

# Using machine learning to identify COVID-19 vaccine-hesitancy predictors in the USA

Enrique M Saldarriaga 

**To cite:** Saldarriaga EM. Using machine learning to identify COVID-19 vaccine-hesitancy predictors in the USA. *BMJ Public Health* 2023;1:e000456. doi:10.1136/bmjph-2023-000456

Received 31 July 2023  
Accepted 16 November 2023

## ABSTRACT

**Introduction** Vaccine hesitancy is complex, multicausative phenomenon that undermines public health efforts to contain the spread of infectious diseases. Improving our understanding of the drivers of vaccine hesitancy might improve our capacity to address it.

**Methods** We used the results of the May 2021 Assistant Secretary for Planning and Evaluation's survey on COVID-19 vaccine hesitancy, which estimated the proportion of adults for every US county that felt either hesitant or unsure and strongly hesitant towards taking the COVID-19 vaccine when it becomes available. We developed a prediction model to identify the most important predictors of vaccine hesitancy. The potential predictors included demographic characteristics, the Centers for Disease Control and Prevention's Social Vulnerability Index and the Republican Party's voting share in the 2020 presidential election as a proxy of political affiliation, all at the county level.

**Results** We found that the main drivers of vaccine hesitancy are income level, marital status, poverty, income, schooling, race/ethnicity, age, health insurance status and political affiliation. While the drivers are shared by both levels of hesitancy, the order changes. Particularly, political affiliation is a more important predictor for strong hesitancy than for hesitancy or unsure.

**Conclusions** These results deepen our understanding of the phenomenon and could help design more targeted interventions to reduce hesitancy in specific subgroups of the population.

## INTRODUCTION

The COVID-19 pandemic has caused over 6 million hospitalisations and over 1 million deaths in the USA.<sup>1</sup> The COVID-19 vaccine was developed in record time<sup>2</sup> and rapidly became the most effective measure to control the spread of the virus and mitigate the severity of COVID-19 illness.<sup>3</sup> Despite the efforts of federal agencies, namely the Food and Drug Administrations and the Centers for Disease Control and Prevention (CDC), to communicate to the public the proven safety and efficacy of the vaccines,<sup>4</sup> vaccine hesitancy has emerged as a major challenge in achieving global vaccination coverage

## WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Vaccine hesitancy is a multicausative, complex, public health challenge that affects the effectiveness of vaccination campaigns across the USA. Previous studies used surveys to link vaccines hesitancy to demographic characteristics.

## WHAT THIS STUDY ADDS

⇒ We used a machine learning algorithm to identify the main predictors of COVID-19 vaccine hesitancy at the county level. We used two hesitancy intensities, hesitant or unsure and strongly hesitant. At both levels, higher proportions of non-married, low-income, unemployed, lower-schooling, uninsured, Republican-affiliated people and higher levels of economic inequality were associated with higher levels of hesitancy.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Future vaccination programme can use our results to identify counties with greater likelihood of hesitancy and use outreach and promotion instruments designed to overcome these vaccination uptake barriers.

and controlling the pandemic.<sup>5</sup> Thus, the efforts of local health departments across the country to administering vaccines were negatively affected.

Vaccine hesitancy refers to the delay in acceptance or refusal of vaccination despite the availability of vaccination services. This is a complex, multicausal phenomena that has arisen as a major public health concern during the last decade.<sup>6</sup> The reasons for vaccine hesitancy include factors related to individual beliefs, attitudes and behaviours, as well as contextual factors such as social, cultural and political factors.<sup>7</sup> Socioeconomic and demographic factors have been identified as important predictors of vaccine hesitancy. Studies have shown that individuals with lower education, income and social status are more likely to be vaccine hesitant.<sup>8,9</sup>



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. Published by BMJ.

### Correspondence to

Dr Enrique M Saldarriaga, University of Washington, Seattle, Washington, USA; emsb@uw.edu

Other demographic factors such as age, gender and race/ethnicity have also been found to be associated with vaccine hesitancy. However, the relationship between socioeconomic and demographic factors and vaccine hesitancy may vary across different populations and contexts.<sup>8 10</sup> In the specific context of the COVID-19 pandemic, misinformation and conspiracy beliefs had played an important role in creating vaccine hesitancy.<sup>11</sup>

Understanding the predictors of vaccine hesitancy is crucial for designing future vaccination campaigns and policies to effectively overcome vaccine hesitancy and increase vaccination coverage. Especially because the societal polarisation around the relevance and safety of public health activities during the COVID-19 pandemic might have spilled over to other diseases.<sup>12</sup> The goal of this study is to identify the most important predictors of COVID-19 vaccine hesitancy in the USA from an ecological perspective through the employment of machine learning methods to increase the accuracy of our predictions. The findings of this study can help identify the population groups with higher likelihood of being reluctant to accept vaccines and furthermore aid in the designing of targeted interventions and policies to promote vaccination uptake.

## MATERIALS AND METHODS

### Study population

A cross-sectional survey was conducted between 26 May 2021 and 7 June 2021 to measure levels of hesitancy towards the COVID-19 vaccine; this was a section within the US Census' Bureau Household Pulse Survey.<sup>13</sup> The survey had a multiple-choice question 'Once a vaccine to prevent COVID-19 is available to you, would you \_\_\_ get a vaccine?', with five possible answers: 'definitely get a vaccine'; 'probably get a vaccine'; 'unsure'; 'probably not get a vaccine'; 'definitely not get a vaccine'. At the time of the survey, the vaccine was under development and would not be approved until 23 August 2021.<sup>14</sup>

The US Department of Health and Human Services' Office of the Assistant Secretary for Planning and Evaluation (ASPE) used the household-level survey results, to create estimates at the county level of the proportion of people within a county that fell into each of the mutually exclusive categories of hesitancy.<sup>15</sup> For the analysis, we classify the proportion that felt strongly against taking the vaccine as strongly hesitant; and the proportion that answered 'unsure' and 'probably not get a vaccine' as hesitant or unsure. Thus, we obtained two metrics of vaccine hesitancy intensity, for all 3142 counties in the USA and the District of Columbia.

