

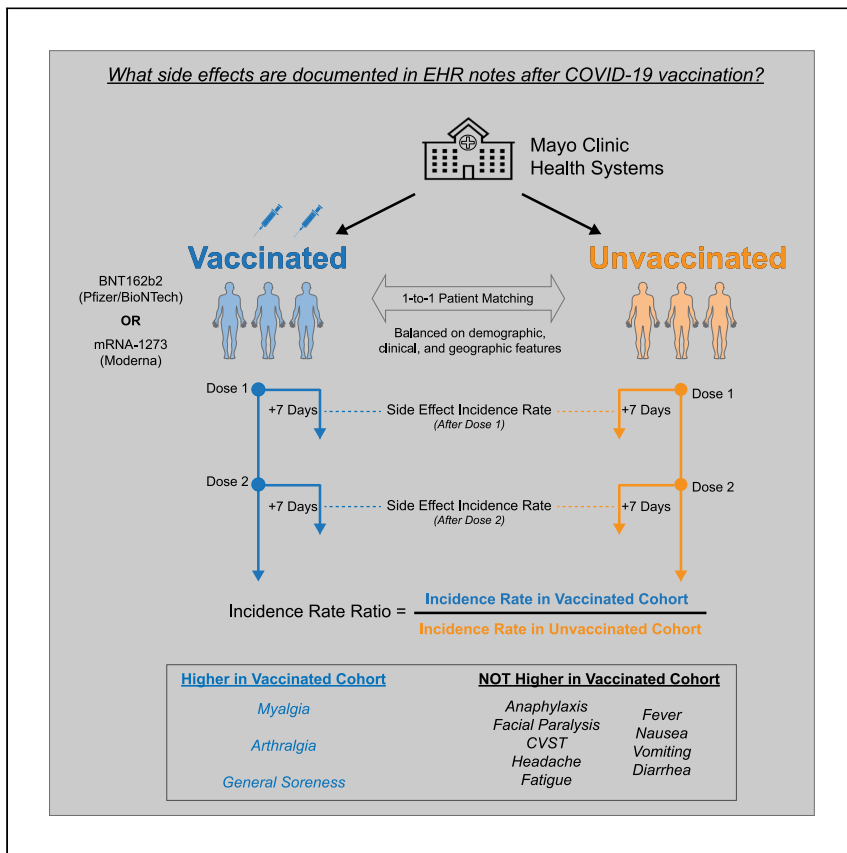


Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Clinical Advances

# Real-time analysis of a mass vaccination effort confirms the safety of FDA-authorized mRNA COVID-19 vaccines



Reid McMurry, Patrick Lenehan, Samir Awasthi, ..., Tyler Wagner, Ajit Rajasekharan, Venky Soundararajan

venky@ference.net

Highlights

Emergent clinical visits are not increased after receiving BNT162b2 or mRNA-1273

Side effect reports are rare in EHR notes compared to clinical trials and V-safe

Myalgia and arthralgia are increased after vaccination with BNT162b2 and mRNA-1273

Severe adverse effects are rare among individuals receiving BNT162b2 or mRNA-1273

McMurry et al. assess the real-world safety of the BNT162b2 and mRNA-1273 COVID-19 vaccines. Using natural language processing, they compare the rates of specified adverse effects between 68,266 vaccinated individuals and 68,266 matched unvaccinated individuals. They find that both vaccines are safe and tolerated in clinical practice.



## Clinical Advances

## Real-time analysis of a mass vaccination effort confirms the safety of FDA-authorized mRNA COVID-19 vaccines

Reid McMurry,<sup>1,4</sup> Patrick Lenehan,<sup>1,4</sup> Samir Awasthi,<sup>1,4</sup> Eli Silvert,<sup>1</sup> Arjun Puranik,<sup>1</sup> Colin Pawlowski,<sup>1</sup> A.J. Venkatakrishnan,<sup>1</sup> Praveen Anand,<sup>2</sup> Vineet Agarwal,<sup>1</sup> John C. O'Horo,<sup>3</sup> Gregory J. Gores,<sup>3</sup> Amy W. Williams,<sup>3</sup> Andrew D. Badley,<sup>3</sup> John Halamka,<sup>3</sup> Abinash Virk,<sup>3</sup> Melanie D. Swift,<sup>3</sup> Katie Carlson,<sup>1</sup> Deeksha Doddahonnaiah,<sup>1</sup> Anna Metzger,<sup>1</sup> Nikhil Kayal,<sup>1</sup> Gabi Berner,<sup>1</sup> Eshwan Ramudu,<sup>1</sup> Corinne Carpenter,<sup>1</sup> Tyler Wagner,<sup>1</sup> Ajit Rajasekharan,<sup>1</sup> and Venky Soundararajan<sup>1,2,5,\*</sup>

## SUMMARY

**Background:** As the coronavirus disease 2019 (COVID-19) vaccination campaign unfolds, it is important to continuously assess the real-world safety of Food and Drug Administration (FDA)-authorized vaccines. Curation of large-scale electronic health records (EHRs) enables near-real-time safety evaluations that were not previously possible.

**Methods:** In this retrospective study, we deployed deep neural networks over a large EHR system to automatically curate the adverse effects mentioned by physicians in over 1.2 million clinical notes between December 1, 2020 and April 20, 2021. We compared notes from 68,266 individuals who received at least one dose of BNT162b2 ( $n = 51,795$ ) or mRNA-1273 ( $n = 16,471$ ) to notes from 68,266 unvaccinated individuals who were matched by demographic, geographic, and clinical features.

**Findings:** Individuals vaccinated with BNT162b2 or mRNA-1273 had a higher rate of return to the clinic, but not the emergency department, after both doses compared to unvaccinated controls. The most frequently documented adverse effects within 7 days of each vaccine dose included myalgia, headache, and fatigue, but the rates of EHR documentation for each side effect were remarkably low compared to those derived from active solicitation during clinical trials. Severe events, including anaphylaxis, facial paralysis, and cerebral venous sinus thrombosis, were rare and occurred at similar frequencies in vaccinated and unvaccinated individuals.

**Conclusions:** This analysis of vaccine-related adverse effects from over 1.2 million EHR notes of more than 130,000 individuals reaffirms the safety and tolerability of the FDA-authorized mRNA COVID-19 vaccines in practice.

**Funding:** This study was funded by nference.

## INTRODUCTION

Following their Emergency Use Authorizations by the Food and Drug Administration (FDA) in December of 2020, more than 280 million doses of BNT162b2 (Pfizer/BioNTech) and mRNA-1273 (Moderna) coronavirus disease 2019 (COVID-19) vaccines have been administered in the United States.<sup>1–3</sup> Phase 3 trials demonstrated strong efficacy and favorable safety profiles for these vaccines in the cohorts studied. Specifically, the trials showed 95.0% (95% confidence interval [CI], 90.3 to 97.6) and

## Context and significance

This is a study of the mRNA COVID-19 vaccines developed by Pfizer/BioNTech and Moderna. Although these vaccines have been shown to be safe and tolerated in clinical trials, it is important to confirm their safety profiles in practice. The results from this study show that individuals receiving these vaccines are likely to experience muscle and joint soreness, but they are not more likely to seek out emergent clinical care or experience severe medical events than unvaccinated individuals. As one of the largest real-world safety studies of COVID-19 vaccines to date, these data reinforce that we should continue expanding efforts to deliver more vaccines with high confidence in their safety.



94.1% (95% CI, 89.3 to 96.8%) efficacy for BNT162b2 and mRNA-1273, respectively. Self-resolving mild to moderate adverse effects were common in vaccinated participants, while serious adverse effects occurred rarely and with a frequency comparable to placebo.<sup>4,5</sup> 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%).<sup>4,5</sup>

Consistent with Centers for Disease Control and Prevention (CDC) recommendations, early vaccination efforts (phase 1a) in the United States targeted healthcare workers along with residents and staff at long-term care facilities, who are at elevated risk to be infected with SARS-CoV-2 or experience severe COVID-19 relative to the general public.<sup>6–9</sup> As the vaccines continue to be administered more broadly, it is critical to continuously evaluate real-world safety and effectiveness 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 trial experience with respect to vaccine efficacy and adverse effects, and it will help to 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 the authorization of both vaccines.<sup>10–14</sup>

As part of the Vaccine Adverse Events Reporting System (VAERS) maintained by the FDA, V-safe has recently been established to specifically monitor the safety profiles of COVID-19 vaccines as they are administered to the public.<sup>15</sup> V-safe is a voluntary program into which vaccinated individuals can enroll. Once enrolled, participants receive frequent reminders to complete surveys which document their side effects electronically. Thus, this program will create an excellent resource of clinical-trial-like safety data derived from significantly larger and more diverse patient populations than those who were enrolled in the phase 1/2 or 3 trials.

