scientific reports

OPEN

Impact of vaccination on the COVID-19 pandemic in U.S. states

Xiao Chen[1], Hanwei Huang[2,3]✉, Jiandong Ju[4], Ruoyan Sun[5]✉ & Jialiang Zhang[4]

Governments worldwide are implementing mass vaccination programs in an effort to end the novel coronavirus (COVID-19) pandemic. Here, we evaluated the effectiveness of the COVID-19 vaccination program in its early stage and predicted the path to herd immunity in the U.S. By early March 2021, we estimated that vaccination reduced the total number of new cases by 4.4 million (from 33.0 to 28.6 million), prevented approximately 0.12 million hospitalizations (from 0.89 to 0.78 million), and decreased the population infection rate by 1.34 percentage points (from 10.10 to 8.76%). We built a Susceptible-Infected-Recovered (SIR) model with vaccination to predict herd immunity, following the trends from the early-stage vaccination program. Herd immunity could be achieved earlier with a faster vaccination pace, lower vaccine hesitancy, and higher vaccine effectiveness. The Delta variant has substantially postponed the predicted herd immunity date, through a combination of reduced vaccine effectiveness, lowered recovery rate, and increased infection and death rates. These findings improve our understanding of the COVID-19 vaccination and can inform future public health policies.

The novel coronavirus (COVID-19) pandemic has had a devastating impact on health and well-being, with more than 131 million cases and 2.8 million deaths across more than 200 countries[1] as of early April 2021. Despite various regional and national non-pharmaceutical interventions[2–4] such as travel restrictions, social distancing measures, stay-at-home orders, and lockdowns, many countries continue to struggle with the growth of COVID-19 cases. It is obvious that a successful COVID-19 vaccination program is needed to end the pandemic and allow a return to normal life[5,6].

By the end of February 2021, two COVID-19 vaccines had been approved in the U.S.: BNT162b2 (Pfizer/BioNTech) and mRNA-1273 (Moderna)[7]. In two large randomized controlled trials (RCTs), the Pfizer vaccine exhibited an efficacy of 95% (95% confidence interval [CI], 90.3%–97.6%)[8] in preventing COVID-19, and the Moderna vaccine showed an efficacy of 94.1% (95% CI, 89.3%–96.8%)[9]. Both are mRNA vaccines that require two doses to complete vaccination and received emergency use authorization by the U.S. Food and Drug Administration in December 2020[10]. Mass vaccination campaigns with these two vaccines have since begun. By early March 2021, more than 121 million doses had been administered across the U.S., with over 43 million individuals (~13% of the population) fully vaccinated with two doses[11].

Although the efficacies of these two vaccines were shown to be high in RCTs, there is limited information on their potential population-level impact on the COVID-19 pandemic. One peer-reviewed study that estimated vaccine effectiveness used data from nationwide mass vaccination in Israel and reported the effectiveness of the Pfizer vaccine to be 46% (95% CI, 40%–51%) after the first dose and 92% (95% CI, 88%–95%) after the second dose for documented infection[12]. Another study that examined the effectiveness of the Pfizer vaccine among U.S. residents in skilled nursing facilities reported an estimation of 63% (95% CI, 33%–79%) after the first dose[13].

In this study, we employed well-established reduced-form econometric techniques[14], commonly used to evaluate the effects of policies or events[15,16], to assess the impact of early-stage vaccination during the ongoing outbreak using data from all 50 U.S. states and the District of Columbia (DC). Although the allocation of vaccines is roughly proportional to state population (Extended Data Fig. 1a), the actual proportion of the vaccinated population differs significantly across states over time (Extended Data Fig. 1b), which provides the key variation to identify the impact of vaccination. Effectively, the observations from each region in the weeks before the vaccination program served as the “control” for the observations after the vaccination program began

[5]Department of Health Care Organization and Policy, School of Public Health, University of Alabama at Birmingham, Birmingham, USA. ✉email: ryansun@uab.edu

