How Supply, Demand Factors Determine State-level COVID-19 Vaccination

Chen Wang, Xuan Li

School of Social Policy & Practice, University of Pennsylvania

MSSP 8970: Applied Linear Modeling

Instructor: Dr. Momin Malik

Policy Research Report 4

December 22, 2022

Executive Summary

To analyze the factors contributing to the state-level administrated vaccination rate in the United States, this study applied cross-sectional regressions, including ordinary least squares regression and iteratively re-weighted least squares to identify the variables associated with monthly administrated vaccination rate. Based on the regression results, it is concluded that supply factors including health expenditure have only positive correlations initially. However, demand factors and state characteristics including democratic margin, vaccination hesitancy, costs obstacles to care, elderly ratio, and mask mandate policy are stable and strong over time. These patterns persist after using monthly vaccination rates as the new dependent variable for robustness check.

Introduction

As COVID-19 spreads across the globe, vaccination has been proven to be one of the most critical measures to defeat the pandemic and achieve immunization. Vaccination coverage differences among countries caused disparate immunization conditions globally, while the difference of states in one country is significant as well. The cumulative vaccination per capita of states in the United States may undermine the broad immunization condition of the country and aggravate inequality in society. This study analyzes the potential factors that associated with the state-level vaccine administration rates in the US to provide policy recommendations and better protect people from the pandemic.

Burgeoning literature discusses the contributing factors of vaccination administration rates. These factors could be summarized into three categories. First, state-level vaccination coverage is positively correlated with healthcare system capacities, such as the number of vaccination sites (Davila-Payan et al., 2014), and healthcare infrastructure (Abubakar et al., 2022). Second, research report a negative association between vaccination rates and economic vulnerability (Brien et al., 2012) and the elderly group who encounter mobility difficulties (Doherty et al., 2018). Finally, evidence suggest that state characteristics such as partisan divergence seems to be linked with vaccination rates (Viswanath et al., 2021).

Methods

Data

We employ the state-level datasets with state-specific characteristics and accumulative vaccination rates of the past nine months in 2021 according to Monrad et al. (2022) and Skinner et al. (2022). The data are extracted from the US Centers for Disease Control (US CDC), the Kaiser Family Foundation (KFF), the US Census Bureau, Americas Health Rankings (AHR), Cook Political, and COVID-19 US State Policy (CUSP) database.

Table 1. Data Sources for Variables included in the model

Data	Description	Year	Source
Vaccines administrated	Vaccines administered per state resident as of any given year	2021	US CDC
COVID-19 Vaccine hesitancy	Percentage who in Jan. 2021 would 'probably' or 'definitely' not get vaccinated	2021	Census Bureau
Elderly population	Percentage of the population above 65 years old	2021	Census Bureau
Population density	Population per square mile	2021	Census Bureau
Population in poverty	Percentage of population living in poverty	2017-2019	Census Bureau
Costs obstacles to care	Percentage reporting not seeing doctor in the past 12 months due to costs	2019	KFF
Healthcare expenditure	Expenditure per capita (in 10,000s USD)	2014	KFF
Public health expenditure	Expenditure per capita (in 100s USD)	2018-2019	AHR
Democratic vote margin	Vote margin in the 2020 elections in percentage points	2020	Cook Political
Face mask mandate	Whether any legal enforcement measures of the face mask mandate are present statewide	2020-2021	CUSP

The analysis is confined to 50 states due to the limitation of datasets (excluding the District of Columbia). Regarding the time range, though public healthcare expenditure is only available in 2014, all other datasets are from 2019 to 2021 Sep. 21st. The end date is chosen to avoid including the booster dose in the analysis (the booster was authorized in 2022 Sep. 22nd). To compare policy changes at separate time points, a cross-sectional analysis will be applied, focusing on vaccination data on three date points: April 1st, June 1st, and August 1st.

The response variable is the number of vaccinations administrated per capita, regardless of the first dose and second dose. Notably, the vaccination administrated rate is our primary variable, rather than the vaccination distribution rate, since the latter only proxies for the supply-side factors. The measurement of 1 vaccination corresponds to the situations in which every resident received 1 dose or half of the residents received 2 doses.

Other variables included are potential drivers for vaccination rates consistent with existing literature and theoretical hypothesis. Demand-side factors include residents' inclinations such as COVID-19 hesitancy and cost obstacles, which we hypothesize block the demand for vaccination. From supply-side factors, healthcare expenditures and public health expenditures are two important indicators of vaccination success. The former indicator is resources dedicated to healthcare services and products, while the latter indicator focuses on disease control and preventive programs. Finally, we include state-specific characteristics, including voting data, population density, poverty ratio, elderly ratio, and legal face mask mandate policy. The face mask mandate variable is categorical, with 1 for implementation of the policy with punishments, and 0 for non-formal implementation or no recommendation.

Analysis

Multivariable linear regression is adopted to examine the effect of supply-side, demand-side and state factors and their fitness level by R^2, which is specified as follows:

$$Y_{it} = \beta_0 + \beta_1 x_{1i} \cdots + \beta_k x_{ki} + \varepsilon_i$$

where Y_{it} is the vaccination per capita for each state i at time t, β_0 is the intercept coefficient, β_1 to β_t represent the coefficient before other regression variables x, and ε is the error term.

Moreover, we will add the interaction effect term, as represented here:

$$Y_{it} = \beta_0 + \beta_1 x_{1i} \cdots + \beta_k x_{ki} + \beta_{k+1} x_{ki} * x_{ji} + \varepsilon_i$$

where x_{ji} is the categorical variable that interacts with another continuous variable, the combination of two variables has a significantly larger effect than the individual variables alone.

Results

Descriptive statistics

The cross-sectional dataset covers 50 observations and 1512 columns. Eventually, 20 variables are chosen for analysis. Fortunately, there do not appear missing values from the dimensions we have chosen.

Table (2) summarizes the descriptive statistics for the full sample.

Table 2. Descriptive Statistics

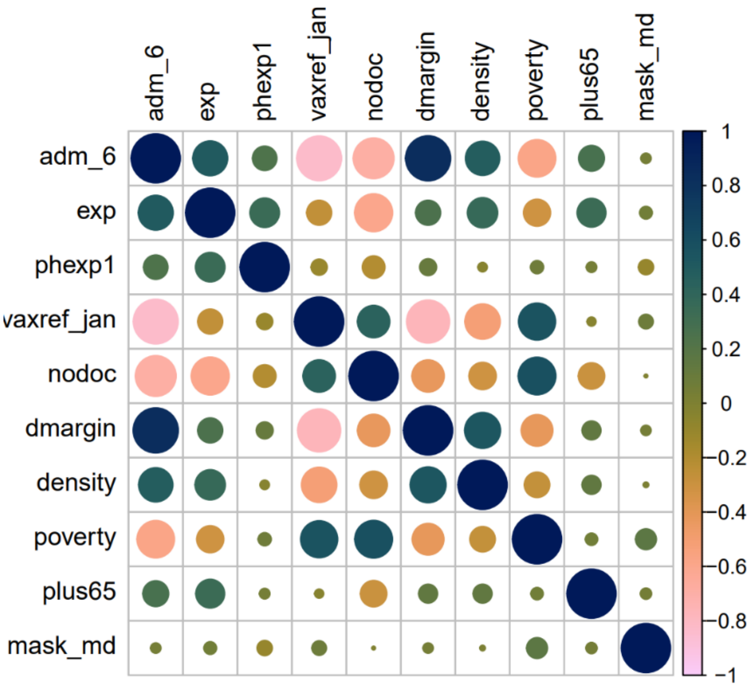
Statistic	N	Mean	St. Dev.	Min	Median	Max
adm_2	50	0.10	0.02	0.07	0.09	0.17
adm_4	50	0.48	0.05	0.36	0.48	0.61
adm_6	50	0.88	0.15	0.59	0.87	1.25
adm_8	50	1.02	0.17	0.72	0.98	1.40
adm_9	50	1.13	0.16	0.86	1.09	1.43
exp	50	0.83	0.12	0.60	0.81	1.11
phexp1	50	1.03	0.46	0.50	0.94	2.89
vaxref_jan	50	22.62	5.81	11	22	36
nodoc	50	12.76	2.63	8.20	12.60	18.80
dmargin	50	-2.34	20.66	-43	-0.5	35
density	50	203.67	266.47	1.27	107.58	1,206.76
poverty	50	11.22	2.84	5.60	10.55	19.10
plus65	50	17.06	1.98	11	17	21
mask_md	50	0.26	0.44	0	0	1

