How do psychobehavioural variables shed light on heterogeneity in COVID-19 vaccine acceptance? Evidence from United States general population surveys on a probability panel and social media

Grace K Charles,1,2 Sofia P Braunstein,1,3 Jessica L Barker,4,1,4,5 Henry Fung,1,5 Lindsay Coome,1 Rohan Kumar,1 Vincent S Huang,1,5 Hannah Kemp,1,5 Eli Grant,1 Drew Bernard,6 Darren Barefoot,7 Sema K Sgaier1,5,8

ABSTRACT

Objectives To (1) understand what behaviours, beliefs, demographics and structural factors predict US adults’ intention to get a COVID-19 vaccination, (2) identify segments of the population (‘personas’) who share similar factors predicting vaccination intention, (3) create a ‘typing tool’ to predict which persona people belong to and (4) track changes in the distribution of personas over time and across the USA.

Design Three surveys: two on a probability-based household panel (NORC’s AmeriSpeak) and one on Facebook.

Setting The first two surveys were conducted in January 2021 and March 2021 when the COVID-19 vaccine had just been made available in the USA. The Facebook survey ran from May 2021 to February 2022.

Participants All participants were aged 18+ and living in the USA.

Outcome measures In our predictive model, the outcome variable was self-reported vaccination intention (0–10 scale). In our typing tool model, the outcome variable was the five personas identified by our clustering algorithm.

Results Only 1% of variation in vaccination intention was explained by demographics, with about 70% explained by psychobehavioural factors. We identified five personas with distinct psychobehavioural profiles: COVID Sceptics (believe at least two COVID-19 conspiracy theories), System Distrusters (believe people of their race/ethnicity do not receive fair healthcare treatment), Cost Anxious (concerns about time and finances), Watchful (prefer to wait and see) and Enthusiasts (want to get vaccinated as soon as possible). The distribution of personas varies at the state level. Over time, we saw an increase in the proportion of personas who are less willing to get vaccinated.

Conclusions Psychobehavioural segmentation allows us to identify why people are unvaccinated, not just who is unvaccinated. It can help practitioners tailor the right intervention to the right person at the right time to optimally influence behaviour.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study carried out one of the most comprehensive surveys of factors likely to influence vaccination intention in US adults.

⇒ A partitioning around medoids clustering algorithm identified five segments of the population with distinct psychobehavioural profiles (vaccine personas), making the complexity of factors more tractable for practitioners to incorporate into intervention design.

⇒ Follow-up surveys let us analyse changes in persona distribution over time (January 2021 to February 2022), although we did not track individuals.

⇒ Two of our data sets are from a probability-based sample (NORC’s AmeriSpeak panel; January 2021 and March 2021) and two from Facebook for increased sample sizes (May 2021 and February 2022), meaning we cannot compare them directly.

⇒ The large sample size from Facebook allows us to estimate persona distribution at the state level, despite potential platform bias due to sampling from Facebook users.

INTRODUCTION

During the first few months of 2021 when the COVID-19 vaccines were first made available to adults in the USA, there was a high degree of variation in individuals’ willingness to get the vaccine, with some people wanting to get vaccinated immediately, some preferring to wait and others not wanting to get vaccinated at all.12 We define the first category as vaccine acceptance, and the latter two as vaccine hesitancy, following the SAGE Immunization Working Group definition.34 Previous studies have proposed some general reasons for people’s hesitancy: cost, convenience, equity, efficacy, perceived health risk from receiving
the vaccine and distrust in the government.\textsuperscript{5–7} However, there is a clear research and actionability gap. First, while these studies offer some ways to characterise heterogeneity in the vaccine-hesitant population based on these reasons for hesitancy, they largely focus on differences along traditional sociodemographic divisions such as race or political party. This sociodemographic approach only tells us who are not vaccinated, but not why they remain unvaccinated. That is, this approach does not provide insights into the underlying beliefs and barriers influencing people’s vaccination intention (where people who show vaccine hesitancy have low intention and people who show vaccine acceptance have high intention).\textsuperscript{5–10} In contrast, we propose that a psychobehavioural approach (encompassing attitudes and beliefs as well as logistical factors that influence behaviour) can show which barriers are the most important to overcome, which enabling factors are most promising to harness and how these barriers and enablers might differ even among people who are demographically similar.\textsuperscript{7,10–11} Second, where studies do address psychobehavioural variables, they do not offer ways to predict which variables are relevant for which people in the population, making it difficult to put this research into practice. The motivation of this study is to illustrate this approach in a nationally representative sample of US adults.