### Measurement of demographic information

We used the American Community Survey 5 years to obtain county-level demographic information.<sup>16</sup> We selected variables with observed association with vaccine hesitancy in the literature.<sup>10 11</sup> The final prediction set included the following categories: race and ethnicity,

marital status, age, biological sex, poverty level respect to federal poverty line, income, internal immigration (people who moved to the county in the last 12 months from within the USA), foreign immigration (people who moved to the county in the last 12 months from outside the USA), schooling, urbanicity, employment and the county's Gini coefficient as a measure of inequality. In addition, we included the CDC's Social Vulnerability Index (SVI), a relative measure of an areas' risk to a hazardous event of any kind.<sup>17</sup>

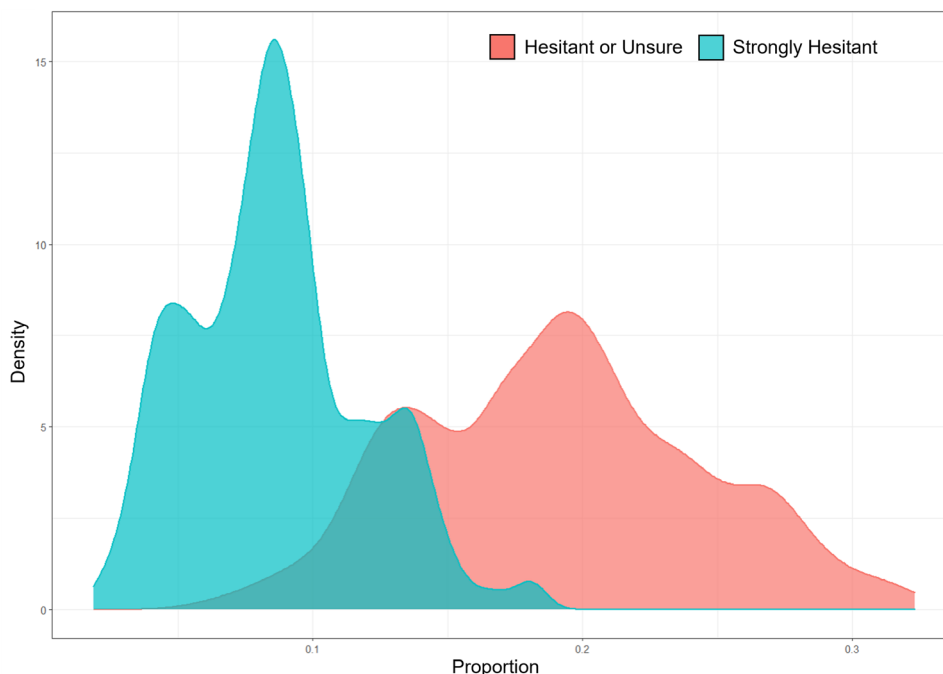
### Political affiliation

Misinformation plays an important role in the generation of vaccine hesitant feelings.<sup>18 19</sup> It has been documented that people whose political preferences lean towards the US Republican Party were more prone to consume wrongful and misleading information about the safety, efficacy and importance of the COVID-19 Vaccine.<sup>20–22</sup> Using a web-based survey, Sargent *et al* found that democratic-leaning people were 2.4 (95% CI 2.2 to 2.7) times more likely to be vaccinated and 1.8 (95% CI 1.5 to 2.2.) times more likely to be receptive towards vaccination, than their Republican-leaning peers.<sup>23</sup> To account for the differences in vaccine hesitancy associated with political affiliation, we included the county-level proportion of votes that the Republican Party obtained in the 2020 Presidential Election. We obtained the data from the MIT Sloan School of Management's repository.<sup>24</sup>

### Statistical analysis

We followed a machine learning algorithm to identify the most important predictors of vaccine hesitancy expressed in two outcomes, hesitant or unsure and strongly hesitant. We used the least absolute shrinkage and selection operator (LASSO) regression to identify the most important predictors for each outcome. LASSO is a type of regularised regression that introduces a penalty in the estimation of a variable's coefficient for collinearity and low explanatory power over the dependent variable.<sup>25</sup> This approach allows us to introduce variables that are seemingly correlated (eg, income and schooling) in the predictor set, and let the model identify their explanatory capacity conditional to all variables included. We used the population of each county as weights for the regression, such that counties with relatively larger populations are represented accordingly in the analytical dataset.

The final data set included only counties with complete information for all potential predictors. We split the sample into a training set equivalent to 80% of the observations randomly selected, and a test set, equivalent to 20%. Thus, we can assess the prediction accuracy of the model out-of-sample. To train the model, we used leave-one-out cross-validation to estimate the optimum penalty parameter (lambda), with the mean absolute error (MAE) as the loss function to minimise. We used the coefficients from the best model to predict in the test set. We report the MAE in the test set for both outcomes.



**Figure 1** Density distribution of the proportion of people feeling hesitant or unsure and strongly hesitant to take the COVID-19 vaccine at the county level.

We show the importance of demographic characteristics in explaining vaccine hesitancy in two ways. First, we used exact postselection inference to estimate CIs via the sandwich SE estimation, to show the level of uncertainty associated with a variable's coefficient.<sup>26 27</sup> This information indicates whether changes in a demographic variable are expected to be associated to changes in the outcome. Second, we plot the values of the coefficients at every value of the log of the LASSO lambda. The higher the lambda, the greater the penalisation and, therefore, the fewer variables will enter the model, remaining only those with the highest explicative power.<sup>28</sup> This plot provides a visual representation of which variables are included at every level of penalisation, allowing for a qualitative description of the variables' importance. This plot shows the variable capacity to explain the variability in the outcome, despite the direction of its association.

All analyses were conducted separately for both vaccine hesitancy outcomes, hesitant or unsure and strongly hesitant. All data management and analyses were conducted in R software.

### Patient and public involvement

It was not appropriate or possible to involve patients or the public in the design, conduct, reporting or dissemination plans of our research.

### RESULTS

Out of 3142 counties, 2489 had complete information for all 32 potential predictors identified. Across counties, the proportion of people who reported feeling hesitant or unsure to take the COVID-19 vaccine ranged from 4.9% to 32.3%, with an average of 18.8%. The proportion

of people who felt strongly hesitant to take the vaccine varied from 1.9% to 18.2%, with a mean of 8.4%. Both outcomes had a multimodal distribution, where strongly hesitant has a greater concentration and a lower range, than hesitant or unsure (figure 1).