An important complementary 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 we have recently applied these methods to assess the real-world effectiveness of both BNT162b2 and mRNA-1273 in over 68,000 individuals receiving these vaccines within the Mayo Clinic health system.<sup>16–20</sup> Here, we expand on these efforts to study the adverse effects experienced by these individuals after COVID-19 vaccination in the real-world setting.

It should be noted that the monitoring of vaccine-associated adverse effects in clinical trials or post-marketing surveillance efforts like V-safe is quite different from such monitoring in routine clinical practice. In clinical trials, participants are aware that they may be receiving an experimental product, and adverse effect reporting is encouraged or solicited. Such solicitation is likely to increase the rate of reported adverse effects, as is demonstrated by the high percentage of placebo-treated participants who report adverse effects. Similarly, as was previously noted, side effect reporting is actively solicited from V-safe participants via frequent surveys. On the other hand, individuals receiving a COVID-19 vaccine during the mass vaccination campaign are informed that they are likely to experience certain adverse effects and can even be discouraged from seeking medical attention unless the symptom

<sup>1</sup>Inference, One Main Street, East Arcade, Cambridge, MA 02142, USA

<sup>2</sup>Inference Labs, 2nd Floor, 22 3rd Cross Rd, Murgesh Pallya, Bengaluru, Karnataka 560017, India

<sup>3</sup>Mayo Clinic, Rochester, MN 55902, USA

<sup>4</sup>These authors contributed equally

<sup>5</sup>Lead contact

\*Correspondence: [venky@inference.net](mailto:venky@inference.net)  
<https://doi.org/10.1016/j.medj.2021.06.006>

is particularly severe. Thus, adverse effects are likely to be captured in EHR notes only for individuals who experience symptoms that 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 recapitulate the adverse effect frequencies reported in prior trials. Instead, it is to establish the rates at which individuals actually report potential vaccine-associated adverse effects to healthcare practitioners (HCPs) and determine whether these rates of adverse effect reporting 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 one-to-one propensity matching procedure to derive a cohort of unvaccinated individuals who are balanced for demographic factors, residential location, history of SARS-CoV-2 and influenza testing, and current residence at long-term care facilities (see [STAR Methods](#)), and we curated their clinical notes over specified time periods to quantify the frequency of the defined symptoms of interest.<sup>21</sup> 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. The results of this study support the safety and tolerability of both BNT162b2 and mRNA-1273 in clinical practice, further strengthening the case for the rapid and broad distribution of these vaccines to the public.

## RESULTS

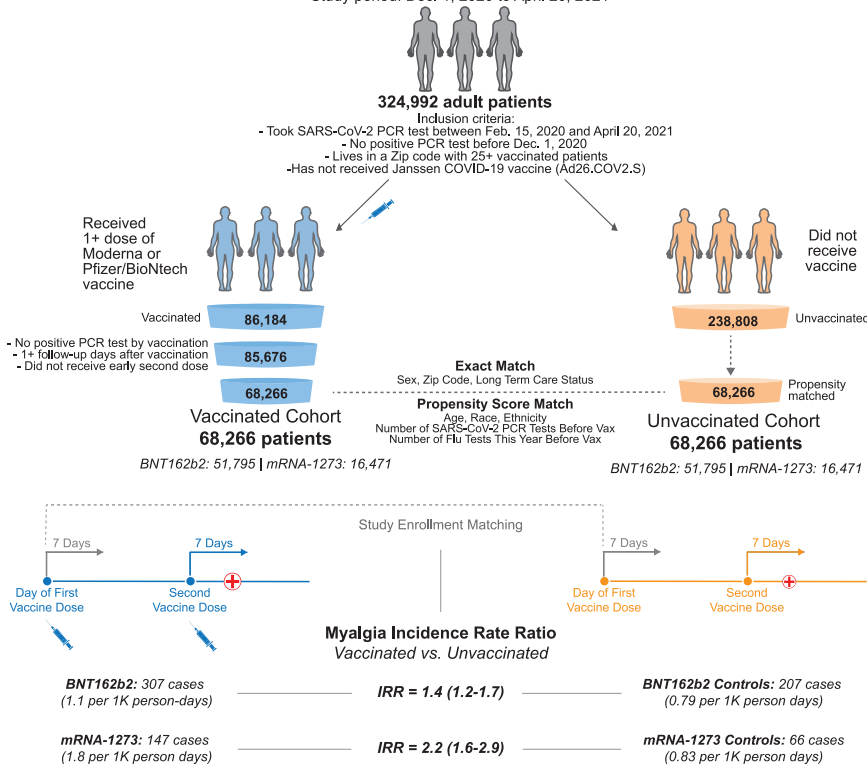
### Vaccinated individuals are more likely to return to clinic after both doses

Of the 324,992 eligible adults in the Mayo Clinic health system, 85,676 had received at least one dose of an mRNA COVID-19 vaccine at the time of this study. We were able to match 68,266 of these vaccinated individuals ( $n_{\text{BNT162b2}} = 51,795$ ;  $n_{\text{mRNA-1273}} = 16,471$ ) to the same number of unvaccinated individuals with similar demographic, clinical, and geographic features ([Figure 1](#); [Tables S1](#) and [S2](#)). 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, 9,998 of the 51,795 (19.3%) individuals receiving BNT162b2 had at least one EHR note compared to 7,457 of 51,795 (14.4%) unvaccinated individuals (odds ratio, 1.42; 95% CI, [1.38–1.47]) ([Table 1](#)). The rates of return to clinic were also moderately higher in BNT162b2-vaccinated individuals within 14 and 21 days of the first dose (odds ratio<sub>14 days</sub>, 1.47 [1.43–1.51]; odds ratio<sub>21 days</sub>, 1.49 [1.45–1.53]) ([Table 1](#)).

Individuals receiving mRNA-1273 were also more likely to return to the clinic after their first dose compared to matched unvaccinated controls. Within 7 days of the first dose, 3,891 of 16,471 (23.62%) vaccinated individuals had at least one EHR note compared to 2,734 of 16,471 (16.6%) unvaccinated individuals (odds ratio, 1.55 [1.47–1.64]) ([Table 1](#)). Similar increases were also observed within 14 and 21 days of receiving the first dose of mRNA-1273 (odds ratio<sub>14 days</sub>, 1.71 [1.63–1.8]; odds ratio<sub>21 days</sub>, 1.81 [1.73–1.89]) ([Table 1](#)).

Among individuals receiving two doses of BNT162b2 ( $n = 39,058$ ), the rates of return to clinic were again moderately higher than those of unvaccinated controls within 7, 14, and 21 days of the second dose (odds ratio<sub>7 days</sub>, 1.46 [1.4–1.51]; odds ratio<sub>14 days</sub>, 1.49 [1.44–1.54]; odds ratio<sub>21 days</sub>, 1.53 [1.49–1.58]) ([Table 2](#)). Individuals

**Real-world safety analysis of the FDA-authorized mRNA COVID-19 vaccines**  
Study period: Dec. 1, 2020 to April 20, 2021



**Figure 1. Schematic illustration of participant selection and study design**

The vaccinated cohort is composed of 68,266 individuals from the Mayo Clinic and associated health systems who received at least one dose of BNT162b2 (Pfizer/BioNTech; n = 51,795) or mRNA-1273 (Moderna; n = 16,471) between December 1, 2020 and April 20, 2021 and did not test positive for SARS-CoV-2 prior to their first vaccination. A control cohort of unvaccinated individuals was generated via a combination of exact matching parameters and one-to-one propensity score matching, yielding 68,266 individuals with similar distributions of age, sex, race, ethnicity, residential location, number of prior influenza and SARS-CoV-2 PCR tests in the past year, and current long-term care residence status. For each cohort, the incidence rates of several adverse effects (e.g., myalgia) were calculated for the 7 days following the first dose and, separately, for the 7 days following the second dose. For a given adverse effect, the incidence rate ratio (IRR) and the corresponding 95% confidence interval (CI) were calculated to determine whether individuals receiving BNT162b2 or mRNA-1273 were more likely to experience the event during these intervals than their matched unvaccinated controls. Incidence rates and IRRs were also calculated for the 14 and 21 days following each vaccine dose.

receiving two doses of mRNA-1273 (n = 11,851) also had higher odds of contributing EHR notes in each time interval after the second dose (odds ratio<sub>7 days</sub>, 1.69 [1.58–1.8]; odds ratio<sub>14 days</sub>, 1.92 [1.81–2.03]; odds ratio<sub>21 days</sub>, 2.08 [1.97–2.2]) (Table 2).