To identify the factors that influence people’s intentions to get the COVID-19 vaccine, we carried out a comprehensive survey in January 2021 on a probability-based panel (NORC’s AmeriSpeak). Our survey instrument was based on the CUBES behavioural framework, which synthesises many models of behaviour change to consider a more comprehensive set of the factors that influence behaviour than is typically addressed in previous studies, including those related to individuals’ cognitive processes (‘perceptual’) and their situation (‘contextual’).\textsuperscript{12–14} This helped ensure we included constructs on the survey that are often overlooked. After performing analysis to rank the factors most strongly predicting vaccination intention, we used a clustering algorithm to find distinct and actionable psychobehavioural segments, or ‘personas’, of adults. These personas turned out to differ in their vaccination intention and their beliefs and behaviours related to COVID-19 vaccination.\textsuperscript{11,15,16} This machine learning technique differs from traditional demographic segmentation by figuring out what segment assignment would maximise similarities among a persona’s members’ behaviours, motivations, beliefs and social norms while maximising dissimilarities between personas, instead of simply dividing people based on age, race, gender or income a priori.\textsuperscript{17} We then demonstrate an approach to select priority segmentation variables to predict which persona an individual is in (a ‘typing tool’), showing how our psychobehavioural segmentation could be used by practitioners seeking to deploy targeted interventions.

As more US adults got vaccinated in the later half of 2021 and early 2022, we posit that the relative distribution (especially among those who remain unvaccinated) of personas also shifted. Such a shift after January 2021 could indicate a change over time in the barriers and enablers people experienced. To investigate this, we carried out a second survey in March 2021 (sampled from the same panel as the first survey), followed by a third larger survey that was fielded from May 2021 to February 2022 on Facebook. The wider reach of the Facebook platform allowed us to scale data collection and estimate state-level variation across the USA.\textsuperscript{18–20}

To summarise, our objectives in this study are to answer the following research questions:

1. Considering a comprehensive range of factors (encompassing behaviours, beliefs, structural factors and demographics), what are the strongest predictors of US adults’ intention to get a COVID-19 vaccination?
2. How do US adults cluster into vaccine personas, defined by psychobehavioural characteristics?
3. Can persona membership be predicted using a reduced set of variables, providing an actionable way for practitioners to use psychobehavioural segmentation to deploy targeted interventions?
4. How does the distribution of personas vary over the course of the first year of vaccine roll-out, and geographically across states?

METHODS

Study design and participants

We conducted three surveys to capture and monitor COVID-19 vaccine acceptance among adults in the USA (table 1; demographics in online supplemental file 1). Participants in all surveys gave informed consent.

We conducted the first survey using the NORC AmeriSpeak panel (a probability-based household panel)\textsuperscript{21} online and over the phone in English and Spanish with a representative sample of the US general population aged 18+. Participants were compensated either with redeemable points or entry into a raffle. We designed this first NORC survey (‘NORC 1’) using the CUBES behavioural framework.\textsuperscript{12–14} The survey took 45 minutes and was intended to elicit a comprehensive picture of the potential barriers to and enablers of vaccination intention. In addition to sociodemographic factors, these include structural barriers and enablers of COVID-19 vaccination (eg, access to healthcare), beliefs and perceptions about the COVID-19 pandemic and COVID-19 vaccines, influencers and sources of information, perceived risk of getting COVID-19, perceived risk of side effects from COVID-19 vaccine, social norms on COVID-19 vaccination, health-seeking and previous vaccination behaviours and knowledge about COVID-19 vaccine and the COVID-19 pandemic. We used NORC 1 to develop descriptive statistics, run a predictive model, use psychobehavioural segmentation to identify personas and develop a typing tool to predict persona membership (see subsections below).

The second NORC survey (‘NORC 2’) used the same methodology but asked a reduced number of questions...
to collect data on selected variables determined by analysis of NORC 1. The purpose was to determine whether the distribution and profiles of the personas changed between January 2021 (when the very first US residents were getting vaccinated) and March 2021 (when eligibility was expanding to more people in the USA).

The third survey was embedded in the ongoing COVID-19 Trends and Impact Survey (CTI Survey) run by the Delphi Group at Carnegie Mellon University in partnership with Facebook. It contained the same reduced set of questions as NORC 2. The sampling frame was Facebook users aged 18+ in the USA who had been active in the last month. Each day, a random sample, stratified by state, received a link to the survey at the top of their Facebook news feed. 1.7% of those who saw the ad clicked on it (median: April 2020 to April 2022; rate declined over time). Although we obtained data every day between 22 May 2021 and 28 February 2022, we restricted our analyses to the first month and the last month. The purpose was to further see how the distribution of personas changed between snapshots of time, rather than carrying out extensive time trend analysis. In addition, the large sample of respondents across the USA allowed us to collect state-level data on the persona distributions.