From the mean value, the categorical variable mask mandate policy (**mask_md**) shows 26% of the state samples did not enforce a mask mandate that only 13 states out of 50 enforced a mask mandate policy during the period. For continuous variables, our primary response variable, cumulative vaccines administrated per capita (**adm_2, adm_4, adm_6, adm_8, adm_9**) on Feb 1st, April 1st, June 1st, Aug 1st, Sep 1st, increases over time but the mean values do not deviate too much from the median values, forming normal distributions. However, the variable of democratic margin (**dmargin**) is highly left-skewed and has an extremely low minimum of -43 which pulls the mean value greatly lower than the median. The variable of density (**density**) is right skewed with an extremely high standard deviation (266.47) and the average is 96.09 per square mile greater than the 50% quartile.

Table 3. variable correlations

	adm_6	exp	phexp1	vaxref_jan	nodoc	dmargin	density	poverty	plus65
adm_6	1	0.81	0.39	-0.9	-0.93	0.97	0.85	-0.94	0.46
exp	0.81	1	0.52	-0.7	-0.93	0.67	0.69	-0.81	0.56
phexp1	0.39	0.52	1	-0.32	-0.46	0.27	0.11	-0.3	0.11
vaxref_jan	-0.97	-0.7	-0.32	1	0.83	-0.99	-0.89	0.91	-0.33
nodoc	-0.93	-0.93	-0.46	0.83	1	-0.83	-0.76	0.9	-0.58
dmargin	0.97	0.67	0.27	-0.99	-0.83	1	0.87	-0.9	0.34
density	0.85	0.69	0.11	-0.89	-0.76	0.87	1	-0.83	0.34
poverty	-0.94	-0.8	-0.3	0.91	0.9	-0.9	-0.83	1	-0.37
plus65	0.46	0.560	0.11	-0.33	-0.58	0.34	0.34	-0.37	1

Figure 1. Correlogram



However, high correlations between independent variables (> 0.5 and < - 0.5) (e.g., nodoc vs exp, poverty vs vaxref_jan) could potentially cause multi-collinearity issues. The VIF test is required to ensure all the VIF values are well below 10.

Base regression model

This study uses cross-sectional models on cumulative vaccinations per capita and only records dates on April 1st, June 1st , and August 1st in Table 4. The change of correlations and significance shows the effect of independent variables change over time.

Table 4. Regression result

	Dependent variable:		
	April 1st	June 1st	August 1st
	(1)	(2)	(3)
exp	0.13**	0.14	0.16*
	(0.004, 0.25)	(-0.03, 0.30)	(-0.01, 0.33)
phexp1	0.03**	0.03	0.01
	(0.01, 0.06)	(-0.01, 0.06)	(-0.02, 0.05)
vaxref_jan	-0.004**	-0.01***	-0.01***
	(-0.01, -0.001)	(-0.01, -0.01)	(-0.02, -0.01)
nodoc	-0.01**	-0.01***	-0.01**
	(-0.01, -0.002)	(-0.02, -0.004)	(-0.02, -0.001)
dmargin	-0.0004	0.003***	0.004***
	(-0.001, 0.0004)	(0.002, 0.004)	(0.003, 0.005)
density	-0.0000	-0.0000	-0.0000
	(-0.0001, 0.0000)	(-0.0001, 0.0000)	(-0.0001, 0.0000)
poverty	0.002	-0.005	-0.002
	(-0.003, 0.01)	(-0.01, 0.003)	(-0.01, 0.01)
plus65	0.001	0.01**	0.01**
	(-0.01, 0.01)	(0.001, 0.02)	(0.0004, 0.02)
mask_md	-0.003	0.02	0.03*
	(-0.03, 0.02)	(-0.01, 0.06)	(-0.003, 0.06)
Constant	0.50***	1.03***	1.14***
	(0.34, 0.66)	(0.81, 1.24)	(0.92, 1.36)
Observations	50	50	50
R^2	0.64	0.92	0.93
Adjusted R^2	0.55	0.91	0.92
Residual Std. Error (df = 40)	0.04	0.05	0.05
F Statistic (df = 9; 40)	7.76***	53.22***	60.66***
Note:			*p**p***p<0.01

Generally, health expenditure (**phexp1**), democratic margin (**dmargin**), population above 65 years old (**plus65**), and mandate mask policy (**mask_md**) are positively correlated with cumulative vaccinations per capita across three models, while vaccine hesitancy and cost obstacles show negative signs. Based on the third model (in August), with 10,000 dollars per capita increasing in health expenditure, the cumulative vaccination per capita increased by 0.16. The democratic margin (**dmargin**) and population above 65 years old (**plus65**) have a minor impact, with coefficients of 0.004 and 0.01 respectively. If the state has a mandate mask policy, the cumulative vaccinations per capita would increase by 0.03. Additionally, both vaccine hesitancy, and cost obstacles to care might hinder the vaccination administration, with both coefficients being -0.01.

Evolving coefficients and significances are also meaningful in the models. Public health expenditure is significant in the first model (in April) but becomes non-significant in the left two. This might be because the public health expenditure was effective at first, but the effect decreased with time. The magnitude of the coefficient population above 65 years old increases from April 1st to June 1st. This is unsurprising that elderly people are vulnerable groups to COVID-19 so they have greater demand for vaccination. The mandate mask policy also has an increasing effect over time perhaps because the policy should be effective after a period of implementation.

Diagnosis tests are conducted for cross-sectional regressions. We only show plots for the third model (in August) here, those of other months are provided in the Appendix.

Multilinearity. The VIF values of the variables are all lower than 10 in Table (5), indicating that there is no multilinearity in the model.

Table 5. VIF of variables

	VIF
exp	2.112478
phexp1	1.440014
vaxref_jan	3.154960
nodoc	2.444040
dmargin	2.789379
density	1.734017
poverty	2.426814
plus65	1.316264

Linearity. The nine continuous variables are displayed in a scatter plot matrix. All variables follow the assumption of linearity, whose fitted lines all fall into the loess confidential interval, though **phep1** and **plus 65** deviate from the linear shapes.

