notes will be written and will mention this phenotype. Thus, our method should identify the symptoms and phenotypes that represent the most serious threats to vaccine safety and tolerability of practical significance. Indeed, this is the central reason why our analysis should be viewed as complementary to the data which has been obtained in the more controlled setting of clinical trials. While the observed adverse effect frequencies in the trial setting are extremely valuable, our assessment specifically

aims to describe the frequencies of adverse effects that receive some form of clinical attention as evidenced by their documentation in the EHR.

The purpose of using propensity matching in this study was to establish an expected frequency for each potential vaccine associated adverse effect in a group of individuals with similar demographic and geographic characteristics. The finding that individuals in the propensity matched unvaccinated cohort were seen in clinic more frequently than those from the vaccinated cohort after the second vaccine dose was surprising, but there are several potential explanations for this. As was previously demonstrated, the incidence rate of testing positive for SARS-CoV-2 was indeed lower in the vaccinated cohort, with a particularly strong effect observed in the time windows beyond the second dose[23]. Thus, we would indeed expect fewer vaccinated individuals to be seen in the clinic for COVID-19. Further, there may be behavioral or psychological differences that set in after receiving the COVID-19 vaccine, wherein individuals who have received both doses feel less compelled to visit the clinic if they experience flu-like symptoms. Finally, there may be a genuine imbalance in the baseline frequency of clinical visits due to the enrichment of healthcare workers in the vaccinated cohort. We addressed these potential confounding factors by performing sub-analyses on only the matched pairs of individuals who both contributed at least one EHR note in each time interval of interest. The fact that adverse effects were still not observed more frequently among vaccinated individuals in this sub-analysis strengthen our overall conclusions regarding the safety and tolerability of these vaccines.

This study does have several important limitations to consider. First, while the analysis was conducted on a population derived from a large healthcare system, the cohort demographics are not representative of the American population. For example, both the vaccinated and unvaccinated cohorts were predominantly Caucasian (>90%) and female (>60%). These biases likely reflect both the populations who receive care at the various Mayo Clinic centers and the populations who have been prioritized in Phase 1a of the vaccine rollout. Second, the BERT model used to curate EHR notes does not imply a direct link between COVID-19 vaccination and the experience of a phenotype. That is, we simply capture the occurrence of an adverse effect without ensuring that the clinical note indeed suggests or confirms that vaccination caused the symptom. This shortcoming is addressed by comparing vaccinated individuals to the unvaccinated control cohort, which establishes a baseline expected frequency for each symptom in the absence of vaccination. Finally, while sentences suggesting the occurrence of anaphylaxis or facial paralysis were manually reviewed to confirm both the positive sentiment and the tense, sentences for the other curated phenotypes were not reviewed. In the future, we will train natural language processing models to discriminate past from present tense, thereby circumventing the need for this manual review.

As the remainder of the US population undergoes COVID-19 vaccination, it will not be feasible to conduct solicited reporting of adverse effects in all vaccinated individuals. However, the use of augmented curation for real world safety monitoring presents a practical solution to this problem. The method demonstrated here represents a scalable approach to continuously monitor serious safety concerns associated with any authorized COVID-19 vaccines. Taken together with our recent study highlighting the real-world efficacy of these vaccines[23], this data reinforces that individuals, providers, and public health officials should proceed rapidly with vaccination efforts, with high confidence in their efficacy and safety.