**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.

**Survey weighting**

For the two NORC surveys, panel-based sampling weights were used to create nationally representative sampling weights. Panel-based sampling weights were calculated as the inverse of the probability of selection for the NORC national frame and were then raked to external population benchmarks based on seven variables: age, gender, census division, race/ethnicity, education, housing tenure and household phone ownership status.

For the CTI Survey, we used survey weights provided by Facebook to create nationally representative sampling weights, correcting for both non-response bias and coverage bias. Panel-based sampling weights were calculated as the inverse of the probability of selection for the Facebook active user base. A post-stratification step was then used to adjust weights to external population benchmarks. CTI weighting is based on state, age and gender. The unweighted sample had more women and people aged 25–64 than did the weighted sample; the weighted sample had more people with higher education levels than does the US population.

### Predictive model

We used a weighted least squares regression model, built from NORC 1 responses (n=2498 complete cases) in R using the ‘glm’ function in the stats package (V.4.0.2) to identify psychobehavioural drivers that have significant associations with self-reported vaccination intention. The response variable was self-reported vaccination intention, with a 0 indicating ‘extremely unlikely’ and 10 indicating ‘extremely likely’ to get the COVID-19 vaccine.

Input features of the model included demographic variables; perceptions of social norms around vaccines; risk perceptions around coronavirus and the COVID vaccine; information channels (eg, how much information respondents are seeking out about coronavirus); and other contextual variables (eg, having a personal care physician or having health insurance). Several groups of related variables were entered as composite variables, such as the level of agreement with conspiratorial statements about COVID-19 vaccines. Numeric (ordinal) independent variables were normalised to ensure that their estimated coefficients were comparable. The parameters of the model were estimated using weighted least squares (to account for sampling weights) with heteroscedastic robust errors. Variables were tested for multicollinearity, and no variables were found to have a variance inflation factor greater than 5. The full list of variables is given in the regression table in online supplemental file 2.

### Table 1  Summary of the surveys used in this study

<table>
<thead>
<tr>
<th>Survey</th>
<th>Sample</th>
<th>Dates</th>
<th>Total n (Unvaccinated n)</th>
<th>Purpose and analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORC 1</td>
<td>AmeriSpeak: probability-based household panel</td>
<td>21 December 2020 to 4 January 2021</td>
<td>2747 (2731)</td>
<td>Predictive model</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Segmentation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Typing tool</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Baseline descriptives</td>
</tr>
<tr>
<td>NORC 2</td>
<td>Random sample of Facebook users</td>
<td>18–25 March 2021</td>
<td>1670 (923)</td>
<td>Tracking national-level change over time: comparison to NORC 1</td>
</tr>
<tr>
<td>CTI</td>
<td>Random sample of Facebook users</td>
<td>The survey ran continuously, but in this article the analyses use data from the first month (May 2021) and the last month (February 2022)</td>
<td></td>
<td>State-level variation</td>
</tr>
<tr>
<td></td>
<td>22–31 May 2021</td>
<td>109 561 (17 990)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1–28 February 2022</td>
<td>281 365 (39 784)</td>
<td></td>
<td>Tracking national-level and state-level changes over time: comparison to May 2021</td>
</tr>
</tbody>
</table>

**Segmentation to create personas**

We used an unsupervised machine learning algorithm called partitioning around medoids clustering, with a Gower distance metric, to identify clusters of individuals that differed on seven variables (online supplemental file 3). This segmentation was performed on the NORC 1 data. The use of Gower distance allowed the inclusion of both binary and numeric variables.

We selected these variables for segmentation based on their relationship to self-reported COVID-19 vaccination intention, as observed in the predictive model and corroborated by the literature; their heterogeneity in the population; and their actionability to identify population groups and effective interventions (online supplemental file 3). After we identified five personas, we profiled them based on their COVID-19 vaccination intention as well as demographics and other characteristics. We used these profiling variables to develop a narrative to describe each persona.

**Typing algorithm and classification of subsequent survey respondents**

To classify survey respondents from NORC 2 (March 2021) and the CTI Survey (May 2021 and February 2022) into the five different vaccine personas identified in NORC 1, we built a classification and regression tree (CART) model based on the results from NORC 1. The purpose of this model is to calculate the likelihood that a given individual falls into each of the personas, based on that individual’s specific attributes, and assign the individual to the most likely persona. That is, each individual is categorised by the typing tool as one of the five personas (membership is mutually exclusive and comprehensively exhaustive). The predictors of the CART model were the features from the segmentation of NORC 1, and the outcome variable was the five personas. To assess the accuracy of the model, we split the CART model into training and testing sets (80% and 20% of the data from NORC 1, respectively).