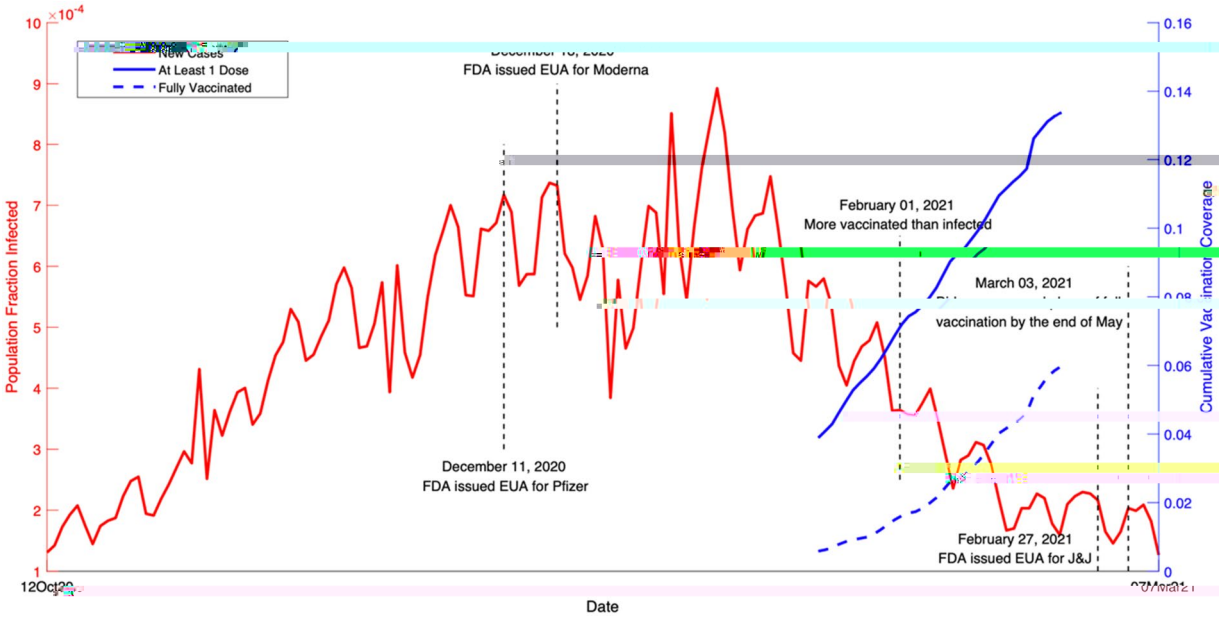


Figure 1. COVID-19 events and vaccination timeline in the U.S. from 12 October 2020 to 7 March 2021. e red curve is the fraction of population infected over time (le y-axis). e solid blue curve is the cumulative vaccination coverage in the population with at least one dose of vaccine (right y-axis). e dashed blue curve is the cumulative vaccination coverage of fully vaccinated individuals in the population (right y-axis).

("treatment"), with variations in the vaccination rates leading to changes in the "treatment intensity." By comparing the outcomes across states before and a er the initiation of vaccination programs, we evaluated the impact of vaccination on the COVID-19 pandemic.

Results

Study design. We collected state-level daily infection and hospitalization data in the U.S. from 12 October 2020 to 7 March 2021. Figure 1 shows a timeline of COVID-19 developments during this period, including important events and vaccination timeline. We aggregated the data to a weekly level in our baseline estimation given the observed weekly cycle[17,18] (see Extended Data Table 3 for results using daily data). e dependent variables used to assess the impact of vaccination on the pandemic are the growth rates of total cases and hospitalizations. Our key independent variables are vaccination rates, including the total number of vaccine doses administered per 100 people (at least one dose) and the total number of second doses administered per 100 people. Without any control variables, Fig. 2 shows the negative correlation between the vaccination rate and the growth rates of total cases and hospitalizations.

We analyzed data in the U.S. from 12 October 2020 to 7 March 2021 for three main reasons. First, we selected similar number of weeks for the pre-treatment and post-treatment periods to balance the sample in our baseline results. Second, two important variables, growth of total hospitalizations and testing, are only available till early March 2021. ird, the Delta variant started to spread since March 2021 and became the dominant strain in the U.S. by July 2021. e presence of the Delta variant has signi cantly changed vaccine e ectiveness, along with infection rate and recovery rate[19]. us, we chose to examine the e ect of early-stage vaccination (till 7 March 2021) in our main text, and leave the analysis of extended data up to 17 November 2021 in the Discussion.

To make the individual states as comparable as possible, we rst accounted for observable factors associated with the COVID-19 pandemic based on previous studies (see Extended Data Table 1). ese time-varying control variables included non-pharmaceutical interventions[5–7], election rallies[20,21] and anti-racism protests[22] that involved mass gatherings, and climate measures of snow depth and temperature[23]. To address the concern that changes in the number of total cases re ect the testing capacity of each state[24], we also controlled for each state's testing capacity. As the proportion of susceptible individuals declines, the infection rate may slow; therefore, we included the share of susceptible individuals in the regressions. We estimated the dependent variables of COVID-19 cases and hospitalizations with a one-week lag to account for the latency period of infection. Finally, we added state xed e ects and time xed e ects to capture spatial and temporal invariants to alleviate omitted-variable bias.

Impact of vaccination. Our data show that the national average weekly growth rate of total cases was 7% (s.e.m. = 5%) between 12 October 2020 and 7 March 2021. At the individual state level, the average growth rate was highest in Vermont (11%) and lowest in Hawaii (4%). e average growth rate of total hospitalizations across the 35 states that reported hospitalization data was 5% (s.e.m. = 4%); the highest growth rate was seen in Montana (8%) and the lowest in New Hampshire (2%).