Figuer 2: Scatter Plot Matrix

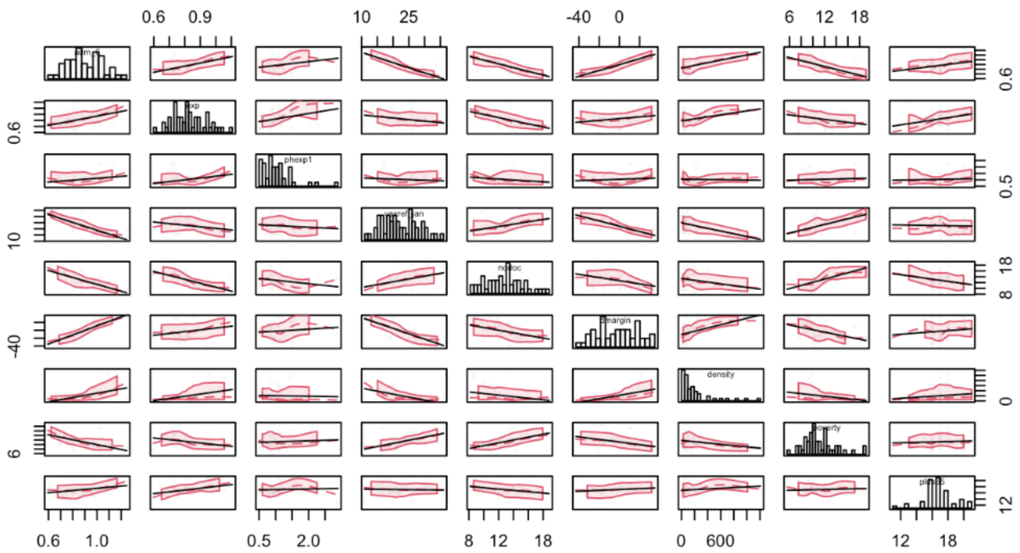
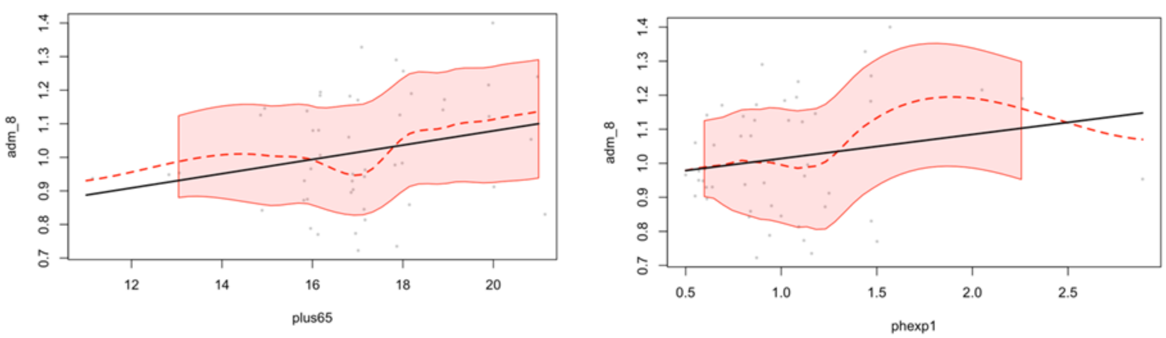
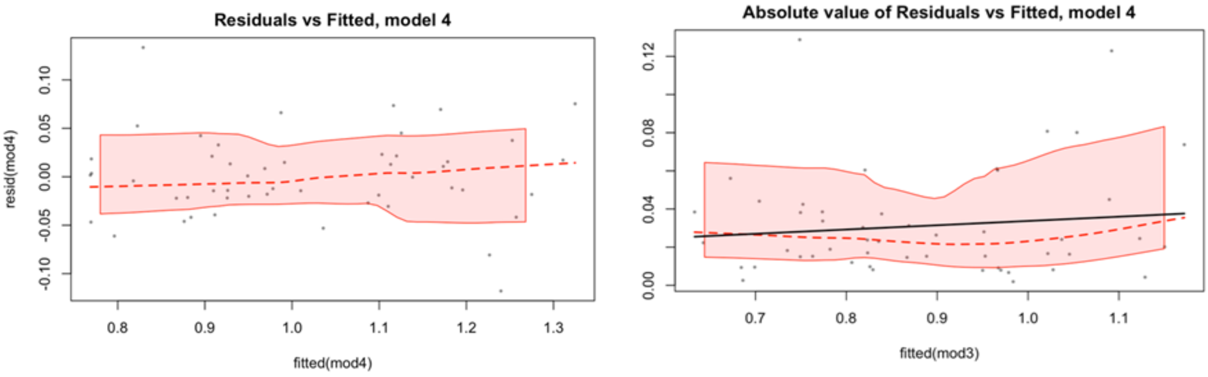


Figure 3. Scatterplot of adm_8 vs phexp1 and adm_8 vs plus 65



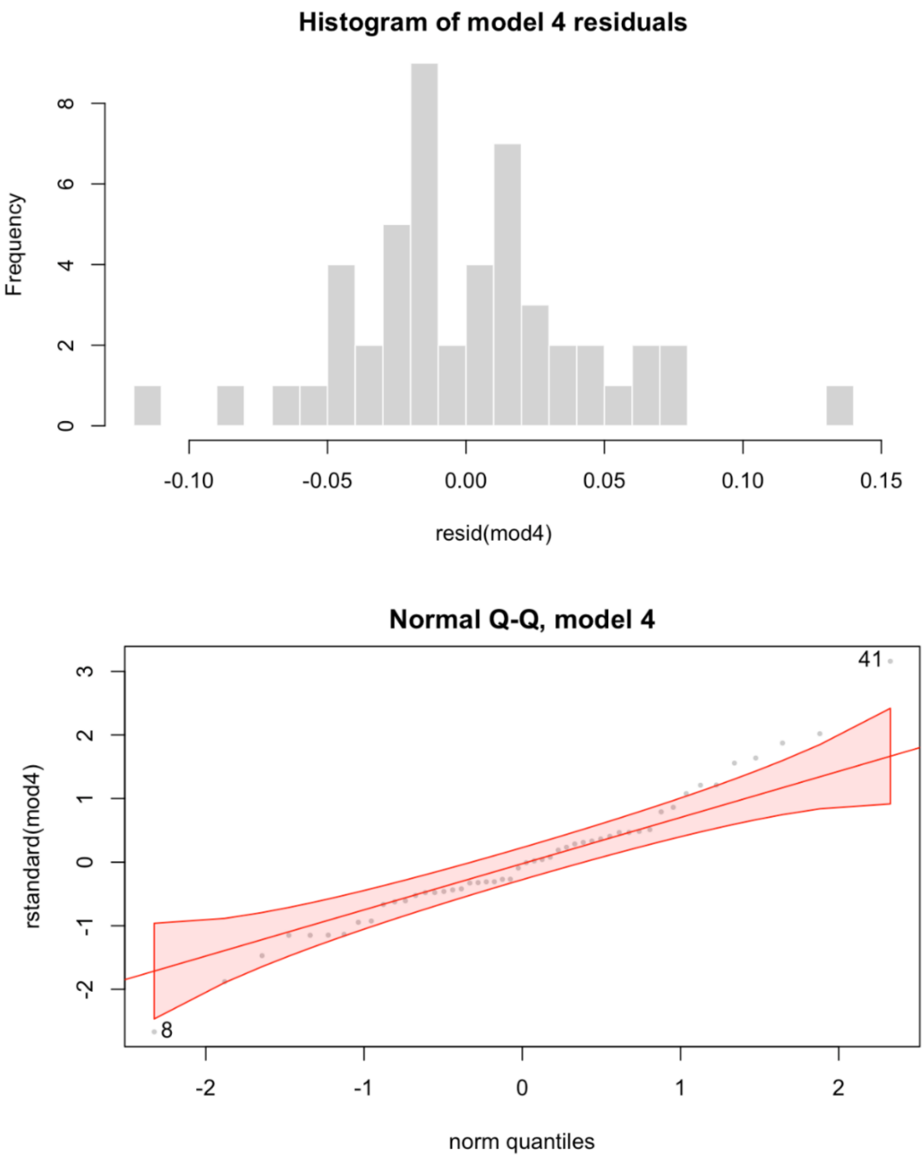
Homoskedasticity. The assumption of homoskedasticity is met by the linear model according to two scatterplots. The spots fall into the confidential interval and the best-fitted line is close to the loess line. It is also confirmed by the Breusch-Pagan test that the p-value is close to 0 and we have sufficient evidence to conclude homoskedasticity.

