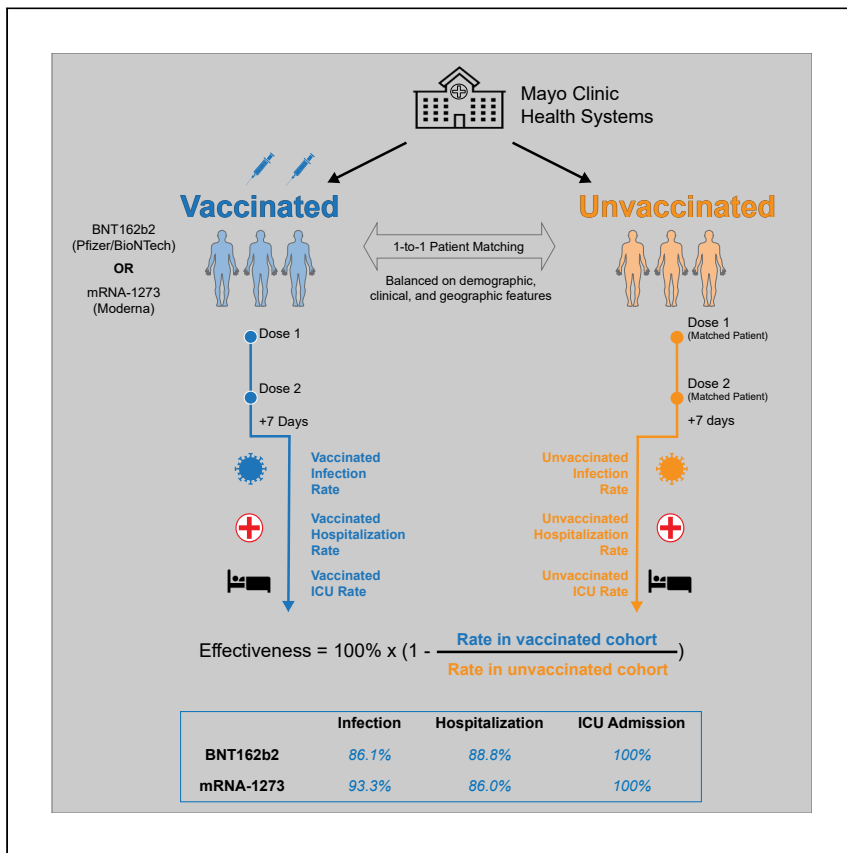


Clinical Advances

# FDA-authorized mRNA COVID-19 vaccines are effective per real-world evidence synthesized across a multi-state health system



Colin Pawlowski, Patrick Lenehan, Arjun Puranik, ..., Andrew D. Badley, John Halamka, Venky Soundararajan  
venky@ference.net

Highlights

BNT162b2 and mRNA-1273 are highly effective in preventing SARS-CoV-2 infection

The real-world effectiveness of BNT162b2 is 86.1% (95% CI: 82.4%–89.1%)

The real-world effectiveness of mRNA-1273 is 93.3% (95% CI: 85.7%–97.4%)

Both vaccines reduce the risk of COVID-19-related hospitalization and ICU admission

Pawlowski et al. assess the real-world effectiveness of the BNT162b2 and mRNA-1273 COVID-19 vaccines among 136,532 individuals. They compare infection, hospitalization, and ICU admission rates between vaccinated and propensity-matched unvaccinated individuals. They find that both vaccines protect against SARS-CoV-2 infection and severe COVID-19.



## Clinical Advances

## FDA-authorized mRNA COVID-19 vaccines are effective per real-world evidence synthesized across a multi-state health system

Colin Pawlowski,<sup>1,3</sup> Patrick Lenehan,<sup>1,3</sup> Arjun Puranik,<sup>1</sup> Vineet Agarwal,<sup>1</sup> A.J. Venkatakrisnan,<sup>1</sup> Michiel J.M. Niesen,<sup>1</sup> John C. O'Horo,<sup>2</sup> Abinash Virk,<sup>2</sup> Melanie D. Swift,<sup>2</sup> Andrew D. Badley,<sup>2</sup> John Halamka,<sup>2</sup> and Venky Soundararajan<sup>1,4,\*</sup>

## SUMMARY

**Background:** Two US Food and Drug Administration (FDA)-authorized coronavirus disease 2019 (COVID-19) mRNA vaccines, BNT162b2 (Pfizer/BioNTech) and mRNA-1273 (Moderna), have demonstrated high efficacy in large phase 3 randomized clinical trials. It is important to assess their effectiveness in a real-world setting.

**Methods:** This is a retrospective analysis of 136,532 individuals in the Mayo Clinic health system (Arizona, Florida, Iowa, Minnesota, and Wisconsin) with PCR testing data between December 1, 2020 and April 20, 2021. We compared clinical outcomes for a vaccinated cohort of 68,266 individuals who received at least one dose of either vaccine ( $n_{\text{BNT162b2}} = 51,795$ ;  $n_{\text{mRNA-1273}} = 16,471$ ) and an unvaccinated control cohort of 68,266 individuals propensity matched based on relevant demographic, clinical, and geographic features. We estimated real-world vaccine effectiveness by comparing incidence rates of positive severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) PCR testing and COVID-19-associated hospitalization and intensive care unit (ICU) admission starting 7 days after the second vaccine dose.

**Findings:** The real-world vaccine effectiveness of preventing SARS-CoV-2 infection was 86.1% (95% confidence interval [CI]: 82.4%–89.1%) for BNT162b2 and 93.3% (95% CI: 85.7%–97.4%) for mRNA-1273. BNT162b2 and mRNA-1273 were 88.8% (95% CI: 75.5%–95.7%) and 86.0% (95% CI: 71.6%–93.9%) effective in preventing COVID-19-associated hospitalization. Both vaccines were 100% effective (95% CI<sub>BNT162b2</sub>: 51.4%–100%; 95% CI<sub>mRNA-1273</sub>: 43.3%–100%) in preventing COVID-19-associated ICU admission.

**Conclusions:** BNT162b2 and mRNA-1273 are effective in a real-world setting and are associated with reduced rates of SARS-CoV-2 infection and decreased burden of COVID-19 on the healthcare system.

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

## INTRODUCTION

To date, there have been over 170 million confirmed cases of coronavirus disease 2019 (COVID-19) and over 3.5 million associated deaths globally.<sup>1</sup> From the moment when severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was identified as the causative agent of COVID-19, efforts were initiated to characterize this virus and develop vaccines against it.<sup>2–4</sup> Within months, several candidate vaccines were shown to be safe and induce robust immune responses against SARS-CoV-2 in a series of early phase trials.<sup>5–8</sup> Subsequently, multiple vaccine candidates showed strong efficacy profiles in large phase 3 clinical trials.<sup>7,9–11</sup> BNT162b2, an

## Context and significance

This is a study of the COVID-19 vaccines developed by Pfizer/BioNTech and Moderna. Although these vaccines have been shown to be effective in clinical trials, it is important to confirm that they work well in practice. The results from this study show that these vaccines are effective in reducing the risk of COVID-19 infection, COVID-19-associated hospitalization, and COVID-19-associated ICU admission. As one of the largest real-world evidence studies of COVID-19 vaccine effectiveness in the United States to date, this study provides strong evidence that COVID-19 vaccines work well in practice.



mRNA vaccine developed by Pfizer/BioNTech, showed 95.0% efficacy (95% confidence interval [CI]: 90.3%–97.6%) in preventing symptomatic COVID-19 with onset 7 or more days after the second dose.<sup>9,10</sup> mRNA-1273, an mRNA vaccine developed by Moderna, showed 94.1% efficacy (95% CI: 89.3%–96.8%) in preventing symptomatic infection with onset at least 14 days after the second dose.<sup>10</sup>

BNT162b2 and mRNA-1273 are now being administered throughout the United States under an Emergency Use Authorization by the US Food and Drug Administration (FDA). The earliest vaccine doses were given to individuals at high risk for becoming infected with SARS-CoV-2 or experiencing severe COVID-19, such as healthcare workers, residents of long-term care facilities, and elderly individuals.<sup>12</sup> Recently these vaccines have been made available to all adults in the United States.<sup>13</sup> To date, 50.3% of the United States population has received at least one dose of a COVID-19 vaccine, and 40.5% are now considered fully vaccinated.<sup>14</sup> Recently, the Centers for Disease Control and Prevention (CDC) revised social distancing guidelines on the basis of vaccination status.<sup>15</sup> As rapid rollout continues and public health policies are adjusted accordingly, it is critical to determine whether the real-world effectiveness of these vaccines mirrors the efficacy observed in the trial settings.

Here we conducted a large-scale real-world interim analysis of COVID-19 vaccination outcomes in the United States. One challenge inherent to such real-world analyses is the lack of a built-in placebo arm, which is essential to establish the expected infection rate during the study period and, thereby, assess vaccine effectiveness. To address this shortcoming, we used 1-to-1 propensity score matching to identify a control arm of unvaccinated individuals who were similar in demographic, clinical, and geographic features to the set of vaccinated individuals. We then compared incidence rates of positive SARS-CoV-2 PCR testing, COVID-19-associated hospitalization, and COVID-19-associated ICU admission during defined time intervals between these vaccinated and unvaccinated cohorts.

## RESULTS

### Generation of cohorts to assess COVID-19 vaccine effectiveness

We identified 86,184 individuals who met the study inclusion criteria and had received at least one dose of BNT162b2 or mRNA-1273 (Figure 1A). As recommended, most second doses were administered 21 days (for BNT162b2) or 28 days (for mRNA-1273) after the first dose (Figure S1). Individuals receiving a second dose 4 or more days prior to the recommended date were excluded from further analysis, as were individuals who had tested positive before their first dose. Using a combination of exact matching and 1-to-1 propensity score matching, 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 individuals who also met the study inclusion criteria and had not yet received a COVID-19 vaccine at the time of this study. These cohorts were adequately balanced for the demographic and clinical variables used in matching (Tables 1 and 2; Figures S2A–S2C). 52,418 of the 68,266 (77%) vaccinated individuals had received two doses, and the median follow-up times after the first and second doses were 70 and 58 days, respectively (Table S1; Figure S3A). Vaccinated individuals were less likely to undergo SARS-CoV-2 testing in the first 3 days after their first dose, but the rates of testing were stably similar between the vaccinated and unvaccinated cohorts at subsequent time points (Figure S4).

