

2842A

**PART IIA PAPER 3 PROJECT**

Q1. Vaccination take-up rates: Inter-State differences in the USA in 2021.

At the end of 2021 the share of the population fully vaccinated against Covid-19 differed widely across US States.

- a) Identify and evaluate the main economic and social factors giving rise to this outcome.
- b) On the basis of your answer to a), suggest how high vaccination rates might be achieved across the entire United States.

(1977 words)

# 1 Introduction

By the end of 2021, only 2% of people wanted a vaccine and had not yet got it (KFF, 2022). Differences in vaccine uptake is hence likely due to personal choice. Existing research has highlighted correlations in vaccine uptake by factors such as race and party affiliation (Viswanath et al., 2021). Causal studies have focused on the effect of media consumption (Pinna, Picard, and Goessmann, 2021), finding that Fox News consumption causally reduces vaccine uptake.

We aim to explain cross-country variation in uptake based on an implicit model of agents forming beliefs on the costs and benefits of vaccination, and then aim to test the impact of financial incentives, to suggest future policy interventions to increase uptake across the US.

## 2 Data and Methods

### 2.1 Data

Regressions were generally done on the county level. Data on vaccination rates across time by county (and by state) were all sourced from the CDC (CDC, 2022). A dataset of time-invariant characteristics of each county was also compiled for 3075 counties in the United States. Data sources and details on excluded counties are reported in Table 1.

### 2.2 Methods

The decision on vaccination can reasonably be broken down into its costs and benefits, and beliefs that agents have about them. We expect that in areas of lower population density, or where COVID would generally spread less quickly, the benefit of vaccination would be lower. As COVID-19’s fatality rate rises sharply with age (Levin et al., 2020), we would expect more vaccination in counties with more elderly (who benefit more). As beliefs on vaccination’s benefits and supposed harms seem to be affected by religious

Table 1: Data Sources

Data	Source	Year
2020 Election Results	MIT Election Data and Science Lab’s Election Returns Dataverse (Lab, 2022)	2020
Religious Affiliation	Public Religion Research Institute 2020 Census of American Religion (PRRI, 2020)	2020
Racial Distribution	2020 Decennial Redistricting Dataset, maintained by the US Census Bureau (‘the Census’) (Census, 2020)	2020
Age Distribution	County Characteristics Resident Population Estimates, maintained by the Census (Census, 2019)	2019
Educational attainment	2015-2019 American Community Survey, maintained by the Census (ERS, 2019b)	Averaged 2015-2019
Poverty rates	As defined by the Census, from the Small Area Income and Poverty Estimates, maintained by the Census (ERS, 2019b)	2019
Rurality	The 2013 Rural-Urban Continuum Code, which classifies how rural counties are based on their level of urbanization and proximity to metropolitan areas (documented at ERS (2019a)), from the same source as above	2013
Income and Cost of Living	Median family income and costs of living obtained from the Economic Policy Institutes Family Budget Calculator (EPI, 2020). Costs of living measure allows us to correct nominal median family income to <i>real</i> income instead.	2020
COVID-19 Cases by 1 December 2020	The CSSE COVID-19 Data Repository maintained by John Hopkins (CSSEGIS, 2022).	2020
County Adjacency	Mapping of every county in the US to every county it borders, obtained from a dataset maintained by the National Bureau of Economic Research (NBER, 2010), originating from a dataset maintained by the Census	N/A

Notes: All datasets tagged each county with their Federal Information Processing Standard Publication (FIPS) codes, which was used to merge them together. All US Territories excluded from the dataset as they lack Presidential voting records. All counties in Alaska and Hawaii excluded due to missing data and different reporting standards (many statistics not reported by county, but other organizational units). Utah’s data excluded due to spotty county-level COVID-19 case data in 2020. Outside of these states, the following counties still assigned FIPS codes also had missing records: Bedford and Clifton Forge City, as both were merged into other counties and were non-existent by this time period, Dukes, Nantucket and Barnstable Counties in Massachusetts, which had missing vaccination records, Ogala Lakota County in North Dakota, a Native American reservation which remains unorganized and generally lacks data, and Yellowstone National Park, which has no resident population. Final dataset of time-invariant characteristics was exactly 3075 counties. Note that, while not all data is available for 2020, and so some are from shortly before, we expect these time invariant characteristics to not dramatically change year-to-year in each county. As explained later, the regressions using these variables do not interpret their coefficients causally given omitted variable bias, but treat them as best linear predictors.

status, race, and political identity, we also expect these variables to be explanatory.