**BNT162b2 and mRNA-1273 are not associated with increased rates of emergent clinical follow-up**

It is important to determine whether the observed vaccine-associated increases in clinical follow-up were due to acute or emergent events. To do so, we first compared the number of emergency department (ED) visits and notes contributed by each group in the 1, 7, 14, and 21 days after each vaccine dose. Individuals receiving either BNT162b2 or mRNA-1273 were not more likely to contribute at least one ED note during these intervals (Tables S3 and S4). For example, within 7 days of the first dose, 567 of 51,795 (1.09%) individuals vaccinated with

**Table 1. Rates of return to clinic in vaccinated and unvaccinated individuals after the first vaccine dose**

Vaccine	Time interval after first actual or assigned dose	Vaccinated individuals with at least one note (%)	Unvaccinated individuals with at least one note (%)	OR (95% CI)
BNT162b2 (51,795 individuals per cohort)	7 days	9,998 (19.3%)	7,457 (14.4%)	1.42 (1.38–1.47)
	14 days	15,162 (29.27%)	11,403 (22.02%)	1.47 (1.43–1.51)
	21 days	18,783 (36.26%)	14,332 (27.67%)	1.49 (1.45–1.53)
mRNA-1273 (16,471 individuals per cohort)	7 days	3,891 (23.62%)	2,734 (16.6%)	1.55 (1.47–1.64)
	14 days	5,956 (36.16%)	4,093 (24.85%)	1.71 (1.63–1.8)
	21 days	7,415 (45.02%)	5,133 (31.16%)	1.81 (1.73–1.89)

We show the numbers and percent of individuals in each group who had at least one phenotype-containing EHR note within 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 the 95% CI was less than 0.91 or the lower bound of the 95% CI was greater than 1.1.

BNT162b2 contributed an ED note compared to 752 of 51,795 (1.45%) unvaccinated individuals (odds ratio, 0.75 [0.67–0.84]). In the same period, 238 of 16,471 (1.44%) individuals receiving mRNA-1273 had at least one ED note versus 219 of 16,471 (1.33%) unvaccinated individuals (odds ratio, 1.09 [0.90–1.31]) (Table S3). These trends were similar within 7 days of the second dose for both BNT162b2 (odds ratio, 0.76 [0.66–0.87]) and mRNA-1273 (odds ratio, 1.14 [0.92–1.41]) (Table S4). Consistent with this, the total number of ED notes contributed by vaccinated individuals was similar or lower than the corresponding number contributed by their unvaccinated controls in each interval for both BNT162b2 and mRNA-1273 (Tables S5 and S6).

### Severe adverse effects are rare and observed at similar rates in vaccinated individuals and unvaccinated controls

Severe adverse events have been rarely observed in individuals who previously received COVID-19 vaccines, including anaphylaxis, facial paralysis, and cerebral venous sinus thrombosis (CVST).<sup>22–24</sup> We next performed augmented curation followed by manual review to identify occurrences of these events in the clinical notes of all vaccinated and unvaccinated individuals within 7, 14, or 21 days of each actual or assigned vaccine dose (see STAR Methods).

We did identify multiple instances of facial paralysis and anaphylaxis in these intervals, but these events were documented at similar or lower frequencies among individuals receiving BNT162b2 or mRNA-1273 compared to each set of unvaccinated controls (Figures 2 and 3; Tables 3, 4, and S7–S10). This supports a recent report showing that facial paralysis was not observed at disproportionately high rates among individuals receiving BNT162b2 or mRNA-1273 compared to individuals receiving influenza vaccines or other viral vaccines more broadly.<sup>25</sup> While there was one case of CVST identified within 14 days of an individual receiving BNT162b2, there were several cases documented among the unvaccinated cohorts (Tables S7–S10). We have also previously shown that the rate of CVST is similar in the 15 and 30 days prior to and after BNT162b2 or mRNA-1273 vaccination in a large cohort at the Mayo Clinic.<sup>26</sup> Taken together with the absence of an increase in emergent clinical visits following each dose, these data suggest favorable tolerability of both vaccines.

**Table 2. Rates of return to clinic in vaccinated and unvaccinated individuals after the second vaccine dose**

Vaccine	Time interval after second actual or assigned dose	Vaccinated individuals with at least one note (%)	Unvaccinated individuals with at least one note (%)	OR (95% CI)
BNT162b2 (39,058 individuals per cohort)	7 days	7,318 (18.74%)	5,342 (13.68%)	1.46 (1.4–1.51)
	14 days	10,784 (27.61%)	7,954 (20.36%)	1.49 (1.44–1.54)
	21 days	13,187 (33.76%)	9,750 (24.96%)	1.53 (1.49–1.58)
mRNA-1273 (11,851 individuals per cohort)	7 days	2,790 (23.54%)	1,828 (15.42%)	1.69 (1.58–1.8)
	14 days	4,261 (35.95%)	2,682 (22.63%)	1.92 (1.81–2.03)
	21 days	5,226 (44.1%)	3,260 (27.51%)	2.08 (1.97–2.2)

We show the numbers and percent of individuals in each group who had at least one phenotype-containing EHR note within 7, 14, or 21 days of the actual or assigned second vaccination date. To assess the magnitude and significance of difference between the follow-up rates, the 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 the 95% CI was less than 0.91 or the lower bound of the 95% CI was greater than 1.1.

**EHR documented frequencies of adverse effects are much lower than their solicited frequencies reported in clinical trials**

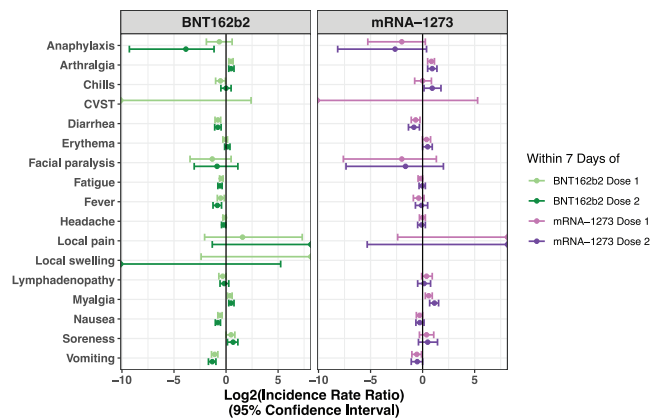
Among 51,732 individuals analyzed (Table S11), the most commonly documented symptoms within 7 days of the first BNT162b2 dose included headache (522 individuals [1.01%]), fatigue (470 [0.91%]), myalgia (381 [0.74%]), arthralgia (360 [0.70%]), nausea (332 [0.64%]), and erythema (273 [0.53%]) (Table 3). These same symptoms were the most commonly documented within 7 days of the first mRNA-1273 dose among the 16,455 analyzed individuals: fatigue (244 [1.48%]), headache (187 [1.14%]), arthralgia (188 [1.14%]), myalgia (165 [1.00%]), nausea (155 [0.94%]), and erythema (124 [0.75%]) (Table 3). These were also the most frequently observed symptoms within 7 days of the second dose for both BNT162b2 and mRNA-1273 (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%], and erythema [10%–15%]) and those captured in V-safe.<sup>4,5,15,27</sup> This is to be expected, as individuals vaccinated outside of the trial or post-marketing surveillance setting are advised that it is normal to experience these adverse effects, and so they are less likely to report them to a healthcare provider. As such, the vaccine-associated adverse effects that are captured in EHR notes are likely to be those that are severe or persistent enough to cause an individual to return to the clinic or otherwise notify their healthcare provider.