The models showed good prediction accuracy for both hesitancy outcomes. For the hesitant or unsure outcome, we found an MAE of 2.8% in the out-of-sample validation set. This means that our model had an average error of 2.8% when comparing the observed versus the predicted proportion of people hesitant or unsure. At the optimum penalisation (ie, the lambda that minimises the in-sample MAE), two variables were excluded from the model due to low explanatory power, namely the proportion of people with some college and proportion of unemployed males. We found that an increase of 1% in the proportion of divorcees in a county is associated with a 0.139% (95% CI 0.292% to 0.147%) increase in the proportion of people who reported feeling hesitant or unsure to take the COVID-19 vaccine (table 1, column B). Other variables associated with an increase in hesitancy were the proportions of black/African-American, other race/ethnicity, people with an annual income between 10 000 and 25 000, people below the federal-level poverty line, internal immigrants, people with a bachelor's degree, unemployed females and uninsured population, higher levels of inequality expressed in the Gini coefficient and a higher vote share for the Republican Party in the 2020 Presidential Election. On the other hand, an increase in the proportion of non-Hispanic Asian population was associated with a 0.349% (95% CI -0.462% to -0.250%) reduction in the proportion of people who reported feeling hesitant or unsure to take the COVID-19 vaccine.

**Table 1** Coefficients from the LASSO regression at the optimum level of penalisation for the hesitant or unsure and strongly hesitant outcomes

	Hesitant or unsure coefficient (% , 95% CI)	Strongly hesitant coefficient (% , 95% CI)
<b>Optimum penalisation (lambda)</b>	0.0001004471	0.0001229605
<b>Variables</b>		
Intercept	-0.061	-0.055
Vote share for the Republican Party	0.116 (0.1 to 0.133)*	0.064 (0.059 to 0.132)*
Race/ethnicity		
Non-Hispanic white	Reference	Reference
Hispanic	-0.053 (-0.068 to -0.039)*	-0.043 (-0.068 to -0.039)*
Non-Hispanic African American	0.137 (0.117 to 0.157)*	0.039 (0.016 to 0.156)*
Non-Hispanic Asian	-0.349 (-0.462 to -0.25)*	-0.218 (-0.461 to -0.2)*
Other	0.139 (0.105 to 0.187)*	0.1 (0.095 to 0.186)*
Marital status		
Never married	Reference	Reference
Married	-0.012 (-0.053 to 0.021)	0.011 (-0.053 to 0.019)
Widowed	-0.276 (-0.413 to -0.147)*	-0.171 (-0.41 to -0.144)*
Divorced	0.371 (0.292 to 0.451)*	0.224 (0.21 to 0.449)*
Age		
50-54	Reference	Reference
Under 18	0.067 (-0.001 to 0.166)	0.027 (0 to 0.163)
18-25	-0.032 (-0.113 to 0.037)	-0.053 (-0.112 to 0.031)
25-39	0.038 (-0.046 to 0.154)	0.015 (-0.038 to 0.151)
55-64	0.123 (-0.007 to 0.316)	0.088 (-0.002 to 0.306)
Over 65	-0.305 (-0.386 to -0.219)*	-0.172 (-0.381 to -0.16)*
Sex		
Female	Reference	Reference
Male	0.086 (-0.014 to 0.198)	0.039 (-0.011 to 0.193)
Poverty, respect to federal threshold		
Above 150%	Reference	Reference
100%-140%	0.046 (-0.017 to 0.124)	0.02 (-0.014 to 0.123)
Below 100%	0.103 (0.042 to 0.175)*	0.042 (0.04 to 0.175)*
Annual income		
Less10K	Reference	Reference
10-25K	0.145 (0.044 to 0.262)*	0.065 (0.042 to 0.256)*
25K-50K	0.016 (-0.064 to 0.149)	OMITTED
50K-65K	-0.155 (-0.31 to 0)	-0.085 (-0.31 to 0)
65+K	-0.134 (-0.226 to 0)	-0.06 (-0.224 to -0.001)*
Internal immigrants	0.176 (0.063 to 0.313)*	0.172 (0.061 to 0.312)*
Foreign immigrants	-0.124 (-0.416 to 0.111)	-0.079 (-0.412 to 0.104)
Schooling		
Less than high school	Reference	Reference
High school degree	0.028 (-0.011 to 0.075)	Omitted
Some college	Omitted	-0.018 (-0.047 to 0.033)
Bachelor's degree	0.139 (0.055 to 0.225)*	0.112 (0.052 to 0.221)*
Graduate degree	-0.013 (-0.116 to 0.043)	-0.027 (-0.113 to 0.034)

Continued



Table 1 Continued

	Hesitant or unsure coefficient (% , 95% CI)	Strongly hesitant coefficient (% , 95% CI)
Living outside city	-0.015 (-0.023 to -0.006)*	-0.009 (-0.023 to -0.005)*
Unemployed male	Omitted	Omitted
Unemployed female	0.069 (0.034 to 0.107)*	0.048 (0.033 to 0.105)*
Uninsured	0.05 (0.001 to 0.099)*	0.045 (0.001 to 0.098)*
Gini coefficient of inequality	0.117 (0.06 to 0.177)*	0.079 (0.059 to 0.176)*
Social Vulnerability Index	-0.025 (-0.037 to -0.014)*	-0.009 (-0.037 to -0.002)*

All variables are expressed in proportions of people with a given characteristic. Therefore, the coefficients represent the percentual increase or decrease in the outcomes associated with a 1% increase in the variables. All coefficients and CI values are rounded to the third decimal point.

Key. 'Reference' indicates variables not included in the analysis to avoid perfect collinearity because these are the complement of the included variables. For example, 'female' is not included because for any given county, this is equal to 100% - 'male'. Omitted: Variables excluded from the analysis at the optimum penalisation for lack of explanatory power over the outcome(s).

\*Indicates variables whose CIs exclude the reference value of 0 and are statistically significant at 95% of CI.

LASSO, least absolute shrinkage and selection operator.

Other variables associated with a decrease in hesitancy are the proportion of widowers, of people of Hispanic ethnicity, of people aged 65 years and over, of people living outside the city and higher levels of SVI (table 1, column B).