**Estimating temporal and spatial heterogeneity**

To investigate how the percentage of people in each persona changed over time, we used a $X^2$ goodness-of-fit test to compare NORC 2 (March 2021) to the reference distribution of NORC 1 (January 2021). As this just tells us whether there is a difference in the overall distribution of personas, we ran post hoc tests to investigate differences between the two time points for each individual persona. We did this by running weighted least squares regressions for each persona, regressing a dummy variable for each persona on time, giving us the OR of a person being in each persona in March 2021 versus January 2021. We followed corresponding steps to compare the first and last months of the CTI Survey (May 2021 and February 2022, respectively). We did not compare NORC and CTI data since the study populations are different (people with a household address vs Facebook users).

To determine the percentage of each vaccine persona by state in May 2021 and February 2022, we first designated a persona for each respondent in the CTI Survey based on the CART typing tool. We then took a weighted count for each persona within each state for each month and divided that count by the CTI Survey weighted total state sample size to arrive at a percentage breakdown of persona for each state. We only report results where a state had 100 unvaccinated respondents in a given month.

**RESULTS**

**Psychobehavioural factors are better predictors of vaccination intention than are demographics**

Analysing NORC 1 data, behaviours and beliefs were the most statistically significant predictors of vaccination intention (model adjusted $R^2=0.78$) (online supplemental file 2). Factors that were most strongly associated with vaccination intention included when respondents wanted to receive the COVID-19 vaccine (ie, within the first 3 months of vaccine availability, or not); concerns around COVID-19 vaccine safety; belief in COVID-19 conspiracy theories; perceptions of COVID-19 risk; perceptions of community norms around mask wearing and vaccination; and having health insurance. Demographic factors such as income and race had statistically significant but small associations with reported vaccination intention.

We used an ANOVA table to examine the percentage of variation in vaccination intention explained by demographic variables versus a subset of psychobehavioural variables that the regression model showed were highly associated with vaccination intention. Most of the explained variation is from psychobehavioural variables, in particular believing that the vaccine is unsafe, which alone explains 45% of the variation (figure 1). In contrast, controlling for psychobehavioural variables, demographic variables together explain less than 1% of the variation in vaccination intention.

**Identification of five personas with distinct profiles**

We used unsupervised cluster analysis on NORC 1 data to explore several cluster solutions (from three to eight clusters) based on seven variables encompassing different barriers and beliefs around the healthcare system, COVID-19 and the COVID-19 vaccine (figure 2 and online supplemental file 3). These variables are: having health insurance, delaying medical care due to cost, believing that the vaccine is unsafe, worrying about catching COVID-19, being an early adopter of the vaccine, believing one’s own race is treated unfairly by the healthcare system and a composite score for belief in conspiracy theories. We selected these seven variables based on their relevance in the literature and predictive power from our regression model, heterogeneity in the population and actionability. We identified a five-cluster solution to maximise separability between and homogeneity within personas in terms of vaccination intention, as well as barriers and beliefs. In addition, personas in...
the five-cluster solution differed from one another based on characteristics that could be easily identified for predicting which persona an individual belongs to, that is, this solution is actionable for use in interventions.

We labelled the five personas Enthusiasts, Watchful, Cost Anxious, System Distrusters and COVID Sceptics in decreasing order of vaccination intention (figure 2). The persona labels indicate characteristics that distinguish the personas from each other, although we note that these characteristics are not necessarily the most important predictors of vaccination intention.

Vaccination intention in the Enthusiast segment was very high (9.2 out of 10), with all individuals willing to get vaccinated in the first 3 months it was available to them. Enthusiasts were the most worried about getting sick from COVID-19, with 77.4% saying they worried a moderate amount or a great deal about catching the disease. This segment had the oldest average age (50.8).

The average vaccination intention for the Watchful segment (5.4 out of 10) was much lower than for the Enthusiasts. Unlike Enthusiasts, none of the individuals in the Watchful segment was willing to get vaccinated in the first 3 months it was available to them. Enthusiasts were the most worried about getting sick from COVID-19, with 77.4% saying they worried a moderate amount or a great deal about catching the disease. This segment had the oldest average age (50.8).

The average vaccination intention for the Watchful segment (5.4 out of 10) was much lower than for the Enthusiasts. Unlike Enthusiasts, none of the individuals in the Watchful segment was willing to get vaccinated in the first 3 months. However, they had fewer structural or perceptual barriers to vaccination than other segments, with none having delayed medical care in the past year because of cost, or believing that the health system treated them unfairly because of their race. 11.4% of individuals in the Watchful segment agreed or strongly agreed that the vaccine is unsafe, more than the Enthusiasts.

The average vaccination intention for the Cost Anxious segment was 3.8 out of 10. All individuals in this segment had delayed medical care in the past year because of cost, and 61.7% had experienced at least one additional barrier to healthcare access, such as transportation, distance, work schedules, childcare or lack of time. In this segment, 20.9% lived in rural areas.

