Gupta S, Cantor J, Simon KI, Bento AI, Wing C, Whaley CM. Vaccinations against COVID-19 may have averted up to 140,000 deaths in the United States. Health Aff (Millwood). 2021;40(9).

APPENDIX

Description of Multivariable Regression Approach

The primary aim of this analysis was to assess the association between within-state variation in state-level COVID-19 vaccine administration and COVID-19 cases, and deaths. To do so, we used data at the state-week level on COVID-19 vaccine administration, COVID-19 cases, and COVID-19 deaths.

Models of COVID-19 Cases and Deaths Throughout, we use $s = 1 \ldots S$ to index states and $t = 1 \ldots T$ to index calendar weeks. Y_{st} is the cumulative number of cases or deaths per 100 adults in state s in week t. We model cumulative cases and death using an exponential conditional mean function and allow for a full set of state fixed effects and week fixed effects. The exponential conditional mean function is an appealing approximation here because it accommodates the restriction that the conditional mean value of cumulative cases and deaths is non-negative in all weeks and because the exponential model is a plausible way to describe changes in the severity of the epidemic over time. Specifically, we estimate the following specification:

$$Y_{st} = E[Y_{st}|V_{st}, \gamma_s, \tau_t] + \epsilon_{st}$$

$$Y_{st} = \exp(\textstyle\sum_{k=0}^{K=4} \delta_k V_{st} + \gamma_s + \tau_t + \varepsilon_{st}) + \epsilon_{st} \qquad (1)$$

In this regression model, Y_{st} represents the dependent variables of interest. In models of the number of COVID-19 cases, Y_{st} is the cumulative number of COVID-19 cases per 100 adults (population aged 16 +) in state s as of week t. In models of the number of COVID-19 deaths, Y_{st}

is the cumulative number of COVID-19 deaths per 100 adults in state s by the end of week t. γ_s represents a fixed effect controls for state and τ_t represents a fixed effect control for week. On the right-hand side, V_{st} represents the cumulative number of vaccine doses administered per 100 adults in state s by the end of week t; it is non-decreasing from week to week in each state as the vaccine rollout progresses. Some states experience more rapid growth in cumulative vaccination than others, and within a state, coverage expands faster in some weeks than in others. Although the contemporaneous vaccine rate is also included in the model, we do not realistically expect that vaccines would have an immediate association with the number of COVID-19 deaths and cases, and our joint F tests always focus on the four lagged values for the number of vaccines per 100 adults in the regressions. We estimate the parameters of the model using the Poisson quasi-likelihood estimator, which allows for an exponential conditional mean function. However, we do not impose the assumption that cases per 100 adults and deaths per 100 adults are distributed according to a Poisson distribution, which would require the restrictive assumption that the conditional mean is equal to the conditional variance of the outcome. Instead, we work with the exponential form for the conditional mean function but relax the remaining distributional assumptions by computing a cluster robust variance matrix that allows for heteroskedasticity, over dispersion, and dependent errors (serial correlation) at the state level. As discussed in the main text, we also estimate models that include leads of the number of vaccines per 100 adults as sensitivity analyses and to understand potential anticipatory effects. We use the delta method to compute standard errors and confidence intervals on transformations of the regression coefficients to obtain estimates of the number of averted COVID-19 deaths and cases by week.

Estimating Vaccine Averted Deaths We use the quasi-Poisson regression model coefficients to estimate the implied number of deaths averted given the estimated associations between the vaccine rollout and cases and deaths. To understand our approach, observe that the model-based estimate of the realized/actual cumulative number of deaths per 100 adults in a state s in a specific week t is given by: $\widehat{Y_{st}^{realized}} = \exp\left(\sum_{k=0}^{K=4} \widehat{\delta_k} V_{s,t-k} + \gamma_s + \tau_t\right)$. The implied cumulative number of deaths is $\widehat{D}_{s^*t^*}^{realized} = \frac{\left(\widehat{Y_{st}^{realized}} \times Pop_s\right)}{100}$, where Pop_s is the number of adults in state s. These realized fitted values from the regression include the associations with current and four weeks of lagged vaccinations.