We first run the following regression on the cross-sectional dataset of 3075 counties:

$$\begin{aligned}
\text{vaccuptake}_i = & \beta_0 + \beta_1 \text{repvotes}_i + \beta_2 \text{whiteevangelical}_i + \beta_3 \text{catholic}_i \\
& + \beta_4 \text{black}_i + \beta_5 \text{poverty}_i + \beta_6 \ln(\text{medincome}_i) + \beta_7 \ln(\text{col}_i) \\
& + \beta_8 \text{pop60to79}_i + \beta_9 \text{above80}_i + \beta_{10} \text{fullcollege}_i + \beta_{11} \text{casespc}_i \\
& + \boldsymbol{\delta} \cdot \text{rural}_i + \boldsymbol{\gamma} \cdot \text{state}_i + u_i
\end{aligned} \tag{1}$$

Where  $\text{vaccuptake}_i$  is a measure of vaccine uptake for a county  $i$ . Definitions for each covariate are listed in Table 2. The model was estimated for a cross-section of vaccine uptake by county as of 31 December 2021, and also for vaccine uptake as of 1 June 2021, to see if the effects of the covariates on uptake had changed over time. We measured vaccine uptake as the percentage of county residents who had taken at least two doses of the vaccine, and also alternatively looked at the percentage of county residents who had taken at least one dose of the vaccine, to check similarity of coefficients.

As different covariates have different units and standard deviations, interpreting their relative effect sizes (which variables are "more important" in explaining variation in vaccine uptake) can be difficult. Hence we estimated the *beta coefficients* of the model. These are calculated by using the z-score of every variable (defined for a covariate  $x_j$ , with a sample standard deviation  $\hat{\sigma}_j$ , as  $z_j := \frac{x_{ij} - \bar{x}_j}{\hat{\sigma}_j}$ ) and noticing then that, if Model (1) holds, then it is also true that

$$z_y = \frac{y_i - \bar{y}}{\hat{\sigma}_y} = \sum_{j=1}^k \frac{\hat{\sigma}_j}{\hat{\sigma}_y} \beta_j z_{ij} \tag{2}$$

Hence, we can calculate, for each covariate  $x_j$ , a beta coefficient  $\hat{b}_j := \frac{\hat{\sigma}_j}{\hat{\sigma}_y} \beta_j$ , interpreted as the standard deviation change in  $y_i = \text{vaccuptake}_i$  from a one standard deviation change in covariate  $x_j$ . This standardizes the units across the covariates and makes their corresponding effect sizes more directly comparable (although, not perfectly so, as some covariates are drawn from more-spread-out distributions).

We also more explicitly estimate any trends in the covariates over time, by estimating

Table 2: Variable Definitions

	Definition	Justification
repvotes	Percentage (0-100) of county vote that Donald Trump received in the 2020 election	Aim to measure effect of political affiliation, which affects news consumption and beliefs on the costs and benefits of vaccination
whiteevangelical	Percentage (0-100) of county identifying as White Evangelical Protestants as of 2020	Aim to measure effect of religion - White Evangelicals are a heavily conservative and Republican-supporting religious group. (Bokemper, Gerber, Omer, and Huber, 2021)
catholic	Percentage (0-100) of county identifying as Catholics as of 202	Aim to measure the effect of religion - the Pope has actively extolled the benefits of vaccination (Gawel et al., 2021), check effect.
black	Percentage (0-100) of county who are African-American as of 2020	Due to historical grievances and mistrust of government, African-Americans are known to be less likely to choose to vaccinate. (Burki, 2021)
poverty	Percentage (0-100) of county under the poverty line as of 2019(as defined by the Census)	Those in poverty may have difficulty finding access to vaccines, or more broadly, access to good-quality information on their costs and benefits.
medincome	Nominal median family income for a family of four, as of 2020	Same reason as above. Poverty variable lets us see if effect is stronger at the low end of the income distribution.
col	Cost of living as of 2020, as measured by the EPI.	Allows coefficient on medincome to be interpreted as the effect of changes in real median income (holding prices/cost of living fixed)
pop60to79	Percentage (0-100) of county aged between 60 and 79 in 2019.	Elderly are at greater risk of death and serious illness from COVID-19
above80	Percentage (0-100) of county aged 80 and above in 2019	See above
fullcollege	Percentage (0-100) of county who completed at least 4 years of college, averaged 2015-2019	Those with college degrees may be less susceptible to misinformation on the supposed harms of vaccination
casespc	Total recorded COVID-19 cases as of 1 December 2020, before the start of the vaccination program.	A proxy for how amenable the area is to COVID-19 spread, which would increase the benefit of vaccination against COVID-19
<b>rural</b>	Vector of dummy variables for each possible score on the 2013 Rural-Urban Continuum Code	More rural areas typically have lower population density and hence less risk of catching COVID-19, and so there are lower benefits to vaccination.
<b>state</b>	Vector of state fixed effects	Allows for variation in state policy responses that affected COVID-19 vaccination rates