Figure 4. Residuals vs Fitted and Absolute value of Residuals vs Fitted



Normality of residuals. A histogram of the residuals shows that the residuals seem to be normally distributed, but the accurate result still needs to be re-assured. From Q-Q plot with confidential intervals, the residuals of this model are generally in normal distribution since the most.

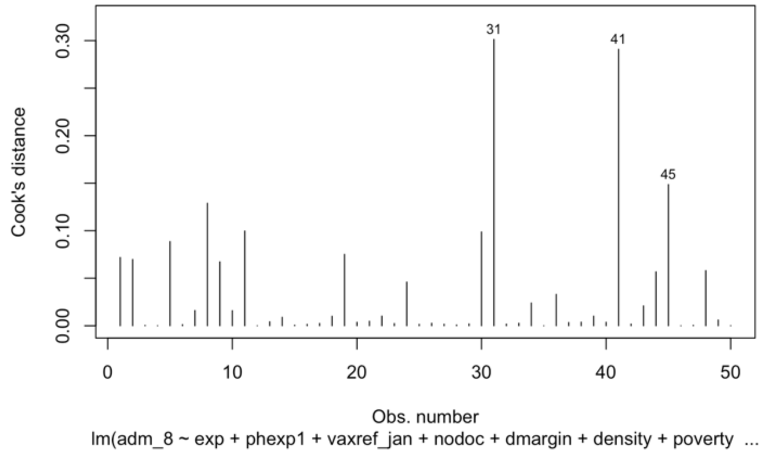
Figure 5. Distribution of residuals (histogram and QQplot)



Outliers. This study applied three methods to identify outliers, namely influence, discrepancy, and leverage.

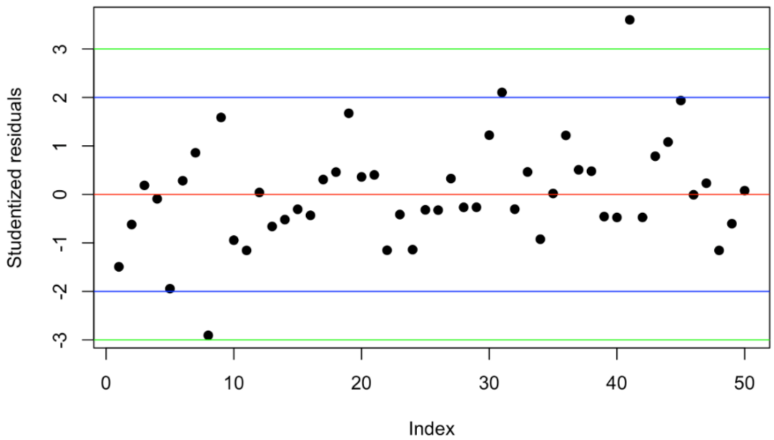
Influence. Cook's distance measures the effect on the residuals of all observations when deleting a given observation. The three potential outliers identified by cook's distance are Alaska, New Mexico, and South Dakota.

Figure 6. Cook's distance values



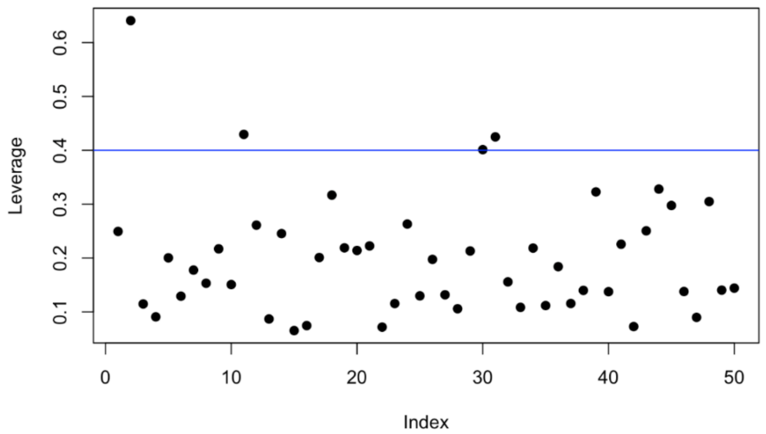
Discrepancy. Looking at the distribution of studentized residuals can give a sense of which observations have large differences between predicted and observed values and consequently might be outliers. In Figure 7, most of the residuals fall in the range of -2 to 2, with only a few observations having values lower than −3 or higher than 3. Observations with studentized residuals greater than ±3 might be outliers. The potential outlier is South Dakota.

Figure 7. Studentized values of residuals



Leverage. To examine the leverage, this analysis uses one rule of thumb. Since there are 50 observations and 9 independent variables in the model, the leverage value is about 0.4. In Figure 8, one observation has a leverage value of more than 0.5, which might be outlier, which is Alaska.

Figure 8. Rule of thumb for leverage



Potential outliers

After comparing with the descriptive statistics, Alaska, with the highest public health expenditure and the lowest population density, is identified as outlier in this model. Thus, Alaska might have the best capacity to distribute vaccines well and achieve the highest vaccination rates among all the states. Thus, this study decided to identify Alaska as an outlier.

Table 6. Potential outliers

state	Alaska	New Mexico	Oklahoma	South Dakota	Vermont
adm_6	0.8375	1.02686	0.7483	0.8774	1.246
exp	1.1064	0.7214	0.7627	0.8933	1.0190
phexp1	2.89	2.26	0.95	1.12	1.57
vaxref_jan	26	17	27	34	19
nodoc	13.5	13.9	16.2	9.8	9.3
dmargin	-10	11	-33	-26	35
density	1.2694	17.3540	58.1740	11.8265	67.6197
poverty	11.8	17.2	12.4	10.6	9.0

plus65	13	18	16	17	20

Corrected Model

Since the linearity of public health expenditure and population above 65 years old in the scatter plots is not completely met, this study examines the quadratic effect of the two variables. However, the results show that the quadratic items of the two variables do not have significant association with administrated vaccination rate. The results are in the Appendix.

To down-weight outliers, we replicate the method used by Monrad et al. (2022), who employed iteratively re-weighted least squares regression (IRLS) using the **rlm()** package in R. The IRLS model is used also because this study only has a small sample size (n = 50), omitting any sample could miss key messages. By viewing the dataset, though Alaska has the highest public health expenditure and the lowest population density, it is still in a reasonable range. Thus, the IRLS model is a more appropriate method than deleting Alaska as an outlier.

The IRLS model retains all the magnitude and significance of variables and undermines the influence of outliers. This study decides to replace the OLS model with the IRLS model.