To compute the implied counterfactual cumulative number of deaths per 100 adults in state s in week t that would have occurred in the absence of the vaccination rollout, we simply set the current and lagged vaccination variables to equal zero in each week. The counterfactual estimate of cumulative COVID-19 deaths per 100 adults is: $\hat{Y}_{st}^{cf} = \exp\left(\gamma_s + \tau_t\right)$ and the counterfactual number of cumulative deaths is $\widehat{D}_{s^*t^*}^{cf} = \frac{\left(\hat{Y}_{st}^{cf} \times Pop_s\right)}{100}$.

Accordingly, the model-based estimate of vaccine averted deaths in state s as of week t are given by $\widehat{AD_{st}} = \widehat{D}_{st}^{cf} - \widehat{D}_{st}^{realized}$. We compute confidence intervals on $\widehat{AD_{st}}, \widehat{D}_{st}^{cf}, and\ \widehat{D}_{st}^{realized}$ using the delta method and the cluster robust variance matrix from the quasi-Poisson regression models.

We form national point estimates of the cumulative number of realized, counterfactual, and averted deaths by aggregating across states. For example, the nationwide number of averted deaths implied by the model is given by $AD_t = \sum_{s=1}^{51} AD_{st}$. We compute confidence intervals on the nationwide totals by summing the lower and upper endpoints of the confidence intervals. For example, if $[L(AD_{st}), U(AD_{st})]$ is the estimated confidence interval on averted deaths in state s in year t then we compute interval on the nationwide number of averted deaths using $L(AD_t) = \sum_{s=1}^{51} L(AD_{st})$ and $U(AD_t) = \sum_{s=1}^{51} U(AD_{st})$.

Additional Results

Sensitivity Analysis for Deaths and Cases

We estimated several alternative models of cumulative COVID-19 deaths and cases to test for the sensitivity of our estimates. First, instead of including state-level fixed effect controls, we estimated a specification that controls for observable state differences. The fixed effects model is our preferred specification because state fixed effects are more effective at controlling for unspecified baseline state characteristics, in which states differ. We obtained state characteristics related to COVID-19 outcomes using data from previous studies.[1] Specifically, we include controls for state-level measures of paid sick leave, Medicaid expansion status by January 1, 2020, minimum wage, population density per square mile, 2018 population, area, number of homeless, 2018 unemployment rate, percentage living under federal poverty limit, percentage at risk for serious illness, 2018 all cause deaths, mental health professionals per 100,000 population in 2019, whether there were any casino(s) in the state (Land and Non-Land based), and whether the state has at least one American Indian/Alaskan Native reservation.

There are of course many more state characteristics that may matter and are more fully accounted for by the state fixed effects in our main model. In results presented in Supplemental Exhibit 5, we find that controlling for state characteristics the result for cases does not change in an appreciable manner—the joint test of significance of the vaccination lags remains statistically significant at the $p<0.10$ level (p value now of 0.08) but with a more limited set of controls for state characteristics the result for COVID-19 deaths is not statistically significant at any conventional levels. This indicates that the estimated association between vaccination rates and COVID-19 deaths rates may be sensitive to unobserved state level factors, reiterating the importance of controlling for other unobserved confounders.

A second potential concern is that the initial vaccine rollout may be higher in states with a higher share of previously-infected individuals. If that occurs, declines in COVID-19 cases and deaths could reflect variations in state-level presence of antibodies even prior to the rollout, rather than vaccine administration. We tested regression models that separately examine responses to vaccine allocations among states above or below the national median cumulative cases per 100 adult population as of December 13, 2020 (cumulative 6.69 cases per 100 [IQR: 5.01- 8.14]), under the assumption that states with above-median case rates have higher baseline exposure to COVID-19. In these specifications reported in Exhibit 6, we find that there does not appear to be an appreciably different pattern of results when states are separated this way; however, when sample sizes are halved by stratification, we find that both for the states that are above the median as well as for the states below the median, the lagged vaccination rates are no longer jointly statistically significant at $p<0.05$ level.

Third, we extended the model in Equation 1 to include controls for three physical distance measures including state policies: stay-at-home orders, mask mandates and overnight business closure mandates - with state-level variations during our study period. In these specifications reported in columns 1 and 2 of Exhibit 7, the results for reported COVID-19 cases and deaths do not change in an appreciable manner—the joint test of significance of the vaccination lags remains statistically significant at the $p<0.05$ level (p value now of 0.002 for cases and 0.04 for deaths).