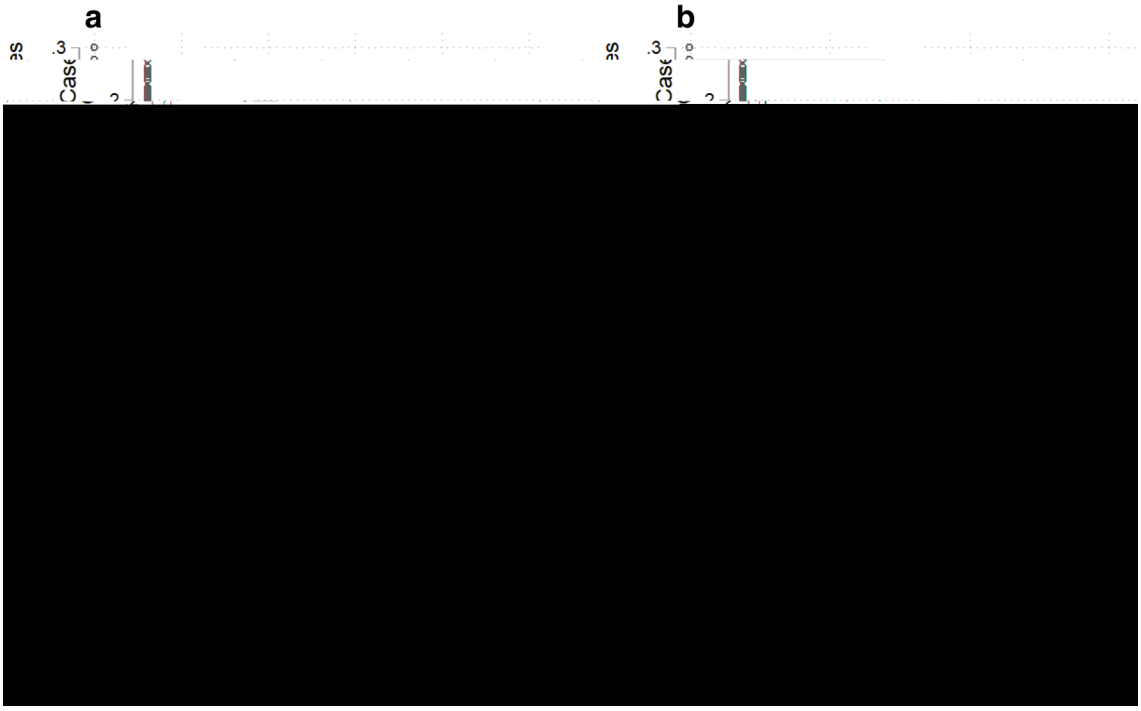


Figure 2. COVID-19 infections (total cases and hospitalizations) and vaccination rate. Vaccination rate is the number of individuals vaccinated per hundred. e solid line in each gure is a tted linear curve between the growth rate of total cases/hospitalizations and vaccination rate. (**a**), Association between the growth rate of total cases and at least 1 dose of vaccination (coe cient = − 0.006, R^2 = 35.3%). (**b**) Association between the growth rate of total cases and 2 doses of vaccination (coe cient = − 0.013, R^2 = 28.6%). (**c**) Association between the growth rate of total hospitalizations and at least 1 dose of vaccination (coe cient = − 0.003, R^2 = 20.8%). (**d**), Association between the growth rate of total hospitalizations and 2 doses of vaccination (coe cient = − 0.007, R^2 = 16.6%).

Vaccination has signi cantly slowed the growth of total COVID-19 cases and hospitalizations in the U.S. Our baseline results (Fig. 3a and Extended Data Table 2) show that one additional vaccinated individual per 100 people (at least 1 dose) reduced the growth rate of total cases by 0.7% (s.e.m. = 0.2%) and the growth rate of total hospitalizations by 0.7% (s.e.m. = 0.2%). e e ects of receiving full vaccination with two doses appear greater, with reductions of 1.1% (s.e.m. = 0.4%) in the growth rate of total cases and 1.1% (s.e.m. = 0.3%) in total hospitalizations. Based on these estimates, vaccination reduced the number of new cases during our study period by 4.4 million (from 33.0 to 28.6 million), which translates into a decrease of 1.34 percentage points in the population infection rate (from 10.10% to 8.76%). Vaccination further reduced the number of hospitalizations by approximately 0.12 million, from 0.89 to 0.78 million (Fig. 3b and Supplementary Methods).

If systematic correlations existed between the pre-vaccination growth rates of infection and hospitalization and the rate of vaccination, our results would have been subject to selection bias. However, this was not the case. We demonstrated that the number of vaccines allocated to each state was proportional to its population size (Extended Data Fig. 1a). More importantly, we found that the pre-vaccination average growth rates of total cases and hospitalizations were not correlated with the average vaccination rate (Extended Data Fig. 2).

Our baseline results focus on the average treatment e ect of vaccination. is e ect may be heterogeneous across states that have di erent characteristics. For example, some evidence shows that the prevalence of COVID-19 di ers across age groups, with older adults bearing the highest risk[25,26]. Because older adults were given priority during the rollout of vaccination, it is intuitive to ask whether this strategy made a di erence. We separated the states into two groups according to their proportion of older adults (at least 65 years of age). Despite the slightly larger point estimate for the states with a share of older adults above the national median, the results do not di er signi cantly from those for the states below the median (Extended Data Fig. 3c). In addition to age, we conducted heterogeneity tests on political a liation, nonpharmaceutical interventions, race, income, and vaccine brand. We found no signi cant heterogeneous e ect of vaccination on any of these characteristics (Extended Data Fig. 3), implying that COVID-19 vaccines have similar e ectiveness across these characteristics.

We conducted a range of sensitivity tests. First, instead of using weekly data, we ran regressions with daily data and obtained results of similar magnitudes (Extended Data Table 3). Second, we used alternative measures to capture the development of the pandemic, including the logarithms of new cases and hospitalizations and the changes in logarithms of total cases and hospitalizations. Again, using these measures, we found that vaccination has signi cantly slowed the pandemic (Extended Data Fig. 4 and Extended Data Table 4). Although the vaccination rollout began on 14 December 2020, our vaccination data begin 11 January 2021; we thus used linear extrapolation to impute the missing data. Our results with the inclusion of imputed data are very similar to the