All respondents in the System Distruster segment disagreed or strongly disagreed that people of their race are treated fairly in the health system. 47.7% identified as Black, and only 15.3% said they were Republican. The average vaccination intention for this segment was 3.2 out of 10, and no one in this segment was willing to get the vaccine in the first 3 months it was available.

Individuals in the COVID Sceptic segment were notable for their low vaccination intention (2.5 out of 10), highest mean conspiratorial belief score (respondents believed an average of 1.7 out of 3 COVID-19 conspiracies), having the lowest level of worry about COVID-19 among the five segments (27.6% worried a moderate or great amount) and the most widespread agreement that the vaccine was unsafe (41.5% agreed). The proportion of COVID Sceptics identifying as Republican was 55.9%, and 22.8% lived in rural areas.

Predicting personas in additional samples with a typing tool

The cluster analysis above assigns individuals from the NORC 1 sample into personas, but a key question is how to predict which persona an individual from another sample belongs to. To do this, we built a CART model, which allows us to identify which of the segmentation variables are the best predictors of persona membership. We trained the model on 80% of the NORC 1 data and used the remainder as a testing set. We found that a model with six variables achieved 90.5% accuracy on the testing set (figure 3). We subsequently included these variables as a six-question ‘typing tool’ in NORC 2 and the CTI Survey, enabling us to classify respondents from these samples.
<table>
<thead>
<tr>
<th>variables</th>
<th>COVID Sceptic</th>
<th>Cost Anxious</th>
<th>Enthusiast</th>
<th>System Distruster</th>
<th>Watchful</th>
</tr>
</thead>
<tbody>
<tr>
<td>unvaccinated US respondents in NORC Survey 1 (%)</td>
<td>18.20</td>
<td>14.30</td>
<td>43.50</td>
<td>8.20</td>
<td>15.80</td>
</tr>
<tr>
<td>mean vaccine intention (scale 0-10)</td>
<td>2.50</td>
<td>3.80</td>
<td>9.20</td>
<td>3.20</td>
<td>5.40</td>
</tr>
</tbody>
</table>

**Segmentation variables**

<table>
<thead>
<tr>
<th></th>
<th>COVID Sceptic</th>
<th>Cost Anxious</th>
<th>Enthusiast</th>
<th>System Distruster</th>
<th>Watchful</th>
</tr>
</thead>
<tbody>
<tr>
<td>health insurance: Yes (%)</td>
<td>91.00</td>
<td>74.00</td>
<td>92.20</td>
<td>72.40</td>
<td>91.70</td>
</tr>
<tr>
<td>delayed medical care due to cost: Yes (%)</td>
<td>0.00</td>
<td>100.00</td>
<td>17.70</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>agree vaccine unsafe: Yes (%)</td>
<td>41.50</td>
<td>30.00</td>
<td>2.90</td>
<td>31.40</td>
<td>11.40</td>
</tr>
<tr>
<td>worry about catching covid: Yes (%)</td>
<td>27.70</td>
<td>56.20</td>
<td>77.40</td>
<td>57.90</td>
<td>71.30</td>
</tr>
<tr>
<td>early adopter: Yes (%)</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>race treated fairly: Yes (%)</td>
<td>99.30</td>
<td>81.40</td>
<td>90.00</td>
<td>0.00</td>
<td>99.50</td>
</tr>
<tr>
<td>mean conspiratorial thinking (scale 0-3)</td>
<td>1.70</td>
<td>0.80</td>
<td>0.30</td>
<td>0.90</td>
<td>0.20</td>
</tr>
</tbody>
</table>