Table 7. Corrected models

	Dependent variable:		
	adm_4	adm_6	adm_8
	(1)	(2)	(3)
exp	0.16***	0.13**	0.16**
	(0.06, 0.26)	(0.01, 0.25)	(0.02, 0.30)
phexp1	0.02**	0.03**	0.01
	(0.0002, 0.04)	(0.002, 0.05)	(-0.02, 0.04)
vaxref_jan	-0.003***	-0.01***	-0.01***
	(-0.01, -0.001)	(-0.01, -0.01)	(-0.02, -0.01)
nodoc	-0.01***	-0.01***	-0.01*
	(-0.01, -0.003)	(-0.01, -0.003)	(-0.01, 0.001)
dmargin	-0.0004	0.003***	0.004***
	(-0.001, 0.0002)	(0.002, 0.004)	(0.003, 0.005)
density	-0.0000	-0.0000	-0.0000
	(-0.0001, 0.0000)	(-0.0001, 0.0000)	(-0.0001, 0.0000)
poverty	0.001	-0.01*	-0.003
	(-0.004, 0.005)	(-0.01, 0.0000)	(-0.01, 0.003)
plus65	0.0004	0.01***	0.01**
	(-0.004, 0.005)	(0.004, 0.01)	(0.002, 0.01)
mask_md	-0.001	0.03***	0.04***
	(-0.02, 0.02)	(0.01, 0.05)	(0.01, 0.06)
Constant	0.48***	1.03***	1.14***
	(0.35, 0.61)	(0.87, 1.19)	(0.96, 1.33)
Observations	50	50	50
Residual Std. Error (df = 40)	0.02	0.03	0.03
Note:			*p **p ***p<0.01

Interaction effects between health expenditure and mask mandate

In Table 7, there is evidence for the effect of mandate mask policy being moderated by health expenditure at 5% significance. Therefore, the impact of mandate mask policy on vaccination rate per capita depends on health expenditure. In August, on average, states that launched the mandate mask policy have a vaccination rate per capita 0.03 percent higher than states not. However, this effect also depends on health expenditure. For every 10,000 dollars per capita increase in health expenditure, states that do not launch the mandate mask policy expect to have an additional 0.44 percent in vaccination rate per capita, while a state that does not launch the policy expects to have 0.12 percent in vaccination rate per capita higher than average. The interactive effect is further evidenced by the ANOVA test displayed below, with the P-value lower than 0.05.

Table 8. Interaction effect of health expenditure

	Dependent variable:		
	adm_4	adm_6	adm_8
	(1)	(2)	(3)
expc	0.16***	0.09	0.12
	(0.06, 0.27)	(-0.06, 0.24)	(-0.05, 0.28)
phexp1	0.02*	0.03**	0.01
	(-0.0004, 0.04)	(0.003, 0.06)	(-0.02, 0.04)
vaxref_jan	-0.003***	-0.01***	-0.01***
	(-0.01, -0.001)	(-0.01, -0.01)	(-0.02, -0.01)
nodoc	-0.01***	-0.01**	-0.01*
	(-0.01, -0.002)	(-0.02, -0.002)	(-0.01, 0.001)
dmargin	-0.0004	0.003***	0.004***
	(-0.001, 0.0003)	(0.002, 0.004)	(0.003, 0.005)
density	-0.0000	-0.0000	-0.0000
	(-0.0001, 0.0000)	(-0.0001, 0.0000)	(-0.0001, 0.0000)
poverty	0.001	-0.01*	-0.003
	(-0.004, 0.01)	(-0.01, 0.001)	(-0.01, 0.004)
plus65	0.0004	0.01***	0.01**
	(-0.004, 0.01)	(0.003, 0.02)	(0.0003, 0.01)

mask_md	-0.001	0.03**	0.03**
	(-0.02, 0.02)	(0.003, 0.06)	(0.003, 0.06)
expc:mask_md	-0.02	0.28**	0.29**
	(-0.20, 0.17)	(0.03, 0.54)	(0.004, 0.58)
Constant	0.61***	1.12***	1.27***
	(0.50, 0.72)	(0.97, 1.27)	(1.11, 1.44)
Observations	50	50	50
Residual Std. Error (df = 39)	0.02	0.03	0.03
Note:			*p**p***p<0.01

Table 9. ANOVA test of the interaction effect of health expenditure

Statistic	N	Mean	St. Dev.	Min	Max
Res.Df	1	40.000		40	40
RSS	2	0.089	0.004	0.087	0.092
Sum of Sq	1	0.006		0.006	0.006

Interaction effects between elderly ratio and mask mandate

The interaction between population above 65 years old and mandate mask policy is significant as well. Table 3 implies that in August, on average, states that ask citizens to wear mask compulsorily have vaccination rate per capita 0.03 percent higher than other states. However, the effect is also influenced by the health expenditure. For every 1-unit increase in the population above 65 years old, states that do not launch the mandate mask policy expect to have additional 0.01 percent in vaccination rate per capita, while a state that forces citizens to wear mask have 0.06 percent in vaccination rate per capita higher than average. The interactive effect is further evidenced by the anova test displayed below, with the P-value lower than 0.005.

Table 10. Interaction effect of population above 65 years old

	Dependent variable:		
	adm_4	adm_6	adm_8
	(1)	(2)	(3)
exp	0.17^{***}	0.11^{*}	0.14^{**}
	(0.07, 0.27)	(-0.02, 0.24)	(0.005, 0.28)
phexp1	0.02^{*}	0.03^{**}	0.01
	(-0.001, 0.04)	(0.002, 0.06)	(-0.02, 0.04)
vaxref_jan	-0.003^{**}	-0.01^{***}	-0.01^{***}
	(-0.01, -0.001)	(-0.01, -0.01)	(-0.02, -0.01)
nodoc	-0.01^{***}	-0.01^{***}	-0.01^{**}
	(-0.01, -0.002)	(-0.02, -0.003)	(-0.01, -0.0000)
dmargin	-0.0004	0.003^{***}	0.004^{***}
	(-0.001, 0.0003)	(0.002, 0.004)	(0.003, 0.005)
density	-0.0000	-0.0000	-0.0000
	(-0.0001, 0.0000)	(-0.0001, 0.0000)	(-0.0001, 0.0000)
poverty	0.0001	-0.01^{*}	-0.003
	(-0.004, 0.005)	(-0.01, 0.001)	(-0.01, 0.003)
plus65c	0.001	0.01^{***}	0.01^{**}
	(-0.004, 0.01)	(0.002, 0.01)	(0.0000, 0.01)
mask_md	-0.0003	0.03^{**}	0.03^{**}
	(-0.02, 0.02)	(0.002, 0.05)	(0.01, 0.06)
plus65c:mask_md	-0.01	0.02^{**}	0.02^{**}
	(-0.02, 0.01)	(0.002, 0.04)	(0.004, 0.04)
Constant	0.48^{***}	1.20^{***}	1.30^{***}
	(0.35, 0.60)	(1.04, 1.35)	(1.14, 1.46)
Observations	50	50	50
Residual Std. Error (df = 39)	0.02	0.03	0.03
Note:			*p**p***p<0.01

Table 11. ANOVA test of the interaction effect of population above 65 years old

Statistic	N	Mean	St. Dev.	Min	Max

Res.Df	1	40.000		40	40
RSS	2	0.092	0.001	0.091	0.092
Sum of Sq	1	0.001		0.001	0.001

Robustness to an alternative measure

Previously, the cross-sectional regressions were based on the cumulative number of vaccinations. However, it might be the case that for the first few months, some states initiate campaigns that drive the vaccination rates, but the states may perform poorly in the subsequent months and the associations persist due to high autocorrelation. We now ran another regression on the increased vaccination administrated in each state for each month to see whether the patterns continue without an autocorrelation over time.