Fourth, we estimated models that include leads of four weeks of cumulative number of vaccines administered per 100 adults to capture any anticipatory changes. In these models reported in columns 3 and 4 of Exhibit 7, a joint test of the significance of leads shows them to be not different from zero (p-values of 0.55 and 0.39 respectively for cases and deaths). To further test the sensitivity of results for our preferred specification we re-examine our main model for COVID-19 cases and deaths using three weeks (Exhibit 8), five weeks (Exhibit 9), and eight weeks lags (instead of four weeks lags) (Exhibit 10). We find that the joint test of the coefficients of three-week lags are statistically significantly different from zero (as they were in the main model); the p-value is less than 0.05 in both cases. However, when we attempted a model with five-week lags (Exhibit 9), we found that the lags were no longer jointly statistically significant for the number of COVID-19 cases (new p-value of 0.118) but remained statistically significant at conventional levels for the number of COVID-19 deaths (new p-value of 0.00009). Likewise, for the model with eight-week lags (Exhibit 10), the lags were not jointly statistically significant for the number of COVID-19 cases (new p-value of 0.40) but remained statistically significant at conventional levels for the number of COVID-19 deaths (p-value of 0.0171). This

suggests that our main specification's use of four-week lags is generally suitable and shows that the results for the number of COVID-19 cases appears more sensitive to this decision than for the number of COVID-19 deaths.

Reference

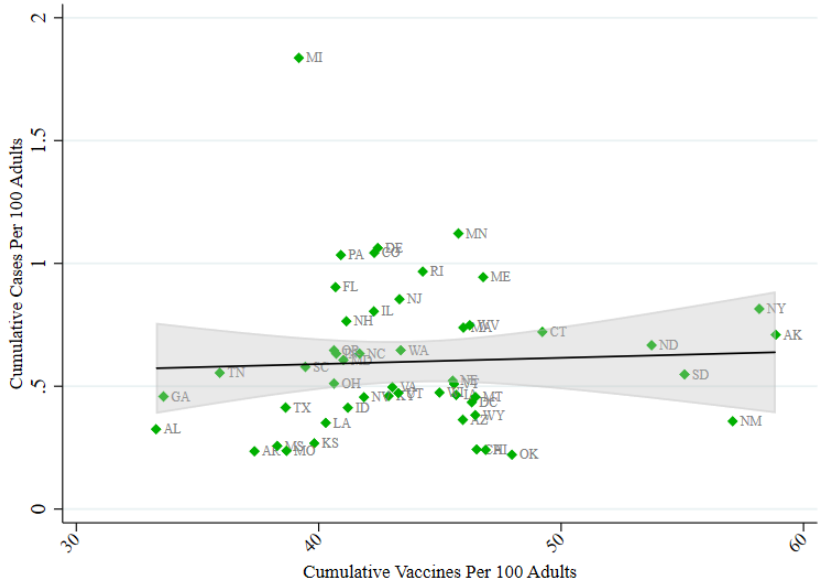
1. Raifman J, Nocka K, Jones D, Bor J, Lipson S, Chan P. COVID-19 US state policy database (CUSP) [Internet]. Google Docs. [cited 2021 Mar 28]; Available from: https://docs.google.com/spreadsheets/d/1zu9qEWI8PsOI_i8nI_S29HDGHlIp2lfVMsGxpQ5tvAQ/edit?usp=sharing&usp=embed_facebook