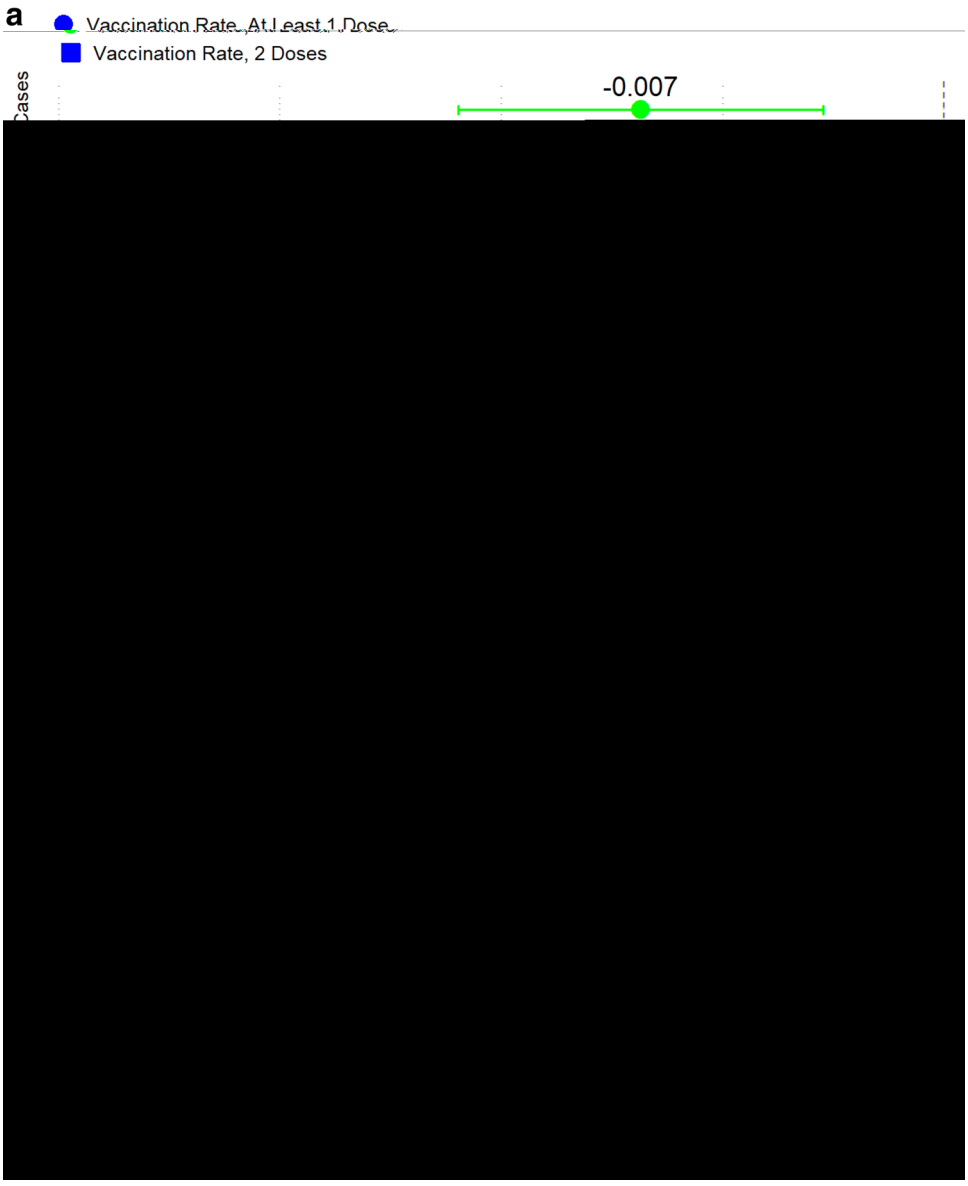
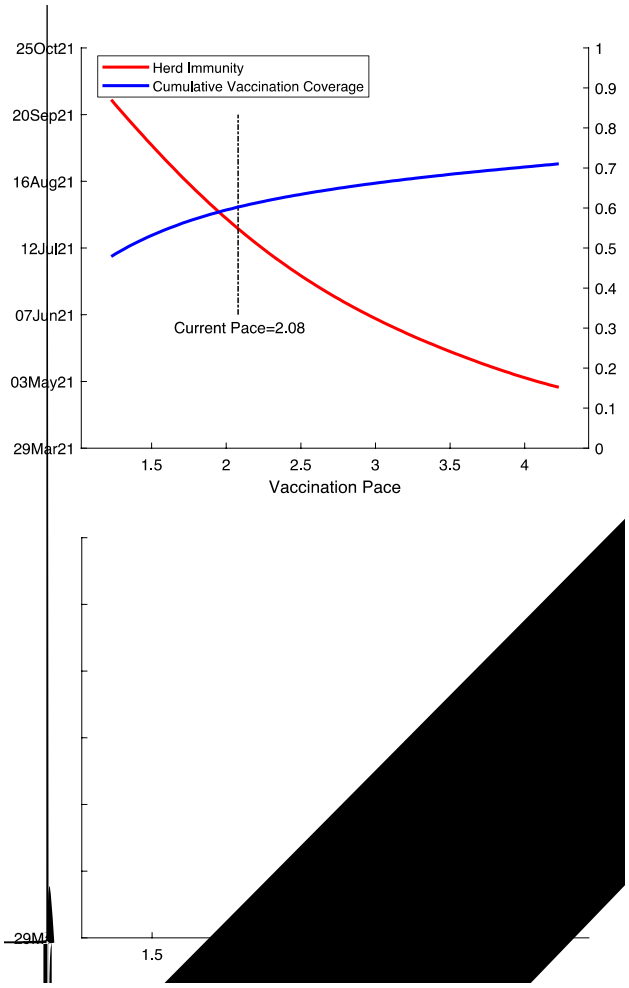


Figure 3. Estimated e ects of vaccination on the COVID-19 pandemic. Blue markers are the estimated e ects of at least 1 dose of vaccine, and red markers are the estimated e ects of 2 doses of vaccine. (**a**) Baseline e ect of vaccination on the growth rates of total cases and hospitalizations. (**b**) Estimated trajectories of total cases and hospitalizations without vaccines (dashed curves) versus actual trajectories of total cases and hospitalizations with vaccines (solid curves).

baseline results (Extended Data Fig. 5). Finally, we selected approximately the same number of weeks for the pre-treatment and post-treatment periods to balance the sample in our baseline results. To check the sensitivity of our results to the sample period, we ran our regressions with varying time windows, and our results remain remarkably stable. We obtained approximately the same coe cients for sample periods from 18 to 45 weeks (Extended Data Fig. 6).

Herd immunity. To predict how the pandemic will develop with vaccines, and especially when herd immunity might be achieved, we built a Susceptible–Infected–Recovered (SIR) model with vaccination and calibrated it to our data. We also aim to identify important factors that substantially a ect the predicted date of herdy immunity. Our model predictions of the infection rate during the study period showed 99.69% correlation with the empirical data at the national level by early March 2021 (Extended Data Fig. 7). Herd immunity is achieved in the model when the real-time basic reproduction number is less than one (Supplementary Methods).