Table 12. Robustness check model

	Dependent variable:		
	adm_r_4	adm_r_6	adm_r_8
	(1)	(2)	(3)
exp	0.03	-0.003	-0.01
	(-0.04, 0.11)	(-0.05, 0.04)	(-0.05, 0.03)
phexp1	-0.005	-0.001	-0.01
	(-0.02, 0.01)	(-0.01, 0.01)	(-0.02, 0.002)
vaxref_jan	-0.004***	-0.002***	-0.001
	(-0.01, -0.002)	(-0.003, -0.0004)	(-0.002, 0.0004)
nodoc	-0.001	0.002	0.004***
	(-0.004, 0.003)	(-0.001, 0.004)	(0.001, 0.01)
dmargin	0.001***	0.001***	-0.0001
	(0.001, 0.002)	(0.0004, 0.001)	(-0.0004, 0.0002)
density	0.0000	-0.0000	0.0000*
	(-0.0000, 0.0000)	(-0.0000, 0.0000)	(-0.0000, 0.0000)
poverty	-0.01***	-0.0001	0.003***
	(-0.01, -0.002)	(-0.002, 0.002)	(0.001, 0.004)
plus65	0.004**	0.001	-0.0002

	(0.001, 0.01)	(-0.001, 0.003)	(-0.002, 0.002)
mask_md	0.02***	0.01*	0.01*
	(0.01, 0.03)	(-0.0002, 0.02)	(-0.0005, 0.02)
Constant	0.32***	0.08***	0.02
	(0.22, 0.41)	(0.02, 0.15)	(-0.03, 0.08)
Observations	50	50	50
Residual Std. Error (df = 40)	0.02	0.01	0.01
Note:			*p$^{<0.1; >**}$p$^{<0.05; >***}$p<0.01

With the new dependent variable, new vaccinations administrated per capita during months in April, June and August, some key patterns persist. Vaccine hesitancy is still negatively correlated with the increased vaccination rates at a similar magnitude with 1% significance, though the significance during August drops. Cost obstacle to care, as predicted, is positively correlated with increased vaccination rate, while only significant during August. Moreover, the democratic margin still shows a significant positive correlation with the response variable, though the magnitude of the coefficient diminishes. Persistent patterns appear for variables of elderly ratio and mask mandate policy, showing positive signs and weaker significance. However, for demand-side factors, both healthcare expenditure and public healthcare expenditure only retain 5% significance in February but lose their significance in the following months. In contrast to the previous model, the poverty factor is correlated with the new vaccination rates, but from April to August, the confusing sign of the correlation changes from negative to positive.

Regarding the interaction effects, the elderly factor does not have a combined effect with the mask mandate policy since the interaction terms are not significant over time. For the interaction term between health expenditure and vaccination rates, the interaction effects are only strong in April.

Conclusion and discussion

In summary, our cross-sectional model captures the changing effects of three-side factors on vaccination rates. Overall, the effect of demand-side factors including vaccination hesitancy, economic obstacles, and democratic voting is stable and growing over time. However, supply-side factors including hospitals, clinics, and COVID information campaigns are only the trigger for vaccination administration in the beginning. After the supply constraints were eliminated, their effects got weaker over time.

Several limitations exist in this replication analysis. As noted by Teperowski et al. (2022), the main limitation of the dataset is to include public health expenditures in 2014, considering relative expenditures across states may have evolved within seven years since then. It also does not explain the strike patterns for non-stats US territories given the data limitation.

Another limitation is the methodology of multivariable regression analysis which only accounts for correlations but not for causal inference. For instance, a bidirectional causal relationship may arise between vaccination rates and demand factors. Driven by increasing vaccination administrated rates within the state, residents may be less hesitant towards receiving vaccinations, implying a negative association. To mitigate this issue, one recommended method is the DID framework that relates the policy changes to real-time vaccination rates. This methodology is beyond our capability and could be explored in the future. In addition, as implied by ecological fallacies, with only results of the aggregated group of states, the individual state is not inferable. For example, different states experienced different challenges in logistic supply, equalizing among communities and developed separate scheduling systems, information campaigns (Tewarson et al., 2021), and incentives mechanisms (Walkey et al., 2021, Brehm et al., 2021).

Finally, our analysis is threatened by omitted variable issues. After including three-fold controls, we attempt to include more variables in the primary model, such as pre-COVID chronic diseases conditions and flu vaccination rate, however, these variables fail to predict the cumulative vaccination rates.

Regarding policy implications, investment in healthcare supplies is indispensable, though its significance diminishes after achieving a threshold. Moreover, more research should explore the reasons behind vaccination hesitancy. Additionally, cost obstacles to healthcare need to be noted. It might seem strange since vaccination is distributed freely. However, some patients are inaccessible to vaccination sites due to heavy traffic costs and corporations do not offer enough benefits for employees to receive vaccination and recover from its side effects. Furthermore, states should have targeted efforts for minority groups, especially those elderly who have mobility issues and are most vulnerable to COVID. Finally, investment in healthcare supplies is indispensable, though the efficiency of input should be measured.

References

Abubakar, I., Dalglish, S. L., Angell, B., Sanuade, O., Abimbola, S., Adamu, A. L., Adetifa, I. M. O., Colbourn, T., Ogunlesi, A. O., Onwujekwe, O., Owoaje, E. T., Okeke, I. N., Adeyemo, A., Aliyu, G., Aliyu, M. H., Aliyu, S. H., Ameh, E. A., Archibong, B., Ezeh, A., & Gadanya, M. A. (2022). The Lancet Nigeria Commission: investing in health and the future of the nation. *The Lancet*, *399*(10330), 1155–1200. https://doi.org/10.1016/S0140-6736(21)02488-0

Brien, S., Kwong, J. C., Charland, K. M., Verma, A. D., Brownstein, J. S., & Buckeridge, D. L. (2012). Neighborhood Determinants of 2009 Pandemic A/H1N1 Influenza Vaccination in Montreal, Quebec, Canada. *American Journal of Epidemiology*, *176*(10), 897–908. https://doi.org/10.1093/aje/kws154

Chatfield, C. (2003). *The Analysis of Time Series*. Chapman and Hall/CRC. https://doi.org/10.4324/9780203491683

Davila-Payan, C., Swann, J., & Wortley, P. M. (2014). System factors to explain H1N1 state vaccination rates for adults in US emergency response to pandemic. *Vaccine*, *32*(25), 3088–3093. https://doi.org/10.1016/j.vaccine.2013.05.069

Doherty, T. M., Connolly, M. P., Del Giudice, G., Flamaing, J., Goronzy, J. J., Grubeck-Loebenstein, B., Lambert, P.-H., Maggi, S., McElhaney, J. E., Nagai, H., Schaffner, W., Schmidt-Ott, R., Walsh, E., & Di Pasquale, A. (2018). Vaccination programs for older adults in an era of demographic change. *European Geriatric Medicine*, *9*(3), 289–300. https://doi.org/10.1007/s41999-018-0040-8