APPENDIX TABLES AND FIGURES

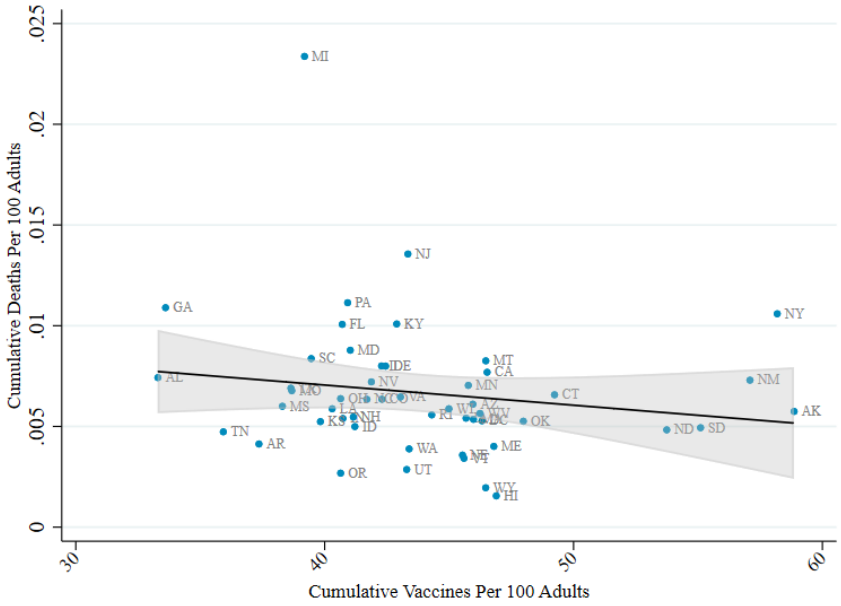
Additional Tables and Figures Supporting Main Results

Exhibit 1: Association Between COVID-19 Vaccinations and COVID-19 Deaths/Cases

(a) COVID-19 Cases



(b) COVID-19 Deaths



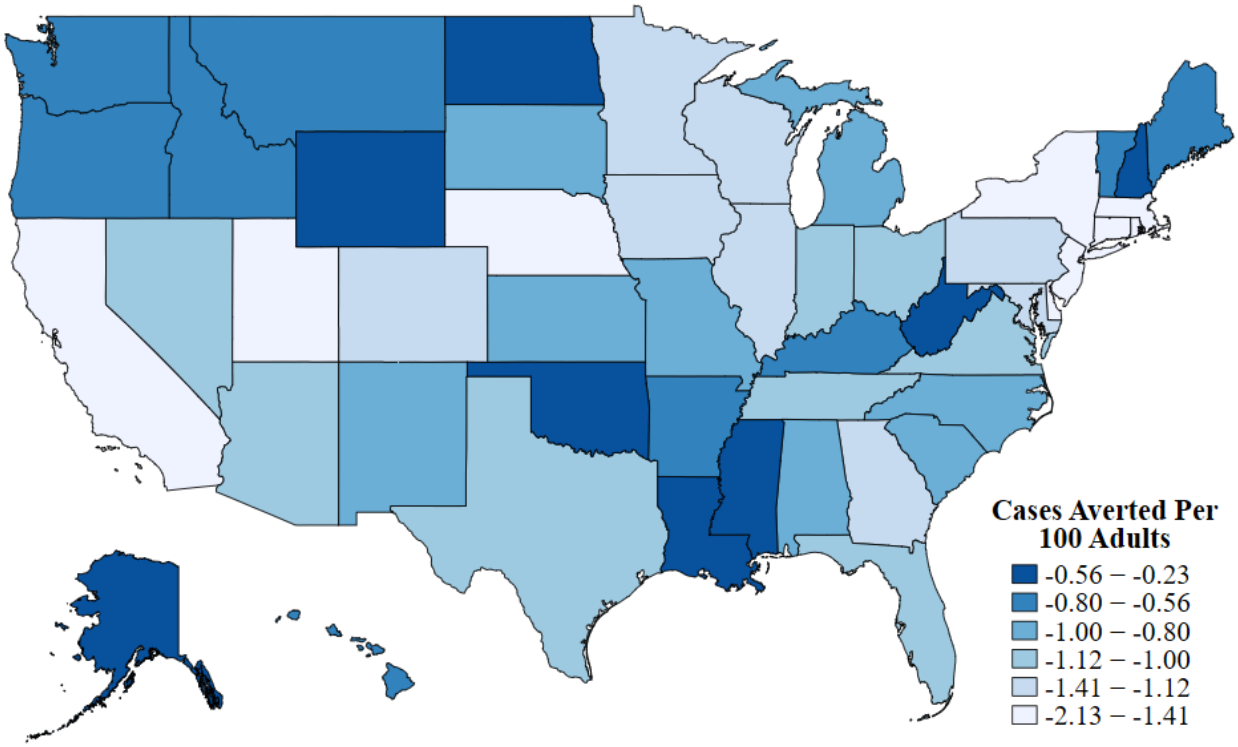
Vaccination data are obtained from The Bloomberg Covid-19 Vaccine Tracker. COVID-19 deaths and cases data are obtained from the New York Times "Coronavirus (Covid-19) Data in the United States" database. The x-axis records the cumulative doses of vaccines administered per 100 population ages 16+ as of March 15th, 2021. The y-axis records the state-level total COVID-19 cases (Panel A) and deaths (Panel B) occurred between April 9th, 2021, and May 9th,2021.