References

1. CDC. COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker/ (2020).
2. FDA authorizes Moderna COVID-19 vaccine. *Med. Lett. Drugs Ther.* **63**, 9–10 (2021).
3. Office of the Commissioner. Pfizer-BioNTech COVID-19 Vaccine. https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/pfizer-biontech-covid-19-vaccine (2021).
4. Baden, L. R. *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N. Engl. J. Med.* **384**, 403–416 (2021).
5. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).
6. [No title]. https://www.fda.gov/media/144413/download.
7. [No title]. https://www.fda.gov/media/144637/download.
8. CDC. CDC's COVID-19 Vaccine Rollout Recommendations. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations.html (2021).
9. Nguyen, L. H. *et al.* Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *Lancet Public Health* **5**, e475–e483 (2020).
10. Mutambudzi, M. *et al.* Occupation and risk of severe COVID-19: prospective cohort study of 120 075 UK Biobank participants. *Occup. Environ. Med.* (2020) doi:10.1136/oemed-2020-106731.
11. Panagiotou, O. A. *et al.* Risk Factors Associated With All-Cause 30-Day Mortality in Nursing Home Residents With COVID-19. *JAMA Intern. Med.* (2021) doi:10.1001/jamainternmed.2020.7968.
12. Adhikari, E. H. & Spong, C. Y. COVID-19 Vaccination in Pregnant and Lactating Women. *JAMA* (2021) doi:10.1001/jama.2021.1658.
13. Wu, K. Boston Doctor Reports Serious Allergic Reaction After Getting Moderna's Covid Vaccine. *New York Times* (2020).
14. CDC COVID-19 Response Team, FDA. *Allergic Reactions Including Anaphylaxis After Receipt of the First Dose of Moderna COVID-19 Vaccine — United States, December 21, 2020–January 10, 2021*. (2021).
15. Covid, C. & Team, R. Allergic reactions including anaphylaxis after receipt of the first dose of Pfizer-BioNTech COVID-19 vaccine—United States, December 14--23, 2020. *MMWR Surveill. Summ.* **70**, 46 (2021).
16. What to Know About Post-Vaccine Deaths and Allergies. *The Washington Post* (2021).
17. Castells, M. C. & Phillips, E. J. Maintaining Safety with SARS-CoV-2 Vaccines. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMra2035343.
18. Pawlowski, C. *et al.* Pre-existing conditions are associated with COVID patients' hospitalization, despite confirmed clearance of SARS-CoV-2 virus. *medRxiv* (2020).
19. Pawlowski, C. *et al.* Enoxaparin Is Associated With Lower Rates of Thrombosis, Kidney Injury, and Mortality Than Unfractionated Heparin in Hospitalized COVID Patients. (2020) doi:10.2139/ssrn.3707421.
20. Pawlowski, C. *et al.* Exploratory analysis of immunization records highlights decreased SARS-CoV-2 rates in individuals with recent non-COVID-19 vaccinations. *medRxiv* (2020).
21. Awasthi, S. *et al.* Plasma IL-6 Levels following Corticosteroid Therapy as an Indicator of

ICU Length of Stay in Critically ill COVID-19 Patients. *medRxiv* (2020).
22. Agarwal, V. *et al.* Quantifying the prevalence of SARS-CoV-2 long-term shedding among non-hospitalized COVID-19 patients. *medRxiv* (2020) doi:10.1101/2020.06.02.20120774.
23. Pawlowski, C. *et al.* FDA-authorized COVID-19 vaccines are effective per real-world evidence synthesized across a multi-state health system. *medRxiv* (2021) doi:10.31219/osf.io/y6pdw.
24. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (2018).
25. Wagner, T. *et al.* Augmented curation of clinical notes from a massive EHR system reveals symptoms of impending COVID-19 diagnosis. *Elife* **9**, (2020).
26. The R Project for Statistical Computing. https://www.R-project.org/.
27. Sahai, H. & Khurshid, A. *Statistics in Epidemiology: Methods, Techniques and Applications*. (CRC Press, 1995).
28. Jackson, L. A. *et al.* An mRNA vaccine against SARS-CoV-2—preliminary report. *N. Engl. J. Med.* (2020).
29. Mulligan, M. J. *et al.* Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature* **586**, 589–593 (2020).

Acknowledgements

The authors thank Murali Aravamudan for the careful review and feedback on this manuscript.

Data Availability

After publication, the data will be made available upon reasonable requests to the corresponding author. A proposal with detailed description of study objectives and the statistical analysis plan will be needed for evaluation of the reasonability of requests. Deidentified data will be provided after approval from the corresponding author and the Mayo Clinic.

Declaration of Interests

RM, PL, ES, AP, SA, CP, VA, AJV, PA, AR, CC, KC, DD, NK, ER, GB, AM, TW, and VS are employees of nference and have financial interests in the company and in the successful application of this research. JCO receives personal fees from Elsevier and Bates College, and receives small grants from nference, Inc, outside the submitted work. ADB is a consultant for Abbvie, is on scientific advisory boards for nference and Zentalis, and is founder and President of Splissen therapeutics. JH, JCO, GJG, AWW, AV, MDS, and ADB are employees of the Mayo Clinic. The Mayo Clinic may stand to gain financially from the successful outcome of the research. This research has been reviewed by the Mayo Clinic Conflict of Interest Review Board and is being conducted in compliance with Mayo Clinic Conflict of Interest policies.

Author Contributions

VS, PL, and SA conceived the study. RM, PL, AJV and VS wrote the manuscript and reviewed the findings. ES, AP, SA, CP, VA, PA, AR, CC, KC, DD, NK, ER, GB, AM, TW contributed methods, analysis, and software. JCOH, GJG, AWW, ADB, MDS, AV, and JH reviewed the study, findings, and the manuscript. All authors revised the manuscript.