Justel, A., Peña, D., & Zamar, R. (1997). A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics & Probability Letters*, *35*(3), 251–259. https://doi.org/10.1016/s0167-7152(97)00020-5

Lim, C.-H., Ryu, J., Choi, Y., Jeon, S. W., & Lee, W.-K. (2020). Understanding global PM2.5 concentrations and their drivers in recent decades (1998–2016). *Environment International*, *144*, 106011. https://doi.org/10.1016/j.envint.2020.106011

Lin, Y., Zou, J., Yang, W., & Li, C.-Q. (2018). A Review of Recent Advances in Research on PM2.5 in China. *International Journal of Environmental Research and Public Health*, *15*(3), 438. https://doi.org/10.3390/ijerph15030438

US EPA. (2016, July 19). *Particulate Matter (PM2.5) Trends*. US EPA. https://www.epa.gov/air-trends/particulate-matter-pm25-trends

Viswanath, K., Bekalu, M., Dhawan, D., Pinnamaneni, R., Lang, J., & McLoud, R. (2021). Individual and social determinants of COVID-19 vaccine uptake. *BMC Public Health*, *21*(1). https://doi.org/10.1186/s12889-021-10862-1

Wang, L., Zhang, N., Liu, Z., Sun, Y., Ji, D., & Wang, Y. (2014). The Influence of Climate Factors, Meteorological Conditions, and Boundary-Layer Structure on Severe Haze Pollution in the Beijing-Tianjin-Hebei Region during January 2013. *Advances in Meteorology*, *2014*, 1–14. https://doi.org/10.1155/2014/685971

Wang, Z., Li, J., Wang, Z., Yang, W., Tang, X., Ge, B., Yan, P., Zhu, L., Chen, X., Chen, H., Wand, W., Li, J., Liu, B., Wang, X., Wand, W., Zhao, Y., Lu, N., & Su, D. (2013). Modeling study of regional severe hazes over mid-eastern China in January 2013 and its implications on pollution prevention and control. *Science China Earth Sciences*, *57*(1), 3–13. https://doi.org/10.1007/s11430-013-4793-0

Wood, L. A. (1970). The use of dew-point temperature in humidity calculations. *Journal of Research of the National Bureau of Standards, Section C: Engineering and Instrumentation*, *74C*(3-4), 117. https://doi.org/10.6028/jres.074c.014

World Health Organization (WHO). (2017). Evolution of WHO air quality guidelines: past, present and future. In *apps.who.int*. World Health Organization Regional Office for Europe. https://apps.who.int/iris/handle/10665/341912

Zhang, Q., Zheng, Y., Tong, D., Shao, M., Wang, S., Zhang, Y., Xu, X., Wang, J., He, H., Liu, W., Ding, Y., Lei, Y., Li, J., Wang, Z., Zhang, X., Wang, Y., Cheng, J., Liu, Y., Shi, Q., & Yan, L. (2019a). Drivers of improved PM2.5 air quality in

China from 2013 to 2017. *Proceedings of the National Academy of Sciences*, *116*(49), 201907956. https://doi.org/10.1073/pnas.1907956116

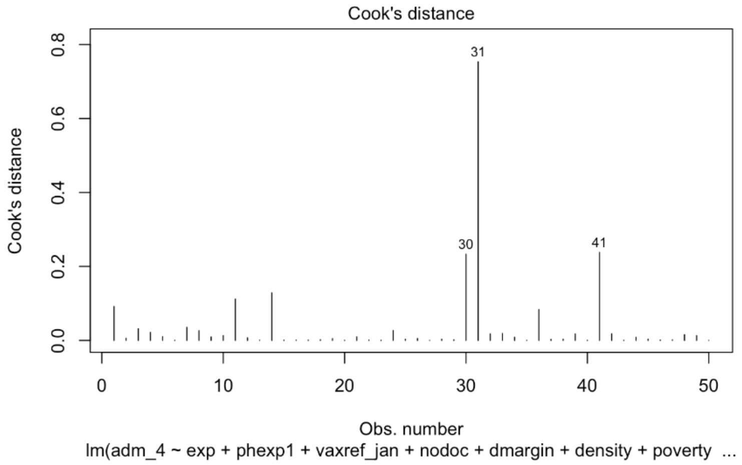
Zhang, Q., Zheng, Y., Tong, D., Shao, M., Wang, S., Zhang, Y., Xu, X., Wang, J., He, H., Liu, W., Ding, Y., Lei, Y., Li, J., Wang, Z., Zhang, X., Wang, Y., Cheng, J., Liu, Y., Shi, Q., & Yan, L. (2019b). Drivers of improved PM2.5 air quality in China from 2013 to 2017. *Proceedings of the National Academy of Sciences*, *116*(49), 201907956. https://doi.org/10.1073/pnas.1907956116

Zhang, S., Guo, B., Dong, A., He, J., Xu, Z., & Chen, S. X. (2017). Cautionary tales on air-quality improvement in Beijing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *473*(2205), 20170457. https://doi.org/10.1098/rspa.2017.0457

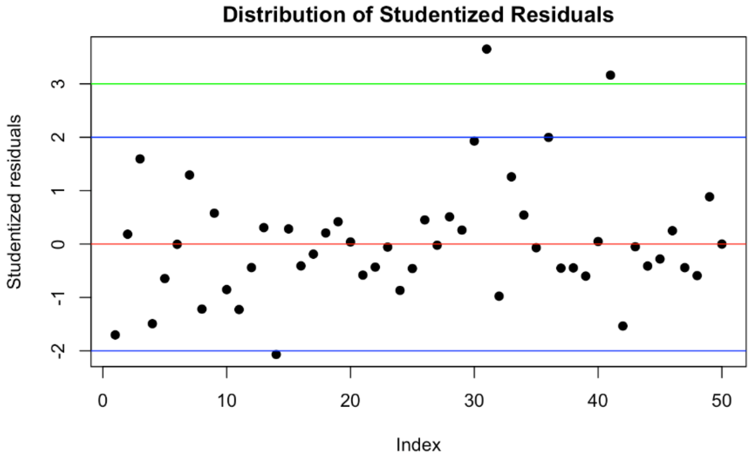
Appendix

Outliers of the first model (April 1st)

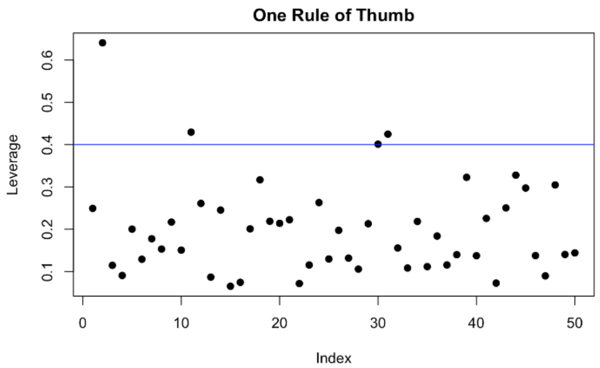
By Cook's D,



By studentized residuals



By rule of thumb



Potential outliers

state	Alaska	New Jersy	New Mexico	South Dakota
adm_4	0.56255	0.5093	0.6109	0.5709
exp	1.1064	0.8859	0.7214	0.8933
phexp1	2.89	0.69	2.26	1.12
vaxref_jan	26	16	17	34
nodoc	13.5	13.1	13.9	9.8
dmargin	-10	16	11	-26
density	1.2694	1206.7609	58.1740	11.8265
poverty	11.8	17.2	12.4	10.6
plus65	13	18	16	17

The quadratic item of public health expenditure and elderly ratiodoes not make much difference on the previous model, so this study does not involve it into the final model.