**Myalgia and arthralgia are documented more frequently in the EHR notes of vaccinated individuals than unvaccinated controls**

We next computed the incidence rate ratio (IRR) for each adverse effect between the vaccinated and propensity-matched unvaccinated cohorts (Figure 1). Within 7 days of the first dose, individuals receiving BNT162b2 were slightly more likely to have reported myalgia (IRR, 1.3 [1.1–1.5]) and arthralgia (IRR, 1.4 [1.2–1.6]), while other side effects were documented at similar rates between the cohorts (Figure 2; Table 3). This profile was similar for mRNA-1273, with only myalgia (IRR, 1.5 [1.2–1.9]) and arthralgia (IRR, 1.8 [1.4–2.2]) documented at moderately higher rates among vaccinated individuals within 7 days of the first dose (Figure 2; Table 3). These symptoms





**Figure 2. IRRs for all surveyed adverse effects in vaccinated versus matched unvaccinated cohorts within 7 days of each vaccine dose**

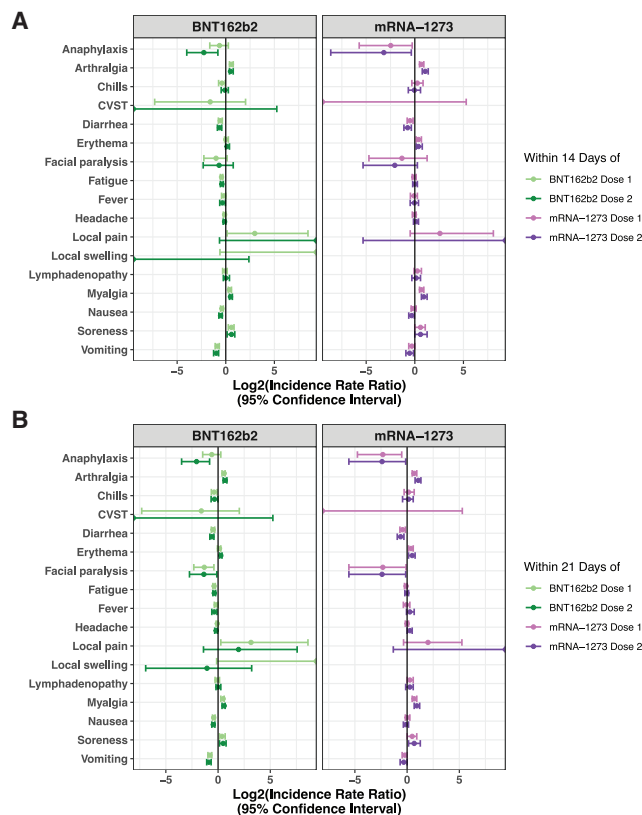
IRRs are shown with their corresponding 95% CIs for BNT162b2 on the left ( $n_{\text{dose 1}} = 51,732$ ;  $n_{\text{dose 2}} = 39,045$ ) and mRNA-1273 on the right ( $n_{\text{dose 1}} = 16,455$ ;  $n_{\text{dose 2}} = 11,849$ ). Data correspond to Tables 3 and 4. A  $\log_2(\text{IRR})$  value greater than 0 indicates that the incidence rate of the given event was higher in the vaccinated cohort, while a value less than 0 indicates that the incidence rate was higher in the unvaccinated cohort.

were also more frequently documented in the vaccinated cohorts in the 7 days following the second dose (Figure 2; Table 4) and in the 14 and 21 days following each dose (Figures 3A and 3B; Tables S7–S10). Other symptoms continued to show similar or lower incidence rates in vaccinated individuals compared to unvaccinated controls in the 14 and 21 days following each dose (Figures 3A and 3B; Tables S7–S10).

## DISCUSSION

This study demonstrates that the two currently FDA-authorized mRNA COVID-19 vaccines, BNT162b2 and mRNA-1273, are safe and tolerated in practice. This conclusion is consistent with the extensive safety and tolerability assessments conducted in phase 1/2 and 3 trials over the past year and the post-marketing surveillance efforts that are currently underway.<sup>4,5,15,28,29</sup> Here, we assessed real-world safety by longitudinally curating the EHR documentation of adverse effects in 68,266 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 more likely to be seen in the clinic, but not in the ED, within 7, 14, or 21 days after the first or second vaccine dose. Myalgia and arthralgia were documented more frequently in the EHR notes of individuals who received BNT162b2 or mRNA-1273 compared to their respective controls during each of these intervals, while the other surveyed symptoms were not. Importantly, severe events were rare and did not occur at unexpectedly high rates in either vaccinated cohort.

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, clinical, and geographic characteristics, akin to a placebo group in a randomized clinical trial. Our finding that EHR notes from vaccinated and propensity-matched unvaccinated individuals record similar rates of potential vaccine-associated adverse effects (other than myalgia and arthralgia) differs from the data obtained in phase 3 trials, wherein vaccinated participants experienced higher rates of several other symptoms than those receiving placebo. Further, the



**Figure 3. IRRs for all surveyed adverse effects in vaccinated versus matched unvaccinated cohorts within 14 or 21 days of each vaccine dose**

IRRs are shown with their corresponding 95% CIs for BNT162b2 ( $n_{\text{dose 1}} = 51,732$ ;  $n_{\text{dose 2}} = 39,045$ ) on the left and mRNA-1273 ( $n_{\text{dose 1}} = 16,455$ ;  $n_{\text{dose 2}} = 11,849$ ) on the right. (A) IRRs within 14 days of each vaccine dose, corresponding to data shown in [Tables S7](#) and [S8](#). (B) IRRs within 21 days of each vaccine dose, corresponding to data shown in [Tables S9](#) and [S10](#). A  $\log_2(\text{IRR})$  value greater than 0 indicates that the incidence rate of the given event was higher in the vaccinated cohort, while a value less than 0 indicates that the incidence rate was higher in the unvaccinated cohort. See also [Tables S7–S10](#).

absolute rates of adverse effects documented in these EHR notes are well below the rates reported in clinical trials or the active post-marketing surveillance effort (V-safe). In V-safe, the most frequently reported systemic adverse effects thus far include headache (26%–47%), fatigue (31%–54%), and myalgia (19%–44%).<sup>27</sup> These discrepancies are likely attributable to differences in the populations analyzed and the methodology of reporting symptoms.

Regarding the populations analyzed, it is important to realize that adverse effects experienced by healthcare workers, who are likely overrepresented in the vaccinated cohorts due to their inclusion in phase 1a of the vaccine rollout, are generally less likely to be documented in EHR notes. At many institutions, healthcare workers experiencing potential vaccine side effects are directed to follow up with Occupational Health Services, which is an entity of the employer and thus does not document the reported symptoms in EHR notes. Several institutions have also established COVID-19 response lines, which will usually direct individuals experiencing nonspecific vaccine side effects toward self-care. Further, healthcare workers may be less likely to report adverse effects at baseline given their clinical expertise and ability to self-assess the severity of their illness.