From the coefficients versus log of lambda plot, we observe that the most important predictor of the hesitant or unsure outcome is income because it has enough explanatory power over the outcome to generate a non-zero coefficient even at the highest penalty level (figure 2). As the penalisation becomes less stringent, other variables enter the model. The remaining top 10 variables include, ranked from highest to lowest: marital status, poverty, employment, schooling, race/ethnicity, political affiliation, age and health insurance coverage.

For the strongly hesitant outcome, we found an MAE of 2.2% in the out-of-sample validation set. At the optimum penalisation, three variables were excluded from the model, proportion of people with an annual income between US\$25 000 and US\$50 000, with a high school degree and proportion of unemployed males. We found that an increase of 1% in the proportion of divorcees in a county is associated with a 0.224% (95% CI 0.21 to 0.449) increase in the proportion of people who reported feeling strongly hesitant to take the COVID-19 vaccine (table 1, column C). Other variables associated with an increase in strong hesitancy were proportion of black/African-American, of other race/ethnicity, of people living below the federal poverty line, of having an annual income between 10000 and 25 000, of internal immigrants, of having a bachelor's degree, of unemployed females, of uninsured, higher levels of economic inequality measured by the Gini coefficient and a higher share of votes casted for the Republican Party in the 2020 Presidential Election. On the other hand, an increase in the proportion of non-white Asian population was associated with a 0.218% (95% CI -0.461% to -0.2%) reduction in the proportion of people who reported feeling

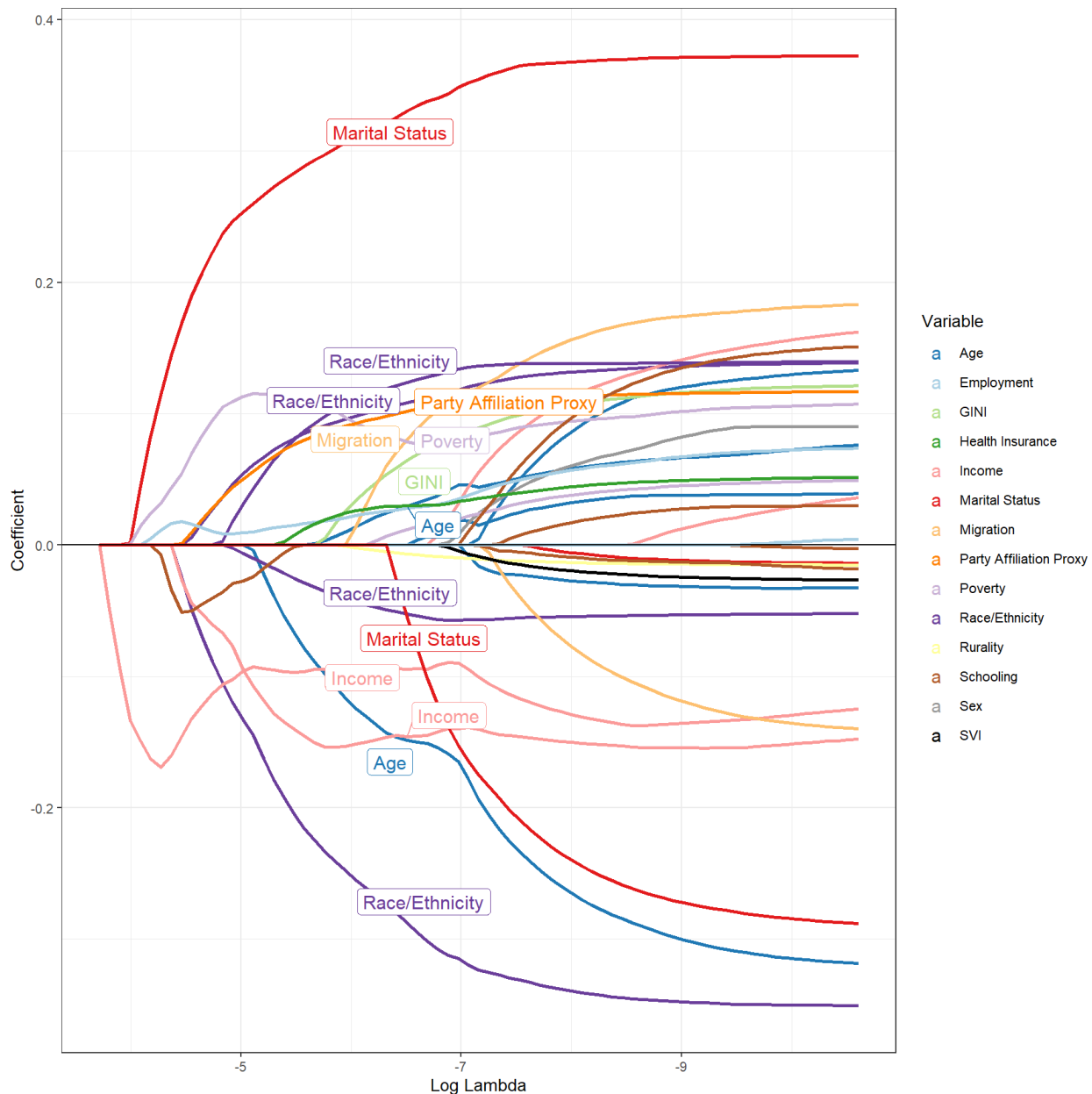
strongly hesitant to take the COVID-19 vaccine. Other variables associated with a decrease in hesitancy are the proportion of Hispanics, of widowers, of people aged 65 years and over, of people with an annual income of US\$65 000 or higher, of people living outside the city and higher levels of SVI (table 1, column C).

From the coefficients versus log of lambda plot, we observe that the most important predictors of the strongly hesitant outcome are income and marital status, which have non-zero coefficients at the highest penalty level (figure 3). The following eight most important predictors include, in order of highest to lowest: employment, political affiliation, poverty, race/ethnicity, health insurance, migration and age.