### BNT162b2 reduces the incidence rate of positive SARS-CoV-2 PCR testing

Starting 7 days after the date of study enrollment, 401 of 50,474 (0.79%) individuals vaccinated with BNT162b2 tested positive for SARS-CoV-2 compared with 1,232 of

<sup>1</sup>Inference, Cambridge, MA 02142, USA

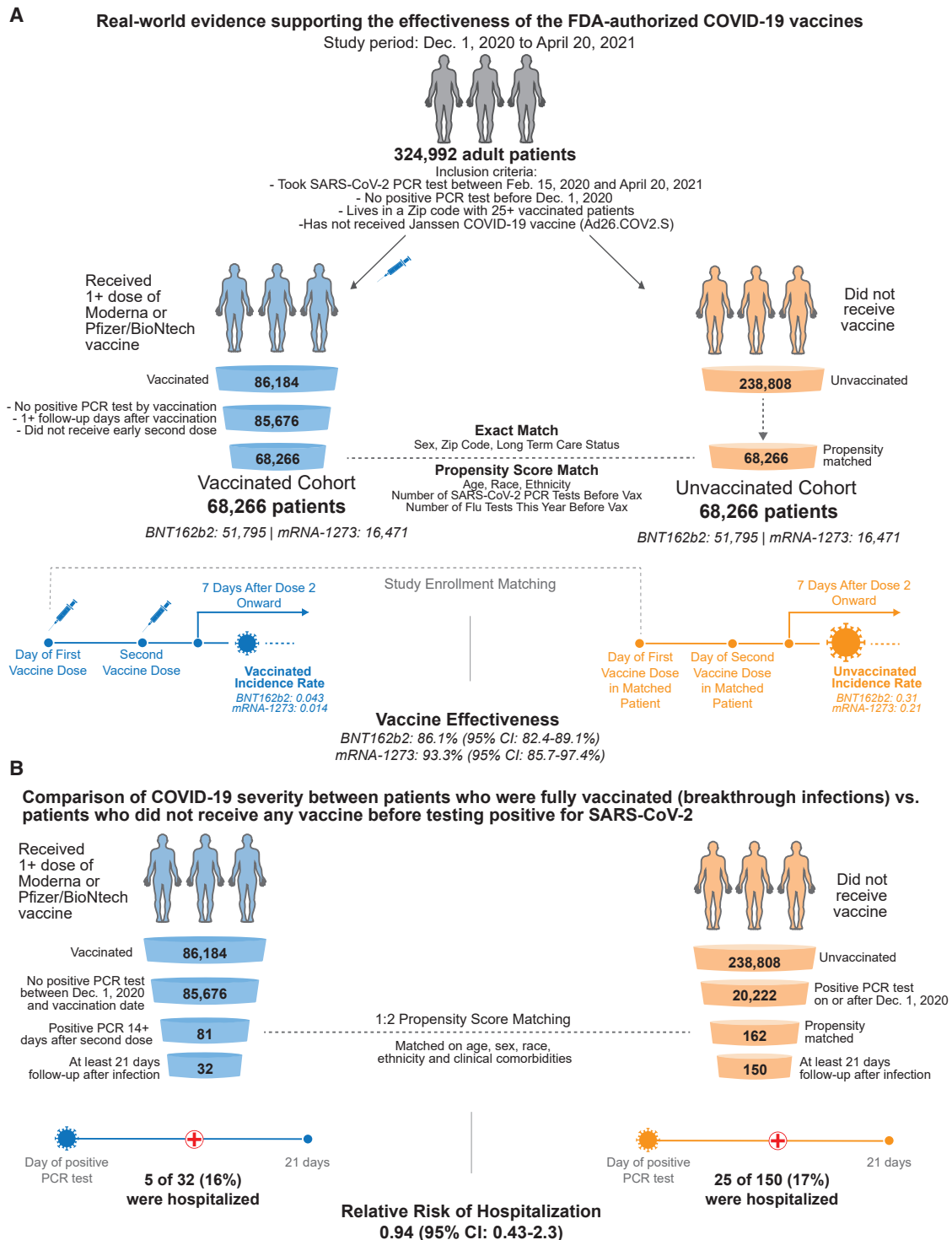
<sup>2</sup>Mayo Clinic, Rochester, MN 55902, USA

<sup>3</sup>These authors contributed equally

<sup>4</sup>Lead contact

\*Correspondence: [venky@inference.net](mailto:venky@inference.net)

<https://doi.org/10.1016/j.medj.2021.06.007>



**Figure 1. Schematic illustrating the algorithms for participant selection and outcome assessment**

(A) Design of the study to compare SARS-CoV-2 infection rates in individuals receiving BNT162b2 or mRNA-1273 vaccination compared with 1-to-1 propensity-matched unvaccinated individuals ( $n_{\text{BNT162b2}} = 51,795$  per group;  $n_{\text{mRNA-1273}} = 16,471$  per group). For each group, incidence rates were calculated to assess the effectiveness of vaccination in preventing a positive SARS-CoV-2 PCR test at least 7 days after the second vaccine dose. Several other time windows were also evaluated for vaccine effectiveness.

their 50,162 (2.46%) matched unvaccinated controls (Table 3). The incidence rates of positive SARS-CoV-2 tests in the vaccinated and unvaccinated cohorts were 0.14 and 0.43 per 1,000 person-days, respectively. This corresponds to an overall vaccine effectiveness of 68.5% (95% CI: 64.7%–71.9%) over the entire study period. For the 401 vaccinated individuals who subsequently tested positive for SARS-CoV-2, the distribution of time from first dose to first positive SARS-CoV-2 PCR test is shown in Figure S5.

BNT162b2 is administered as a two-dose regimen, with one dose expected to confer partial immunity but both doses required to achieve full vaccine effectiveness. In the interval starting 7 days after the date of study enrollment and extending to the date of the second dose, 293 of 50,474 (0.58%) individuals vaccinated with BNT162b2 tested positive for SARS-CoV-2 compared with 534 of their 50,162 (1.06%) matched unvaccinated controls. The incidence rates of positive SARS-CoV-2 tests in the vaccinated and unvaccinated cohorts were 0.38 and 0.70 per 1,000 person-days, respectively, corresponding to a single dose effectiveness of 45.5% (95% CI: 37.1%–52.9%) starting 7 days after vaccination. A log rank test indicates that the hazard rate is significantly lower in the vaccinated cohort over this time interval, with the curves beginning to separate noticeably about 14 days after vaccination ( $p = 2.3 \times 10^{-17}$ ; Figure 2A). As expected, the estimated effectiveness of a single vaccination was higher (61.0%; 95% CI: 50.8%–69.2%) when considering infections with onset at least 14 days after study enrollment.

Starting 7 days after the second dose, 82 of 35,990 (0.23%) vaccinated individuals had a positive SARS-CoV-2 test compared with 563 of 35,011 (1.61%) eligible unvaccinated individuals. This corresponds to incidence rates of 0.043 and 0.31 per 1,000 person-days, respectively, and a vaccine effectiveness of 86.1% (95% CI: 82.4%–89.1%). Consistent with this, a log rank test indicates that the hazard rate is significantly lower in the vaccinated cohort over this time interval ( $p = 7.3 \times 10^{-85}$ ; Figure 2B). Starting 14 days after the second dose, 59 of 33,963 (0.17%) vaccinated individuals had a positive SARS-CoV-2 test compared with 468 of 32,910 (1.42%) eligible unvaccinated individuals. This corresponds to incidence rates of 0.035 and 0.29 per 1,000 person-days, respectively, and a vaccine effectiveness of 88.0% (95% CI: 84.2%–91.0%).

### **mRNA-1273 reduces the incidence rate of positive SARS-CoV-2 PCR testing**

Starting 7 days after the date of study enrollment, 97 of 16,369 (0.59%) individuals vaccinated with mRNA-1273 tested positive for SARS-CoV-2 compared with 303 of their 16,309 (1.86%) matched unvaccinated controls (Table 3). The incidence rates of positive SARS-CoV-2 tests in the vaccinated and unvaccinated cohorts were 0.10 and 0.33 per 1,000 person-days, respectively. This corresponds to an overall vaccine effectiveness of 68.6% (95% CI: 60.5%–75.3%) over the entire study period. For the 97 vaccinated individuals who subsequently tested positive for SARS-CoV-2, the distribution of time from first dose to first positive SARS-CoV-2 PCR test is shown in Figure S5.

Like BNT162b2, a single dose of mRNA-1273 is expected to confer partial immunity, and both doses are needed to achieve full vaccine effectiveness. In the interval

---

(B) Design of the study to compare COVID-19 disease severity in individuals who were fully vaccinated at least 14 days prior to diagnosis with COVID-19 ( $n = 81$ ) and had at least 21 days of follow-up after diagnosis ( $n = 32$ ) versus 1-to-2 propensity-matched unvaccinated individuals ( $n = 162$ ) with at least 21 days of follow-up ( $n = 150$ ). Hospitalization and ICU admission were assessed within 21 days of PCR diagnosis, and mortality was assessed within 28 days of PCR diagnosis.