**Table 3. Incidence rates of adverse effects in the 7 days following the date of the first BNT162b2 or mRNA-1273 dose**

Vaccine	Adverse effect	Vaccinated incidence rate, cases/ person-days [cases per 1,000 person-days]	Unvaccinated incidence rate, cases/ person-days [cases per 1,000 person-days]	IRR (95% CI)
BNT162b2 (51,795 individuals per cohort)	anaphylaxis	11/361,515 [0.03]	17/359,538 [0.047]	0.64 (0.27, 1.5)
	arthralgia	360/360,052 [1]	261/358,536 [0.73]	1.4 (1.2, 1.6)
	chills	67/361,310 [0.19]	96/359,211 [0.27]	0.69 (0.5, 0.96)
	CVST	0/361,565 [0]	2/359,600 [0.0056]	0 (0, 5.3)
	diarrhea	183/360,859 [0.51]	312/358,329 [0.87]	0.58 (0.48, 0.7)
	erythema	273/360,470 [0.76]	283/358,459 [0.79]	0.96 (0.81, 1.1)
	facial paralysis	4/361,554 [0.011]	10/359,560 [0.028]	0.4 (0.091, 1.4)
	fatigue	470/359,712 [1.3]	638/356,984 [1.8]	0.73 (0.65, 0.82)
	fever	128/361,104 [0.35]	181/358,857 [0.5]	0.7 (0.56, 0.89)
	headache	522/359,506 [1.5]	571/357,234 [1.6]	0.91 (0.81, 1)
	local pain	3/361,553 [0.0083]	1/359,604 [0.0028]	3 (0.24, 160)
	local swelling	2/361,560 [0.0055]	0/359,610 [0]	inf (0.19, inf)
	lymphadenopathy	114/361,137 [0.32]	141/359,015 [0.39]	0.8 (0.62, 1)
	myalgia	381/360,002 [1.1]	294/358,405 [0.82]	1.3 (1.1, 1.5)
	nausea	332/360,302 [0.92]	495/357,565 [1.4]	0.67 (0.58, 0.77)
	soreness	116/361,049 [0.32]	85/359,252 [0.24]	1.4 (1, 1.8)
vomiting	128/361,057 [0.35]	270/358,495 [0.75]	0.47 (0.38, 0.58)	
mRNA-1273 (16,471 individuals per cohort)	anaphylaxis	2/114,990 [0.017]	8/114,619 [0.07]	0.25 (0.026, 1.2)
	arthralgia	188/114,237 [1.6]	107/114,227 [0.94]	1.8 (1.4, 2.2)
	chills	29/114,900 [0.25]	28/114,538 [0.24]	1 (0.59, 1.8)
	CVST	0/114,994 [0]	1/114,643 [0.0087]	0 (0, 39)
	diarrhea	80/114,685 [0.7]	127/114,163 [1.1]	0.63 (0.47, 0.84)
	erythema	124/114,467 [1.1]	95/114,251 [0.83]	1.3 (0.99, 1.7)
	facial paralysis	1/114,992 [0.0087]	4/114,637 [0.035]	0.25 (0.0051, 2.5)
	fatigue	244/114,010 [2.1]	277/113,561 [2.4]	0.88 (0.74, 1)
	fever	62/114,781 [0.54]	80/114,328 [0.7]	0.77 (0.54, 1.1)
	headache	187/114,286 [1.6]	188/113,908 [1.7]	0.99 (0.81, 1.2)
	local pain	2/114,983 [0.017]	0/114,644 [0]	inf (0.19, inf)
	local swelling	0/114,994 [0]	0/114,644 [0]	N/A
	lymphadenopathy	80/114,688 [0.7]	60/114,429 [0.52]	1.3 (0.94, 1.9)
	myalgia	165/114,290 [1.4]	111/114,241 [0.97]	1.5 (1.2, 1.9)
	nausea	155/114,400 [1.4]	188/113,947 [1.6]	0.82 (0.66, 1)
	soreness	42/114,845 [0.37]	32/114,524 [0.28]	1.3 (0.81, 2.1)
vomiting	69/114,744 [0.6]	101/114,299 [0.88]	0.68 (0.49, 0.93)	

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 7-day study period or until 4 days prior to a positive SARS-CoV-2 test result. For example, we see that 470 cases of fatigue were recorded in the BNT162b2-vaccinated cohort over a total of 359,712 person-days, corresponding to an incidence rate of 1.3 cases per 1,000 person-days. N/A, not applicable; inf, infinity.

Regarding reporting methods, 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. V-safe similarly solicits adverse effect reports from its voluntary participants on a regular basis. Such methods that rely on active solicitation can be prone to overestimating the true frequency of vaccine-induced side effects, as evidenced by the relatively high fraction of individuals reporting side effects after receiving placebo.<sup>4,5</sup> 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.

**Table 4. Incidence rates of adverse effects in the 7 days following the date of the second BNT162b2 or mRNA-1273 dose**

Vaccine	Adverse effect	Vaccinated incidence rate, cases/person-days [cases per 1,000 person-days]	Unvaccinated incidence rate, cases/person-days [cases per 1,000 person-days]	IRR (95% CI)
BNT162b2 (39,058 individuals per cohort)	anaphylaxis	1/273,178 [0.0037]	14/263,580 [0.053]	0.069 (0.0016, 0.45)
	arthralgia	269/272,056 [0.99]	186/262,819 [0.71]	1.4 (1.2, 1.7)
	chills	69/272,926 [0.25]	66/263,371 [0.25]	1 (0.71, 1.4)
	CVST	0/273,185 [0]	0/263,641 [0]	N/A
	diarrhea	135/272,624 [0.5]	224/262,684 [0.85]	0.58 (0.47, 0.72)
	erythema	242/272,184 [0.89]	212/262,757 [0.81]	1.1 (0.91, 1.3)
	facial paralysis	4/273,167 [0.015]	7/263,612 [0.027]	0.55 (0.12, 2.2)
	fatigue	321/271,819 [1.2]	462/261,725 [1.8]	0.67 (0.58, 0.77)
	fever	82/272,907 [0.3]	142/263,007 [0.54]	0.56 (0.42, 0.74)
	headache	373/271,605 [1.4]	423/261,896 [1.6]	0.85 (0.74, 0.98)
	local pain	3/273,177 [0.011]	0/263,641 [0]	inf (0.4, inf)
	local swelling	0/273,185 [0]	1/263,634 [0.0038]	0 (0, 38)
	lymphadenopathy	103/272,784 [0.38]	112/263,181 [0.43]	0.89 (0.67, 1.2)
	myalgia	307/271,929 [1.1]	207/262,746 [0.79]	1.4 (1.2, 1.7)
	nausea	233/272,224 [0.86]	386/261,988 [1.5]	0.58 (0.49, 0.69)
	soreness	93/272,805 [0.34]	57/263,400 [0.22]	1.6 (1.1, 2.2)
vomiting	92/272,827 [0.34]	224/262,687 [0.85]	0.4 (0.31, 0.51)	
mRNA-1273 (11,851 individuals per cohort)	anaphylaxis	1/82,935 [0.012]	6/79,821 [0.075]	0.16 (0.0035, 1.3)
	arthralgia	137/82,410 [1.7]	68/79,584 [0.85]	1.9 (1.4, 2.6)
	chills	39/82,809 [0.47]	20/79,762 [0.25]	1.9 (1.1, 3.4)
	CVST	0/82,942 [0]	0/79,839 [0]	N/A
	diarrhea	51/82,737 [0.62]	87/79,530 [1.1]	0.56 (0.39, 0.81)
	erythema	107/82,507 [1.3]	73/79,511 [0.92]	1.4 (1, 1.9)
	facial paralysis	1/82,941 [0.012]	3/79,822 [0.038]	0.32 (0.0061, 4)
	fatigue	182/82,268 [2.2]	179/79,135 [2.3]	0.98 (0.79, 1.2)
	fever	51/82,752 [0.62]	53/79,657 [0.67]	0.93 (0.62, 1.4)
	headache	120/82,502 [1.5]	123/79,373 [1.5]	0.94 (0.72, 1.2)
	local pain	1/82,941 [0.012]	0/79,839 [0]	inf (0.025, inf)
	local swelling	0/82,942 [0]	0/79,839 [0]	N/A
	lymphadenopathy	52/82,731 [0.63]	46/79,664 [0.58]	1.1 (0.72, 1.7)
	myalgia	147/82,330 [1.8]	66/79,573 [0.83]	2.2 (1.6, 2.9)
	nausea	112/82,478 [1.4]	130/79,313 [1.6]	0.83 (0.64, 1.1)
	soreness	28/82,821 [0.34]	19/79,750 [0.24]	1.4 (0.76, 2.7)
vomiting	50/82,749 [0.6]	69/79,546 [0.87]	0.7 (0.47, 1)	

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 7-day study period or until 4 days prior to a positive SARS-CoV-2 test result. For example, we see that 321 cases of fatigue were recorded in the BNT162b2-vaccinated cohort over a total of 271,819 person-days, corresponding to an incidence rate of 1.2 cases per 1,000 person-days. N/A, not applicable; inf, infinity.

That said, serious safety concerns requiring medical care are indeed 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 that have 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.