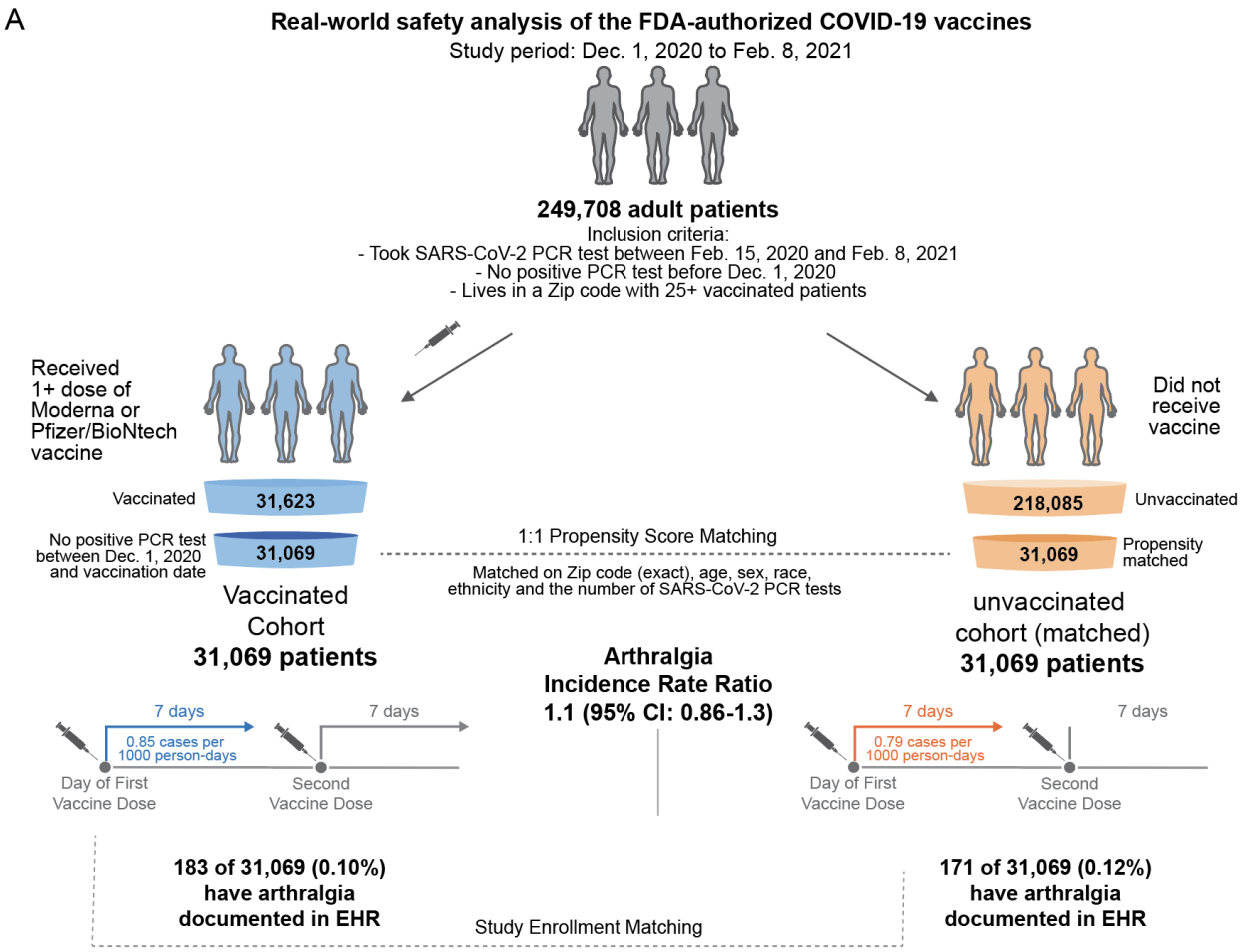

Figure 1. Schematic illustration of participant selection and study design. The vaccinated cohort is composed 31,069 individuals from the Mayo Clinic and associated health systems who received at least one dose of BNT162b2 (Pfizer/BioNTech) or mRNA-1273 (Moderna) between December 1, 2020 and February 8, 2021 and did not test positive for SARS-CoV-2 prior to their first vaccination. A control cohort of unvaccinated individuals was generated via 1:1 propensity score matching, yielding 31,069 individuals with similar distributions of age, sex, race, ethnicity, residential location, and number of prior SARS-CoV-2 PCR tests in the past year. For each cohort, the incidence rates of several adverse effects (e.g., arthralgia) were calculated for the seven days following the first dose and, separately, for the seven days following the second dose. For a given adverse effect, the incidence rate ratio (and the corresponding 95% CI) was calculated to determine whether one cohort was more likely to experience the event than the other. Incidence rates and incidence rate ratios were also calculated for the 14 and 21 days following each vaccine dose.