the model:

$$\text{vaccuptake}_{it} = \beta_0 + \boldsymbol{\alpha} \cdot \mathbf{x}_i + \boldsymbol{\beta} \cdot \mathbf{x}_i \cdot t + \boldsymbol{\gamma} \cdot \mathbf{x}_i \cdot t^2 + u_i \quad (3)$$

on a panel dataset of observations of vaccine uptake at the start of every month (from May 2021 to March 2021), where  $\mathbf{x}_i$  is a vector of covariates (including the state fixed effects), using clustered standard errors by county, so as to correct for serial correlation across time in each county. For any covariate  $x_j$ , we have that

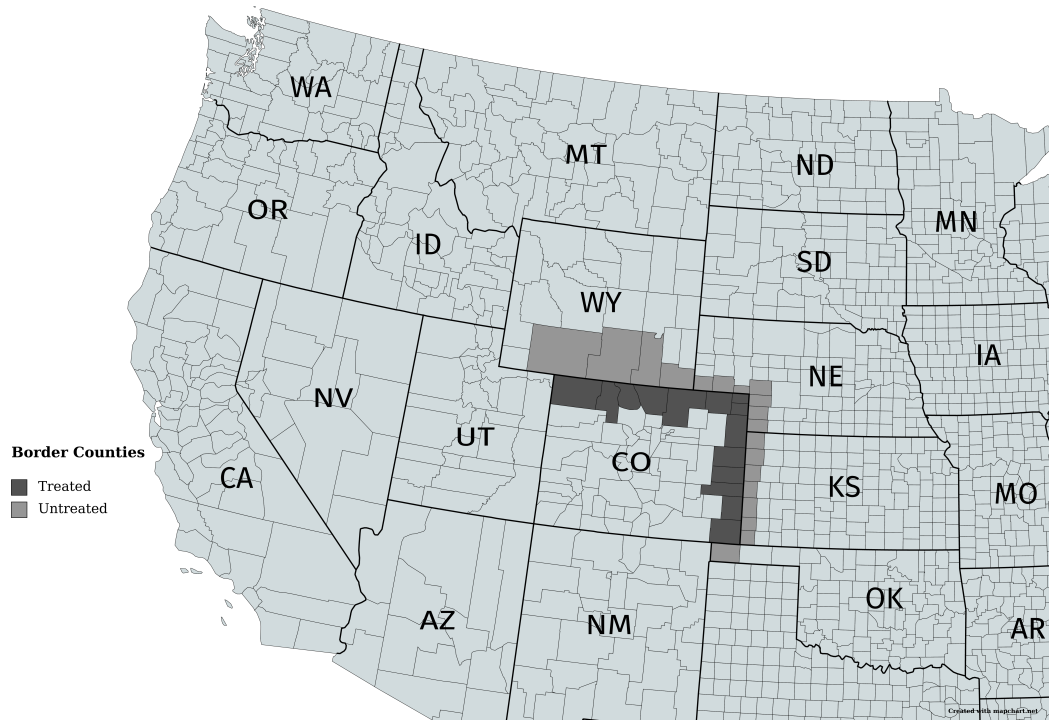
$$\frac{\partial \text{vaccineuptake}_{it}}{\partial x_{ij}} = \alpha_j + \beta_j t + \gamma_j t^2 \quad (4)$$

Hence,  $\beta_j$  and  $\gamma_j$  measure a quadratic trend in the effect-size of the covariate on vaccine uptake over time.

We cannot interpret any of the coefficients estimated above as estimates of the *causal impact* of the covariates on vaccine uptake. That would assume exogeneity (and strict exogeneity in the panel Model (4)), which is unlikely to be satisfied; for example, we cannot conclude, if the coefficient on  $\ln(\text{medincome})$  is positive, that increasing incomes would increase vaccine uptake. Higher incomes may simply correlate with being more informed, an unobserved variable, and so being less susceptible to vaccine misinformations. This would lead to omitted variable bias, biasing the coefficient on  $\ln(\text{medincome})$  upwards. Coefficients simply reflect correlations in the data, which are at best suggestive.