Table 12. Quadratic effect of public health expenditure

	Dependent variable:	
	adm_8	
	OLS	robust linear
	(1)	(2)
exp	0.16*	0.16**
	(-0.01, 0.33)	(0.01, 0.31)
phexp1	0.01	
	(-0.02, 0.05)	
phexp1c2		0.002
		(-0.03, 0.04)
phexp1c		0.01
		(-0.03, 0.05)
vaxref_jan	-0.01***	-0.01***
	(-0.02, -0.01)	(-0.02, -0.01)
nodoc	-0.01**	-0.01*
	(-0.02, -0.001)	(-0.01, 0.001)
dmargin	0.004***	0.004***
	(0.003, 0.005)	(0.003, 0.005)
density	-0.0000	-0.0000
	(-0.0001, 0.0000)	(-0.0001, 0.0000)
poverty	-0.002	-0.003
	(-0.01, 0.01)	(-0.01, 0.004)
plus65	0.01**	0.01**
	(0.0004, 0.02)	(0.001, 0.02)
mask_md	0.03*	0.04***
	(-0.003, 0.06)	(0.01, 0.07)
Constant	1.14***	1.15***
	(0.92, 1.36)	(0.96, 1.35)
Observations	50	50
R2	0.93	
Adjusted R2	0.92	
Residual Std. Error	0.05 (df = 40)	0.03 (df = 39)
F Statistic	60.66*** (df = 9; 40)	
Note:	*p**p***p<0.01	

Table 13. Quadratic effect of population above 65 years old

	Dependent variable:	
	adm_8	
	OLS	robust linear
	(1)	(2)
exp	0.16*	0.18**
	(-0.01, 0.33)	(0.04, 0.32)
phexp1	0.01	0.001
	(-0.02, 0.05)	(-0.03, 0.03)
vaxref_jan	-0.01***	-0.01***
	(-0.02, -0.01)	(-0.02, -0.01)
nodoc	-0.01**	-0.01*
	(-0.02, -0.001)	(-0.01, 0.0003)
dmargin	0.004***	0.004***
	(0.003, 0.005)	(0.003, 0.005)
density	-0.0000	-0.0000
	(-0.0001, 0.0000)	(-0.0001, 0.0000)
poverty	-0.002	-0.002
	(-0.01, 0.01)	(-0.01, 0.004)
plus65	0.01**	
	(0.0004, 0.02)	
plus65c2		0.001
		(-0.0004, 0.003)
plus65c		0.01***
		(0.003, 0.02)
mask_md	0.03*	0.04***
	(-0.003, 0.06)	(0.01, 0.06)
Constant	1.14***	1.26***
	(0.92, 1.36)	(1.09, 1.42)
Observations	50	50
R2	0.93	
Adjusted R2	0.92	
Residual Std. Error	0.05 (df = 40)	0.03 (df = 39)
F Statistic	60.66*** (df = 9; 40)	
Note:	*p**p***p<0.01	

The interaction items of public health expenditure and elderly ratio are not significant, so this study does not involve them into the final model.

Table 14. Robustness check model for expenditure vs mask mandate

	Dependent variable:		
	april_adm	june_adm	august_adm
	(1)	(2)	(3)
expc	0.03	0.02	-0.01
	(-0.05, 0.11)	(-0.03, 0.07)	(-0.07, 0.05)
phexp1	-0.01	-0.01	-0.01
	(-0.02, 0.01)	(-0.02, 0.004)	(-0.02, 0.003)
vaxref_jan	-0.004^{***}	-0.001^{**}	-0.0004
	(-0.01, -0.002)	(-0.003, -0.0004)	(-0.002, 0.001)
nodoc	-0.0004	0.002^{**}	0.004^{**}
	(-0.004, 0.003)	(0.0002, 0.005)	(0.001, 0.01)
dmargin	0.001^{***}	0.001^{***}	-0.0000
	(0.001, 0.002)	(0.0004, 0.001)	(-0.0004, 0.0004)
density	-0.0000	-0.0000	0.0000
	(-0.0000, 0.0000)	(-0.0000, 0.0000)	(-0.0000, 0.0000)
poverty	-0.01^{***}	0.0001	0.002^{*}
	(-0.01, -0.002)	(-0.002, 0.002)	(-0.0002, 0.005)
plus65	0.005^{**}	0.0004	-0.001
	(0.001, 0.01)	(-0.002, 0.003)	(-0.004, 0.002)
mask_md	0.02^{**}	0.01^{*}	0.01
	(0.001, 0.03)	(-0.001, 0.02)	(-0.003, 0.02)
expc:mask_md	0.08	0.01	-0.08
	(-0.06, 0.22)	(-0.07, 0.09)	(-0.18, 0.03)
Constant	0.34^{***}	0.08^{***}	0.02
	(0.26, 0.42)	(0.03, 0.13)	(-0.04, 0.09)
Observations	50	50	50
R^2	0.91	0.74	0.60
Adjusted R^2	0.89	0.68	0.49
Residual Std. Error (df = 39)	0.02	0.01	0.02
F Statistic (df = 10; 39)	39.36^{***}	11.27^{***}	5.80^{***}
Note:			*p$^{<0.1; >**}$p$^{<0.05; >***}$p<0.01

Table 14. Robustness check model for elderly ratio vs mask mandate

	Dependent variable:		
	april_adm	june_adm	august_adm
	(1)	(2)	(3)
exp	0.04	0.02	-0.02
	(-0.04, 0.12)	(-0.03, 0.07)	(-0.08, 0.04)
phexp1	-0.01	-0.01	-0.01
	(-0.02, 0.01)	(-0.02, 0.004)	(-0.02, 0.004)
vaxref_jan	-0.004***	-0.002***	-0.0002
	(-0.01, -0.002)	(-0.003, -0.0004)	(-0.002, 0.001)
nodoc	-0.001	0.002*	0.004***
	(-0.004, 0.003)	(0.0000, 0.004)	(0.001, 0.01)
dmargin	0.001***	0.001***	-0.0000
	(0.001, 0.002)	(0.0004, 0.001)	(-0.0004, 0.0003)
density	-0.0000	-0.0000	0.0000*
	(-0.0000, 0.0000)	(-0.0000, 0.0000)	(-0.0000, 0.0000)
poverty	-0.01***	0.0003	0.002
	(-0.01, -0.002)	(-0.002, 0.002)	(-0.001, 0.004)
plus65c	0.004**	0.0001	-0.0004
	(0.001, 0.01)	(-0.002, 0.002)	(-0.003, 0.002)
mask_md	0.02**	0.01	0.01
	(0.001, 0.03)	(-0.002, 0.02)	(-0.002, 0.02)
plus65c:mask_md	0.004	0.003	-0.01
	(-0.01, 0.01)	(-0.003, 0.01)	(-0.01, 0.001)
Constant	0.39***	0.08***	0.02
	(0.30, 0.48)	(0.02, 0.13)	(-0.05, 0.08)
Observations	50	50	50
R^2	0.91	0.75	0.60
Adjusted R^2	0.89	0.68	0.50
Residual Std. Error (df = 39)	0.02	0.01	0.02
F Statistic (df = 10; 39)	38.74***	11.62***	5.89***
Note:			*p<0.1; >**p<0.05; >***p<0.01