## DISCUSSION

We sought to identify the most important predictors of COVID-19 vaccine hesitancy in the USA from a population-level perspective. Hesitancy was measured in two intensities, hesitant or unsure and strongly hesitant.<sup>15</sup> We selected potential predictors based on the literature review and inclusive of socioeconomic and demographic characteristics, and political preferences. We trained a statistical model on a random selection of 80% of the 2489 counties with complete information, using a LASSO regression whose penalty parameter was tuned via leave-one-out cross validations. We validated the prediction accuracy of our models in the remaining 20% of the sample. We found that the model had an MEA of 2.82% for the hesitant or unsure outcome, and 2.21% for the strongly hesitant. The slight improvement in performance observed in the latter outcome might be due to a smoother distribution that more closely resembles a normal distribution than the hesitant or unsure counterpart (see figure 1).

We found great similitude in the characteristics associated to increases and decrements in vaccine hesitancy

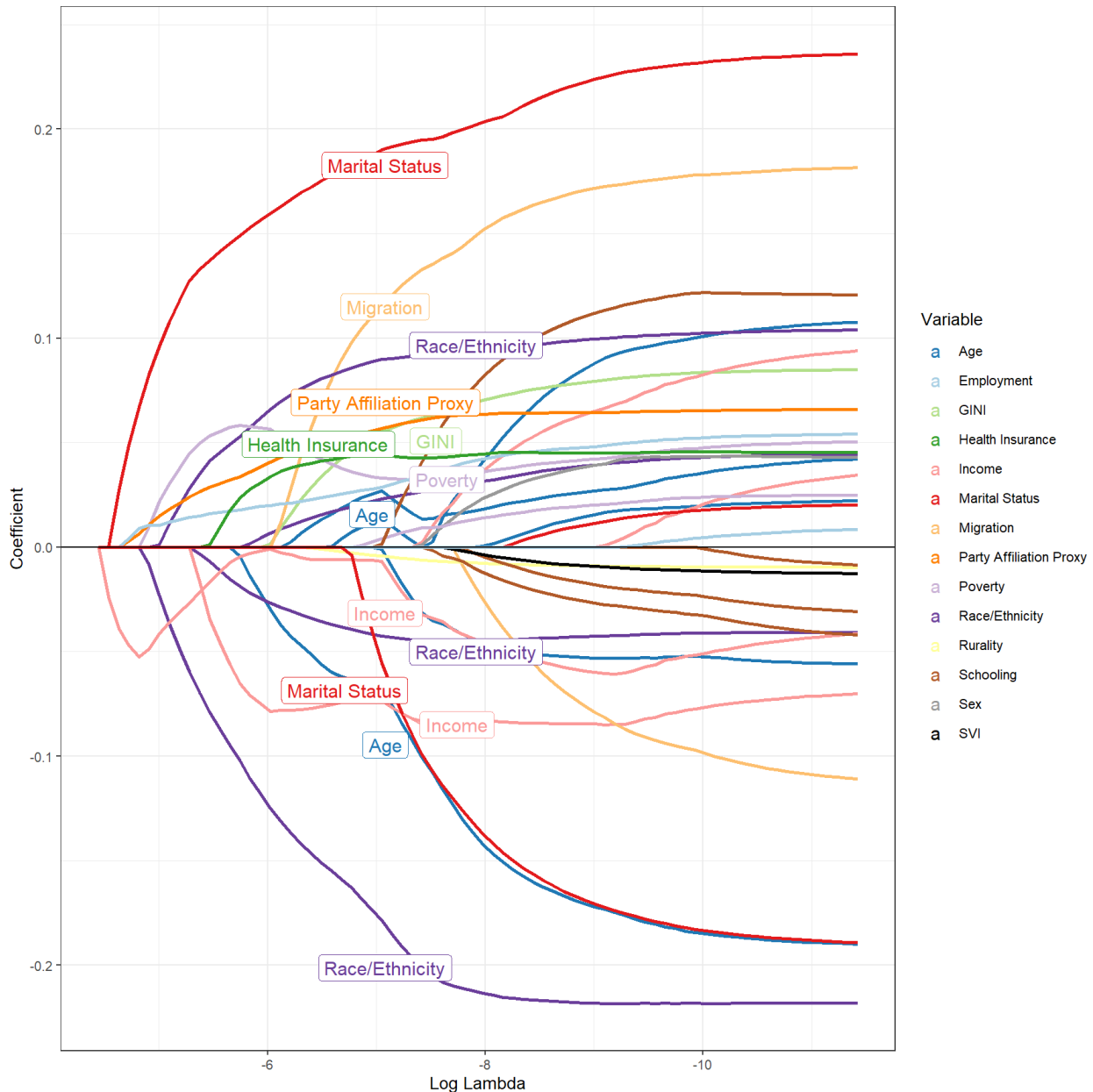


**Figure 2** Hesitant or unsure outcome. Trajectory of the coefficients, colour-coded and labelled by variable, at different levels of penalisation from the LASSO regression. Lines represent the coefficients of a specific metric (eg, non-Hispanic African American) while colour and label represent variable (eg, race/ethnicity). LASSO, least absolute shrinkage and selection operator; SVI, Social Vulnerability Index.

across both intensities studied. Greater proportions of people living in poverty, with lower income, and unemployment were associated with a higher proportion of people feeling hesitant or unsure and strongly hesitant towards the COVID-19 vaccines. These results are consistent with previous studies that found associations between lower-income individuals and greater vaccine hesitancy.<sup>11 29</sup> One interesting finding was that higher levels of economic inequality, measured by the Gini coefficient, are associated with greater vaccine hesitancy. While we found no evidence of previous studies reporting a direct association between this metric and hesitancy, the presence of larger economic disparities is consistent with higher proportions of people living under lower

economic conditions, which has been documented as a risk factor for vaccine hesitancy.<sup>6</sup>

Regarding race and ethnicity, we found a clear divide between the effects associated with black/African Americans on one hand, and non-Hispanic Asian and Hispanic on the other one. We found that increases in the proportion of the former group were associated with higher levels of hesitancy. The distrust of the African American community towards vaccination has been well documented across therapeutic areas, making it a deeply rooted phenomenon in the USA.<sup>30-32</sup> On the other hand, we found that increases in the proportion of non-Hispanic Asian and Hispanic populations are associated with reductions in vaccine hesitancy. Previous