	Vaccinated individuals with at least one note (%) [n = 31,069]	Unvaccinated individuals with at least one note (%) [n = 31,069]	Odds Ratio (95% CI)
Within 7 days of first actual or assigned vaccination date	6,121 (19.7%)	5,789 (18.6%)	1.07 (1.03-1.12)
Within 14 days of first actual or assigned vaccination date	8,783 (28.3%)	8,126 (26.2%)	1.11 (1.07-1.15)
Within 21 days of first actual or assigned vaccination date	10,478 (33.7%)	9,673 (31.1%)	1.13 (1.09-1.16)

Table 1. Rates of return to clinic in vaccinated and unvaccinated individuals after first vaccination. The first two columns give the numbers and percent of individuals in each group who had at least one phenotype-containing EHR note with 7, 14, or 21 days of the actual or assigned first vaccination date. To assess the magnitude and significance of difference between the follow-up rates, the odds ratio (OR) and corresponding 95% CI are shown. With the null hypothesis that the OR falls between 0.91 and 1.1, a difference was considered significant if the upper bound of 95% CI was less than 0.91 or the lower bound of the 95% CI was greater than 1.1.

	Vaccinated individuals with at least one note (%) [n = 17,067]	Unvaccinated individuals with at least one note (%) [n = 17,067]	Odds Ratio (95% CI)
Within 7 days of second actual or assigned vaccination date	2,513 (14.7%)	2,778 (16.3%)	0.89 (0.84-0.94)
Within 14 days of second actual or assigned vaccination date	3,324 (19.5%)	3,681 (21.6%)	0.88 (0.83-0.93)
Within 21 days of second actual or assigned vaccination date	3,722 (21.8%)	4,080 (23.9%)	0.89 (0.84-0.93)

Table 2. Rates of return to clinic in vaccinated and unvaccinated individuals after second vaccination. The first two columns give the numbers and percent of individuals in each group who had at least one phenotype-containing EHR note with 7, 14, or 21 days of the actual or assigned first vaccination date. To assess the difference in follow-up rates between groups, the odds ratio (OR) and corresponding 95% CI are shown. With the null hypothesis that the OR falls between 0.91 and 1.1, a difference was considered significant if the upper bound of 95% CI was less than 0.91 or the lower bound of the 95% CI was greater than 1.1.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	9 / 216,929 *[0.041]*	14 / 216,570 *[0.065]*	0.64 (0.25, 1.6)
Vomiting	0 / 216,955 *[0]*	2 / 216,599 *[0.0092]*	0 (0, 5.3)
Anaphylaxis	1 / 216,949 *[0.0046]*	0 / 216,604 *[0]*	0 (0.026, Inf)
Myalgia	125 / 216,475 *[0.58]*	135 / 216,128 *[0.62]*	0.92 (0.72, 1.2)
Lymphadenopathy	66 / 216,745 *[0.3]*	119 / 216,197 *[0.55]*	0.55 (0.4, 0.75)
Erythema	159 / 216,375 *[0.73]*	184 / 215,984 *[0.85]*	0.86 (0.69, 1.1)
Diarrhea	179 / 216,331 *[0.83]*	266 / 215,678 *[1.2]*	0.67 (0.55, 0.81)
Arthralgia	183 / 216,287 *[0.85]*	171 / 216,046 *[0.79]*	1.1 (0.86, 1.3)
Headache	11 / 216,923 *[0.051]*	9 / 216,588 *[0.042]*	1.2 (0.46, 3.3)
Facial paralysis	1 / 216,948 *[0.0046]*	3 / 216,593 *[0.014]*	0.33 (0.0063, 4.1)
Fever	85 / 216,732 *[0.39]*	198 / 216,048 *[0.92]*	0.43 (0.33, 0.55)

Table 3. Incidence rates of adverse effects in the seven days following the date of first vaccination. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts (n = 31,069 each) as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 7-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 85 cases of fever were recorded over a total of 216,732 person-days (contributed by 31,069 vaccinated individuals), corresponding to an incidence rate of 0.39 cases per 1000 person-days.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	0 / 119,409 *[0]*	5 / 119,144 *[0.042]*	0 (0, 1.1)
Vomiting	0 / 119,409 *[0]*	0 / 119,162 *[0]*	NA
Anaphylaxis	0 / 119,409 *[0]*	1 / 119,155 *[0.0084]*	0 (0, 39)
Myalgia	58 / 119,191 *[0.49]*	62 / 118,931 *[0.52]*	0.93 (0.64, 1.4)
Lymphadenopathy	23 / 119,324 *[0.19]*	56 / 118,930 *[0.47]*	0.41 (0.24, 0.68)
Erythema	53 / 119,197 *[0.44]*	93 / 118,820 *[0.78]*	0.57 (0.4, 0.8)
Diarrhea	56 / 119,180 *[0.47]*	120 / 118,745 *[1.0]*	0.46 (0.33, 0.64)
Arthralgia	67 / 119,124 *[0.56]*	81 / 118,840 *[0.68]*	0.83 (0.59, 1.2)
Headache	5 / 119,385 *[0.042]*	6 / 119,136 *[0.05]*	0.83 (0.2, 3.3)
Facial paralysis	0 / 119,409 *[0]*	1 / 119,161 *[0.0084]*	0 (0, 39)
Fever	31 / 119,293 *[0.26]*	109 / 118,774 *[0.92]*	0.28 (0.18, 0.43)