To answer part (b), however, we need to find out what *causes* vaccine uptake to increase. We hence turn our attention to a policy many states attempted in 2021: vaccine lotteries. We will focus in particular on Colorado’s ”Colorado Comeback Cash” program, where all those who had received at least one dose of the vaccine were eligible for a weekly \$1 million dollar lottery (Murphy, 2021). The program began on May 25, and ran to June 30 (Thirumurthy, Milkman, Volpp, Buttenheim, and Pope, 2022).

To evaluate the effect of the vaccine lottery, we use a design inspired by Dube, Lester, and Reich (2010). We pick counties in Colorado which border counties in states which did not have vaccine lotteries across this period (see Figure 1). We reason that these counties are similar to counties they share a border with - most things which affect one county in any period  $t$  will probably also affect a paired county across the state border. Table 3 compares the selected border counties on some selected covariates; within each



Oklahoma, Kansas, Nebraska and Wyoming did not have vaccine lotteries over this period (Thirumurthy, Milkman, Volpp, Bottenheim, and Pope, 2022). New Mexico did, and hence counties bordering New Mexico are excluded from this regression. Utah is excluded from the dataset due to missing data on some covariates. Each county-pair in the pair design dataset consists of one county in Colorado (in dark grey), and one county in an untreated state (in light grey). County-pairs were matched using the NBER county adjacency dataset.

Figure 1: Counties in Pair Design

Table 3: Summary Statistics

	(1)	(2)	(3)	(4)
	Treated	Untreated	Pair Differences	Nationwide SD
repvotes	72.87	80.46	8.65	16.03
black	0.82	0.72	0.67	14.79
fullcollege	24.75	24.51	7.12	9.54
casespc	0.04	0.06	0.03	0.02
whiteevangelical	31.79	32.67	6.65	12.64
catholic	13.79	15.89	4.45	10.03
poverty	12.81	11.74	3.13	5.78
medinc	68923.79	69379.73	11087.17	16720.16
pop60to79	21.72	21.87	3.61	4.58
above80	5.37	5.98	1.75	1.50
$N$	14	18	62	3075

Columns 1 and 2 report the average value of the predictive covariates in the treated and untreated groups, corresponding to the Colorado border counties and their neighbors respectively (see Figure 1). Column 3 contains the average absolute difference in covariates between two counties along the Colorado border which are connected. Column 4 is the standard deviation of the covariates across the entire nationwide dataset. Selected covariates are those with the largest standardized effect sizes with respect to vaccine uptake from estimating Model 2, with results in Table 4.

county-pair, differences in covariate values are generally within one standard deviation of the covariate.

However, only one member of each pair (the county within Colorado) is affected by the lottery. We can hence use the paired non-treated county as a control to estimate the causal impact of the vaccine lottery. We construct a panel of consisting of every cross-border county-pair, from 10 days before the start of the lottery, to 10 days after. We then estimate the model:

$$\text{vaccuptake}_{ipt} = \alpha + \beta \text{treated}_{it} + \phi_i + \tau_{pt} + u_{ipt} \quad (5)$$

$i$  is a specific county,  $p$  is a county-pair it belongs to, and  $t$  is the day in the panel.  $\tau_{pt}$  is a pair-specific time fixed effect (shocks beside the lottery happen to both members of a pair), and  $\phi_i$  is a county fixed account (accounting for initial differences).  $\text{treated}_{it} = 1$  if the county  $i$  is in Colorado, and  $t$  is after 25 May.  $\beta$ , the effect of the vaccine lottery on vaccine uptake, is consistently estimated if  $\text{Cov}(\text{treated}_{it}, u_{itp}) = 0$  - ie, if no shocks systematically occurred to Colorado counties without affecting the paired counties, after 25 May. We also estimate an alternative model including a  $\gamma t \cdot \text{treated}_{it}$  term, allowing



the effect of the vaccine lottery to change linearly with time over the 10 days post-treatment.

Finally, to further check the robustness of this causal estimate, we utilize a Synthetic Control design, as detailed in (Abadie, 2021) and used in prior studies on Ohio’s vaccine lottery (Lang, Esbenshade, and Willer, 2022). We collapse our 3075 county level dataset into a state-level dataset, and exclude the 23 other states who had vaccine lotteries in this period (Thirumurthy et al., 2022). We construct a synthetic Colorado by taking a weighted average of other states to closely match Colorado on some predictive covariates (we use the results of estimating Model (2) to gauge which covariates are most important in explaining variation in vaccine uptake). We use these weights to see how vaccine uptake would have evolved in the synthetic Colorado (made of untreated states), and take the difference between the synthetic Colorado and the real Colorado’s vaccine uptake on day  $t$  as the causal impact of the vaccine lottery on day  $t$ , and test its significance.