Exhibit 2. Main Regression Results: Association between state COVID-19 vaccinations and COVID-19 cases and deaths.

	(1) Cumulative COVID-19 Cases per 100 Adults	(2) Cumulative COVID-19 Deaths per 100 Adults
Vaccine doses per 100 adults	0.00601***	-0.00273
	(0.002)	(0.003)
Vaccine doses per 100 adults 1 week ago	-0.000584	0.00319***
	(0.001)	(0.001)
Vaccine doses per 100 adults 2 weeks ago	-0.00116	-0.000231
	(0.001)	(0.001)
Vaccine doses per 100 adults 3 weeks ago	0.000341	0.000356
	(0.002)	(0.001)
Vaccine doses per 100 adults 4 weeks ago	-0.00515**	-0.00344
	(0.002)	(0.003)
Observations	969	969
Baseline DV mean	6.100	0.100
Joint F-stat Lags	16.14	10.96
Joint F-test p-value	0.0028	0.0270

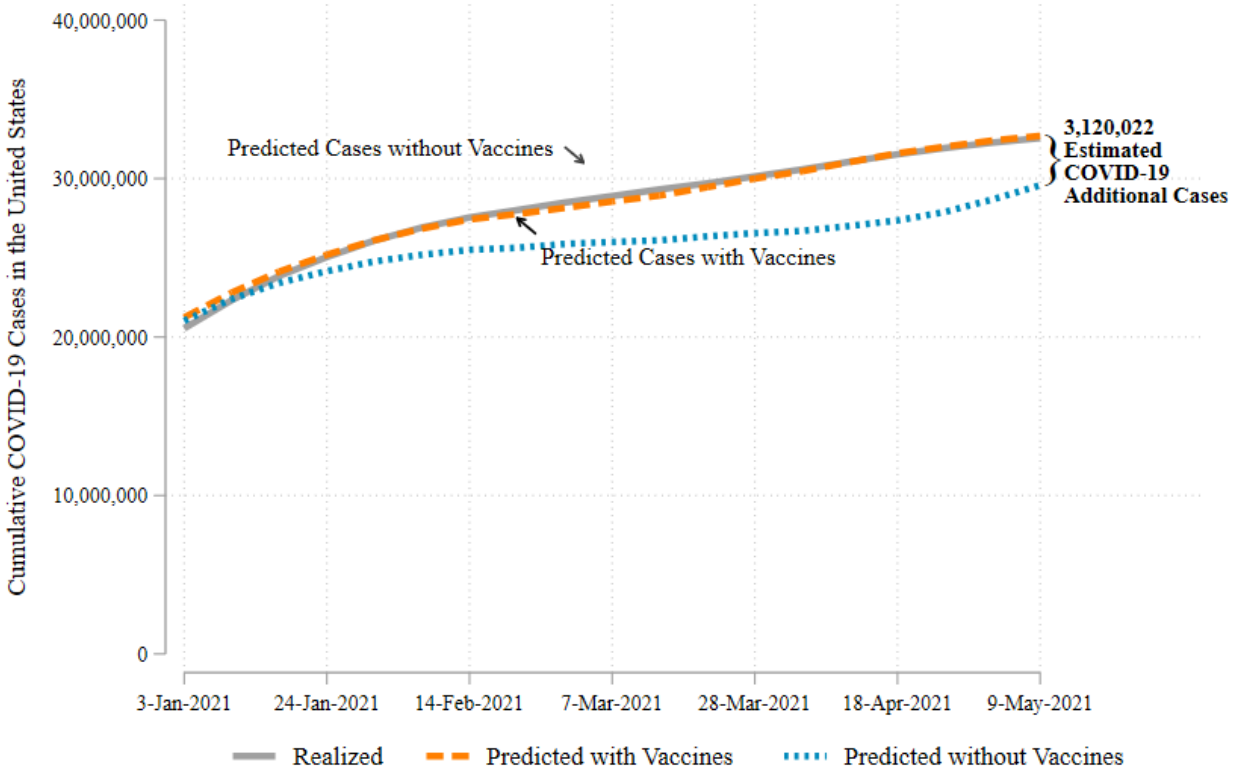
Standard errors clustered at state-level in parentheses * p<0.10, ** p<0.05, *** p<0.01. The quasi-Poisson regression controls for calendar week and state fixed effects.