**Figure 3** Strongly hesitant outcome. Trajectory of the coefficients, colour-coded and labelled by variable, at different levels of penalisation from the LASSO regression. Lines represent the coefficients of a specific metric (eg, non-Hispanic African American) while colour and label represent variable (eg, race/ethnicity). LASSO, least absolute shrinkage and selection operator; SVI, Social Vulnerability Index.

studies have found that Asians in the USA had higher levels of trust in the healthcare system than their racial/ethnic minorities counterparts.<sup>32</sup> However, the literature has mixed evidence regarding the level of trust among the Hispanic population. One study found higher level of vaccine hesitancy among Hispanic healthcare workers, compared with non-Hispanics whites,<sup>33</sup> while another one found no association after accounting for sociodemographic characteristics such as education and income,<sup>34</sup> and several others suggest that this population is very diverse to find effects above and beyond their educational and economical variability.<sup>30 35 36</sup> Our results suggest that after accounting for socioeconomic determinants (including schooling and in employment), higher

proportions of Hispanics in a county are associated with lower levels of hesitancy. However, these decrements are much lower than the magnitude of effect associated to Asians and African Americans.

We found no significant associations with sex in either of the hesitancy outcomes. However, we did find a direct effect associated with unemployment among females. This is consistent with previous studies that found that hesitancy is more prevalent among females than males.<sup>29</sup> Regarding marital status there was a clear divide. Higher levels of divorced people were associated with higher levels of hesitancy, and the opposite was true for higher levels of widowers. Previous studies have found that single people are more likely to be hesitant<sup>29</sup> and on the contrary,

married people tend to be less hesitant.<sup>34</sup> Studies have shown that people directly impacted by a disease tend to be less reluctant to take preventive measures to protect their health.<sup>34 37</sup> This could be the case of widowers in our study, especially considering that the ASPE survey took place in May 2021, after half a million of people have died in the USA.<sup>1</sup> Further, this could explain why our results indicate that a higher proportion of people aged 65 years and above is associated with lower levels of vaccine hesitancy. We did not find any other significant association with age. We found that higher levels of SVI, the CDC's metric of relative community stress, were associated with lower hesitancy, likely because they may have already experienced a greater burden of COVID-19 or other hazardous events.

Political affinity to the Republican Party, measured as the vote share of the party in the 2020 Presidential Elections, was associated to higher levels of vaccine hesitancy. One of the most important dimensions in the frameworks that explain the existence and persistence of distrust in vaccines and, more broadly, the healthcare system, is the consumption and believe of misinformation.<sup>38</sup> Further, previous studies have shown that people with low health literacy, prior beliefs regarding the authenticity of the COVID-19 pandemic as a health emergency, and conspiracy beliefs, are important predictors of COVID-19 vaccine hesitancy.<sup>11 29 39</sup> It has been documented how the information sources most frequently consulted and trusted by individuals with affinity for the Republican Party, contained untruthful and misleading information about the importance of the COVID-19 vaccine, and its safety and efficacy.<sup>40–42</sup> Previous studies have linked a higher affinity to the Republican Party with higher levels of vaccine hesitancy at an individual<sup>37</sup> and population<sup>43</sup> levels, as well as higher mortality rates during the COVID-19 pandemic among Republican-registered voters compared with Democratic counterparts.<sup>44</sup> In this study, we found that party affiliation was an important predictor of vaccine hesitancy in both outcomes, but while it ranked eighth for the hesitant or unsure, it ranked fourth for the strongly hesitant. Thus, the explanatory power of party affiliation increases as the intensity of the hesitancy does.

Vaccine hesitancy has long-lasting effects that hinder the effectiveness of measures to prevent and control the spreading of diseases due to its contribution to broader mistrust in public health interventions and authorities.<sup>45</sup> For example, the coverage of required vaccines among children in kinder garden dropped 1% for the 2020–2021 school year<sup>46</sup> and has not recovered 2 years after, according to the last estimation for the 2022–2023.<sup>47</sup> While there are potentially several causes for this decline, previous studies have shown that unvaccinated children more likely belong to families that intentionally refuse vaccination.<sup>48 49</sup> The success of vaccination programmes frequently relies on high coverage rates; therefore, vaccine hesitancy and subsequent refusal importantly reduce the possibility to achieve protection for individuals who are too young to be vaccinated, cannot receive the vaccine due to health conditions or do not develop a sufficient immunological response to the

vaccine.<sup>50 51</sup> Beyond the increased vulnerability to diseases, vaccine hesitancy increases the economic burden of vaccine-preventable diseases either through increases in out-of-pocket expenditures, lengthier public health campaigns, loss of productivity or labour force due to disability or death and overall social disruption.<sup>45 52</sup>

The possibility to overcome vaccine hesitancy and generalised distrust depends on the public health system's capacity to understand and acknowledge its association with systemic (racial, economic, education and healthcare access) inequity,<sup>53</sup> and design communication materials and delivery strategies that are tailored to the concerns of specific populations.<sup>54 55</sup> Our results can be leveraged to increase our understanding of the drivers of hesitancy but even more importantly, to create more detailed profiles of population groups at a higher risk of rejecting vaccines and other preventive measures.

This study has limitations. First, to preserve the consistency of the explanatory variables, we were limited in the diversity of information we included as potential predictors. While we believe all socioeconomic and demographic domains are well represented in the analysis, it would have been interesting to explore other religious, psychological, occupational, family structure related, caretaker status and participation in economic aid variables that have been associated with vaccine hesitancy.<sup>11</sup> Furthermore, the specific status of the COVID-19 epidemic at the moment of the survey might have influenced the respondents' perceptions, which would have warranted the addition of the case rate or other epidemiological data. However, the differences in reporting and data-sharing agreements across states could have introduced bias to the estimates. Second, the nature of our hesitancy estimates and our methodological design led us to quantify associations at a population level, which allowed us to better understand the drivers of vaccine hesitancy. However, because it is aggregated data, we could not observe the potential heterogeneity that occurs at more granular levels. Third, it is plausible that people with greater levels of distrust towards public health activities at large would have been reluctant to answer the survey, in which case the sample might be lacking some of the more extreme sentiments. It is unclear, however, how prevalent this self-selection bias could have been and if it could have been enough to bias our results.