### 3 Results and Discussion

Results from the estimation of Model (1) are presented in Table 4. Our estimates (not to be interpreted causally, but reflecting correlations within the data) match prior work and have the expected sign - Republicans and black individuals are less likely to get vaccinated, and this is reflected on the county level. Counties with older people, or more college graduates, are more likely to get vaccinated. Poverty and income have the expected signs, and are jointly significant (see caption), although individually insignificant likely due to multicollinearity. Places which had more cases per person in 2020 had higher vaccination rates in 2021, as hypothesized that the benefits of vaccination would be higher and more clear when this was true.

By comparing beta coefficients, we can identify that by far the factor associated with the biggest effect size on vaccine uptake is the republican vote share in the 2020 election, followed by the number of blacks. These effect sizes seem to grow over time, possibly suggesting an increasing hardening of attitudes towards the vaccine.

Catholicism was not associated with higher vaccination rates in June, but was by December. The estimation of Model (4), presented in Table 5 affirms this, as in either

Table 4: Cross-Section Regression

	December 31 Sample			June 1 Sample	
	(1) Two Doses	(2) Two Doses Standardized	(3) One Dose	(4) Two Doses	(5) Two Doses Standardized
repvotes	-0.52*** (0.03)	-0.65	-0.60*** (0.04)	-0.31*** (0.03)	-0.35
whiteevangelical	-0.03 (0.06)	-0.03	-0.05 (0.07)	-0.04 (0.05)	-0.03
catholic	0.11* (0.04)	0.09	0.14** (0.05)	-0.02 (0.04)	-0.02
black	-0.25*** (0.03)	-0.29	-0.27*** (0.04)	-0.16*** (0.03)	-0.17
poverty	-0.15 (0.08)	-0.07	-0.14 (0.10)	-0.13* (0.06)	-0.05
lmedincome	2.22 (2.96)	0.04	1.33 (3.46)	4.01 (2.28)	0.07
lcol	4.19 (3.41)	0.04	6.37 (4.19)	0.41 (3.01)	0.00
pop60to79	0.18* (0.08)	0.06	0.20* (0.10)	0.31*** (0.06)	0.10
above80	0.73*** (0.21)	0.08	0.43 (0.26)	0.53** (0.17)	0.06
fullcollege	0.10* (0.04)	0.07	0.11* (0.05)	0.11*** (0.03)	0.08
casespc	57.27*** (15.05)	0.11	55.28** (16.87)	52.05** (16.25)	0.09
Constant	2.87 (41.54)		4.21 (50.46)	-11.61 (33.07)	
rural Wald Statistic	0.72		1.59	1.08	
$N$	3075	3075	3075	3075	3075
$R^2$	0.67	0.67	0.69	0.80	0.80
White Test F-Stat	1.274		52.24***	5.16**	

Notes: The dependent variable in Columns 1 and 4 was the percentage of people in each county who had received at least two doses of the vaccine. The dependent variable in Column 3 was the percentage of people who had received at least one dose of the vaccine. Columns 1-3 consider vaccine up take as of 31 December 2021, whilst columns 4-5 consider vaccine uptake as of 1 June 2021. The one-dose regression is not repeated for 1 June for brevity, as its results are shown to be very similar to the two-dose regression in the 31 December sample. Beta coefficients for the two dose regressions are presented in Columns 2 and 5 without standard errors (their significance is the same as in Columns 1 and 4 respectively). The F statistic of the special variant of the White test for heteroskedasticity, ie estimating the model  $\hat{u}_i^2 = \beta_0 + \beta_1 \hat{y}_i + \hat{y}_i^2 + e_i$  and testing for overall significance, is reported at the bottom. As there is evidence of heteroskedasticity in 2 out of 3 regressions, heteroskedasticity-robust standard errors are reported in parantheses for **all** coefficients for convenience (robust standard errors are still consistent in the absense of heteroskedasticity). State fixed effects are excluded for brevity, and the effect of the rural dummies are excluded as they are jointly insignificant, and their F-statistics are reported instead. The regressors poverty, lmedincome and lcol are jointly significant, with Wald statisitics of 6.46 (p=0.00), 3.80 (p=0.01), and 12.34 (p=0.0) in Columns 1,3 and 4 respectively.  $R^2$  is relatively high; much of the variation in vaccine uptake explainable by these factors.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 5: Coefficient Change Over Time