Table 4. Incidence rates of adverse effects in the seven days following the date of second vaccination. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts (n = 17,067 each) as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 7-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 31 cases of fever were recorded over a total of 119,293 person-days (contributed by 17,067 vaccinated individuals), corresponding to an incidence rate of 0.26 cases per 1000 person-days.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	1 / 9,949 *[0.1]*	2 / 9,938 *[0.2]*	0.5 (0.0085, 9.6)
Vomiting	0 / 9,955 *[0]*	1 / 9,936 *[0.1]*	0 (0, 39)
Anaphylaxis	0 / 9,955 *[0]*	0 / 9,941 *[0]*	NA
Myalgia	33 / 9,825 *[3.4]*	29 / 9,829 *[3]*	1.1 (0.67, 1.9)
Lymphadenopathy	19 / 9,884 *[1.9]*	31 / 9,835 *[3.2]*	0.61 (0.33, 1.1)
Erythema	37 / 9,812 *[3.8]*	52 / 9,776 *[5.3]*	0.71 (0.45, 1.1)
Diarrhea	30 / 9,847 *[3.0]*	64 / 9,770 *[6.6]*	0.47 (0.29, 0.73)
Arthralgia	50 / 9,782 *[5.1]*	41 / 9,780 *[4.2]*	1.2 (0.79, 1.9)
Headache	1 / 9,955 *[0.1]*	5 / 9,927 *[0.5]*	0.2 (0.0042, 1.8)
Facial paralysis	0 / 9,955 *[0]*	0 / 9,941 *[0]*	NA
Fever	23 / 9,901 *[2.3]*	39 / 9,813 *[4]*	0.58 (0.33, 1)

Table 5. Incidence rates of adverse effects in the seven days following the date of first vaccination for all individuals with at least one clinical note recorded during this time. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts (n = 1,432 each) as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 7-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 23 cases of fever were recorded over a total of 9,901 person-days (contributed by 1,432 vaccinated individuals), corresponding to an incidence rate of 2.3 cases per 1000 person-days.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	0 / 6,595 *[0]*	2 / 6,529 *[0.31]*	0 (0, 5.3)
Vomiting	0 / 6,595 *[0]*	0 / 6,534 *[0]*	NA
Anaphylaxis	0 / 6,595 *[0]*	1 / 6,527 *[0.15]*	0 (0, 39)
Myalgia	19 / 6,514 *[2.9]*	7 / 6,513 *[1.1]*	2.7 (1.1, 7.6)
Lymphadenopathy	6 / 6,572 *[0.91]*	13 / 6,479 *[2.0]*	0.46 (0.14, 1.3)
Erythema	13 / 6,543 *[2.0]*	16 / 6,467 *[2.5]*	0.8 (0.36, 1.8)
Diarrhea	15 / 6,532 *[2.3]*	24 / 6,441 *[3.7]*	0.62 (0.3, 1.2)
Arthralgia	20 / 6,533 *[3.1]*	17 / 6,467 *[2.6]*	1.2 (0.58, 2.4)
Headache	2 / 6,588 *[0.3]*	0 / 6,534 *[0]*	Inf (0.19, inf)
Facial paralysis	0 / 6,595 *[0]*	0 / 6,534 *[0]*	NA
Fever	11 / 6,555 *[1.7]*	22 / 6,475 *[3.4]*	0.49 (0.22, 1.1)

Table 6. Incidence rates of adverse effects in the seven days following the date of second vaccination for all individuals with at least one clinical note recorded during this time. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts (n = 944 each) as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 7-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 11 cases of fever were recorded over a total of 6,555 person-days (contributed by 944 vaccinated individuals), corresponding to an incidence rate of 1.7 cases per 1000 person-days.

Supplemental Material

Time Interval	Matched Pairs With At Least One EHR Note After First Dose	Matched Pairs With At Least One EHR Note After First Dose
Within 7 Days	1,432	944
Within 14 Days	2,690	1,843
Within 21 Days	3,606	2,532