**Profiling variables: demographic**

<table>
<thead>
<tr>
<th></th>
<th>COVID Sceptic</th>
<th>Cost Anxious</th>
<th>Enthusiast</th>
<th>System Distruster</th>
<th>Watchful</th>
</tr>
</thead>
<tbody>
<tr>
<td>race: Black (%)</td>
<td>6.50</td>
<td>12.60</td>
<td>7.40</td>
<td>47.70</td>
<td>11.00</td>
</tr>
<tr>
<td>race: Hispanic (%)</td>
<td>14.80</td>
<td>20.60</td>
<td>14.30</td>
<td>24.60</td>
<td>18.90</td>
</tr>
<tr>
<td>race: White (%)</td>
<td>75.20</td>
<td>57.30</td>
<td>67.30</td>
<td>22.80</td>
<td>63.30</td>
</tr>
<tr>
<td>race: Other (%)</td>
<td>3.50</td>
<td>9.60</td>
<td>11.10</td>
<td>5.00</td>
<td>6.90</td>
</tr>
<tr>
<td>gender: Man (%)</td>
<td>53.50</td>
<td>56.70</td>
<td>44.00</td>
<td>57.70</td>
<td>60.50</td>
</tr>
<tr>
<td>gender: Woman (%)</td>
<td>46.50</td>
<td>43.30</td>
<td>56.00</td>
<td>42.40</td>
<td>39.50</td>
</tr>
<tr>
<td>political affiliation: Democrat (%)</td>
<td>12.70</td>
<td>30.60</td>
<td>45.70</td>
<td>49.50</td>
<td>38.90</td>
</tr>
<tr>
<td>political affiliation: Independent (%)</td>
<td>31.00</td>
<td>38.80</td>
<td>31.00</td>
<td>35.20</td>
<td>35.70</td>
</tr>
<tr>
<td>political affiliation: Republican (%)</td>
<td>56.30</td>
<td>30.60</td>
<td>23.30</td>
<td>15.30</td>
<td>25.40</td>
</tr>
<tr>
<td>urbanicity: Rural (%)</td>
<td>24.90</td>
<td>20.90</td>
<td>12.10</td>
<td>10.20</td>
<td>10.60</td>
</tr>
<tr>
<td>mean age</td>
<td>47.20</td>
<td>40.70</td>
<td>50.80</td>
<td>40.70</td>
<td>48.70</td>
</tr>
<tr>
<td>income: $100,000 or more (%)</td>
<td>18.50</td>
<td>13.10</td>
<td>30.40</td>
<td>16.50</td>
<td>21.70</td>
</tr>
<tr>
<td>income: $30,000 to $60,000 (%)</td>
<td>23.40</td>
<td>32.90</td>
<td>24.00</td>
<td>23.50</td>
<td>26.70</td>
</tr>
<tr>
<td>income: $60,000 to under $100,000 (%)</td>
<td>25.90</td>
<td>21.70</td>
<td>25.40</td>
<td>19.50</td>
<td>26.70</td>
</tr>
<tr>
<td>income: Less than $30,000 (%)</td>
<td>32.20</td>
<td>32.40</td>
<td>20.30</td>
<td>40.50</td>
<td>24.90</td>
</tr>
<tr>
<td>education: Bachelor's degree (%)</td>
<td>17.60</td>
<td>15.30</td>
<td>22.00</td>
<td>20.00</td>
<td>25.40</td>
</tr>
<tr>
<td>education: High school (%)</td>
<td>36.70</td>
<td>29.00</td>
<td>24.60</td>
<td>30.30</td>
<td>24.60</td>
</tr>
<tr>
<td>education: Less than high school (%)</td>
<td>12.40</td>
<td>13.70</td>
<td>7.30</td>
<td>12.40</td>
<td>6.00</td>
</tr>
<tr>
<td>education: Post grad study/professional degree (%)</td>
<td>5.70</td>
<td>11.20</td>
<td>20.30</td>
<td>9.80</td>
<td>15.30</td>
</tr>
<tr>
<td>education: Some college/ associates (%)</td>
<td>27.60</td>
<td>30.80</td>
<td>25.90</td>
<td>27.40</td>
<td>28.80</td>
</tr>
</tbody>
</table>

**Profiling variables: other**

<table>
<thead>
<tr>
<th></th>
<th>COVID Sceptic</th>
<th>Cost Anxious</th>
<th>Enthusiast</th>
<th>System Distruster</th>
<th>Watchful</th>
</tr>
</thead>
<tbody>
<tr>
<td>social responsibility: Yes (%)</td>
<td>16.20</td>
<td>27.90</td>
<td>86.30</td>
<td>26.90</td>
<td>49.60</td>
</tr>
<tr>
<td>anticipated regret: Yes (%)</td>
<td>25.00</td>
<td>43.40</td>
<td>91.10</td>
<td>40.10</td>
<td>63.80</td>
</tr>
<tr>
<td>trust vaccine process: Yes (%)</td>
<td>56.80</td>
<td>68.70</td>
<td>94.20</td>
<td>60.30</td>
<td>85.40</td>
</tr>
<tr>
<td>believes covid dangerous to community: Yes (%)</td>
<td>34.50</td>
<td>62.00</td>
<td>77.70</td>
<td>63.20</td>
<td>76.40</td>
</tr>
<tr>
<td>believes majority of community take vaccine: Yes (%)</td>
<td>31.00</td>
<td>42.00</td>
<td>88.10</td>
<td>42.20</td>
<td>66.10</td>
</tr>
<tr>
<td>expected chance long term effect from vaccine: High (%)</td>
<td>38.80</td>
<td>31.00</td>
<td>2.70</td>
<td>38.70</td>
<td>18.60</td>
</tr>
</tbody>
</table>

*Figure 2* Characteristics of the five personas. Segmentation variables were used in a partitioning around medoids clustering algorithm (with a Gower distance metric). Darker colours indicate higher row-wise values in comparison to other columns. Data from NORC 1.
into the five personas using a much shorter survey instrument and a high degree of accuracy.