	(1) Linear Trends	(2) Quadratic Trend
t · repvotes	-0.0323*** (0.00321)	-0.0489*** (0.00558)
t · whiteevangelical	0.00238 (0.00515)	0.0148 (0.00792)
t · catholic	0.0198*** (0.00415)	0.0131* (0.00651)
t · black	-0.0112*** (0.00266)	-0.00437 (0.00460)
t · poverty	-0.00860 (0.00704)	-0.0691*** (0.0118)
t · lmedincome	-0.223 (0.277)	-1.213** (0.401)
t · lcol	1.087*** (0.238)	2.204*** (0.466)
t · pop60to79	-0.0158* (0.00664)	-0.0426*** (0.0115)
t · above80	0.0135 (0.0190)	-0.00855 (0.0300)
t · fullcollege	-0.00415 (0.00414)	0.0173** (0.00643)
t · casespc	0.417 (1.271)	3.850* (1.787)
t <sup>2</sup> · repvotes		0.00139** (0.000496)
t <sup>2</sup> · whiteevangelical		-0.00104 (0.000745)
t <sup>2</sup> · catholic		0.000559 (0.000658)
t <sup>2</sup> · black		-0.000566 (0.000459)
t <sup>2</sup> · poverty		0.00504*** (0.00108)
t <sup>2</sup> · lmedincome		0.0825* (0.0413)
t <sup>2</sup> · lcol		-0.0931* (0.0398)
t <sup>2</sup> · pop60to79		0.00224* (0.000975)
t <sup>2</sup> · above80		0.00183 (0.00274)
t <sup>2</sup> · fullcollege		-0.00179** (0.000605)
t <sup>2</sup> · casespc		-0.286 (0.156)
N	33825	33825

Notes: Standard errors, clustered by county, in brackets. Dependent variable was percentage of people with at least two doses of the COVID-19 vaccine. **rural** omitted as it was insignificant in Table 1, state fixed effects and non-interactions omitted for brevity. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 6: Effect of Colorado Vaccine Lottery

	(1) Two Doses	(2) One Dose	(3) $\Delta$ Two Doses	(4) $\Delta$ One Dose
treated	0.0315 (0.116)	1.696 (1.814)	0.00712 (0.0239)	0.0332 (0.0187)
treated	0.0195 (0.167)	2.329 (2.453)	-0.00707 (0.0331)	0.0308 (0.0396)
$t \cdot$ treated	-0.00219 (0.0139)	0.115 (0.112)	-0.00258 (0.00655)	-0.000432 (0.00742)
$N$	1178	1178	1178	1065

Notes: Standard errors in parentheses, clustered by state and county-pair. Column 1 shows the estimated treatment effect on the percentage of people who received two doses of the vaccine, within seven days of the start of the treatment. Column 2 shows the estimated treatment effect on the percentage of people who received at least one dose of the vaccine. Columns 3 and 4 measure the effect of the treatment on the daily change in the percentage of people who received both doses and one dose of the vaccine respectively. County fixed effects and pair-time fixed effects are not reported, for brevity. The second section estimates the alternative model, with the treatment effect varying linearly over time.  $t = 0$  10 days from 25 May at the end of the panel, so the coefficient on treated is the treatment effect by the end of the panel. It is worth noting that, as would be expected, the effect of the lottery on the uptake of first doses is substantially higher the effect on the uptake of second doses, given that only the former enters you into the lottery and we consider only the first 10 days after the lottery began, before those who just got their first doses for the lottery would get a second dose. However, the effects are still not statistically different from 0. Note that column 4 has less observations, as due to very low variation in  $\Delta OneDose$ , with some counties seeing 0 changes in some days, there is some issue of perfect collinearity. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

specification there is an increase in the effect size of Catholicism over time. While we do not have an adequate design to demonstrate this, this may suggest that the Pope’s messaging on vaccination (see Gawel et al. (2021)) has successfully changed attitudes over time. This suggests one way to achieve high vaccination rates across the United States: engaging local religious leaders to get them to urge their worshippers to get vaccinated, and to educate them against COVID-19 vaccine related misinformation.

Table 5 also affirms that the Republican and black aversion to vaccination is only getting stronger with time (although, at a declining rate for Republicans, by the coefficient on  $t^2 \cdot \text{repvotes}$ ). This is again suggestive of polarization, and further highlights the need to break into the Republican and black social networks with pro-vaccine messaging.

Regardless of dependent variable chosen, there is no statistically significant effect of the Colorado vaccine lottery on vaccination rates. This conclusion is supported by our alternative method using a Synthetic Control. Across all days the vaccine lottery was in effect, the gap between Colorado’s actual vaccine uptake and the vaccine uptake in the synthetic control was not statistically different from 0, regardless of what measure of

vaccine uptake was used.