Table S1. Number of matched pairs in which both the vaccinated individual and their propensity matched unvaccinated individual had at least one clinical note in the specified time intervals. The first column gives the time interval: within 7, 14, or 21 days. The second column gives the number of matched pairs with at least one clinical note in the defined time interval after the first actual or assigned vaccination date. The third column gives the number of matched pairs with at least one clinical note in the defined time interval after the second actual or assigned vaccination date.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases/Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	13 / 433,168 *[0.03]*	23 / 431,695 *[0.053]*	0.56 (0.26, 1.2)
Vomiting	0 / 433,261 *[0]*	2 / 431,807 *[0.0046]*	0 (0, 5.3)
Anaphylaxis	1 / 433,248 *[0.0023]*	0 / 431,826 *[0]*	Inf (0.026, inf)
Myalgia	215 / 431,687 *[0.5]*	224 / 430,256 *[0.52]*	0.96 (0.79, 1.2)
Lymphadenopathy	117 / 432,474 *[0.27]*	185 / 430,446 *[0.43]*	0.63 (0.5, 0.8)
Erythema	261 / 431,351 *[0.61]*	305 / 429,701 *[0.71]*	0.85 (0.72, 1)
Diarrhea	298 / 431,156 *[0.69]*	417 / 428,948 *[0.97]*	0.71 (0.61, 0.83)
Arthralgia	294 / 431,037 *[0.68]*	268 / 429,945 *[0.62]*	1.1 (0.92, 1.3)
Headache	14 / 433,151 *[0.032]*	11 / 431,746 *[0.025]*	1.3 (0.54, 3.1)
Facial paralysis	2 / 433,241 *[0.0046]*	4 / 431,794 *[0.0093]*	0.5 (0.045, 3.5)
Fever	159 / 432,489 *[0.37]*	333 / 430,131 *[0.77]*	0.47 (0.39, 0.58)

Table S2. Incidence rates of adverse effects in the 14 days following the date of first vaccination. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts (n = 31,069 each) as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 14-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 159 cases of fever were recorded over a total of 432,489 person-days (contributed by 31,069 vaccinated individuals), corresponding to an incidence rate of 0.37 cases per 1000 person-days.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	1 / 238,784 *[0.0042]*	8 / 237,924 *[0.034]*	0.12 (0.0028, 0.93)
Vomiting	0 / 238,786 *[0]*	0 / 237,980 *[0]*	NA
Anaphylaxis	0 / 238,786 *[0]*	1 / 237,966 *[0.0042]*	0 (0, 39)
Myalgia	75 / 238,150 *[0.31]*	87 / 237,293 *[0.37]*	0.86 (0.62, 1.2)
Lymphadenopathy	33 / 238,523 *[0.14]*	77 / 237,331 *[0.32]*	0.43 (0.27, 0.65)
Erythema	71 / 238,162 *[0.3]*	136 / 236,945 *[0.57]*	0.52 (0.38, 0.7)
Diarrhea	75 / 238,137 *[0.31]*	184 / 236,656 *[0.78]*	0.41 (0.31, 0.53)
Arthralgia	88 / 238,004 *[0.37]*	121 / 237,042 *[0.51]*	0.72 (0.54, 0.96)
Headache	7 / 238,720 *[0.029]*	7 / 237,919 *[0.029]*	1 (0.3, 3.3)
Facial paralysis	0 / 238,786 *[0]*	1 / 237,972 *[0.0042]*	0 (0, 39)
Fever	39 / 238,494 *[0.16]*	155 / 236,936 *[0.65]*	0.25 (0.17, 0.36)

Table S3. Incidence rates of adverse effects in the 14 days following the date of second vaccination. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts (n = 17,067 each) as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 14-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 39 cases of fever were recorded over a total of 238,494 person-days (contributed by 17,607 vaccinated individuals), corresponding to an incidence rate of 0.16 cases per 1000 person-days.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	15 / 649,105 *[0.023]*	32 / 645,889 *[0.05]*	0.47 (0.23, 0.89)
Vomiting	0 / 649,272 *[0]*	3 / 646,114 *[0.0046]*	0 (0, 2.4)
Anaphylaxis	1 / 649,252 *[0.0015]*	1 / 646,151 *[0.0015]*	1 (0.013, 78)
Myalgia	276 / 646,212 *[0.43]*	308 / 643,007 *[0.48]*	0.89 (0.76, 1.1)
Lymphadenopathy	156 / 647,675 *[0.24]*	229 / 643,521 *[0.36]*	0.68 (0.55, 0.83)
Erythema	345 / 645,406 *[0.53]*	407 / 641,933 *[0.63]*	0.84 (0.73, 0.98)
Diarrhea	373 / 645,246 *[0.58]*	556 / 640,495 *[0.87]*	0.67 (0.58, 0.76)
Arthralgia	386 / 644,814 *[0.6]*	354 / 642,331 *[0.55]*	1.1 (0.94, 1.3)
Headache	18 / 649,075 *[0.028]*	16 / 645,986 *[0.025]*	1.1 (0.54, 2.3)
Facial paralysis	2 / 649,238 *[0.0031]*	5 / 646,089 *[0.0077]*	0.4 (0.038, 2.4)
Fever	207 / 647,778 *[0.32]*	451 / 642,830 *[0.7]*	0.46 (0.38, 0.54)