**Table 1. Clinical characteristics of individuals vaccinated with BNT162b2 and their 1-to-1 propensity-matched unvaccinated cohorts**

Clinical covariate	BNT162b2-vaccinated cohort	1-to-1 propensity-matched unvaccinated cohort	Standardized mean difference (SMD)
Total number of individuals	51,795	51,795	
Age, mean (SD)	53.83 (18.32)	53.5 (18.02)	0.02***
<b>Age groups in years</b>			
18–24	2,419 (4.7%)	2,526 (4.9%)	0.01***
25–34	7,576 (14.6%)	7,550 (14.6%)	0.00***
35–44	8,367 (16.2%)	8,503 (16.4%)	0.01***
45–54	7,901 (15.3%)	7,764 (15.0%)	0.01***
55–64	10,546 (20.4%)	10,303 (19.9%)	0.01***
65–74	7,404 (14.3%)	8,929 (17.2%)	0.08***
75+	7,582 (14.6%)	6,220 (12.0%)	0.08***
<b>Sex</b>			
Female	31,099 (60.0%)	31,099 (60.0%)	0.00***
Male	20,695 (40.0%)	20,695 (40.0%)	0.00***
Unknown	1 (0.0%)	1 (0.0%)	0.00***
<b>Race</b>			
Asian	1,568 (3.0%)	1,602 (3.1%)	0.00***
Black/African American	1,156 (2.2%)	1,519 (2.9%)	0.04***
Native American	127 (0.2%)	126 (0.2%)	0.00***
White/Caucasian	47,270 (91.3%)	46,853 (90.5%)	0.03***
Other	1,185 (2.3%)	1,253 (2.4%)	0.01***
Unknown	489 (0.9%)	442 (0.9%)	0.01***
<b>Ethnicity</b>			
Hispanic or Latino	1,676 (3.2%)	1,462 (2.8%)	0.02***
Not Hispanic or Latino	48,941 (94.5%)	49,304 (95.2%)	0.03***
Unknown	1,178 (2.3%)	1,029 (2.0%)	0.02***
<b>Mean number of prior SARS-CoV-2 PCR tests</b>			
February 1–May 30, 2020	0.2282 (0.5282)	0.2152 (0.5171)	0.02***
June 1–Aug 31, 2020	0.4562 (0.7652)	0.4547 (0.7775)	0.00***
September 1–November 30, 2020	0.6488 (0.8929)	0.6585 (0.8835)	0.01***
<b>Mean number of prior influenza tests</b>			
Feb 1 - May 30, 2020	0.2122 (0.7284)	0.2324 (0.772)	0.03***
Jun 1 - Aug 31, 2020	0.01658 (0.2355)	0.0219 (0.2695)	0.02***
Sep 1 - Nov 30, 2020	0.04334 (0.3588)	0.04864 (0.3759)	0.01***
<b>State</b>			
Arizona	3,572 (6.9%)	3,572 (6.9%)	0.00***
Florida	5,834 (11.3%)	5,834 (11.3%)	0.00***
Iowa	133 (0.3%)	133 (0.3%)	0.00***
Minnesota	30,021 (58.0%)	30,021 (58.0%)	0.00***
Wisconsin	12,235 (23.6%)	12,235 (23.6%)	0.00***
Long-term care resident	67 (0.1%)	67 (0.1%)	0.00***

Covariates for matching include (1) demographics (age, sex, race, ethnicity); (2) number of prior SARS-CoV-2 PCR tests before December 1, 2020; (3) number of influenza tests between February 1 and December 1, 2020; (4) residential location (zip code); and (5) long-term care facility status. Sex, zip code, and long-term care status are matched exactly between the two cohorts, so the proportions of individuals with each feature in these categories are identical. Highly balanced covariates with SMD < 0.1 are indicated by \*\*\*. See also [Table S1](#).

starting 7 days after the date of study enrollment and extending to the date of the second dose, 88 of 16,369 (0.54%) individuals vaccinated with mRNA-1273 tested positive for SARS-CoV-2 compared with 181 of their 16,309 (1.11%) matched unvaccinated controls. The incidence rates of positive SARS-CoV-2 tests in the vaccinated and unvaccinated cohorts were 0.24 and 0.49 per 1,000 person-days, respectively, corresponding to a single dose effectiveness of 51.7% (95% CI: 37.3%–63.0%) starting 7 days after vaccination. A log rank test indicates that the hazard rate is significantly lower in the vaccinated cohort over this time interval, with the curves

**Table 2. Clinical characteristics of individuals vaccinated with mRNA-1273 and their 1-to-1 propensity-matched unvaccinated cohorts**

Clinical covariate	mRNA-1273-vaccinated cohort	1-to-1 propensity-matched unvaccinated cohort	Standardized mean difference (SMD)
Total number of individuals	16,471	16,471	
Age, mean (SD)	63 (16.14)	62.23 (16.72)	0.05***
<b>Age groups in years</b>			
18–24	375 (2.3%)	388 (2.4%)	0.01***
25–34	926 (5.6%)	1,074 (6.5%)	0.04***
35–44	1,285 (7.8%)	1,484 (9.0%)	0.04***
45–54	1,598 (9.7%)	1,601 (9.7%)	0.00***
55–64	3,402 (20.7%)	3,436 (20.9%)	0.01***
65–74	5,598 (34.0%)	5,121 (31.1%)	0.06***
75+	3,287 (20.0%)	3,367 (20.4%)	0.01***
<b>Sex</b>			
Female	8,758 (53.2%)	8,758 (53.2%)	0.00***
Male	7,713 (46.8%)	7,713 (46.8%)	0.00***
Unknown	0 (0.0%)	0 (0.0%)	N/A
<b>Race</b>			
Asian	378 (2.3%)	463 (2.8%)	0.03***
Black/African American	530 (3.2%)	437 (2.7%)	0.03***
Native American	44 (0.3%)	30 (0.2%)	0.02***
White/Caucasian	15,088 (91.6%)	15,141 (91.9%)	0.01***
Other	273 (1.7%)	278 (1.7%)	0.00***
Unknown	158 (1.0%)	122 (0.7%)	0.02***
<b>Ethnicity</b>			
Hispanic or Latino	522 (3.2%)	408 (2.5%)	0.04***
Not Hispanic or Latino	15,583 (94.6%)	15,834 (96.1%)	0.07***
Unknown	366 (2.2%)	229 (1.4%)	0.06***
<b>Mean number of prior SARS-CoV-2 PCR tests</b>			
February 1–May 30, 2020	0.2189 (0.5214)	0.2168 (0.5282)	0.00***
June 1–August 31, 2020	0.5114 (0.8292)	0.4902 (0.8364)	0.03***
September 1–November 30, 2020	0.5724 (0.8154)	0.6195 (0.9051)	0.05***
<b>Mean number of prior influenza tests</b>			
February 1–May 30, 2020	0.266 (0.85)	0.2498 (0.8491)	0.02***
June 1–Aug 31, 2020	0.03967 (0.3932)	0.03291 (0.4238)	0.02***
September 1–November 30, 2020	0.08333 (0.5536)	0.08354 (0.5692)	0.02***
<b>State</b>			
Arizona	1,896 (11.5%)	1,896 (11.5%)	0.00***
Florida	5,453 (33.1%)	5,453 (33.1%)	0.00***
Iowa	64 (0.4%)	64 (0.4%)	0.00***
Minnesota	6,299 (38.2%)	6,299 (38.2%)	0.00***
Wisconsin	2,759 (16.8%)	2,759 (16.8%)	0.00***
Long-term care resident	50 (0.3%)	50 (0.3%)	0.00***

Covariates for matching include (1) demographics (age, sex, race, ethnicity); (2) number of prior SARS-CoV-2 PCR tests before December 1, 2020; (3) number of influenza tests between February 1 and December 1, 2020; (4) residential location (zip code); and (5) long-term care facility status. Sex, zip code, and long-term care status are matched exactly between the two cohorts, so the proportions of individuals with each feature in these categories are identical. Highly balanced covariates with SMD < 0.1 are indicated by \*\*\*. See also [Table S1](#).

beginning to separate noticeably between 14 and 21 days after vaccination ( $p = 1.0 \times 10^{-8}$ ; [Figure 2C](#)). Consistent with this, the estimated effectiveness of one mRNA-1273 dose was higher (66.6%; 95% CI: 51.9%–77.3%) when considering infections with onset at least 14 days after study enrollment.

Starting 7 days after the second dose, 7 of 11,612 (0.060%) vaccinated individuals had a positive SARS-CoV-2 test compared with 101 of 11,332 (0.89%) eligible unvaccinated individuals. This corresponds to incidence rates of 0.014 and 0.21 per 1,000

**Table 3. Incidence rates of positive SARS-CoV-2 PCR tests in vaccinated and 1-to-1 propensity-matched unvaccinated cohorts and corresponding vaccine effectiveness**