Our results here are consistent with the broader literature which generally finds that such lotteries had either no effect or very small effects on vaccination. If non-vaccination is largely driven by false beliefs on supposed extreme health dangers of the vaccine, or mistrust of the government, offering financial rewards may not overcome these perceived costs of vaccination.

## 4 Conclusion

Variation in vaccine uptake across the United States is explainable, largely, by variation in political and religious affiliations and race; We have also found that racial, religious and political polarization is only increasing in time; we argue that this is suggestive of the impact of social influence on vaccine uptake. While data limitations prevent us from directly testing the importance of social messaging, this is line with prior literature.

We were able to construct two designs to plausibly consistently estimate the treatment effect of the vaccine lottery program in Colorado, and find no significant effect. Given that financial incentives seem ineffective at raising vaccine uptake, focus needs to be on social messaging; the increasing vaccine uptake in counties with more Catholics over time may suggest that messaging from religious leaders can be effective.

Table 7: Synthetic Control Results

		(1) Two Doses	(2) One Dose	(3) $\Delta$ Two Doses	(4) $\Delta$ One Dose
Gap between real and synthetic Colorado by day					
18 May (pre-treatment)		2.41	2.35	0.00506	.000720
25 May		2.28 (0.462)	2.58 (0.423)	-0.00004367 (0.999)	.000292 (0.269)
1 June		2.17 (0.500)	2.94 0.346	-0.0000253 (0.923)	0.000432 (0.115)
8 June		2.35 (0.500)	3.01 (0.308)	0.000737 (0.653)	.0000947 (0.231)
15 June		2.56 (0.500)	3.02 (0.269)	0.000330 (0.538)	0.000111 (0.269)
22 June		2.76 (0.462)	3.11 (0.269)	0.000501 (0.538)	0.000151 (0.269)
29 June		2.93 (0.462)	3.20 (0.307)	0.000199 (0.538)	0.000133 (.346)
Covariates	Real Colorado				
repvotes	42.1	43.1	42.4	42.1	42.53
black	4.64	14.5	16.2	16.9	16.6
fullcollege	41.0	37.3	38.4	37.9	38.1
casespc	0.0402	0.0329	0.0359	0.0354	0.0358
whiteevangelical	16.4	13.5	13.3	13.7	13.6
catholic	18.2	22.2	20.01	20.5	19.9
poverty	9.43	11.2	11.4	11.5	11.4
medincome	93200	92300	93600	92300	92700
pop0to4	5.75	5.95	6.03	6.04	6.03
pop5to9	6.01	6.05	6.03	6.07	6.05
pop10to14	6.31	6.07	5.95	6.02	5.99
pop15to19	6.35	6.37	6.22	6.30	6.25
pop60to64	6.05	6.19	6.12	6.05	6.10
pop65to69	5.16	5.10	5.09	5.01	5.06
pop70to74	3.93	3.98	3.94	3.90	3.93
pop75to79	2.47	2.64	2.61	2.49	2.60
pop80to84	1.54	1.67	1.66	1.64	1.65
above85	1.57	1.79	1.79	1.76	1.78
State Weight					
Washington, DC		0.172	0.246	0.218	0.232
Georgia				0.035	0.029
New Hampshire		0.341	0.254	0.264	0.251
Texas		0.403	0.304	0.358	0.311
Virginia		0.007			
Wyoming		0.078	0.196	0.124	0.176

Notes: **p-values**, not standard errors, in parantheses. Estimates for the gap between Synthetic (untreated) Colorado and real Colorado were available everyday from 15 April to 30 June, but only select few were reported for brevity. All post-treatment gaps are statistically insignificant from 0. Covariates of real and each synthetic Colorado (constructed for each possible dependent variable) in the second section of the table. Covariates were selected based on what Table 1 revealed to be most predictive - with finer controls for age group; a version of Model 1 was estimated with finer age controls and found the number of children and elderly to be highly significant, although this was omitted for brevity. Synthetic Colorado was constructed, ultimately, using only a few states as weights, in all designs, and their weights are listed in the third section of the table. Exact estimation done using method explained in Wiltshire (2022).