According to our model predictions, at the national average vaccination pace of 2.08 doses per 100 people per week between January and early March of 2021, the U.S. would achieve herd immunity around the last week of July 2021, with a cumulative vaccination coverage rate of 60.2% and a cumulative infection rate of 13.3%. To understand how the speed of vaccination rollout would a ect the time needed to reach herd immunity, we



simulated herd immunity dates by varying vaccination pace (Fig. 4). We observed a general trend that a faster vaccination pace would allow the U.S. to achieve herd immunity sooner, but with a greater number of total vaccine doses administered and a lower cumulative infection rate. is result can be explained as more individuals gaining immunity from vaccines than from infections if the vaccination pace increases.

Our predictions of herd immunity assume a continuation of vaccine uptake. In reality, however, a few potential factors could a ect uptake. A certain proportion of the population might not receive the vaccination due to vaccine hesitancy. Studies have shown that vaccine hesitancy is a common phenomenon in the U.S.[27,28], where some individuals are reluctant to receive vaccines due to the perceived risks versus bene ts, certain religious beliefs, and a lack of trust in government[28]. Another issue is the e ectiveness of vaccines against new coronavirus variants[29]. For example, vaccine e ectiveness is lower against the Delta variant and it remains unclear how the vaccine is e ective at preventing the Omicron variant[30,31].

To examine how vaccine hesitancy and changes in vaccine e ectiveness could a ect our predictions for herd immunity, we incorporated in our model a range of potential vaccine hesitancy and vaccine e ectiveness estimates. We assumed that if x% of the population is hesitant, then cumulative vaccination coverage in each state will stop when (1 − x%) of the population is vaccinated. Table 1 shows that a higher percentage of vaccine-hesitant individuals will lead to lower vaccination coverage with more individuals infected with COVID-19 at herd immunity. We also tested a range of vaccine e ectiveness values and presented the results in Table 1. In general, higher vaccine hesitancy and lower vaccine e ectiveness postpone the model-predicted herd immunity date. We further discuss the potential impact of the Delta variant on herd immunity date in the Discussion with updated data.

	Herd Immunity Date[a]	Vaccination Coverage[b]	Fraction Infected[b]
Pace = 1.5			
Vaccine Hesitancy[c]			
10%	06 Sep 2021	53.6%	15.2%
30%	06 Sep 2021	53.3%	15.2%
50%	06 Sep 2021	48.4%	16.3%
Vaccine E ectiveness[d]			
60%	27 Sep 2021	58.1%	16.7%
80%	23 Aug 2021	50.6%	14.6%
100%	26 Jul 2021	44.6%	13.3%
Pace = 2.08 (national average pace between January and early March in 2021)			
Vaccine Hesitancy			
10%	26 Jul 2021	61.4%	13.4%
30%	26 Jul 2021	60.9%	13.9%
50%	30 Aug 2021	50.0%	14.9%
Vaccine E ectiveness			
60%	16 Aug 2021	67.6%	14.7%
80%	12 Jul 2021	57.2%	12.9%
100%	14 Jun 2021	48.9%	11.8%
Pace = 2.5			
Vaccine Hesitancy			
10%	28 Jun 2021	63.6%	12.6%
30%	28 Jun 2021	63.1%	13.2%
50%	20 Sep 2021	50.0%	14.3%
Vaccine E ectiveness			
60%	26 Jul 2021	73.5%	13.8%
80%	14 Jun 2021	58.6%	12.1%
100%	24 May 2021	51.1%	11.1%

Table 1. Predicted herd immunity with di erent vaccination pace, vaccine hesitancy, and vaccine e ectiveness estimates. [a]Estimated week when herd immunity is achieved. e date mentioned in each row marks the rst day of the week. [b]Cumulative values when herd immunity is achieved. [c]Percentage of the population who are hesitant to get the vaccine. [d]Population-level e ectiveness of the rst dose of COVID vaccines.

Discussion

To examine whether our main study results hold in later stages of the vaccination program, we extended our empirical analysis to 14 November 2021. Due to data limitations, we can only update one outcome measure, the growth of COVID-19 cases, but not hospitalization. As shown by Extended Data Fig. 8, vaccination (both at least one dose and two doses) is always negatively associated with the growth of total cases, but the magnitude of the estimated e ect declines and the statistical signi cance gradually disappears with extended study time. is nding is not unexpected. First, there is evidence that the protection o ered by vaccines against COVID decreases over time[19,32]. Second, the Delta variant became the dominant strain in the U.S. by mid-summer 2021. Studies have shown that the vaccines have lower e ectiveness against the Delta variant[30,31].

We also incorporated updated data on vaccine hesitancy and vaccine e ectiveness in our SIR model to predict herd immunity. e weighted rst-dose vaccine e ectiveness is reduced to 52.28% against the Delta variant[30,31]. Based on vaccination coverage data in November, around 70% of the U.S. population received at least one dose. We thus approximated vaccine hesitancy in the population to be 30%. Additionally, researchers estimated the infectiousness of the Delta variant to be 40–60% higher than previous variants[33], along with longer median duration (18 vs 13 days) and lower recovery rate (calculated as 1/duration)[34]. With these updated parameter estimates, our SIR model predicts the new herd immunity date to be around May 2022 (140% infection rate and 70% removal rate). We tested a range of possible values of the infection rate and removal rate (recovery rate + death rate) and show our results in Extended Data Table 4. In general, a lower removal rate tends to delay the herd immunity date, but the e ect of a higher infection rate is ambiguous as agents can also get immunity via faster infection.

Our model has a few limitations. First, due to the lack of valid COVID recovery data from a few states, we imputed the missing removal rate in those states to be the national median. To address the problem of missing values, we conducted robustness checks on the removal rate. e results indicated that the national median value, which is our baseline, ts the data better, and the herd immunity date is not very sensitive to variations in removal rate (Extended Data Fig. 9). Second, our model does not consider some recent developments of the pandemic, given that it is designed to model the early stage of the vaccination program. It does not take into account new variants such as the Delta or the Omicron. Our SIR model also assumes that only susceptible individuals undergo

vaccination. However, in real life, many individuals who recovered from COVID later received vaccines. e biggest limitation is the inherent assumption of an SIR model, permanent immunity, which is not true in the long run due to decreased COVID vaccine protection and the appearance of new variants. Modifying our model to an SIRS model may better capture the temporary immunity brought by COVID vaccines. at being said, our model provides valid predictions based on early-stage vaccination trends.