Vaccine	Time period	Vaccinated incidence rate: Cases/person-days [per 1,000 person-days] (no. of individuals contributing)	Unvaccinated incidence rate: Cases/person-days [per 1,000 person-days] (no. of individuals contributing)	Incidence rate ratio (95% CI)	Vaccine effectiveness (95% CI)
BNT162b2	on or after 7 days following first dose	401/2,942,986. [0.14] (n = 50,474)	1,232/2,851,069. [0.43] (n = 50,162)	0.32 (0.28, 0.35)	68.5% (64.7%, 71.9%)
	on or after 14 days following first dose	211/2,595,549. [0.081] (n = 48,305)	960/2,506,231. [0.38] (n = 47,863)	0.21 (0.18, 0.25)	78.8% (75.3%, 81.8%)
	on or after 7 days following first dose and prior to second dose	293/766,978. [0.38] (n = 50,474)	534/761,511. [0.7] (n = 50,162)	0.54 (0.47, 0.63)	45.5% (37.1%, 52.9%)
	on or after 14 days following first dose and prior to second dose	103/419,541. [0.25] (n = 48,305)	262/416,673. [0.63] (n = 47,863)	0.39 (0.31, 0.49)	61.0% (50.8%, 69.2%)
	on or after 7 days following the second dose	82/1,914,500. [0.043] (n = 35,990)	563/1,828,466. [0.31] (n = 35,011)	0.14 (0.11, 0.18)	86.1% (82.4%, 89.1%)
	on or after 14 days following the second dose	59/1,670,896. [0.035] (n = 33,963)	468/1,591,900. [0.29] (n = 32,910)	0.12 (0.09, 0.16)	88.0% (84.2%, 91.0%)
mRNA-1273	on or after 7 days following first dose	97/946,890. [0.1] (n = 16,369)	303/927,716. [0.33] (n = 16,309)	0.31 (0.25, 0.4)	68.6% (60.5%, 75.3%)
	on or after 14 days following first dose	49/833,098. [0.059] (n = 15,985)	241/814,413. [0.3] (n = 15,896)	0.2 (0.14, 0.27)	80.1% (72.9%, 85.7%)
	on or after 7 days following first dose and prior to second dose	88/369,195. [0.24] (n = 16,369)	181/366,881. [0.49] (n = 16,309)	0.48 (0.37, 0.63)	51.7% (37.3%, 63.0%)
	on or after 14 days following first dose and prior to second dose	40/255,403. [0.16] (n = 15,985)	119/253,578. [0.47] (n = 15,896)	0.33 (0.23, 0.48)	66.6% (51.9%, 77.3%)
	on or after 7 days following the second dose	7/495,550. [0.014] (n = 11,612)	101/478,322. [0.21] (n = 11,332)	0.067 (0.026, 0.14)	93.3% (85.7%, 97.4%)
	on or after 14 days following the second dose	6/417,664. [0.014] (n = 10,610)	75/402,416. [0.19] (n = 10,318)	0.077 (0.027, 0.18)	92.3% (82.4%, 97.3%)

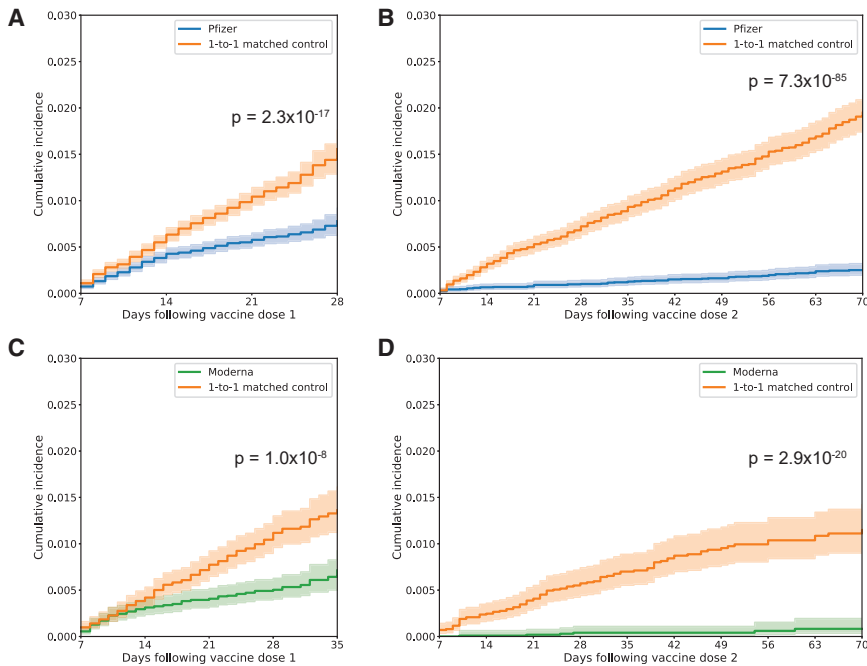
Incidence is calculated as the number of individuals with at least one positive SARS-CoV-2 PCR test per 1,000 person-days. Time period: time period relative to the first vaccine dose for the vaccinated cohort or study enrollment day for the unvaccinated cohort. Vaccinated incidence rate: number of individuals with positive PCR tests in the vaccinated cohort in the time period, divided by the number of at-risk person-days for the vaccinated cohort in the time period; in brackets, the number of cases per 1,000 person-days. Unvaccinated incidence rate: number of individuals with positive PCR tests in the propensity-matched unvaccinated cohort in the time period, divided by the number of at-risk person-days for the propensity-matched unvaccinated cohort in the time period; in brackets, the number of cases per 1,000 person-days. Incidence rate ratio: vaccinated incidence rate divided by unvaccinated incidence rate along with the exact 95% CI.<sup>16</sup> Vaccine effectiveness: 100% × (1 – incidence rate ratio) along with the 95% CI. See also [Figure S5](#).

person-days, respectively, and an effectiveness of 93.3% (95% CI: 85.7%–97.4%). A log rank test also indicates that the hazard rate is significantly lower in the vaccinated cohort over this time interval ( $p = 2.9 \times 10^{-20}$ ; [Figure 2D](#)). Starting 14 days after the second dose, 6 of 10,610 (0.057%) vaccinated individuals had a positive SARS-CoV-2 test compared with 75 of 10,318 (0.73%) eligible unvaccinated individuals. This corresponds to incidence rates of 0.014 and 0.19 per 1,000 person-days, respectively, and an effectiveness of 92.3% (95% CI: 82.4%–97.3%).

### BNT162b2 and mRNA-1273 protect against severe COVID-19

Although the data above demonstrate that both mRNA vaccines reduce the risk of testing positive for SARS-CoV-2, it is important to confirm that vaccination also reduces the risk of severe illness in individuals with COVID-19. To do so, we first compared the incidence rates of COVID-19-associated hospitalization in the vaccinated and unvaccinated cohorts. BNT162b2 was 88.8% (95% CI: 75.5%–95.7%)





**Figure 2. Kaplan-Meier analyses to assess cumulative incidence of SARS-CoV-2 infection between vaccinated and unvaccinated cohorts**

Cumulative incidence at time  $t$  is the estimated proportion of individuals who had a positive SARS-CoV-2 PCR test on or before time  $t$  (i.e., 1 minus the standard Kaplan-Meier survival estimate).

(A) For BNT162b2, cumulative incidence of positive SARS-CoV-2 testing from 7 days after the date of the first dose until the date of the second dose, the date of censoring, or the end of the study period (whichever occurs first for a given matched pair). A log rank test rejects the null hypothesis of equal hazard rates ( $p = 2.3 \times 10^{-17}$ ).

(B) For BNT162b2, cumulative incidence of positive SARS-CoV-2 testing from 7 days after the second dose until the end of the study period. A log rank test rejects the null hypothesis of equal hazard rates ( $p = 7.3 \times 10^{-85}$ ).

(C) Same as in (A) but for mRNA-1273. A log rank test rejects the null hypothesis of equal hazard rates ( $p = 1.0 \times 10^{-8}$ ).

(D) Same as in (B) but for mRNA-1273. A log-rank test rejects the null hypothesis of equal hazard rates ( $p = 2.9 \times 10^{-20}$ ).

In all cases, the cumulative incidence is compared between vaccinated individuals and their matched unvaccinated controls over the same time period.

effective in preventing this outcome at least 7 days after the second dose, and mRNA-1273 was 86.0% (95% CI: 71.6%–93.9%) effective (Table 4). We similarly compared the incidence rates of COVID-19-associated intensive care unit (ICU) admission in the vaccinated and unvaccinated cohorts. BNT162b2 and mRNA-1273 were 100% effective (95% CI<sub>BNT162b2</sub>: 51.4%–100%; 95% CI<sub>mRNA-1273</sub>: 43.3%–100%) in preventing this outcome 7 or more days after the second dose (Table 5).

### Breakthrough cases have similar rates of hospitalization, ICU admission, and mortality as unvaccinated individuals with COVID-19

Vaccinated individuals can be subsequently infected with SARS-CoV-2, and it is important to understand whether these “breakthrough cases” are more or less likely to progress to severe illness. To assess this, we compared the rates of hospitalization, ICU admission, and mortality in individuals with COVID-19 who were infected at least 14 days after a second vaccine dose (breakthrough cases,  $n = 81$ ) versus 1:2 propensity-matched unvaccinated individuals with COVID-19 ( $n = 162$ ) (Figure 1B). These cohorts were balanced for the demographic and clinical variables used in the

**Table 4. Incidence rates of hospitalization within 21 days of SARS-CoV-2 PCR test in vaccinated and 1-to-1 propensity-matched unvaccinated cohorts and corresponding vaccine effectiveness against COVID-19-related hospitalization**

Vaccine	Time period	Vaccinated incidence rate: Cases/person-days [per 1,000 person-days] (no. of individuals contributing)	Unvaccinated incidence rate: Cases/person-days [per 1,000 person-days] (no. of individuals contributing)	Incidence rate ratio (95% CI)	Vaccine effectiveness (95% CI)
BNT162b2	on or after 7 days following the second dose	7/1,915,615. [0.0037] (n = 35,990)	60/1,837,276. [0.033] (n = 35,011)	0.11 (0.043, 0.25)	88.8% (75.5%, 95.7%)
	on or after 14 days following the second dose	6/1,671,628. [0.0036] (n = 33,963)	49/1,599,076. [0.031] (n = 32,910)	0.12 (0.041, 0.27)	88.3% (72.6%, 95.9%)
mRNA-1273	on or after 7 days following the second dose	9/948,311. [0.0095] (n = 16,369)	63/932,315. [0.068] (n = 16,309)	0.14 (0.061, 0.28)	86.0% (71.6%, 93.9%)
	on or after 14 days following the second dose	5/833,681. [0.006] (n = 15,985)	52/817,970. [0.064] (n = 15,896)	0.094 (0.029, 0.23)	90.6% (76.5%, 97.1%)

Incidence rate is calculated as the number of individuals who were hospitalized within 21 days of their first positive PCR test per 1,000 person-days. Time period: time period relative to the first vaccine dose for the vaccinated cohort or study enrollment day for the unvaccinated cohort. Vaccinated incidence rate: number of individuals experiencing the outcome in the vaccinated cohort in the time period, divided by the number of at-risk person-days for the vaccinated cohort in the time period; in brackets, the number of cases per 1,000 person-days. Unvaccinated incidence rate: number of individuals experiencing the outcome in the propensity-matched unvaccinated cohort in the time period, divided by the number of at-risk person-days for the propensity-matched unvaccinated cohort in the time period; in brackets, the number of cases per 1,000 person-days. Incidence rate ratio: vaccinated incidence rate divided by unvaccinated incidence rate along with the exact 95% CI.<sup>16</sup> Vaccine effectiveness:  $100\% \times (1 - \text{incidence rate ratio})$  along with the 95% CI.

matching procedure (Table S2; Figures S2D and S2E), and their distributions of follow-up time from the date of COVID-19 diagnosis are shown in Figure S3B.