Exhibit 3. Model-based Estimated Cumulative Cases Averted (Per 100 Population aged 16+) by state.



We estimate averted cases in each state using the model defined in the Methods section. Averted cases are the difference between predicted values of cases at prevailing levels of vaccinations and predicted values of cases if no vaccinations had occurred during the study period. The models use state fixed effects, week fixed effects and a four weekly lags of the vaccination doses per 100 population ages 16+. Vaccination data are obtained from The Bloomberg Covid-19 Vaccine Tracker. COVID-19 cases data are obtained from the New York Times database.

Exhibit 4. Estimated COVID-19 cases averted by vaccination by week.



We estimate averted cases using the model in the methods section. Averted cases are the difference between predicted values of cases at prevailing levels of vaccinations and predicted values of cases if no vaccinations had occurred during the study period. The models use state fixed effects, week fixed effects and a four weekly lags of the vaccination doses per 100 population ages 16+. The grey solid line shows actual cumulative number of cases by week in the U.S. The orange dashed line shows estimated values of cumulative cases from our regression model at prevailing levels of vaccination, indicating a close fit to the actual data. The blue dotted line shows predicted values of cases from the model, had no vaccinations occurred. Vaccination data are obtained from The Bloomberg Covid-19 Vaccine Tracker. COVID-19 cases data are obtained from the New York Times database.

Sensitivity Analyses of COVID-19 Deaths and Cases Estimations

Exhibit 5. Association between state COVID-19 vaccinations and COVID-19 cases and deaths, controlling for observable state characteristics

	(1) Cumulative COVID-19 Cases per 100 Adults	(2) Cumulative COVID-19 Deaths per 100 Adults
Vaccine doses per 100 adults	-0.00401	0.00170
	(0.005)	(0.006)
Vaccine doses per 100 adults 1 week ago	0.00721**	0.00227
	(0.003)	(0.003)
Vaccine doses per 100 adults 2 weeks ago	-0.00153	-0.000518
	(0.002)	(0.002)
Vaccine doses per 100 adults 3 weeks ago	0.00488	0.00234
	(0.005)	(0.004)
Vaccine doses per 100 adults 4 weeks ago	-0.00412	0.00200
	(0.005)	(0.005)
Observations	969	969
Baseline DV mean	6.100	0.100
Joint F-stat Lags	8.430	1.833
Joint F-test p-value	0.0770	0.766

Standard errors clustered at state level in parentheses * p<0.10 ** p<0.05 *** p<0.01. The quasi-Poisson regression controls for calendar week fixed effects and state level measures of paid sick leave, Medicaid expansion status by January 1, 2020, minimum wage, population density per square mile, 2018 population, area, number of homeless, 2018 unemployment rate, percentage living under federal poverty limit, percentage at risk for serious illness, 2018 all cause deaths, Mental health professionals per 100,000 population in 2019, whether there were any casino(s) in state (Land and Non-Land based) and whether the state has at least one American Indian/Alaskan Native reservation. State characteristics data are taken from Raifman et al (2020).