Some additional limitations include the discontinuation of non-pharmaceutical interventions and changes in individual attitudes/behaviors towards the pandemic. Our model assumes a continuation of the non-pharmaceutical interventions in place in early March. Relaxation of these policies would likely increase the time needed to reach herd immunity. Another issue is moral hazard, that is, whether vaccinated individuals will change their behaviors and undertake more social interaction[35]. is change could result in higher risks of infection and a delay in reaching herd immunity.

Our study provides strong evidence that vaccination has signi cantly decreased COVID-19 cases and hospitalizations in the U.S. Following the vaccination trends between January and early March in 2021, our model predicts that herd immunity can be achieved earlier with faster vaccination pace and lower vaccination hesitancy. However, a few factors, such as moral hazard and variants of the SARS-CoV-2 virus, could lead to changes and cast doubt as to whether herd immunity can be achieved a er all.

Methods

Data collection and processing. A summary is provided of the data used in our analysis. Our supplementary notes give further details, including a summary statistics table for all variables.

Epidemiological and vaccination data. We collected our state-level epidemiological data (total COVID-19 cases, hospitalization, and tests) from the COVID Tracking Project[36], a commonly cited source[37–39]. e vaccination data across states were obtained from the U.S. Centers for Disease Control and Prevention's (CDC) COVID data tracker[40], where "people vaccinated" re ects the total number of people who have received at least one vaccine dose, and "people fully vaccinated" re ects the number who have received both doses prescribed by the vaccination protocol. We downloaded the CDC vaccination data from an open-source GitHub project by Our World in Data[41]. Both the BNT162b2 (P zer/BioNTech) vaccine and the mRNA-1273 (Moderna) vaccine require two doses[9]. In addition, the CDC shares data on COVID-19 vaccine distribution allocations by state for both the P zer[42] and Moderna[43] vaccines, as provided by the O ce of the Assistant Secretary for Public A airs under the U.S. Department of Health & Human Services.

Nonpharmaceutical interventions. In addition to epidemiological data, we obtained information on nonpharmaceutical intervention policies. We adopted the policy stringency index constructed by the Oxford COVID-19 Government Response Tracker[44], which systematically collects information on various policy responses implemented by various governments in response to the pandemic. We focused on the policy category of "containment and closure," which covers eight policies: school closings, workplace closings, cancelation of public events, restrictions on gathering sizes, cessation of public transportation, stay-at-home requirements, restrictions on internal movement, and restrictions on international travel. is stringency index is a weighted score across these eight containment and closure policies and is scaled between 0 and 100. A detailed explanation of these measures was given by Hale et al. (2021)[45]. We determined the stringency index for each state on a weekly basis by averaging the daily data.

Meteorological data. Another set of important independent variables included in this study regarded the local climate. We obtained station-level hourly weather data provided by the National Centers for Environmental Information[46]. ese station-level weather data were then matched with the station location and corresponding state provided by the Global Historical Climatology Network Daily[47]. We calculated the average values from these weather reports for each week across all stations within each state. Given the lack of humidity data, temperature and snow depth were used as our climate measures.

Election rallies and black lives matter (BLM) demonstrations. Several large-scale mass gatherings for political campaigns and protests also occurred during our study period. We constructed binary measures for election rallies[48]. For states with a rally during week t, this binary measure takes the value of 1 for week t and for the week a er ($t+1$). Our BLM data from Elephrame o ered detailed information (date, location, etc.) about each demonstration from news reports[49], which were extracted using a Web scraper. We then calculated the total number of demonstrations that occurred across all cities within each state for each week.

Sociodemographic data. We also collected the sociodemographic characteristics of each state's population using 2019 estimates from the U.S. Census Bureau[50,51]. Speci cally, we downloaded data on age, race, and income. We constructed each of our sociodemographic variables to be binary, above or below the national median. We derived the proportion of individuals 65 years of age and older in the population, the proportion of the white population, and the income for each state to calculate a national median. Finally, our data for the 2020 Electoral College results were obtained from the National Archives[52]. We classi ed the states into those won by Joe Biden and those by won by Donald Trump.

Econometric analysis. *Reduced-form analysis.* e following reduced-form empirical model was used to estimate the impact of vaccination on the pandemic:

Here, y_{it} is the dependent variable that measures the growth of either total cases or total hospitalizations in state i at period t. Our baseline measure is the growth rate, which is de ned as $\frac{C_{i,t}-C_{i,t-1}}{C_{i,t-1}}$ for total cases and $\frac{H_{i,t}-H_{i,t-1}}{H_{i,t-1}}$ for total hospitalizations, where $C_{i,t}$ and $H_{i,t}$ are the cumulative numbers of cases and hospitalizations. Alternative outcome measures were also used in the sensitivity analysis (Extended Data Fig. 4).

Our key independent variable, $Vaccination_{i,t-1}$, is the rate of vaccination of state i in period t-1, and a_1 is the coe cient of interest. We used two measures of vaccination rate: the number of vaccinated people (i.e., those who had received at least one dose of vaccine) per hundred and the number of fully vaccinated people (i.e., those who had received two doses of vaccine) per hundred. As the proportion of susceptible individuals in the total population decreases over time, the growth rate of infection may also decline. To deal with this intrinsic dynamic, $_{i,t-1}/L_i$ was included in the regression model to control for the stock of susceptible individuals $S_{i,t-1}$ in the total population We measured ccm8(es a(e g8.5(m-5.1(id w)8.5(e)-4.6(e)4.5(k).5( )]TJ /Spa12:150 Td ( )xt(þÿ

are highly correlated with the empirical data. For each individual state, our model estimates reached a median correlation of 99.04% (range, 86.37% to 99.95%) (Extended Data Fig. 7).