# References

- Abadie, A. (2021, June 1). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2), 391–425. doi:10.1257/jel.20191450
- Bokemper, S. E., Gerber, A. S., Omer, S. B., & Huber, G. A. (2021, December 7). Persuading US white evangelicals to vaccinate for COVID-19: Testing message effectiveness in fall 2020 and spring 2021. *Proceedings of the National Academy of Sciences*, 118(49), e2114762118. Publisher: Proceedings of the National Academy of Sciences. doi:10.1073/pnas.2114762118
- Burki, T. (2021, November 1). Increasing COVID-19 vaccine uptake in black americans. *The Lancet Infectious Diseases*, 21(11), 1500–1501. Publisher: Elsevier. doi:10.1016/S1473-3099(21)00637-X
- CDC. (2022). COVID-19 vaccinations in the united states,county — data — centers for disease control and prevention [Data.CDC.gov]. Retrieved May 17, 2022, from <https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh>
- Census. (2019). County population by characteristics: 2010-2019 [Census.gov]. Retrieved May 16, 2022, from <https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html>
- Census. (2020). P1: RACE - census bureau table. Retrieved May 17, 2022, from <https://data.census.gov/cedsci/table?q=P1%3A%20RACE%20county&tid=DECENNIALPL2020.P1>
- CSSEGIS. (2022, May 16). COVID-19 data repository by the center for systems science and engineering (CSSE) at johns hopkins university. Retrieved May 16, 2022, from <https://github.com/CSSEGISandData/COVID-19>
- Dube, A., Lester, T. W., & Reich, M. (2010, November 1). Minimum wage effects across state borders: Estimates using contiguous counties. *The Review of Economics and Statistics*, 92(4), 945–964. doi:10.1162/REST\_a.00039
- EPI. (2020). Family budget map [Economic policy institute]. Retrieved May 16, 2022, from <https://www.epi.org/resources/budget/budget-map/>
- ERS, U. (2019a). USDA ERS - documentation. Retrieved May 16, 2022, from <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/>

- ERS, U. (2019b). USDA ERS - download data. Retrieved May 16, 2022, from <https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/>
- Gaweł, A., Mańdziuk, M., Żmudziński, M., Gosek, M., Krawczyk-Suszek, M., Pisarski, M., ... Cyganik, W. (2021, December 16). Effects of pope francis' religious authority and media coverage on twitter user's attitudes toward COVID-19 vaccination. *Vaccines*, 9(12), 1487. doi:10.3390/vaccines9121487
- KFF. (2022, May 4). KFF COVID-19 vaccine monitor dashboard [KFF]. Retrieved May 17, 2022, from <https://www.kff.org/coronavirus-covid-19/dashboard/kff-covid-19-vaccine-monitor-dashboard/>
- Lab, M. E. D. Š. (2022, March 18). County presidential election returns 2000-2020. Type: dataset. Harvard Dataverse. doi:10.7910/DVN/VOQCHQ
- Lang, D., Esbenshade, L., & Willer, R. (2022, February 7). Did ohio's vaccine lottery increase vaccination rates? a pre-registered, synthetic control study. *Journal of Experimental Political Science*, 1–19. Publisher: Cambridge University Press. doi:10.1017/XPS.2021.32
- Levin, A. T., Hanage, W. P., Owusu-Boaitey, N., Cochran, K. B., Walsh, S. P., & Meyerowitz-Katz, G. (2020, December 1). Assessing the age specificity of infection fatality rates for COVID-19: Systematic review, meta-analysis, and public policy implications. *European Journal of Epidemiology*, 35(12), 1123–1138. doi:10.1007/s10654-020-00698-1
- Murphy, C. (2021, May 17). Colorado vaccine drawing isn't a lottery, and you might already be entered — colorado public radio. Retrieved May 17, 2022, from <https://www.cpr.org/2021/05/25/colorado-vaccine-lottery-no-its-a-drawing-how-to-enter/>
- NBER. (2010). County adjacency. National Bureau of Economic Research. Retrieved May 16, 2022, from <https://www.nber.org/research/data/county-adjacency>
- Pinna, M., Picard, L., & Goessmann, C. (2021, July 20). *Cable news and COVID-19 vaccine compliance* (SSRN Scholarly Paper No. 3890340). Social Science Research Network. Rochester, NY. doi:10.2139/ssrn.3890340
- PRRI. (2020). 2020 census of american religion [Public religion research institute]. Retrieved May 16, 2022, from <https://www.prri.org/research/2020-census-of-american-religion/>



- Thirumurthy, H., Milkman, K. L., Volpp, K. G., Buttenheim, A. M., & Pope, D. G. (2022, March 30). Association between statewide financial incentive programs and COVID-19 vaccination rates. *PLOS ONE*, *17*(3), e0263425. doi:10.1371/journal.pone.0263425
- Viswanath, K., Bekalu, M., Dhawan, D., Pinnamaneni, R., Lang, J., & McLoud, R. (2021, April 28). Individual and social determinants of COVID-19 vaccine uptake. *BMC Public Health*, *21*(1), 818. doi:10.1186/s12889-021-10862-1
- Wiltshire, J. C. (2022). Allsynth: (stacked) synthetic control bias-correction utilities for stata, 40.