Among individuals with at least 21 days of follow-up since diagnosis (n = 32 vaccinated, n = 150 unvaccinated), the 21-day hospitalization rates were similar between these cohorts (16% versus 17%; relative risk = 0.94; 95% CI: 0.43–2.3) (Table S3). No vaccinated patients with breakthrough infections were admitted to the ICU, but the difference in 21-day ICU admission rates was not statistically significant (0% versus 2%; relative risk = 0; 95% CI: 0–12). Consistent with this, Kaplan-Meier analyses from the date of the first positive PCR test indicate that vaccinated and unvaccinated individuals had similar hazard rates for hospitalization (p = 0.50) and ICU admission (p = 0.25) (Figure S6). Similarly, among individuals with at least 28 days of follow-up since diagnosis (n = 26 vaccinated, n = 148 unvaccinated), there were no deaths in the vaccinated group, but the difference in 28-day mortality rates between these cohorts was not statistically significant (0% versus 3.4%; relative risk = 0; 95% CI: 0–8.8). Stratified analyses for each vaccine indicate that neither BNT162b2 nor mRNA-1273 was associated with statistically significant reductions in hospitalization, ICU admission, or mortality during these intervals. It is important to continue monitoring these outcomes as the number of cases expands in the future.

## DISCUSSION

Recent phase 3 trials have led to authorization of three COVID-19 vaccines in the United States.<sup>9–11</sup> Along with other recent real-world analyses, this study provides strong further evidence supporting the use of vaccination to prevent COVID-19.<sup>17,18</sup> In general, real-world analyses of vaccine effectiveness are complicated by the challenge of ascertaining an adequately balanced unvaccinated cohort that can serve as a proxy for the placebo group in a randomized controlled trial. To address this, we used propensity matching to generate cohorts of vaccinated and unvaccinated individuals who are balanced for demographic, geographic, clinical,

**Table 5. Incidence rates of ICU admission within 21 days of SARS-CoV-2 PCR test in vaccinated and 1-to-1 propensity-matched unvaccinated cohorts and corresponding vaccine effectiveness against COVID-19-related ICU admission**

Vaccine	Time period	Vaccinated incidence rate: Cases/person-days [per 1,000 person-days] (no. of individuals contributing)	Unvaccinated incidence rate cases/person-days [per 1,000 person-days] (no. of individuals contributing)	Incidence rate ratio (95% CI)	Vaccine effectiveness (95% CI)
BNT162b2	on or after 7 days following the second dose	0/1,915,733. [0] (n = 35,990)	9/1,838,072. [0.0049] (n = 35,011)	0 (0, 0.49)	100.0% (51.4%, 100%)
	on or after 14 days following the second dose	0/1,671,730. [0] (n = 33,963)	6/1,599,732. [0.0038] (n = 32,910)	0 (0, 0.81)	100.0% (18.7%, 100%)
mRNA-1273	on or after 7 days following the second dose	0/495,630. [0] (n = 11,612)	8/479,979. [0.017] (n = 11,332)	0 (0, 0.57)	100% (43.3%, 100%)
	on or after 14 days following the second dose	0/417,743. [0] (n = 10,610)	6/403,589. [0.015] (n = 10,318)	0 (0, 0.82)	100% (17.9%, 100%)

Incidence rate is calculated as the number of individuals who were admitted to the ICU within 21 days of their first positive PCR test per 1,000 person-days. Time period: time period relative to the first vaccine dose for the vaccinated cohort or study enrollment day for the unvaccinated cohort. Vaccinated incidence rate: number of individuals experiencing the outcome in the vaccinated cohort in the time period, divided by the number of at-risk person-days for the vaccinated cohort in the time period; in brackets, the number of cases per 1,000 person-days. Unvaccinated incidence rate: number of individuals experiencing the outcome in the propensity-matched unvaccinated cohort in the time period, divided by the number of at-risk person-days for the propensity-matched unvaccinated cohort in the time period; in brackets, the number of cases per 1,000 person-days. Incidence rate ratio: vaccinated incidence rate divided by unvaccinated incidence rate along with the exact 95% CI.<sup>16</sup> Vaccine effectiveness: 100% × (1 – incidence rate ratio) along with the 95% CI.

and social variables and then evaluated the effect of vaccination on the rate of positive SARS-CoV-2 PCR testing between these cohorts.

When administered as two serial doses, BNT162b2 and mRNA-1273 were 86.1% (95% CI: 82.4%–89.1%) and 93.3% (95% CI: 85.7%–97.4%) effective in preventing PCR-confirmed SARS-CoV-2 infection with onset at least 7 days after the second dose, respectively. These results are in line with the previously reported efficacies for BNT162b2 (95.0%; 95% CI: 90.3%–97.6%) and mRNA-1273 (94.1%; 95% CI: 89.3%–96.8%) in preventing symptomatic COVID-19 with onset at least 7 or 14 days after the second dose, respectively.<sup>9,10</sup> It should be noted that our effectiveness estimates of BNT162b2 and mRNA-1273 cannot be compared directly because different matched control arms were utilized for each estimate. This was an intentional aspect of our study design because the demographics of individuals receiving these vaccines in the Mayo Clinic health system were clearly distinct from each other. For example, 70% of individuals receiving BNT162b2 were below age 65, and this was true for only 46% of individuals receiving mRNA-1273. Further, 60% of individuals receiving BNT162b2 were female compared with only 53% of those receiving mRNA-1273.

Our finding that BNT162b2 and mRNA-1273 are highly protective against severe COVID-19 (here defined as a positive PCR test followed by hospitalization or ICU admission within 21 days) is consistent with prior clinical trials and real-world studies.<sup>9,10,18</sup> We additionally sought to assess whether prior vaccination affects the risk of experiencing severe illness after an individual tests positive for SARS-CoV-2. This is a relevant question for clinical practice because it must be determined whether vaccination should be considered a beneficial feature in risk stratification algorithms for individuals already diagnosed with COVID-19. Our results suggest that when an individual has been diagnosed with COVID-19, prior vaccination may not robustly protect against progression to severe illness. That said, these conclusions are derived from a relatively small number of cases. With the observed 3.2% mortality among the analyzed cohort of unvaccinated individuals with COVID-19, sample

sizes approximately five times larger would have been required to have an 80% chance to detect a statistically significant 50% risk reduction. In light of that, it is promising that no individuals who contracted COVID-19 at least 2 weeks after their second dose (and have adequate follow-up) in our cohort have been admitted to the ICU or died, and it is possible that a more robust signal for vaccine-induced mitigation of disease severity will emerge over time.

Further investigation of reinfection following COVID-19 vaccination is warranted. In this study, we summarized the clinical outcomes of 81 vaccinated individuals with “breakthrough cases” who tested positive for SARS-CoV-2 at least 14 days after their second dose (Pfizer/BioNTech: 73 individuals, Moderna: 8 individuals). As vaccine rollout continues and the broader population becomes vaccinated, it will be critical to perform holistic analyses of electronic health records (EHR) databases to determine whether features exist that are predictive of infection risk among vaccinated individuals. Such insights could be quite useful at the personal and societal levels with respect to the implementation of post-vaccination behavioral adjustments and guidelines, respectively.

Our data demonstrate a strong real-world effect of both authorized mRNA COVID-19 vaccines on par with the results reported in previous randomized trials. Importantly, these data confirm that these vaccines are highly effective in a population that is enriched with individuals who are at highest risk for acquiring COVID-19 or experiencing severe illness (e.g., healthcare workers and older individuals). Further, this study builds upon the clinical trial results by affirming that vaccination reduces the rate of documented SARS-CoV-2 infection as defined by a positive PCR test alone rather than a positive PCR test in conjunction with symptoms. In summary, we emphasize that BNT162b2 and mRNA-1273 should continue to be administered as broadly and rapidly as possible to the public and that the real-world effectiveness of these and other COVID-19 vaccines should continue to be transparently monitored in the coming months.

### Limitations of study

There are several limitations of this study. First, although the cohorts were even larger than those studied in phase 3 trials, they are certainly not representative of the overall United States population. For example, over 90% of individuals in each vaccinated and unvaccinated cohort were white, approximately 60% of individuals vaccinated with BNT162b2 were female, and over 50% individuals vaccinated with mRNA-1273 were age 65 or older. These features likely represent enrichments among individuals who receive care at the Mayo Clinic or who were prioritized for early vaccination. Related to this point, the variability in vaccine rollout may have affected the study findings. This includes factors such as when the vaccines were made available to different populations, how rapidly the vaccines were administered (e.g., number of doses given per day), and the landscape of SARS-CoV-2 variants circulating in the population at a given time. Although this study does account for geographic variability by matching each vaccinated individual to an unvaccinated individual from the same zip code, these conclusions should continue to be tested longitudinally across diverse populations.