We assessed herd immunity based on our model estimates of the real-time basic reproduction number for each state, $R'_{i,t} = \frac{\beta_{i,t} S_{i,t}}{\gamma_i}$; that is, the number of cases directly caused by an infected individual throughout his or her infectious period. e model achieves herd immunity when $R'_{i,t}$ falls below 1 in 49 states (except for Maryland and Kentucky; see Supplementary Methods for details).

For each given vaccination pace, we ran the simulation forward and projected the future dynamic of the pandemic across the U.S., assuming that no changes are made in nonpharmaceutical interventions. We then computed the time required for every state to achieve herd immunity and calculated the share of the U.S. population vaccinated when herd immunity is achieved. In addition, we conducted a sensitivity analysis regarding herd immunity with variations in vaccine e ectiveness and with the addition of vaccine hesitancy. We incorporated vaccine hesitancy into our model by assuming that if x% of the population is hesitant, the cumulative vaccination coverage in each state will stop when (1 − x%) of the population is vaccinated.

Data and Code availability

e datasets and code used for the analyses are available at https://github.com/huntabaobao007/US-COVID-19-vaccination.

Published online: 28 January 2022

References

1. WHO. WHO Coronavirus Disease (COVID-19) Pandemic. https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed 10 April 2021).
2. Pinto Neto, O. *et al.* Mathematical model of COVID-19 intervention scenarios for Sao Paulo-Brazil. *Nat. Commun.* **12**, 1–13 (2021).
3. Hsiang, S. *et al.* e e ect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* **584**, 262–267 (2020).
4. Lancet, T. India under COVID-19 lockdown. *Lancet* **395**, 1315 (2020).
5. Kissler, S. M. *et al.* Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868 (2020).
6. van Riel, D. & de Wit, E. Next-generation vaccine platforms for COVID-19. *Nat. Mater.* **19**, 810–812 (2020).
7. Creech, B. C., Walker, S. C. & Samuels, R. J. SARS-CoV-2 vaccines. *JAMA* **325**, 1318–1320 (2021).
8. Polack, F. P. *et al.* Safety and e cacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).
9. Baden, L. R. *et al.* E cacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N. Engl. J. Med.* **384**, 403–416 (2021).
10. FDA. P zer-BioNTech COVID-19 Vaccine. https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/p zer-biontech-covid-19-vaccine (accessed 10 April 2021).
11. Carlsen, A., Huang, P., Levitt, Z., & Wood, D. How is the COVID-19 vaccination campaign going in your state? https://www.npr.org/sections/health-shots/2021/01/28/960901166/how-is-the-covid-19-vaccination-campaign-going-in-your-state (accessed 10 April 2021).
12. Dagan, N. *et al.* BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. *N. Engl. J. Med.* **384**, 1412–1423 (2021).
13. Britton, A. *et al.* E ectiveness of the P zer-BioNTech COVID-19 vaccine among residents of two skilled nursing facilities experiencing COVID-19 outbreaks—Connecticut, December 2020–February 2021. *Morb. Mortal. Wkly. Rep.* **70**, 396–401 (2021).
14. Angrist, J. D. & Pischke, J.-S. *Mostly harmless econometrics: an empiricist's companion* (Princeton Univ, 2008).
15. Romer, C. D. & Romer, D. H. e macroeconomic e ects of tax changes: Estimates based on a new measure of scal shocks. *Am. Econ. Rev.* **100**, 763–801 (2010).
16. Burke, M., Hsiang, S. M. & Miguel, E. Global non-linear e ect of temperature on economic production. *Nature* **527**, 235–239 (2015).
17. Ricon-Becker, I., Tarrasch, R., Blinder, P. & Ben-Eliyahu, S. A seven-day cycle in COVID-19 infection and mortality rates: Are inter-generational social interactions on the weekends killing susceptible people?. Preprint at *medRxiv.* https://doi.org/10.1101/2020.05.03.20089508 (2020).
18. Pavlíček, T., Rehak, P. & Král, P. Oscillatory dynamics in infectivity and death rates of COVID-19. *mSystems.* **5**, e00700-e720 (2020).
19. Cohn, B. A. *et al.* SARS-CoV-2 vaccine protection and deaths among US veterans during 2021. *Science*. https://doi.org/10.1126/science.abm0620 (2021).
20. Dave, D. *et al.* Did President Trump's Tulsa Rally reignite Covid-19? Indoor events and o setting community e ects. *NBER Work. Pap. Ser*. w27522 (2020).
21. Bernheim, B. D. *et al.* e e ects of large group meetings on the spread of COVID-19: e case of Trump rallies. *Stanford Institute for Economic Policy Research (SIEPR)* (2020).
22. Neyman, G. & Dalsey, W. Black Lives Matter protests and COVID-19 cases: Relationship in two databases. *J. Public Health (Oxf).* fdaa212 (2020).
23. Mecenas, P. *et al.* E ects of temperature and humidity on the spread of COVID-19: A systematic review. *PLoS ONE* **15**, e0238339 (2020).
24. Kaashoek, J. & Santillana, M. COVID-19 positive cases, evidence on the time evolution of the epidemic or an indicator of local testing capabilities? A case study in the United States. Preprint at *SSRN*. https://doi.org/10.2139/ssrn.3574849 (2020).
25. Davies, N. G. *et al.* Age-dependent e ects in the transmission and control of COVID-19 epidemics. *Nat. Med.* **26**, 1205–1211 (2020).
26. Mueller, A. L., McNamara, M. S. & Sinclair, D. A. Why does COVID-19 disproportionately a ect older people?. *Aging* **12**, 9959–9981 (2020).
27. Kempe, A. *et al.* Parental hesitancy about routine childhood and in uenza vaccinations: A national survey. *Pediatrics* **146**, e20193852 (2020).
28. Kestenbaum, L. A. & Feemster, K. A. Identifying and addressing vaccine hesitancy. *Pediatr. Ann.* **44**, e71–e75 (2015).
29. Mahase, E. Covid-19: Where are we on vaccines and variants?. *BMJ* **372**, n597 (2021).
30. Bernal, J. L. *et al.* E ectiveness of Covid-19 vaccines against the B. 1.617. 2 (Delta) variant. *N Engl. J. Med.* **385**, 585–594 (2021).
31. Nasreen S., *et al.* E ectiveness of COVID-19 vaccines against variants of concern, Canada. Preprint at *medRxiv*. https://doi.org/10.1101/2021.06.28.21259420 (2021).
32. Chemaitelly, H. *et al.* Waning of BNT162b2 vaccine protection against SARS-CoV-2 infection in Qatar. *N. Engl. J. Med.* **385**, e83 (2021).