In light of this, it is worth noting that the rate of severe side effects was actually similar among participants receiving vaccine or placebo in the Pfizer/BioNTech and Moderna phase 3 trials at 2% or less, with the exception of severe fatigue in 3.8% of participants after their second dose of BNT162b2.<sup>4,5</sup> This is consistent with our observations that vaccinated and unvaccinated individuals have similar EHR documentation rates for each surveyed potential vaccine-associated symptom. Our finding that vaccinated individuals are not more likely to receive emergent care provides further orthogonal support for the safety of these vaccines.

As the remainder of the US population undergoes COVID-19 vaccination, it will not be feasible to solicit reports of adverse effects from all vaccinated individuals. It is important that clinicians are aware of VAERS as a centralized referral source for the documentation of serious adverse effects experienced by patients after vaccination and that patients are aware of the opportunity to enroll in voluntary directed surveillance efforts like V-safe.<sup>15,30</sup> Augmented curation of EHR notes for real-world safety monitoring is a practical solution to this large-scale challenge that can be deployed to complement these existing surveillance methods. The method demonstrated here represents a scalable approach to continuously monitor serious safety concerns associated with any authorized COVID-19 vaccine. Taken together with our recent study highlighting the real-world effectiveness of these vaccines, these data reinforce that individuals, providers, and public health officials should proceed rapidly with vaccination efforts with high confidence in their safety.<sup>21</sup>

### Limitations of study

There are a few limitations of this study. 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, over 90% of individuals in each vaccinated and unvaccinated cohort were Caucasian, ~60% of individuals who received BNT162b2 were female, and over 50% individuals who received mRNA-1273 were age 65 years or older (Tables S1 and S2). These biases may limit the generalizability of our study, as populations receiving the vaccines in later phases of the rollout could exhibit or report distinct side effects profiles due to differences in underlying demographic or clinical characteristics. Related to this point, the likely enrichment of healthcare workers in the vaccinated cohorts (due to their inclusion in phase 1a of the rollout) likely leads to an underestimation of the rates of return to clinic due to factors previously discussed, such as access to an Occupational Health Services office and other institution-specific COVID-19 response centers. Unfortunately, we are not able to extract healthcare worker status or utilization of such resources from the EHR database to improve our propensity score matching or enable stratified analyses, but future studies investigating this bias are certainly warranted.

Second, the bidirectional encoder representations from transformers (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, facial paralysis, or CVST 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 this need for manual review.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Human subjects
- **METHOD DETAILS**
  - Study design, setting, and population
  - Definition of time intervals for safety analyses
  - Definition of adverse effects of interest
  - Curation of adverse effects from clinical notes
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Evaluating rates of return to clinic, including emergent visits, after vaccination
  - Evaluating rates of adverse effects between vaccinated and matched unvaccinated cohorts

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.medj.2021.06.006>.

### ACKNOWLEDGMENTS

The authors thank Murali Aravamudan for the careful review and feedback on this manuscript. This study was funded by nference.

### AUTHOR CONTRIBUTIONS

V.S., P.L., and S.A. conceived the study. R.M., P.L., A.J.V., and V.S. wrote the manuscript and reviewed the findings. E.S., A.P., S.A., C.P., V.A., P.A., A.R., C.C., K.C., D.D., N.K., E.R., G.B., A.M., and T.W. contributed methods, analysis, and software. J.C.O.H., G.J.G., A.W.W., A.D.B., M.D.S., A.V., and J.H. reviewed the study, findings, and manuscript. All authors revised the manuscript.

### DECLARATION OF INTERESTS

R.M., P.L., E.S., A.P., S.A., C.P., V.A., A.J.V., P.A., A.R., C.C., K.C., D.D., N.K., E.R., G.B., A.M., T.W., and V.S. are employees of nference and have financial interests in the company and in the successful application of this research. R.M. is a student at Boston University School of Medicine. P.L. is a student at Harvard Medical School. J.C.O. receives personal fees from Elsevier and Bates College and small grants from nference outside the submitted work. A.D.B. is a consultant for AbbVie, is on scientific advisory boards for nference and Zentalis, and is founder and president of Splissen Therapeutics. J.H., J.C.O., G.J.G., A.W.W., A.V., M.D.S., and A.D.B. are employees of the Mayo Clinic. The Mayo Clinic may stand to gain financially from the successful outcome of the research. nference collaborates with Janssen and other bio-pharmaceutical companies on data science initiatives unrelated to this study. These collaborations had no role in study design, data collection and

analysis, decision to publish, or preparation of the manuscript. 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.