## Conclusions

To our knowledge, this is the first study that aims to identify the most important predictors of COVID-19 vaccine hesitancy at a population level in the USA. We found that economic living conditions such as income and poverty, marital status, race/ethnicity and political party affiliation are the most important drivers of hesitancy. These results can help improve our understanding of the populations at higher risk of rejecting healthcare preventive measures. Therefore, they can be leveraged to better target interventions to increase the overall uptake of vaccines in future health emergencies.



**Acknowledgements** The author wishes to thank Dr. Gui Lui, Merck and Dr. Shuxian Chen, Meta, for their thoughtful feedback during the writing process.

**Contributors** EMS is the sole author and guarantor of this study.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Ethics approval** This research was considered not human subjects' research by the University of Washington IRB, and therefore, did not need a human subject determination review.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository. Data are publicly available. Outcome: US Census Bureau. Census.gov. (cited 3 November 2021). Household Pulse Survey: Measuring Social and Economic Impacts during the Coronavirus Pandemic. Available from: <https://www.census.gov/householdpulse> Demographic variables: US Census Bureau. The US Census Bureau. (cited 18 May 2021). American Community Survey (ACS). Available from: <https://www.census.gov/programs-surveys/acs> Political party affiliation proxy: MIT Election Data and Science Lab. County Presidential Election Returns 2000-2020 (Internet). Harvard Dataverse; 2022 (cited 29 June 2023). Available from: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VOQCHQ>.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID ID

Enrique M Saldarriaga <http://orcid.org/0000-0001-8879-6354>

## REFERENCES

- CDC. Centers for disease control and prevention. COVID data tracker; 2020. Available: <https://covid.cdc.gov/covid-data-tracker> [Accessed 29 Jun 2023].
- Kaur SP, Gupta V. COVID-19 vaccine: a comprehensive status report. *Virus Res* 2020;288:198114.
- Suthar AB, Wang J, Seffren V, et al. Public health impact of covid-19 vaccines in the US: observational study. *BMJ* 2022;377:e069317.
- Commissioner O of the. FDA. FDA takes key action in fight against COVID-19 by issuing emergency use authorization for first COVID-19 vaccine. 2020. Available: <https://www.fda.gov/news-events/press-announcements/fda-takes-key-action-fight-against-covid-19-issuing-emergency-use-authorization-first-covid-19>
- Razai MS, Chaudhry UAR, Doerholt K, et al. Covid-19 vaccination hesitancy. *BMJ* 2021;373:1138.
- Dubé E, Laberge C, Guay M, et al. Vaccine hesitancy: an overview. *Hum Vaccin Immunother* 2013;9:1763-73.
- MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: definition, scope and determinants. *Vaccine* 2015;33:4161-4.
- Larson HJ, Jarrett C, Eckersberger E, et al. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007-2012. *Vaccine* 2014;32:2150-9.
- Kumar D, Chandra R, Mathur M, et al. Vaccine hesitancy: understanding better to address better. *Isr J Health Policy Res* 2016;5:2.
- Dubé E, Gagnon D, Nickels E, et al. Mapping vaccine hesitancy—country-specific characteristics of a global phenomenon. *Vaccine* 2014;32:6649-54.
- Anakpo G, Mishi S. Hesitancy of COVID-19 vaccines: rapid systematic review of the measurement, predictors, and preventive strategies. *Hum Vaccin Immunother* 2022;18:2074716.
- Zhang V, Zhu P, Wagner AL. Spillover of vaccine hesitancy into adult COVID-19 and influenza: the role of race, religion, and political affiliation in the United States. *Int J Environ Res Public Health* 2023;20:3376.
- US Census Bureau. Census.gov. Household pulse survey: measuring social and economic impacts during the Coronavirus pandemic. 2021. Available: <https://www.census.gov/householdpulse> [Accessed 03 Nov 2021].
- Commissioner O of the. FDA. FDA approves first COVID-19 vaccine. FDA; 2021. Available: <https://www.fda.gov/news-events/press-announcements/fda-approves-first-covid-19-vaccine> [Accessed 29 Jun 2023].
- Office of the Assistant Secretary for Planning and Evaluation (ASPE). Vaccine hesitancy for COVID-19: state, county, and local estimates. Available: <https://aspe.hhs.gov/reports/vaccine-hesitancy-covid-19-state-county-local-estimates> [Accessed 27 Oct 2021].
- US Census Bureau, The United States Census Bureau. American community survey (ACS). Available: <https://www.census.gov/programs-surveys/acs> [Accessed 18 May 2021].
- CDC. CDC SVI Documentation 2020 | Place and Health | ATSDR. 2020. Available: [https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/SVI\\_documentation\\_2020.html](https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/SVI_documentation_2020.html) [Accessed 29 Jun 2023].
- Poland GA, Spier R. Fear, misinformation, and innumerates: how the Wakefield paper, the press, and advocacy groups damaged the public health. *Vaccine* 2010;28:2361-2.
- Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the internet. *Vaccine* 2010;28:1709-16.
- Blake A. Washington Post. Analysis | How badly unvaccinated republicans are misinformed, in one stat, Available: <https://www.washingtonpost.com/politics/2021/09/27/misinformation-epidemic-among-unvaccinated-republicans-one-stat/> [Accessed 29 Jun 2023].
- Brumfiel G. NPR. Inside the growing alliance between anti-vaccine activists and pro-Trump Republicans. Available: <https://www.npr.org/2021/12/06/1057344561/anti-vaccine-activists-political-conference-trump-republicans> [Accessed 29 Jun 2023].
- Lerer L. How Republican vaccine opposition got to this point. The New York Times; 2021. Available: <https://www.nytimes.com/2021/07/17/us/politics/coronavirus-vaccines-republicans.html> [Accessed 29 Jun 2023].
- Sargent RH, Laurie S, Moncada L, et al. Masks, money, and mandates: a national survey on efforts to increase COVID-19 vaccination intentions in the United States. *PLoS One* 2022;17:e0267154.
- MIT Election Data and Science Lab. County presidential election returns 2000-2020. Harvard Dataverse; 2022. Available: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VOQCHQ> [Accessed 29 Jun 2023].
- Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Methodol* 1996;58:267-88.
- Lee JD, Sun DL, Sun Y, et al. 2016 Exact post-selection inference, with application to the lasso. *Ann Statist*;44.
- Taylor J, Tibshirani R. Post-selection inference for L1-penalized likelihood models. *arXiv* 2016.
- Murphy KP. Machine learning: a probabilistic perspective. In: *Adaptive computation and machine learning series*. Cambridge, MA: MIT Press, 2012: 1067. Available: <https://probml.github.io/probml-book/book0.html>
- Fajar JK, Sallam M, Soegiarto G, et al. Global prevalence and potential influencing factors of COVID-19 vaccination hesitancy: a meta-analysis. *Vaccines (Basel)* 2022;10:1356.
- Ojha RP, Stallings-Smith S, Flynn PM, et al. The impact of vaccine concerns on racial/ethnic disparities in influenza vaccine uptake among health care workers. *Am J Public Health* 2015;105:e35-41.
- Hirth JM, Fuchs EL, Chang M, et al. Variations in reason for intention not to vaccinate across time, region, and by race/ethnicity, NIS-Teen (2008-2016). *Vaccine* 2019;37:595-601.
- Bagasra AB, Doan S, Allen CT. Racial differences in institutional trust and COVID-19 vaccine hesitancy and refusal. *BMC Public Health* 2021;21:2104.
- Momplaisir FM, Kuter BJ, Ghadimi F, et al. Racial/ethnic differences in COVID-19 vaccine hesitancy among health care workers in 2 large academic hospitals. *JAMA Netw Open* 2021;4:e2121931.
- Badr H, Zhang X, Oluoyomi A, et al. Overcoming COVID-19 vaccine hesitancy: insights from an online population-based survey in the United States. *Vaccines (Basel)* 2021;9:1100.
- Bikaki A, Machiorlatti M, Clark LC, et al. Factors contributing to SARS-CoV-2 vaccine hesitancy of Hispanic population in Rio Grande valley. *Vaccines (Basel)* 2022;10:1282.
- Moran MB, Chatterjee JS, Frank LB, et al. Individual, cultural and structural predictors of vaccine safety confidence and influenza vaccination among Hispanic female subgroups. *J Immigr Minor Health* 2017;19:790-800.
- Gerretsen P, Kim J, Caravaggio F, et al. Individual determinants of COVID-19 vaccine hesitancy. *PLoS ONE* 2021;16:e0258462.