Exhibit 6. Association between state COVID-19 vaccinations and COVID-19 cases and deaths, stratified by COVID-19 case rates prior to vaccine rollout

	(1) Cumulative Cases per 100 Adults Low Cases Dec 2020	(2) Cumulative Cases per 100 Adults High Cases Dec 2020	(3) Cumulative Deaths per 100 Adults Low Cases Dec 2020	(4) Cumulative Deaths per 100 Adults High Cases Dec 2020
Vaccine doses per 100 adults	0.00300*	0.000912	-0.00656**	-0.00445
	(0.002)	(0.002)	(0.003)	(0.004)
Vaccine doses per 100 adults 1 week ago	-0.00101	0.00170**	0.00335**	0.00435**
	(0.001)	(0.001)	(0.001)	(0.002)
Vaccine doses per 100 adults 2 weeks ago	0.000245	-0.00163**	0.000698	-0.000719
	(0.001)	(0.001)	(0.001)	(0.001)
Vaccine doses per 100 adults 3 weeks ago	0.00147	-0.000801	-0.000139	-0.000456
	(0.002)	(0.001)	(0.001)	(0.001)
Vaccine doses per 100 adults 4 weeks ago	-0.00258	-0.000847	0.000896	-0.00161
	(0.002)	(0.001)	(0.002)	(0.003)
Observations	494	475	494	475
Baseline DV mean	4.240	8.030	0.100	0.100
Joint F-stat Lags	2.599	7.251	5.708	6.883
Joint F-test p-value	0.627	0.123	0.222	0.142

Standard errors in parentheses * p<0.10 ** p<0.05 *** p<0.01. The quasi-Poisson regression controls for calendar week and state fixed effects. Column (1) and (3) estimates are based on the subset of states with *below* median cumulative cases per 100 adults as of December 13, 2020. Column (2) and (4) estimates are based on the subset of states with *above* median cumulative cases per 100 adults as of December 13, 2020.

Exhibit 7. Association between state COVID-19 vaccinations and COVID-19 cases and deaths, <u>controlling for state-level COVID-19 mitigation policies and vaccination leads</u>

	(1) Cumulative COVID-19 Cases per 100 Adults	(2) Cumulative COVID-19 Deaths per 100 Adults	(3) Cumulative COVID-19 Cases per 100 Adults	(4) Cumulative COVID-19 Deaths per 100 Adults
Vaccine doses per 100 adults	0.00532***	-0.00404	-0.00132	0.000431
	(0.001)	(0.002)	(0.001)	(0.001)
Vaccine doses per 100 adults 1 week ago	-0.000401	0.00324***	-0.000831	0.00128
	(0.001)	(0.001)	(0.001)	(0.001)
Vaccine doses per 100 adults 2 weeks ago	-0.000776	0.000347	-0.00209	-0.0000400
	(0.001)	(0.001)	(0.001)	(0.001)
Vaccine doses per 100 adults 3 weeks ago	0.00000965	0.000314	-0.00102	-0.00180*
	(0.002)	(0.001)	(0.001)	(0.001)
Vaccine doses per 100 adults 4 weeks ago	-0.00493**	-0.00313	0.000505	-0.00364
	(0.002)	(0.003)	(0.002)	(0.003)
Mask Mandate in Effect	0.0672*	0.0278	0.0680**	0.0313
	(0.036)	(0.053)	(0.033)	(0.048)
Stay-At-Home Orders in Effect	0.0644***	-0.243***	0.0262	-0.231***
	(0.018)	(0.034)	(0.016)	(0.034)
Overnight Business Closure Mandate	-0.0504***	-0.0482	-0.0362**	-0.0489
	(0.018)	(0.035)	(0.015)	(0.035)
Vaccine doses per 100 adults 1 week later			0.000958	-0.0000584
			(0.001)	(0.001)
Vaccine doses per 100 adults 2 weeks later			-0.000811	-0.00177*
			(0.001)	(0.001)
Vaccine doses per 100 adults 3 weeks later			-0.00123	0.00187*
			(0.001)	(0.001)
Vaccine doses per 100 adults 4 weeks later			0.00449***	-0.00275
			(0.002)	(0.002)

Observations	969	969	765	765
Baseline DV mean	6.100	0.100	6.100	0.100
Joint F-stat Lags	16.36	10.02	3.050	4.130
Joint F-test p-value lags	0.00257	0.0402	0.550	0.390
Joint F-stat Leads			16.19	7.834
Joint F-test p-value leads			0.00278	0.0979

Standard errors clustered at state-level in parentheses * p<0.10 ** p<0.05 *** p<0.01. The quasi-Poisson regression controls for calendar week and state fixed effects. Estimates in all columns now include controls for three social distance related state policies – stay-at-home orders, mask mandates and overnight business closure mandates (data taken from Raifman et al (2020).[1] Columns (3)-(4) additionally controls for 4 week leads of vaccination doses per 100 adults (age 16+).