**The distribution of personas changed from January 2021 to February 2022, reflecting increasing scepticism and distrust among the remaining unvaccinated population**

We found that the distribution of personas among unvaccinated people changed from January 2021 to March 2021 (i.e., the two NORC surveys; $\chi^2=81.025$, df=4, $p<0.001$), and from May 2021 to February 2022 (i.e., the start and end of the CTI Survey; $\chi^2=37.417$, df=4, $p<0.001$), as shown in figure 4.

Compared with respondents in January 2021, unvaccinated respondents in March 2021 were significantly less likely to be Watchful (odds ratio (OR)=0.50), and more likely to be COVID Sceptics (OR=1.69). Differences for the other three personas are significant but not substantial (OR between 0.96 and 0.99). Compared with May 2021, unvaccinated people in February 2022 were significantly less likely to be Enthusiasts (OR=0.51) or Watchful (OR=0.83), and more likely to be Cost Anxious (OR=1.07), COVID Sceptics (OR=1.28) or System Distrusters (OR=1.36). That is, we broadly see an increase in personas who are less willing to get vaccinated and a decrease in those who are more willing to get vaccinated.

Reflecting this, in January 2021 respondents on average rated their intention to get vaccinated at 6.11 on a 0–10 scale (95% CI 5.92–6.29). By March, mean self-reported vaccination intention among unvaccinated respondents (60% of all respondents) in NORC 2 was 5.25 (95% CI 4.89–5.60), significantly lower than in January ($t=2.27$, df=1514.6, $p=0.023$). The vaccination intention question was not included in the CTI Survey.

The demographic profiles of each persona generally remained stable over time. While we found that interaction between time and each of the demographic variables did have a significant effect on the likelihood of being in any given vaccine persona, none of the differences were substantial (OR between 0.8 and 1.1).

**The distribution of personas was heterogeneous at the state level and changed from May 2021 to February 2022**

The distribution of personas varied among states, with personas that are dominant in some states having a much lower prevalence in others (figure 5). In May 2021, the state with the highest percentage of each persona had twice or almost twice as high a percentage as the state with the lowest (COVID Sceptics: 40.5% in Montana vs 19.9% in Maryland; System Distrusters: 16.9% in Connecticut vs 6.7% in Maine; Cost Anxious: 32.2% in Missouri vs 17.6% in New Jersey; Watchful: 15.1% in Missouri vs 6.1% in Virginia; Enthusiasts: 33.6% in New Jersey vs 12.4% in Idaho).

There were changes between May 2021 and February 2022 in which states had the highest and lowest percentages of each persona, and in February 2022, there was still substantial variation among states. Comparing the states with the highest and lowest percentages of each persona...
persona, again the ratio was 2:1 for System Distrusters, Cost Anxious and Watchful personas (System Distrusters: 22.2% in Nevada vs 10.8% in Maine; Cost Anxious: 35.4% in Wyoming vs 17.7% in Massachusetts; Watchful: 13.7% in New Hampshire vs 5.6% in Montana). This range was smaller for COVID Sceptics (42.3% in Iowa vs 30.2% in New Mexico) and larger for Enthusiasts (20.5% in Connecticut vs 4.9% in Wyoming).

**DISCUSSION**

**Principal findings and strengths of the study**

The first main contribution of the study is to quantify the association between psychobehavioural factors and vaccination intention, relative to demographics and vaccination intention (our first objective). This study’s predictive model revealed that vaccination intention in the USA is more strongly associated with beliefs, perceptions and barriers (ie, psychobehavioural factors) than with demographics. Despite this, many polls of COVID-19 vaccine acceptance focus simply on differences between demographic groups, meaning that we not only miss a more nuanced picture of who is and is not getting vaccinated, but more crucially do not gain an understanding of why certain people are more willing to get vaccinated. Those studies that do address these reasons why typically do so in a specific sociodemographic group (eg, in a single state, a single race or a single profession) and also consider a more limited range of psychobehavioural variables than we investigated. This breadth is the second contribution of our study. Where there is overlap between our study and others, there is generally agreement in the variables most strongly associated with vaccination intention in the USA, including fear of negative effects and feelings of social responsibility.