- 38 Betsch C, Schmid P, Heinemeier D, *et al*. Beyond confidence: development of a measure assessing the 5C psychological antecedents of vaccination. *PLOS ONE* 2018;13:e0208601.
- 39 Kricorian K, Civen R, Equils O. COVID-19 vaccine hesitancy: misinformation and perceptions of vaccine safety. *Hum Vaccin Immunother* 2022;18:1950504.
- 40 Rothwell J, Desai S. How misinformation is distorting COVID policies and behaviors. Center for Economic Security and Opportunity, Brookings Institute; 2020. Available: <https://www.brookings.edu/articles/how-misinformation-is-distorting-covid-policies-and-behaviors/> [Accessed 05 Jul 2023].
- 41 Bolsen T, Palm R. Politicization and COVID-19 vaccine resistance in the U.S. *Prog Mol Biol Transl Sci* 2022;188:81–100.
- 42 Simonov A, Sacher SK, Dubé JPH, *et al*. The persuasive effect of fox news: non-compliance with social distancing during the COVID-19 pandemic. National Bureau of economic research Working Paper Series; 2020. Available: <https://www.nber.org/papers/w27237> [Accessed 15 Nov 2023].
- 43 King WC, Rubinstein M, Reinhart A, *et al*. Time trends, factors associated with, and reasons for COVID-19 vaccine hesitancy: a massive online survey of US adults from January-May 2021. *PLoS One* 2021;16:e0260731.
- 44 Wallace J, Goldsmith-Pinkham P, Schwartz JL. Excess death rates for republicans and democrats during the COVID-19 pandemic. National Bureau of Economic Research; 2022. Available: <https://www.nber.org/papers/w30512> [Accessed 16 Nov 2022].
- 45 Salmon DA, Dudley MZ, Glanz JM, *et al*. Vaccine hesitancy. *Am J Prev Med* 2015;49:S391–8.
- 46 Seither R, Laury J, Mugerwa-Kasujja A, *et al*. Vaccination coverage with selected vaccines and exemption rates among children in kindergarten — United States, 2020–21 school year. *MMWR Morb Mortal Wkly Rep* 2022;71:561–8.
- 47 Seither R, Yusuf OB, Dramann D, *et al*. Coverage with selected vaccines and exemption from school vaccine requirements among children in kindergarten — United States, 2022–23 school year. *MMWR Morb Mortal Wkly Rep* 2023;72:1217–24.
- 48 Luman ET, McCauley MM, Shefer A, *et al*. Maternal characteristics associated with vaccination of young children. *Pediatrics* 2003;111:1215–8.
- 49 Allred NJ, Wooten KG, Kong Y. The association of health insurance and continuous primary care in the medical home on vaccination coverage for 19- to 35-month-old children. *Pediatrics* 2007;119 Suppl 1:S4–11.
- 50 Omer SB, Salmon DA, Orenstein WA, *et al*. Mandatory immunization, and the risks of vaccine-preventable diseases. *N Engl J Med* 2009;360:1981–8.
- 51 Siddiqui M, Salmon DA, Omer SB. Epidemiology of vaccine hesitancy in the United States. *Vaccines* 2013;9:2643–8.
- 52 Lo NC, Hotez PJ. Public health and economic consequences of vaccine hesitancy for measles in the United States. *JAMA Pediatr* 2017;171:887–92.
- 53 Ojikutu BO, Bogart LM, Dong L. Empowerment, and structural change: lessons we should be learning from COVID-19. *Am J Public Health* 2022;112:401–4.
- 54 Salmon D, Opel DJ, Dudley MZ, *et al*. Reflections on governance, communication, and equity: challenges and opportunities in COVID-19 vaccination. *Health Affairs* 2021;40:419–25.
- 55 Weintraub RL, Subramanian L, Karlage A, *et al*. COVID-19 vaccine to vaccination: why leaders must invest in delivery strategies now. *Health Affairs* 2021;40:33–41.