Exhibit 8. Association between state COVID-19 vaccinations and COVID-19 cases and deaths; 3-week lag specification

	(1) Cumulative COVID-19 Cases per 100 Adults	(2) Cumulative COVID-19 Deaths per 100 Adults
Vaccine doses per 100 adults	0.00756***	-0.00209
	(0.002)	(0.003)
Vaccine doses per 100 adults 1 week ago	-0.00205**	0.00220*
	(0.001)	(0.001)
Vaccine doses per 100 adults 2 weeks ago	-0.0000503	0.000475
	(0.002)	(0.001)
Vaccine doses per 100 adults 3 weeks ago	-0.00563***	-0.00328
	(0.002)	(0.003)
Observations	1020	1020
Baseline DV mean	6.100	0.100
Joint F-stat Lags	13.36	12.24
Joint F-test p-value	0.00392	0.00660

Standard errors clustered at state-level in parentheses * p<0.10 ** p<0.05 *** p<0.01. The quasi-Poisson regression controls for calendar week and state fixed effects.

Exhibit 9. Association between state COVID-19 vaccinations and COVID-19 cases and deaths; 5-week lag specification

	(1) Cumulative COVID-19 Cases per 100 Adults	(2) Cumulative COVID-19 Deaths per 100 Adults
Vaccine doses per 100 adults	-0.00223	0.00549***
	(0.002)	(0.002)
Vaccine doses per 100 adults 1 week ago	0.00196*	-0.00110
	(0.001)	(0.001)
Vaccine doses per 100 adults 2 weeks ago	0.000925	0.0000152
	(0.001)	(0.001)
Vaccine doses per 100 adults 3 weeks ago	-0.000156	-0.000327
	(0.001)	(0.001)
Vaccine doses per 100 adults 4 weeks ago	-0.00139*	-0.00239**
	(0.001)	(0.001)
Vaccine doses per 100 adults 5 weeks ago	-0.00183	-0.00216
	(0.002)	(0.003)
Observations	918	918
Baseline DV mean	0.100	6.100
Joint F-stat Lags	8.795	20.78
Joint F-test p-value	0.118	0.000890

Standard errors clustered at state-level in parentheses * p<0.10 ** p<0.05 *** p<0.01. The quasi-Poisson regression controls for calendar week and state fixed effects.

Exhibit 10. Association between state COVID-19 vaccinations and COVID-19 cases and deaths; 8-week lag specification

	(1) Cumulative COVID-19 Cases per 100 Adults	(2) Cumulative COVID-19 Deaths per 100 Adults
Vaccine doses per 100 adults	-0.00122	0.00521***
	(0.002)	(0.001)
Vaccine doses per 100 adults 1 week ago	0.000805	-0.00107
	(0.001)	(0.001)
Vaccine doses per 100 adults 2 weeks ago	0.000475	-0.000460
	(0.001)	(0.001)
Vaccine doses per 100 adults 3 weeks ago	0.000370	0.000269
	(0.001)	(0.001)
Vaccine doses per 100 adults 4 weeks ago	-0.00131*	-0.00230**
	(0.001)	(0.001)
Vaccine doses per 100 adults 5 weeks ago	0.000560	-0.00108
	(0.001)	(0.001)
Vaccine doses per 100 adults 6 weeks ago	-0.0000528	-0.00151
	(0.001)	(0.001)
Vaccine doses per 100 adults 7 weeks ago	-0.000619	-0.0000554
	(0.001)	(0.001)
Vaccine doses per 100 adults 8 weeks ago	-0.000566	0.00196
	(0.002)	(0.001)
Observations	765	765
Baseline DV mean	0.100	6.100
Joint F-stat Lags	8.295	18.61
Joint F-test p-value	0.405	0.0171

Standard errors clustered at state-level in parentheses * p<0.10 ** p<0.05 *** p<0.01. The quasi-Poisson regression controls for calendar week and state fixed effects.