Received: March 5, 2021

Revised: May 4, 2021

Accepted: June 15, 2021

Published: July 1, 2021

## REFERENCES

- Centers for Disease Control and Prevention (CDC) (2020). COVID Data Tracker. <https://covid.cdc.gov/covid-data-tracker/>.
- FDA authorizes Moderna COVID-19 vaccine. *Med. Lett. Drugs Ther.* 63, 9–10.
- Office of the Commissioner (2021). Pfizer-BioNTech COVID-19 Vaccine. <https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/pfizer-biontech-covid-19-vaccine>.
- Baden, L.R., El Sahly, H.M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S.A., Rouphael, N., Creech, C.B., et al.; COVE Study Group (2021). Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N. Engl. J. Med.* 384, 403–416.
- Polack, F.P., Thomas, S.J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J.L., Pérez Marc, G., Moreira, E.D., Zerbini, C., et al.; C4591001 Clinical Trial Group (2020). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* 383, 2603–2615.
- Centers for Disease Control and Prevention (CDC) (2021). CDC's COVID-19 Vaccine Rollout Recommendations. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations.html>.
- Nguyen, L.H., Drew, D.A., Graham, M.S., Joshi, A.D., Guo, C.G., Ma, W., Mehta, R.S., Warner, E.T., Sikavi, D.R., Lo, C.H., et al.; COronavirus Pandemic Epidemiology Consortium (2020). Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *Lancet Public Health* 5, e475–e483.
- Mutambudzi, M., Niedwiedz, C., Macdonald, E.B., Leyland, A., Mair, F., Anderson, J., Celis-Morales, C., Cleland, J., Forbes, J., Gill, J., et al. (2020). Occupation and risk of severe COVID-19: prospective cohort study of 120 075 UK Biobank participants. *Occup. Environ. Med.* Published online December 9, 2020. <https://doi.org/10.1136/oemed-2020-106731>.
- Panagiotou, O.A., Kosar, C.M., White, E.M., Bantisi, L.E., Yang, X., Santostefano, C.M., Feifer, R.A., Blackman, C., Rudolph, J.L., Gravenstein, S., and Mor, V. (2021). Risk Factors Associated With All-Cause 30-Day Mortality in Nursing Home Residents With COVID-19. *JAMA Intern. Med.* 181, 439–448. <https://doi.org/10.1001/jamainternmed.2020.7968>.
- Wu, K. (2020). Boston Doctor Reports Serious Allergic Reaction After Getting Moderna's Covid Vaccine. *The New York Times*, December 26, 2020. A6. <https://www.nytimes.com/2020/12/25/health/covid-moderna-vaccine-allergies.html>.
- Allergic Reactions Including Anaphylaxis After Receipt of the First Dose of Moderna COVID-19 Vaccine — United States, December 21, 2020–January 10, 2021. *MMWR Morb Mortal Wkly Rep* 2021 70, 125–129. <https://doi.org/10.15585/mmwr.mm7004e1>
- CDC COVID-19 Response Team; Food and Drug Administration (2021). Allergic reactions including anaphylaxis after receipt of the first dose of Pfizer-BioNTech COVID-19 vaccine—United States, December 14–23, 2020. *MMWR Morb. Mortal. Wkly. Rep.* 70, 46–51.
- Lauerman, J., and Gale, J. What to Know About Post-Vaccine Deaths and Allergies, *The Washington Post*, February 10, 2021. [www.washingtonpost.com/business/what-to-know-about-post-vaccinedeaths-and-allergies/2021/02/05/4cb4bdf6-67f0-11eb-bab8-707f8769d785\\_story.html](http://www.washingtonpost.com/business/what-to-know-about-post-vaccinedeaths-and-allergies/2021/02/05/4cb4bdf6-67f0-11eb-bab8-707f8769d785_story.html).
- Castells, M.C., and Phillips, E.J. (2021). Maintaining Safety with SARS-CoV-2 Vaccines. *N. Engl. J. Med.* 384, 643–649.
- United States Department of Health and Human Services (2021). Department of Health and Human Services (DHHS), Public Health Service (PHS), Food and Drug Administration (FDA)/Centers for Disease Control (CDC). Vaccine Adverse Event Reporting System (VAERS) (CDC WONDER Online Database).
- Pawlowski, C., Venkatakrisnan, A.J., Ramudu, E., Kirkup, C., Puranik, A., Kayal, N., Berner, G., Anand, A., Barve, R., O'Horo, J.C., et al. (2020). Pre-existing conditions are associated with COVID patients' hospitalization, despite confirmed clearance of SARS-CoV-2 virus. *medRxiv*. <https://doi.org/10.1101/2020.10.28.20221655>.
- Pawlowski, C., Venkatakrisnan, A.J., Kirkup, C., Berner, G., Puranik, A., O'Horo, J.C., Badley, A.D., Soundaratan, V., et al. (2020). Enoxaparin Is Associated With Lower Rates of Thrombosis, Kidney Injury, and Mortality Than Unfractionated Heparin in Hospitalized COVID Patients. *medRxiv*. <https://doi.org/10.1101/2020.10.06.20208025>.
- Pawlowski, C., Bandi, H., Venkatakrisnan, A.J., Agarwal, V., Kennedy, R., O'Horo, J.C., Gores, G.J., Williams, A.W., Halamka, J., Badley, A.D., et al. (2020). Exploratory analysis of immunization records highlights decreased SARS-CoV-2 rates in individuals with recent non-COVID-19 vaccinations. *medRxiv*. <https://doi.org/10.1101/2020.07.27.20161976>.
- Awasthi, S., Wagner, T., Venkatakrisnan, A.J., Puranik, A., Hurchik, M., Agarwal, V., Conrad, I., Kirkup, C., Arunachalam, R., O'Horo, J., et al. (2020). Plasma IL-6 Levels following Corticosteroid Therapy as an Indicator of ICU Length of Stay in Critically ill COVID-19 Patients. *medRxiv*. <https://doi.org/10.1101/2020.07.02.20144733>.
- Agarwal, V., Venkatakrisnan, A.J., Puranik, A., Kirkup, C., Lopez-Marquez, A., Challener, D.W., O'Horo, J.C., Binnicker, M.J., Kremers, W.K., Faubion, W.A., Jr., et al. (2020). Quantifying the prevalence of SARS-CoV-2 long-term shedding among non-hospitalized COVID-19 patients. *medRxiv*. <https://doi.org/10.1101/2020.06.02.20120774>.
- Pawlowski, C., Lenehan, P., Puranik, A., Agarwal, V., Venkatakrisnan, A.J., Neisen, M.J.M., O'Horo, J.C., Badley, A.D., Halamka, J., et al. (2021). FDA-authorized COVID-19 vaccines are effective per real-world evidence synthesized across a multi-state health system. *Med 2*. Published online June 29, 2021. <https://doi.org/10.1016/j.medj.2021.06.007>.
- Ozonoff, A., Nanishi, E., and Levy, O. (2021). Bell's palsy and SARS-CoV-2 vaccines. *Lancet Infect. Dis.* 21, 450–452.
- Shimabukuro, T.T., Cole, M., and Su, J.R. (2021). Reports of Anaphylaxis After Receipt of mRNA COVID-19 Vaccines in the US—December 14, 2020–January 18, 2021. *JAMA* 325, 1101–1102.
- Cines, D.B., and Bussell, J.B. (2021). SARS-CoV-2 Vaccine-Induced Immune Thrombotic Thrombocytopenia. *N. Engl. J. Med.* 384, 2254–2256.
- Renoud, L., Khouri, C., Revol, B., Lepelley, M., Perez, J., Roustit, M., and Cracowski, J.L. (2021). Association of Facial Paralysis With mRNA COVID-19 Vaccines: A Disproportionality Analysis Using the World Health Organization Pharmacovigilance Database. *JAMA Intern. Med.* Published online April 21, 2021. <https://doi.org/10.1001/jamainternmed.2021.2219>.
- Pawlowski, C., Rincón-Hekking, J., Awasthi, S., Pandey, V., Lenehan, P., Venkatakrisnan, A.J., Bade, S., O'Horo, C., Virk, A., Swift, M.D., et al. (2021). Cerebral venous sinus thrombosis (CVST) is not significantly linked to COVID-19 vaccines or non-COVID vaccines in a large multi-state US health system. *medRxiv*. <https://doi.org/10.1101/2021.04.20.21255806>.
- Chapin-Bardales, J., Gee, J., and Myers, T. (2021). Reactogenicity Following Receipt of mRNA-Based COVID-19 Vaccines. *JAMA* 325, 2201–2202.

28. Jackson, L.A., Anderson, E.J., Roupael, N.G., Roberts, P.C., Makhene, M., Coler, R.N., McCullough, M.P., Chappell, J.D., Denison, M.R., Stevens, L.J., et al. (2020). An mRNA vaccine against SARS-CoV-2—preliminary report. *N. Engl. J. Med.* 383, 1920–1931.
29. Mulligan, M.J., Lyke, K.E., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Neuzil, K., Raabe, V., Bailey, R., Swanson, K.A., et al. (2020). Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature* 586, 589–593.
30. Centers for Disease Control and Prevention (CDC) (2021). V-safe after vaccination health checker. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/vsafe.html>.
31. Seabold, S., and Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. In Proceedings of the 9th Python in Science Conference (SciPy).
32. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Association for Computational Linguistics), pp. 4171–4186.
33. Wagner, T., Shweta, F., Murugadoss, K., Awasthi, S., Venkatakrishnan, A.J., Bade, S., Puranik, A., Kang, M., Pickering, B.W., O'Horo, J.C., et al. (2020). Augmented curation of clinical notes from a massive EHR system reveals symptoms of impending COVID-19 diagnosis. *eLife* 9, e58227.
34. R Core Team (2020). R: A language and environment for statistical computing (R Foundation for Statistical Computing). <https://www.R-project.org/>.
35. Allaire, J. (2012). RStudio: integrated development environment for R (R Studio Team).
36. Sahai, H., and Khurshid, A. (1995). *Statistics in Epidemiology: Methods, Techniques and Applications* (CRC Press).



## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Data and code</b>		
Number of clinical notes and individuals contributing clinical notes after the first and second doses of BNT162b2		<a href="#">Data S1</a>
Number of clinical notes and individuals contributing clinical notes after the first and second doses of mRNA-1273		<a href="#">Data S2</a>
R scripts for statistical analysis of clinical follow-up in vaccinated and unvaccinated individuals	This study	<a href="#">Data S3</a>
Python scripts for statistical analysis of adverse event frequencies in vaccinated and unvaccinated individuals	This study	<a href="#">Data S4</a>
<b>Software and algorithms</b>		
Python software package: statsmodels v0.10.0		<a href="https://www.statsmodels.org">https://www.statsmodels.org</a>
R software package: stats v4.0.3		<a href="https://www.rdocumentation.org/packages/stats/versions/3.6.2">https://www.rdocumentation.org/packages/stats/versions/3.6.2</a>
RStudio v1.3.959		<a href="https://www.rstudio.com/">https://www.rstudio.com/</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for information should be directed to and will be fulfilled by the lead contact, Venky Soundararajan ([venky@nference.net](mailto:venky@nference.net)).

#### Materials availability

This study did not generate new reagents.