As there are many psychobehavioural factors to consider—for example, our regression contained over 30 psychobehavioural variables—providing a way to characterise and predict which are most important and for which people in which locations is a key step in making our findings more actionable. This is the third key contribution and was a main objective of our study. We used an unsupervised clustering algorithm to identify five ‘vaccine personas’—COVID Sceptics, System Distrusters, Cost Anxious, Watchful, Enthusiasts—who vary from each other in their intention to get vaccinated and in the beliefs and barriers that they experience. By answering six short questions in the typing tool, a person can easily be classified into a persona with a high degree of accuracy. Overall, our findings contribute to both the academic literature and practitioners’ toolkits by providing an
efficient framework for assessing the diversity of reasons why people may be hesitant to getting the COVID-19 vaccine, and determining how to focus intervention efforts given limited time and resources. An open question is the extent to which psychobehavioural personas developed from equivalent data in other countries would be similar to those in the USA, given different cultural norms, attitudes and beliefs.

Our first survey was fielded less than a month after the first COVID-19 vaccines had been made available in the USA, raising the question of whether it is still important to address the same psychobehavioural barriers to and enablers of vaccination that we identified in January 2021. Our second and third surveys provide data from March 2021, and the period spanning May 2021 and February 2022, and highlight how the distribution of personas among unvaccinated people has changed: a key difference is the increasing proportion of COVID Sceptics, as more people from personas with higher vaccination intention have got vaccinated. At the time of publishing this article, just over 2 years on from universal vaccine eligibility for US adults, vaccination rates in the USA have plateaued: almost one-fifth of the population has not received a dose of the COVID-19 vaccine, and just over two-thirds have finished their primary series. This means that efforts to achieve high vaccination rates, requiring widespread vaccine acceptance and access, are certainly not over—and indeed increased and more innovative efforts may be needed as the remaining unvaccinated people are those who are more hesitant.

**Implications for practitioners**

There has been a huge number and diversity of interventions to promote COVID-19 vaccine uptake in the USA, from messaging campaigns to state-run lotteries. However, because different individuals may have very different reasons for not getting vaccinated, a given intervention may work well for one person but not another. Psychobehavioural segmentation provides an effective lens for policymakers and other practitioners to assess the key barriers and enablers of their populations towards COVID-19 vaccination. This approach has previously been leveraged for health behaviour change in other contexts, allowing delivery of the right intervention to the right person at the right time to optimally influence behaviour.

The main recommendation based on the findings from our psychobehavioural segmentation is to tailor interventions to the specific barriers and enablers relevant to each persona. For example, encouraging the Watchful could be achieved by emphasising social norms through communication campaigns that show members of their community having safe, positive vaccination experiences. In contrast, the Cost Anxious may respond to initiatives making vaccines more easily accessible to people, free childcare, paid time off work for vaccination

---

Figure 5  State-level breakdown of vaccine personas among unvaccinated US Facebook users (CTI Survey) in May 2021 and February 2022. States with n<100 unvaccinated respondents in a given month are not shown. CTI, COVID-19 Trends and Impact.
Our study has some limitations. First, our psychosocial segmentation does not capture several structural barriers to COVID-19 vaccination, such as distance to vaccine clinics, appointment scheduling or lack of access to information on the COVID-19 vaccine. Second, our main outcome measure was self-reported vaccination intention, which may not always translate into the action of getting vaccinated. Third, although we present data from four time points, our study did not track individuals over time. A key unanswered question is the extent to which unvaccinated people transition between different personas; for example, whether COVID Sceptics become Watchful and then Enthusiasts before they decide to get vaccinated. Fourth, we are unable to directly compare data from our first two time points (January and March 2021) to the latter two (May 2021 and February 2022): as the latter data were collected on Facebook, there is a potential platform effect. While weighting can correct for demographic differences (respondents on Facebook tend to skew more white and female relative to the US population), it cannot account for potential unknown differences in other factors between Facebook users and the general population. Fifth, we make the assumption that the personas developed from the single NORC 1 data set are generalisable to any other. This is plausible because NORC 1 is a probability-based representative sample of the US population, but we have not empirically validated the personas’ generalisability. Sixth, while data collection on Facebook provided a large enough sample to calculate state-level estimates of the distribution of personas, to tailor interventions in practice requires a higher level of geographic granularity. Reflecting this, health officials have emphasised the need for local data to inform precise interventions. An important future direction is to continue efforts to collect data at the county or even zip code or census tract level.

Throughout the development of the personas and typing tool, we have emphasised actionability. Therefore, a motivation and our goal for this study is that the typing tool could be trialled in an online experiment where unvaccinated participants answer the typing tool questions and are randomised to receive either a generic or persona-specific message about COVID-19 vaccines, and would then state their intention to get vaccinated. Finally, the effects on actual vaccine uptake, rather than intentions, could be tested in the field, for example, by a healthcare provider sending text messages to patients.
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
Jessica I. Barker http://orcid.org/0000-0002-8369-4823

REFERENCES