Table S4. Incidence rates of adverse effects in the 21 days following the date of first vaccination. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts (n = 31,069 each) as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 21-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 207 cases of fever were recorded over a total of 647,778 person-days (contributed by 31,609 vaccinated individuals), corresponding to an incidence rate of 0.32 cases per 1000 person-days.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	1 / 358,136 *[0.0028]*	11 / 356,501 *[0.031]*	0.09 (0.0021, 0.62)
Vomiting	0 / 358,145 *[0]*	0 / 356,598 *[0]*	NA
Anaphylaxis	0 / 358,145 *[0]*	1 / 356,577 *[0.0028]*	0 (0, 39)
Myalgia	83 / 356,988 *[0.23]*	107 / 355,268 *[0.3]*	0.77 (0.57, 1)
Lymphadenopathy	39 / 357,634 *[0.11]*	88 / 355,416 *[0.25]*	0.44 (0.29, 0.65)
Erythema	87 / 356,976 *[0.24]*	153 / 354,666 *[0.43]*	0.56 (0.43, 0.74)
Diarrhea	82 / 356,986 *[0.23]*	229 / 353,988 *[0.65]*	0.36 (0.27, 0.46)
Arthralgia	105 / 356,711 *[0.29]*	148 / 354,786 *[0.42]*	0.71 (0.54, 0.91)
Headache	7 / 358,030 *[0.02]*	9 / 356,491 *[0.025]*	0.77 (0.25, 2.3)
Facial paralysis	0 / 358,145 *[0]*	1 / 356,583 *[0.0028]*	0 (0, 39)
Fever	42 / 357,615 *[0.12]*	178 / 354,690 *[0.5]*	0.23 (0.16, 0.33)

Table S5. Incidence rates of adverse effects in the 21 days following the date of second vaccination. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts (n = 17,067 each) as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 21-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 42 cases of fever were recorded over a total of 357,615 person-days (contributed by 17,607 vaccinated individuals), corresponding to an incidence rate of 0.12 cases per 1000 person-days.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	3 / 37,230 *[0.081]*	10 / 36,886 *[0.27]*	0.3 (0.053, 1.2)
Vomiting	0 / 37,249 *[0]*	1 / 36,924 *[0.027]*	0 (0, 39)
Anaphylaxis	1 / 37,236 *[0.027]*	0 / 36,936 *[0]*	0 (0.025, Inf)
Myalgia	68 / 36,704 *[1.9]*	75 / 36,426 *[2.1]*	0.9 (0.64, 1.3)
Lymphadenopathy	45 / 36,945 *[1.2]*	68 / 36,477 *[1.9]*	0.65 (0.44, 0.97)
Erythema	78 / 36,693 *[2.1]*	97 / 36,243 *[2.7]*	0.79 (0.58, 1.1)
Diarrhea	68 / 36,783 *[1.8]*	135 / 36,091 *[3.7]*	0.49 (0.36, 0.67)
Arthralgia	106 / 36,529 *[2.9]*	84 / 36,370 *[2.3]*	1.3 (0.93, 1.7)
Headache	5 / 37,226 *[0.13]*	6 / 36,883 *[0.16]*	0.83 (0.2, 3.2)
Facial paralysis	0 / 37,249 *[0]*	1 / 36,925 *[0.027]*	0 (0, 39)
Fever	50 / 37,000 *[1.4]*	104 / 36,362 *[2.9]*	0.47 (0.33, 0.67)

Table S6. Incidence rates of adverse effects in the 14 days following the date of first vaccination for all individuals with at least one clinical note recorded during this time. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts (n = 2,690 each) as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 14-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 50 cases of fever were recorded over a total of 37,000 person-days (contributed by 2,690 vaccinated individuals), corresponding to an incidence rate of 1.4 cases per 1000 person-days.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	1 / 25,755 *[0.039]*	4 / 25,446 *[0.16]*	0.25 (0.005, 2.5)
Vomiting	0 / 25,757 *[0]*	0 / 25,475 *[0]*	NA
Anaphylaxis	0 / 25,757 *[0]*	1 / 25,461 *[0.039]*	0 (0, 39)
Myalgia	34 / 25,466 *[1.3]*	24 / 25,312 *[0.95]*	1.4 (0.81, 2.5)
Lymphadenopathy	12 / 25,652 *[0.47]*	23 / 25,262 *[0.91]*	0.51 (0.23, 1.1)
Erythema	21 / 25,575 *[0.82]*	31 / 25,247 *[1.2]*	0.67 (0.37, 1.2)
Diarrhea	20 / 25,575 *[0.78]*	55 / 25,066 *[2.2]*	0.36 (0.2, 0.6)
Arthralgia	30 / 25,476 *[1.2]*	31 / 25,204 *[1.2]*	0.96 (0.56, 1.6)
Headache	4 / 25,729 *[0.16]*	3 / 25,462 *[0.12]*	1.3 (0.22, 9)
Facial paralysis	0 / 25,757 *[0]*	0 / 25,475 *[0]*	NA
Fever	17 / 25,632 *[0.66]*	43 / 25,193 *[1.7]*	0.39 (0.21, 0.7)