#### Data and code availability

- Data: The summarized data used to assess rates of clinical follow-up (including emergent follow-up) are provided in [Data S1](#) and [Data S2](#). Other datasets supporting this study have not been deposited in a public repository because they contain personally identifiable information from human subjects which are protected by national privacy regulations, but this data may be made available from the lead contact on request. A proposal with a detailed description of study objectives and a statistical analysis plan will be needed to evaluate the reasonability of requests. Deidentified data will be provided after approval from the lead contact and the Mayo Clinic's standard IRB process for such requests.
- Code: Original code from this analysis is available in [Data S3](#) and [Data S4](#).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### Human subjects

The total cohort included 136,532 individuals. Each individual was part of one of the following four cohorts, on the basis of whether they had or had not received at least one dose of an mRNA COVID-19 vaccine: (1) BNT162b2 vaccinated (n = 51,795 individuals), (2) BNT162b2 matched unvaccinated (n = 51,795 individuals), (3) mRNA-1273 vaccinated (n = 16,471 individuals), or (4) mRNA-1273 matched unvaccinated (n = 16,471 individuals). More details describing the participant selection algorithm are provided in the [method details](#) and are illustrated in [Figure 1](#). Demographic and clinical characteristics of the analyzed cohorts (including age, sex, race, ethnicity,

and number of prior flu and SARS-CoV-2 diagnostic tests in the past year) are provided in [Tables S1](#) and [S2](#).

This study was reviewed and approved by the Mayo Clinic Institutional Review Board (IRB 20-003278) as a minimal risk study. Subjects were excluded if they did not have a research authorization on file. The IRB approved was titled: Study of COVID-19 patient characteristics with augmented curation of Electronic Health Records (EHR) to inform strategic and operational decisions with the Mayo Clinic. The study was deemed exempt by the Mayo Clinic Institutional Review Board and waived from consent. The following resource provides further information on the Mayo Clinic Institutional Review Board and adherence to basic ethical principles underlying the conduct of research, and ensuring that the rights and well-being of potential research subjects are adequately protected (<https://www.mayo.edu/research/institutional-review-board/overview>).

## METHOD DETAILS

### 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.

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."<sup>21</sup> In total, there were 572,291 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; (4) has no record of receiving the Johnson & Johnson (J&J) COVID-19 vaccine (Ad26.COV2.S). This population included 324,992 individuals, of whom 86,184 have received BNT162b2 or mRNA-1273 and 238,808 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, as were individuals with zero follow-up days after vaccination (i.e., those who received the first vaccine dose on the last date of data collection). Individuals who had received their second vaccine dose four or more days earlier than recommended (17 or fewer days after the first dose for BNT162b2; 24 or fewer days after the first dose for mRNA-1273) were also excluded, leaving 85,676 eligible individuals for the final vaccinated cohort.

A propensity matched unvaccinated cohort was selected from the previously derived set of 238,808 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 baseline, 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. A detailed description of the matching procedure is given in the related vaccine effectiveness manuscript.<sup>21</sup> Briefly, for each vaccinated individual, we attempted to identify an unvaccinated individual who (i) resides in the same location (per zip code), (ii) is of the same sex, (iii) has the same current status regarding long term care facility residence, and (iv) has a

similar propensity score accounting for age, race, ethnicity, and history of testing for SARS-CoV-2 and influenza during the past year. Propensity scores were calculated for all eligible individuals (both vaccinated and unvaccinated) by training a logistic regression model to predict vaccination status using the statsmodels v0.10.0 package in Python.<sup>31</sup>

We were able to identify valid matches for 68,266 of the 85,676 eligible vaccinated individuals. Thus, our final overall vaccinated and unvaccinated cohorts each contained 68,266 individuals ( $n = 51,795$  each for BNT162b2, and  $n = 16,471$  each for mRNA-1273). Demographic and clinical characteristics of these cohorts are provided in [Tables S1](#) and [S2](#), and the age distributions of the cohorts before and after matching are shown in [Figure S1](#); note that these tables and this figure are identical to those in the related manuscript that utilizes the same cohorts.<sup>21</sup> Of the 51,795 individuals who received at least one dose of BNT162b2, 39,058 received two doses; of the 16,471 individuals who received at least one dose of mRNA-1273, 11,851 received two doses. The distribution of time between doses for BNT162b2 and mRNA-1273 is shown in [Figure S2](#), which is also duplicated from the previous manuscript.<sup>21</sup> The selection algorithm and its associated counts are summarized in [Figure 1](#), and additional cohort characterization details can be found in the related manuscript.<sup>21</sup>

#### 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, including fatigue, fever, chills, myalgia, arthralgia, headache, lymphadenopathy, erythema, diarrhea, vomiting, and local pain and swelling.<sup>4,5</sup> We also included anaphylaxis, facial paralysis (Bell’s palsy), and cerebral venous sinus thrombosis (CVST), 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 various ways that a given phenotype could 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 to classify the sentiment for the phenotypes (described above) mentioned in the clinical notes.<sup>32</sup> 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%.<sup>33</sup>

Here we performed curation on a total of 1,279,292 EHR notes, including 650,137 contributed by vaccinated individuals and 629,155 contributed by unvaccinated individuals. These notes document clinical interactions on the full spectrum of acuity from routine (e.g., pre-scheduled) appointments to emergency department (ED) visits; approximately 3% of the notes analyzed were derived from ED visits. The

EHR also captures both in-person and telehealth interactions, which is relevant given the increased frequency of remote patient-provider interactions during the COVID-19 pandemic.

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 severe phenotypes of anaphylaxis, facial paralysis, and cerebral venous sinus thrombosis (CVST), 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 versus present). Only sentences manually confirmed as a present diagnosis were used to count these severe events.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Evaluating rates of return to clinic, including emergent visits, 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  $V_2$  (Data S1 and S2). The fraction of individuals with clinical follow-up was calculated as the number of individuals with at least one clinical note in the time window divided by the total number of individuals in each group (for BNT162b2:  $n_{\text{Dose } 1} = 51,795$ ;  $n_{\text{Dose } 2} = 39,058$ ; for mRNA-1273:  $n_{\text{Dose } 1} = 16,471$ ;  $n_{\text{Dose } 2} = 11,851$ ). The difference in clinical follow-up rates was assessed by calculating the odds ratio (OR) along with its corresponding 95% confidence interval (CI). 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.

To evaluate emergent clinical follow-up, we compared the number and percentage of vaccinated and unvaccinated individuals who contributed at least one ED note in the 1, 7, 14, or 21 days after each actual or assigned vaccine administration date (Data S1 and S2). The odds ratio and corresponding 95% CI were calculated for each time window, with the null hypothesis stating that the OR falls between 0.91 and 1.1 (as above). We also determined the total number of ED notes from vaccinated and unvaccinated individuals in these same time windows, along with the fraction of ED notes relative to all clinical notes (Data S1 and S2). The difference in the fraction of ED notes was assessed by computing an odds ratio and corresponding 95% CI, with the null hypothesis stating that the OR falls between 0.91 and 1.1. ED notes included those under the following categories in the EHR system: ED Notes, ED Procedure Note, ED Provider Notes, ED Provider Triage Note, ED Re-evaluation Note, and ED Triage Notes.

All odds ratios and 95% confidence intervals were calculated using the `fisher.test` function from the `stats` package in R (version 4.0.3).<sup>34</sup> These analyses were performed in RStudio (version 1.3.959).<sup>35</sup> Code corresponding to these analyses are provided in Data S3.

### Evaluating rates of adverse effects between vaccinated and matched unvaccinated cohorts

To evaluate adverse effects associated with receiving a COVID-19 vaccine in the clinical setting, we compared the vaccinated and matched unvaccinated populations

described above and summarized in [Figure 1](#). Specifically, we compared (i) 51,795 individuals with follow-up who received at least one dose of BNT162b2 to 51,795 matched individuals who have never received a COVID-19 vaccine, and (ii) 16,471 individuals with follow-up who received at least one dose of mRNA-1273 to 16,471 matched individuals who have never received a COVID-19 vaccine.

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. 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.<sup>36</sup> The IRR was considered to be statistically significant if the 95% CI did not include 1. This process was repeated to evaluate adverse effects which were documented in clinical notes within 14 or 21 days of the first or second vaccine dose.

For each individual, at-risk person-days are defined as the number of days from the start of the time period to the day on which the individual experienced the adverse effect or died, or four days prior to testing positive for SARS-CoV-2. This four day buffer was included to reduce the likelihood of mistaking early COVID-19 symptoms for vaccine side effects. Because some infections occurred within 4 days of either dose and because the rate of infection differed between the vaccinated and unvaccinated groups, the cohorts from which incidence rates were calculated differ slightly from the total cohorts of vaccinated and unvaccinated patients described above. The sizes of the cohorts considered in these analyses are provided in [Table S3](#), and code corresponding to these analyses are provided in [Data S4](#).