Second, there are several potential confounding factors not considered that may have affected the matching and statistical analysis. The variability in SARS-CoV-2 variants may have affected the disease severity analysis of vaccine breakthrough cases, which were not propensity matched on geographic factors because of the small sample sizes. Socioeconomic factors were not available in the dataset and,

thus, were not included in the matching. In addition, more specific comorbidities, such as disease stage and type of therapy for individuals with cancer, may have influenced vaccine responses and disease severity in COVID-19 breakthrough cases but are not currently captured. It is also possible that the likelihood of seeking out a SARS-CoV-2 PCR test was different between vaccinated and propensity-matched unvaccinated individuals, which would introduce bias into our estimates of vaccine effectiveness. Vaccinated individuals may feel less compelled to undergo subsequent PCR testing, reducing the number of positive tests recorded in this group. However, our data suggest that this is likely not a strong confounding factor during our observation period (i.e., starting 7 days after the first vaccine dose). Although there is a tendency for vaccinated individuals to receive fewer tests directly prior to and after vaccination, the numbers of daily PCR tests are similar over time for the vaccinated and unvaccinated cohorts by the fourth day after the first vaccine dose.

Third, we did not compare the clinical symptomatology of COVID-19 infection between vaccinated and unvaccinated individuals. As mentioned previously, this deviates from the clinical trial analyses that specifically evaluated the rates of symptomatic COVID-19 between individuals receiving a vaccine or placebo. Finally, there is potential for miscategorization of individuals as vaccinated or unvaccinated. However, because the Mayo Clinic COVID-19 vaccination records are linked to their corresponding state registries, it is relatively unlikely that an individual would be classified incorrectly as unvaccinated unless they were vaccinated in a different state.

## 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
  - Matching to select the unvaccinated cohort for vaccine effectiveness analysis
  - Evaluation of vaccine effectiveness in preventing positive SARS-CoV-2 testing
  - Evaluation of vaccine effectiveness in preventing severe disease
  - Propensity score matching to construct a control cohort for breakthrough cases of COVID-19
  - Comparison of outcomes for breakthrough cases and propensity-matched controls
  - Deep neural networks to identify comorbidities from clinical notes
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.medj.2021.06.007>.

## ACKNOWLEDGMENTS

We thank Murali Aravamudan for his mentorship and insightful discussions regarding study design. We thank the peer reviewers for their thoughtful reviews and feedback, which greatly improved the quality of the final manuscript. In addition, we thank the commentators on medRxiv who provided feedback on the initial preprint for this study. Finally, we thank the healthcare workers at Mayo Clinic Health Systems and around the United States who have administered COVID-19 vaccines to millions of people, which has made this real-world effectiveness analysis possible. This study was funded by nference.

## AUTHOR CONTRIBUTIONS

C.P. and A.P. performed the statistical analysis and had unrestricted access to all data. C.P., P.L., A.P., M.J.M.N., A.V., and V.S. prepared the first draft of the manuscript, which was reviewed and edited by all authors. All authors agreed to submit the manuscript, read and approved the final draft, and take full responsibility of its content, including the accuracy of the data and statistical analysis.

## DECLARATION OF INTERESTS

C.P., P.L., A.P., V.A., A.V., M.J.M.N., and V.S. are employees of nference and have financial interests in the company and in the successful application of this research. J.C.O. receives personal fees from Elsevier and Bates College and small grants from nference, Inc., outside of 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., M.D.S., A.V., 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 5, 2021

Accepted: June 17, 2021

Published: June 29, 2021

## REFERENCES

1. Johns Hopkins Coronavirus Resource Center (2021). COVID-19 Dashboard. <https://coronavirus.jhu.edu/map.html>.
2. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220.
3. Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581, 221–224.
4. Venkatakrishnan, A.J., Puranik, A., Anand, A., Zemmour, D., Yao, X., Wu, X., Chilaka, R., Murakowski, D.K., Standish, K., Raghunathan, B., et al. (2020). Knowledge synthesis of 100 million biomedical documents augments the deep expression profiling of coronavirus receptors. (2020. *eLife* 9, e58040.
5. Jackson, L.A., Anderson, E.J., Rouphael, N.G., Roberts, P.C., Makhene, M., Coler, R.N., McCullough, M.P., Chappell, J.D., Denison, M.R., Stevens, L.J., et al.; mRNA-1273 Study Group (2020). An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. *N. Engl. J. Med.* 383, 1920–1931.
6. 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.
7. Voysey, M., Clemens, S.A.C., Madhi, S.A., Weckx, L.Y., Folegatti, P.M., Aley, P.K., Angus, B., Baillie, V.L., Barnabas, S.L., Bhorat, Q.E., et al.; Oxford COVID Vaccine Trial Group (2021). Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* 397, 99–111.
8. Sadoff, J., Le Gars, M., Shukarev, G., Heerwegh, D., Truyers, C., de Groot, A.M., Stoop, J., Tete, S., Van Damme, W., Leroux-Roels, I., et al. (2021). Interim Results of a Phase 1-2a Trial of Ad26.COV2.S Covid-19 Vaccine. *N. Engl. J. Med.* 384, 1824–1835.
9. 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.
10. Baden, L.R., El Sahly, H.M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S.A., Roupshael, 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.
  11. Sadoff, J., Gray, G., Vandebosch, A., Cárdenas, V., Shukarev, G., Grinsztejn, B., Goepfert, P.A., Truyers, C., Fennema, H., Spiessens, B., et al.; ENSEMBLE Study Group (2021). Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. *N. Engl. J. Med.* 384, 2187–2201.
  12. CDC (2021). COVID-19 Vaccine Rollout Recommendations. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations.html>.
  13. CDC (2021). How CDC is making COVID-19 vaccine recommendations. [https://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations-process.html?CDC\\_AA\\_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fvaccines%2Frecommendations.html](https://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations-process.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fvaccines%2Frecommendations.html).
  14. CDC (2021). COVID Data Tracker. <https://covid.cdc.gov/covid-data-tracker/>.
  15. CDC (2021). When you've been fully vaccinated. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/fully-vaccinated.html>.
  16. Sahai, H., and Khurshid, A. (1995). *Statistics in Epidemiology: Methods, Techniques and Applications* (CRC Press).
  17. Dagan, N., Barda, N., Kepten, E., Miron, O., Perchik, S., Katz, M.A., Hernán, M.A., Lipsitch, M., Reis, B., and Balicer, R.D. (2021). BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting. *N. Engl. J. Med.* 384, 1412–1423.
  18. CDC (2021). Fully vaccinated adults 65 and older are 94% less likely to be hospitalized with COVID-19. <https://www.cdc.gov/media/releases/2021/p0428-vaccinated-adults-less-hospitalized.html>.
  19. Austin, P.C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav. Res.* 46, 399–424.
  20. Seabold, S., and Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. In *Proceedings of the 9th Python in Science Conference*, pp. 92–96.
  21. Austin, P.C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Stat. Med.* 33, 1057–1069.
  22. Austin, P.C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat. Med.* 28, 3083–3107.
  23. Stuart, E.A., Lee, B.K., and Leacy, F.P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *J. Clin. Epidemiol.* 66, S84.
  24. Bland, J.M., and Altman, D.G. (2004). The logrank test. *BMJ* 328, 1073.
  25. CDC (2020). COVID-19 and Your Health. <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/evidence-table.html>.
  26. CDC (2021). COVID-19 Breakthrough Case Investigations and Reporting. <https://www.cdc.gov/vaccines/covid-19/health-departments/breakthrough-cases.html>.
  27. Luque Fernandez, M.A. (2020). Delta Method in Epidemiology: An Applied and Reproducible Tutorial. <https://migariane.github.io/DeltaMethodEpiTutorial.nb.html>.
  28. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*, arXiv:1810.04805. <https://arxiv.org/abs/1810.04805>.
  29. 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.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Code</b>		
Python script for propensity score matching to generate the overall vaccinated and unvaccinated cohorts used for effectiveness analyses in this manuscript	This study	<a href="#">Data S1</a>
Python script for overall vaccine effectiveness computations	This study	<a href="#">Data S2</a>
Python script for propensity score matching to generate the vaccinated and unvaccinated COVID-19 patient cohorts for analysis of breakthrough cases	This study	<a href="#">Data S3</a>
Python script for comparison of outcomes in breakthrough cases versus matched unvaccinated COVID-19 patients	This study	<a href="#">Data S4</a>
<b>Software and algorithms</b>		
Python version 3.8.8		<a href="https://www.python.org/">https://www.python.org/</a>
Python software package: statsmodels v0.10.0		<a href="https://www.statsmodels.org">https://www.statsmodels.org</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 datasets supporting the current 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 corresponding author on request. A proposal with detailed description of study objectives and statistical analysis plan will be needed for evaluation of 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: All original code is available in this paper's supplemental information as [Data S1](#). This includes the Python scripts which were used for the statistical analyses including: propensity score matching, analysis of vaccine effectiveness, and analysis of breakthrough cases.
- 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

Vaccine effectiveness analyses 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 1A](#). Demographic summaries of the analyzed cohorts, including age, sex, race, and ethnicity, are provided in [Tables 1](#) and [2](#). Clinical outcomes for these cohorts including incidence of positive PCR tests, hospitalization, and ICU admission are provided in [Tables 3](#), [4](#), and [5](#). Kaplan-Meier curves showing incidence of positive PCR tests for these cohorts are provided in [Figure 2](#). Follow-up testing, dosing, and PCR test information for these cohorts is provided in [Table S1](#). The distribution of the time intervals between first and second dose for both the BNT162b2 and mRNA-1273 vaccinated cohorts are shown in [Figure S1](#).

Analyses of COVID-19 severity as a function of prior vaccination included 243 individuals. Each individual was part of one of the following two cohorts, on the basis of whether they had received two doses of BNT162b2 or mRNA-1273 at least 14 days before COVID-19 diagnosis versus never having received any COVID-19 vaccine: (1) breakthrough COVID-19 patients ( $n = 81$  individuals), or (2) unvaccinated COVID-19 patients ( $n = 162$  individuals). More details describing the participant selection algorithm are provided in the [Method details](#) and are illustrated in [Figure 1B](#). A demographic summary of these cohorts, including age, sex, race, and ethnicity, is provided in [Table S2](#). Clinical outcomes for these cohorts including hospitalization, ICU admission, and mortality rates are provided in [Table S3](#).