33. e Scienti c Pandemic In uenza Group on Modelling, Operational sub-group (SPI-M-O). Consensus statement on COVID-19, 2nd June 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/ le/993321/S1267_SPI-M-O_Consensus_Statement.pdf (accessed 10 December 2021)
34. Ong, S.W.X., *et al.* Clinical and Virological Features of SARS-CoV-2 Variants of Concern: A Retrospective Cohort Study Comparing B. 1.1. 7 (Alpha), B. 1.315 (Beta), and B. 1.617. 2 (Delta). Preprint at *SSRN*. https://doi.org/10.2139/ssrn.3861566 (2021).
35. Moghtaderi, A. & Dor, A. Immunization and moral hazard: e HPV vaccine and uptake of cancer screening. *Med. Care Res. Rev.* **78**, 125–137 (2021).
36. e COVID Tracking Project. e data. https://covidtracking.com/data (accessed 14 March 2021).
37. Erfani, P. *et al.* COVID-19 testing and cases in immigration detention centers, April-August 2020. *JAMA* **325**, 182–184 (2021).
38. Lyu, W. & Wehby, G. L. Comparison of estimated rates of coronavirus disease 2019 (COVID-19) in border counties in Iowa without a stay-at-home order and border counties in Illinois with a stay-at-home order. *JAMA Netw. Open.* **3**, e2011102 (2020).
39. Murray, C. J. *et al.* Digital public health and COVID-19. *Lancet Public Health.* **5**, e469–e470 (2020).
40. CDC. COVID-19 vaccinations in the United States. https://covid.cdc.gov/covid-data-tracker/#vaccinations (accessed 14 March 2021).
41. Gavrilov, D. *et al.* Our World in Data: COVID-19-data. https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations (accessed 14 March 2021).
42. CDC. COVID-19 vaccine distribution allocations by jurisdiction – P zer. http://data.cdc.gov/Vaccinations/COVID-19-Vaccine-Distribution-Allocations-by-Juris/saz5–9hgg (accessed 14 March 2021).
43. CDC. COVID-19 vaccine distribution allocations by jurisdiction – Moderna. http://data.cdc/gov/Vaccinations/COVID-19-Vaccine-Distribution-Allocations-by-Juris/b7pe-5nws (accessed 14 March 2021).
44. University of Oxford. COVID-19 government response tracker. http://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker (accessed 14 March 2021).
45. Hale, L. *et al.* Variation in US states responses to COVID-19 2.0. Blavatnik School of Government Working Paper. http://www.bsg.ox.ac.uk/covidtracker (accessed 14 March 2021).
46. NOAA. Climate data online: Dataset discovery. https://www.ncdc.noaa.gov/cdo-web/datasets (accessed 14 March 2021).
47. NOAA. Global Historical Climatology Network Daily (GHCND). https://www.ncdc.noaa.gov/ghcnd-data-access (accessed 14 March 2021).
48. Wikipedia. List of post-2016 election Donald Trump rallies. https://en.wikipedia.org/wiki/List_of_post-2016_election_Donald_Trump_rallies (accessed 14 March 2021).
49. Elephrame. List of 5,788 Black Lives Matter demonstrations. https://elephrame.com/ (accessed 14 March 2021).
50. U.S. Census Bureau. State population by characteristics: 2010–2019. https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html (accessed 14 March 2021).
51. U.S. Census ACS survey. Median Household Income by State. https://worldpopulationreview.com/state-rankings/median-household-income-by-state (accessed 11 March 2021).
52. National Archives. 2020 Electoral College results. https://www.archives.gov/electoral-college/2020 (accessed 14 March 2021).
53. Cameron, A. C., Gelbach, J. B. & Miller, D. L. Robust inference with multiway clustering. *J. Bus. Econ. Stat.* **29**, 238–249 (2011).

Acknowledgements

We thank K.E. Warner and S. Mennemeyer for their feedback.

Author contributions

All authors designed the analyses, interpreted the results, and designed the gures, and are listed alphabetically. X.C., H.H., R.S., and J.Z. contribute equally to the paper. H.H. and R.S. collected the data. J.Z. conducted the reduced-form empirical analysis. X.C. conducted the analysis with the SIR model. H.H., J.J., and R.S. wrote the paper.

Funding

e Funding was provided by the National Natural Science Foundation of China (72003026), the Fundamental Research Funds for the Central Universities in UIBE (19QD01), startup grant of City University of Hong Kong (7200689), and the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 11501121).

Competing interests

e authors declare no competing interests.

Additional information

Supplementary Information e online version contains supplementary material available at https://doi.org/10.1038/s41598-022-05498-z.

Correspondence and requests for materials should be addressed to H.H. or R.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional a liations.