Table S7. Incidence rates of adverse effects in the 14 days following the date of second vaccination for all individuals with at least one clinical note recorded during this time. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts (n = 1,843 each) as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 14-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 17 cases of fever were recorded over a total of 25,632 person-days (contributed by 1,843 vaccinated individuals), corresponding to an incidence rate of 0.66 cases per 1000 person-days.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	3 / 74,692 *[0.04]*	14 / 73,719 *[0.19]*	0.21 (0.039, 0.76)
Vomiting	0 / 74,725 *[0]*	1 / 73,818 *[0.014]*	0 (0, 39)
Anaphylaxis	1 / 74,705 *[0.013]*	1 / 73,837 *[0.014]*	0.99 (0.013, 78)
Myalgia	102 / 73,604 *[1.4]*	106 / 72,729 *[1.5]*	0.95 (0.72, 1.3)
Lymphadenopathy	68 / 73,953 *[0.92]*	92 / 72,791 *[1.3]*	0.73 (0.52, 1)
Erythema	127 / 73,446 *[1.7]*	157 / 72,262 *[2.2]*	0.8 (0.62, 1)
Diarrhea	114 / 73,643 *[1.5]*	194 / 72,001 *[2.7]*	0.57 (0.45, 0.73)
Arthralgia	143 / 73,061 *[2.0]*	128 / 72,499 *[1.8]*	1.1 (0.87, 1.4)
Headache	6 / 74,666 *[0.08]*	10 / 73,719 *[0.14]*	0.59 (0.18, 1.8)
Facial paralysis	0 / 74,725 *[0]*	1 / 73,819 *[0.014]*	0 (0, 39)
Fever	77 / 74,126 *[1.0]*	155 / 72,595 *[2.1]*	0.49 (0.37, 0.64)

Table S8. Incidence rates of adverse effects in the 21 days following the date of first vaccination for all individuals with at least one clinical note recorded during this time. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts (n = 3,606 each) as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 21-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 77 cases of fever were recorded over a total of 74,126 person-days (contributed by 3,606 vaccinated individuals), corresponding to an incidence rate of 1.0 cases per 1000 person-days.

Adverse effect	**Vaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Unvaccinated Incidence Rate** Cases / Person-Days *[Cases Per 1000 Person-Days]*	**Incidence Rate Ratio** (95% CI)
Fatigue	1 / 53,057 *[0.019]*	4 / 52,356 *[0.076]*	0.25 (0.005, 2.5)
Vomiting	0 / 53,066 *[0]*	0 / 52,406 *[0]*	NA
Anaphylaxis	0 / 53,066 *[0]*	1 / 52,385 *[0.019]*	0 (0, 38)
Myalgia	42 / 52,493 *[0.8]*	40 / 51,918 *[0.77]*	1 (0.66, 1.6)
Lymphadenopathy	15 / 52,871 *[0.28]*	35 / 51,925 *[0.67]*	0.42 (0.21, 0.79)
Erythema	31 / 52,654 *[0.59]*	50 / 51,793 *[0.97]*	0.61 (0.38, 0.97)
Diarrhea	36 / 52,580 *[0.68]*	86 / 51,471 *[1.7]*	0.41 (0.27, 0.61)
Arthralgia	44 / 52,412 *[0.84]*	44 / 51,831 *[0.85]*	0.99 (0.64, 1.5)
Headache	4 / 53,010 *[0.075]*	6 / 52,341 *[0.11]*	0.66 (0.14, 2.8)
Facial paralysis	0 / 53,066 *[0]*	0 / 52,406 *[0]*	NA
Fever	21 / 52,790 *[0.4]*	60 / 51,787 *[1.2]*	0.34 (0.2, 0.57)

Table S9. Incidence rates of adverse effects in the 21 days following the date of second vaccination for all individuals with at least one clinical note recorded during this time. For each adverse effect, incidence rates were calculated for the vaccinated and propensity matched unvaccinated cohorts as the number of positive cases divided by the total number of at-risk person days during this time period. Individuals were considered at risk for developing an adverse effect from their actual or assigned date of first vaccination until they experienced the event, died, or reached the end of the 21-day study period, or until four days prior to a positive SARS-CoV-2 test. For example, in the last row, we show that 21 cases of fever were recorded over a total of 52,790 person-days (contributed by 2,532 vaccinated individuals), corresponding to an incidence rate of 0.4 cases per 1000 person-days.