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 to the Mayo health system. 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 April 20, 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 Janssen 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. Vaccination status was determined from the Mayo Clinic EHR, which is linked to the state immunization registries of Arizona, Florida, Iowa, Minnesota, and Wisconsin.

Vaccinated individuals who had tested positive for SARS-CoV-2 by PCR between December 1, 2020 and the date of their first vaccine dose (inclusive) 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). As expected, the median and mode of days between first and second dose were 21 days for BTN162b2 and 28 days for mRNA-1273. 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. Balanced unvaccinated cohorts for analyses of vaccine effectiveness were selected from the previously derived set of 238,808 unvaccinated individuals. More details on the matching procedure for the vaccine effectiveness analysis are provided in the next section.

We conducted a similar matched analysis to assess the impact of vaccination upon COVID-19 disease severity. For this analysis, we considered the 81 vaccinated patients who tested positive for SARS-CoV-2 at least 14 days following their second vaccine dose during the study period ("breakthrough infections"), and 20,222 unvaccinated patients who tested positive for SARS-CoV-2 during the study period. For each of the 81 patients with breakthrough infections, we selected 2 controls from the unvaccinated cohort using 1-to-2 propensity score matching. More details on the propensity score matching procedures for the disease severity analysis are provided below.

#### Matching to select the unvaccinated cohort for vaccine effectiveness analysis

We used a combination of exact matching and 1-to-1 propensity score matching to construct an unvaccinated cohort similar to the vaccinated cohort regarding key risk factors for SARS-CoV-2 infection.<sup>19</sup> 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>20</sup> The features included in this model were:

- Demographic features: age, sex, race, ethnicity.
- Records of SARS-CoV-2 PCR testing: number of negative PCR tests taken in three intervals between February 1, 2020 and November 30, 2020: February 1 to May 31, June 1 to August 31, and September 1 to November 30. This feature is intended to serve as a proxy for longitudinal access to and likelihood of seeking out SARS-CoV-2 diagnostic testing.
- Records of diagnostic influenza testing: number of influenza tests (PCR or antigen detection) taken in the same three time intervals as described above for SARS-CoV-2 PCR tests. Unlike SARS-CoV-2 tests, which can be performed for routine surveillance, these tests are typically performed in the context of a symptomatic clinical presentation. This feature is intended to balance the cohorts with respect to their prior experience of influenza-like illness, thereby addressing to some degree the distinction between symptom-driven versus routine or required asymptomatic SARS-CoV-2 testing.
- Long term care (LTC) facility resident: binary variable capturing whether an individual is currently a resident of a LTC facility or nursing home. This feature is included because LTC residents were included in the Phase 1a population for the vaccine rollout due to their elevated risk of acquiring COVID-19 and experiencing more severe disease.

We then attempted to match each of the 85,676 vaccinated individuals (derived above) with one out of the 238,808 unvaccinated individuals using the following steps:

1. Exact match on sex: For a given male vaccinated individual, only the male unvaccinated individuals were considered for matching. For a given female vaccinated individual, only the female unvaccinated individuals were considered for matching.
2. Exact match on geography: For a given vaccinated individual, only unvaccinated individuals with the same zip code were considered for matching. This match helps to account for variability in the vaccine rollout process (i.e., timeline and definition of eligible populations) between and within states.
3. Exact match on LTC facility status: Vaccinated individuals who lived in LTC facilities were matched with unvaccinated individuals who lived in LTC facilities. Similarly, vaccinated individuals who did not live in LTC facilities were matched with unvaccinated individuals who did not live in LTC facilities.
4. Bucketed match on SARS-CoV-2 PCR testing history: All individuals were classified as having 0, 1, or multiple SARS-CoV-2 PCR tests before December 1, 2020, as well as between December 1, 2020 and the date of their study enrollment. To be considered as a possible match, an unvaccinated individual had to exactly match these two bucketed classifications for the given vaccinated individual.
5. Propensity score matching: Among the unvaccinated individuals meeting the exact or bucketed match criteria listed above, one individual was selected by using greedy nearest-neighbor matching without replacement, with a standard caliper of  $0.2 \times$  pooled standard deviation of the logit propensity score.<sup>21</sup> That is, the remaining unvaccinated individual with a propensity score closest to that of the given vaccinated individual was selected.

If an unvaccinated individual met the above criteria but had tested positive for SARS-CoV-2 on or before the date of study enrollment, then that individual was returned to the pool and we attempted to identify a new matched unvaccinated individual. If no unvaccinated individuals met the criteria to be considered as a possible match for a given vaccinated individual, then that vaccinated individual was excluded from further analysis. From the 85,676 eligible vaccinated individuals, we were able to identify valid matches for 68,266. Thus, our final vaccinated and unvaccinated cohorts each contained 68,266 individuals ( $n = 51,795$  each for BTN162b2, and  $n = 16,471$  each for mRNA-1273). We ensured that the resulting cohorts were balanced by assessing the standardized mean differences (SMD) of their clinical covariates.<sup>22,23</sup> Overall, there is no substantial difference between the two cohorts in any of the clinical covariates that were included in propensity score matching (with  $SMD < 0.1$  for all covariates) (see [Tables 1](#) and [2](#)). In [Figures S2A–S2C](#), we show the age distributions before and after propensity score matching, demonstrating how the procedure is effective in balancing this covariate for both the BNT162b2 and mRNA-1273 comparisons. The code used to perform this matching is provided in [Data S1](#).

### Evaluation of vaccine effectiveness in preventing positive SARS-CoV-2 testing

To evaluate the effectiveness of the FDA-authorized mRNA COVID-19 vaccines in a real-world clinical setting, we compared the populations described above. The code used to perform these effectiveness analyses is provided in [Data S2](#). For a given matched pair, the following time points were defined for various effectiveness analyses:

- *Day D<sub>1</sub>*: for a given matched pair, this corresponds to the date of study enrollment. This is defined as the date of the first vaccine dose for the vaccinated individual.

- *Day D<sub>2</sub>*: for a given matched pair, this corresponds to the date of the second vaccine dose for the vaccinated individual. If the vaccinated individual has only received one dose, then this time point is not defined for a given matched pair.
- *Day E*: this corresponds to the last date of data collection (April 20, 2021).
- *Day C*: for a given matched pair, this corresponds to the date of censoring which is defined only if the vaccinated individual did not receive a second vaccine dose by one week after the recommended date. For a matched pair in which the vaccinated individual received BNT162b2, this is defined as Day D<sub>1</sub> + 28 days. For a matched pair in which the vaccinated individual received mRNA-1273, this is defined as Day D<sub>1</sub> + 35 days.
- *Day S*: for a given matched pair, this corresponds to the final day that was eligible for inclusion as an at-risk person day for the single-dose effectiveness analyses. This is defined as the earliest day among Day *D<sub>2</sub>*, Day *E*, and Day *C*.
- *Day F*: for a given matched pair, this corresponds to the final day that was eligible for inclusion as an at-risk person day for the overall effectiveness analyses. This is defined as the earliest day among Day *E* and Day *C*.

The outcome of interest was a positive SARS-CoV-2 PCR test. In [Figure S3A](#), we show the distribution of follow-up time during which this outcome was measured for the vaccinated cohorts. The distributions of follow-up time for the propensity matched unvaccinated cohorts are identical due to the study design which exactly matches on the first vaccine dose date as the index date. It should be noted that the experience of this outcome depends on an individual seeking out and obtaining the test (e.g., as a result of experiencing symptoms), as individuals were not routinely or randomly tested in this study. Thus, this outcome does not measure absolute infection rates but instead likely approximates the rates of symptomatic and/or self-reported SARS-CoV-2 infection in these cohorts. Importantly, by including the number of SARS-CoV-2 PCR tests taken prior to the date of study enrollment in our matching procedure, we intended to derive cohorts with individuals who were similarly likely to have access to and seek out such testing. We found that even after this balancing, vaccinated individuals were significantly less likely to undergo testing during the first three days after vaccination, which may be due to the confusion of COVID-19 symptoms for vaccine associated side effects ([Figure S4](#)). However, within 7 days after the first vaccine dose, the daily testing rates were again stably similar between the vaccinated and unvaccinated cohorts. We thus conservatively decided to exclude the first week after study enrollment from all effectiveness analyses.

Cumulative incidence of SARS-CoV-2 infection was compared between vaccinated and unvaccinated individuals by Kaplan Meier analysis. Cumulative incidence at time *t* is the estimated proportion of individuals who experience the outcome on or before time *t* (i.e., 1 minus the standard Kaplan-Meier survival estimate). To analyze the effectiveness of a single vaccine dose, we considered cumulative incidence from 7 days after Day *D*<sub>1</sub> through Day *S*. To analyze the effectiveness of full vaccination, we considered the cumulative incidence from 7 days after Day *D*<sub>2</sub> onward among matched pairs in which the vaccinated individual received their second dose no more than 7 days after the recommended time (i.e., no more than 28 days after the first dose for BNT162b2, or no more than 35 days after the first dose for mRNA-1273). Statistical significance was assessed with the log rank test.<sup>24</sup>

Effectiveness was also assessed during defined intervals by computing the incidence rate ratio (IRR) of the vaccinated and unvaccinated cohorts. Effectiveness was defined as 100% × (1 - IRR). For each cohort in a given time period, incidence rates

were calculated as the number of individuals testing positive for SARS-CoV-2 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 in the time period in which the individual has not yet tested positive for SARS-CoV-2 or died. The IRR was calculated as the incidence rate of the vaccinated cohort divided by the incidence rate of the unvaccinated cohort, and its 95% confidence interval was computed using an exact approach.<sup>16</sup>

To evaluate the overall effectiveness of vaccination, we computed incidence rates (i) from 7 days after Day  $D_1$  onward (through Day  $F$ ), and (ii) from 14 days after Day  $D_1$  onward (through Day  $F$ ). To evaluate single dose effectiveness, we computed incidence rates (i) from 7 days after Day  $D_1$  through Day  $S$ , and (ii) from 14 days after Day  $D_1$  through Day  $S$ . To evaluate the effectiveness of full vaccination, we computed incidence rates (i) from 7 days after Day  $D_2$  onward and (ii) from 14 days after Day  $D_2$  onward among matched pairs in which the vaccinated individual received their second dose no more than 7 days after the recommended time (i.e., no more than 28 days after the first dose for BNT162b2, or no more than 35 days after the first dose for mRNA-1273). In [Figure S5](#), we show the distribution of the time from first vaccine dose to first positive PCR test for all of the vaccinated individuals with subsequent positive PCR tests.

#### Evaluation of vaccine effectiveness in preventing severe disease

To evaluate the effectiveness of full vaccination (i.e., two doses) with BNT162b2 and mRNA-1273 in preventing severe COVID-19, we followed a similar framework to the one described in the previous section. However, here we defined two new outcomes of interest to replace the previous one (i.e., a positive SARS-CoV-2 test): (i) hospitalization occurring within 21 days of the first positive SARS-CoV-2 PCR test ("COVID-19 associated hospitalization") and (ii) ICU admission occurring within 21 days of the first positive SARS-CoV-2 PCR test ("COVID-19 associated ICU admission"). For each of these outcomes, we computed incidence rates, IRRs, and effectiveness as described in the previous section for two time intervals that capture the effect of full vaccination: (i) from 7 days after Day  $D_2$  onward and (ii) from 14 days after Day  $D_2$  onward. Note that, as described for the full vaccination effectiveness calculations above, only matched pairs in which the vaccinated individual received their second dose no more than 7 days after the recommended time (i.e., no more than 28 days after the first dose for BNT162b2, or no more than 35 days after the first dose for mRNA-1273) were included for this analysis. The code used to perform these analyses is provided in [Data S2](#).

#### Propensity score matching to construct a control cohort for breakthrough cases of COVID-19

To understand whether prior vaccination impacts the risk of progressing to severe COVID-19 once an individual tests positive for SARS-CoV-2, we applied 1:2 propensity score matching to construct a SARS-CoV-2 positive unvaccinated cohort similar in baseline clinical covariates to the cohort of patients who tested positive for SARS-CoV-2 at least 14 days after their second COVID-19 vaccine dose, also known as "breakthrough cases."<sup>19</sup> In particular, we used propensity score matching to match approximately based upon demographic features (age, sex, race, ethnicity) and comorbidities (asthma, cancer, cardiomyopathy, chronic kidney disease, chronic obstructive pulmonary disease, coronary artery disease, heart failure, hypertension, obesity, pregnancy, severe obesity, sickle cell disease, solid organ transplant, stroke / cerebrovascular disease, type 2 diabetes mellitus). This list of comorbidities was derived from the list of risk factors for severe COVID-19 illness provided by the

Centers of Disease Control and Prevention.<sup>25</sup> We used deep neural networks to automatically identify comorbidities from the clinical notes, which are described in the next section. To obtain the propensity scores, we trained a regularized logistic regression model with these features using the software package statsmodels v0.10.0 in Python.<sup>20</sup>

Based on these propensity scores, we matched each of the individuals that tested positive for SARS-CoV-2 after full vaccination ( $n = 81$ ) with 2 individuals that tested positive for SARS-CoV-2 and were not vaccinated. As in the previous propensity score matching procedure, we used greedy nearest-neighbor matching without replacement.<sup>21</sup> In order to be considered a valid match, the date of the first positive PCR test for an unvaccinated individual was required to be within 28 days of the first positive PCR test for the corresponding vaccinated individual. This step was taken to ensure that important temporal aspects of COVID-19 severity (e.g., landscape of viral variants, available treatments or established standards of care) were generally shared between the three members of any given matched set. We ensured that the resulting cohorts were balanced by assessing the standardized mean differences (SMD) of their clinical covariates.<sup>22,23</sup> Overall, there is no substantial difference between the two cohorts in any of the clinical covariates that were included in propensity score matching (with  $SMD < 0.1$  for all covariates) (see [Table S2](#)). In [Figures S2D](#) and [S2E](#), we show the age distributions before and after propensity score matching, demonstrating how the procedure is effective in balancing this covariate. The code used to perform this matching is provided in [Data S3](#).

### Comparison of outcomes for breakthrough cases and propensity-matched controls

We compared three clinical outcomes between vaccinated COVID-19 patients (“breakthrough cases”) and propensity-matched unvaccinated COVID-19 patients in order to evaluate the impact of vaccination upon the disease severity: (i) 21-day hospitalization rate, (ii) 21-day ICU admission rate, and (iii) 28-day mortality rate. Vaccinated individuals were only included if their first positive SARS-CoV-2 PCR test occurred at least 14 days after their second vaccine dose ( $n = 81$ ), per the CDC definition of breakthrough infections.<sup>26</sup> In [Figure S3B](#), we show the distribution of follow-up time during which this outcome was measured for the vaccinated cohort. The distribution of follow-up time for the propensity matched unvaccinated cohort is identical due to the study design which exactly matches on the PCR diagnosis date as the index date. The code used to perform these analyses is provided in [Data S4](#).

Hospitalization and ICU admission rates were assessed among individuals with at least 21 days of follow-up after their first positive SARS-CoV-2 PCR test ( $n = 32$  vaccinated,  $n = 150$  unvaccinated). Mortality rates were assessed among individuals with at least 28 days of follow-up after their first positive test ( $n = 26$  vaccinated, 148 unvaccinated). For each outcome, we report the relative risk (rate in the vaccinated cohort divided by the rate in the matched unvaccinated cohort), 95% confidence interval for the relative risk, and the Fisher exact test p value.<sup>27</sup>

Hospital-free and ICU-free survival from the date of COVID-19 diagnosis (defined by the first positive SARS-CoV-2 PCR test) were also compared via Kaplan-Meier analysis, with statistical significance assessed via the log rank test.<sup>24</sup> For this analysis, all 81 fully vaccinated and 162 unvaccinated COVID-19 patients were analyzed (i.e., there was no filtering based on the amount of follow-up time after COVID-19 diagnosis). In [Figure S6](#), Kaplan-Meier curves showing hospitalization and ICU rates over time for the matched cohorts are provided.

### Deep neural networks to identify comorbidities from clinical notes

In order to identify the comorbidities from the electronic health record for each patient, we used a BERT-based neural network model to classify the sentiment for the phenotypes that appeared in the clinical notes.<sup>28</sup> In particular, we applied a phenotype sentiment classification model that had been trained on 18,500 sentences which achieves an out-of-sample accuracy of 93.6% with precision and recall scores above 95%.<sup>29</sup> This classification model predicts four classes, including: (1) “Yes”: confirmed diagnosis (2) “No”: ruled-out diagnosis, (3) “Maybe”: possibility of disease, and (4) “Other”: alternate context (e.g., family history of disease). For each patient, we applied the sentiment model to the clinical notes in the Mayo Clinic electronic health record from December 1, 2015 to November 30, 2020. For each comorbidity phenotype, if a patient had at least one mention of the phenotype during the time period with a confidence score of 90% or greater, then the patient was labeled as having the phenotype.

### QUANTIFICATION AND STATISTICAL ANALYSIS

We assessed the quality of balancing achieved by our patient matching procedure by calculating the standardized mean differences (SMD) between cohorts for each of the clinical covariates.<sup>22,23</sup> For each covariate, the cohorts were assumed to be adequately balanced if the SMD was no greater than 0.1.

Vaccine effectiveness in preventing a given clinical outcome (i.e., positive SARS-CoV-2 test, COVID-19 associated hospitalization, or COVID-19 associated ICU admission) was quantified by computing the incidence rate of the given outcome in vaccinated individuals ( $IR_{\text{vaccinated}}$ ) and the incidence rate in propensity matched unvaccinated individuals ( $IR_{\text{unvaccinated}}$ ). The incidence rate ratio (IRR) was calculated as  $IRR = IR_{\text{vaccinated}} / IR_{\text{unvaccinated}}$ . The IRR 95% confidence interval was computed using an exact approach.<sup>16</sup> Vaccine effectiveness (VE) was defined as  $VE = 100\% \times (1 - IRR)$ . Vaccination was considered to be effective if the lower bound of the 95% confidence interval for effectiveness was at least 30%.

Cumulative incidence of a given outcome (e.g., SARS-CoV-2 infection, hospitalization, or ICU admission) was compared between vaccinated and unvaccinated individuals by Kaplan Meier analysis. Cumulative incidence at time  $t$  is the estimated proportion of individuals who experience the outcome on or before time  $t$  (i.e., 1 minus the standard Kaplan-Meier survival estimate). Statistical significance was assessed with the log rank test.<sup>24</sup> The difference in cumulative incidence was considered significant if  $p < 0.05$ .

To compare outcomes in COVID-19 patients who were previously vaccinated (“breakthrough infections”) versus COVID-19 patients who were not previously vaccinated, we calculated the fraction of patients in each cohort with adequate follow-up who were hospitalized with 21 days of COVID-19 diagnosis, admitted to the ICU within 21 days of COVID-19 diagnosis, or died within 28 days of COVID-19 diagnosis. Relative risk (RR) was calculated as  $RR = \text{Fraction experiencing outcome}_{\text{vaccinated}} / \text{Fraction experiencing outcome}_{\text{unvaccinated}}$ . To assess the statistical significance, we calculated the Fisher exact test  $p$  value and the 95% confidence interval for the relative risk using the delta method.<sup>27</sup> The relative risk was considered significant if  $p < 0.05$  and the 95% confidence interval did not include 1. Kaplan Meier analyses to assess the cumulative incidence of hospitalization and ICU admission in these cohorts was performed as described above.