Unraveling Attributes of COVID-19 Vaccine Hesitancy in the U.S.: A Large Nationwide Study

Sean D. McCabe[1,2,†], Elizabeth A. Hammershaimb[1,3,4†], David Cheng[1], Andy Shi[1,2], Derek Shyr[1,2], Shuting Shen[1,2], Lyndsey D. Cole[5], Jessica R. Cataldi[5,6], William Allen[1,8], Ryan Probasco[1], Ben Silbermann[1], Feng Zhang[1,9,10,11,12], Regan Marsh[13,14,15], Mark A. Travassos[1,3,4*], Xihong Lin[1,2,7,16*]

† These authors contributed equally
* Corresponding authors: Xihong Lin (xlin@hsph.harvard.edu, 0000-0001-7067-7752),
Mark A. Travassos (*mtravass@som.umaryland.edu*),

[1]The How We Feel Project, USA
[2]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[3]Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD, USA
[4]Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD, USA
[5] Department of Pediatrics, University of Colorado Anschutz Medical Campus
[6] Adult and Child Consortium for Health Outcomes Research and Delivery Science, University of Colorado Anschutz Medical Campus and Children's Hospital Colorado, Aurora, CO
[7]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[8]Society of Fellows, Harvard University, Cambridge, MA, USA
[9]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[10]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA
[11]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[12]Howard Hughes Medical Institute, Chevy Chase, MD, USA
[13] Department of Emergency Medicine, Brigham and Women's Hospital, Boston, MA, USA
[14] Department of Emergency Medicine, Harvard Medical School, Boston, MA, USA
[15] Partners in Health, Boston, MA, USA
[16]Department of Statistics, Harvard University, Cambridge, MA, USA

Abstract

SARS-CoV-2 vaccines are powerful tools to combat the COVID-19 pandemic, but vaccine hesitancy threatens these vaccines' effectiveness. To address COVID-19 vaccine hesitancy and ensure equitable distribution, understanding the extent of and factors associated with vaccine hesitancy is critical. We report the results of a large nationwide study conducted December 2020-January 2021 of 34,470 users from COVID-19-focused smartphone-based app How We Feel on their willingness to receive a COVID-19 vaccine. Nineteen percent of respondents expressed vaccine hesitancy, the majority being undecided. Vaccine hesitancy was significant among females, younger people, minority and low-income communities, healthcare and essential workers, rural residents, geographical regions with higher COVID-19 burden, those who did not use protective measures, and those who did not receive COVID-19 tests. Our findings support the need for targeted efforts to develop education and outreach programs to overcome vaccine hesitancy and improve equitable access, diversity, and inclusion in the national response to COVID-19.

Main

The emergence in late 2019 of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as a novel human pathogen and causative agent of the global coronavirus disease 2019 (COVID-19) pandemic [1] fueled an unprecedented effort to rapidly develop a vaccine [2]. While the successful development of several effective SARS-CoV-2 vaccines has given many hope, the defining challenge now to effectively combat COVID-19 is ensuring equitable vaccine distribution and high vaccine uptake.

At least 70% of the U.S. population needs immunity to SARS-CoV-2 to end the COVID-19 pandemic [3]. Public opinion polling in early 2020 suggested that as many as 72% of U.S. adults were willing to receive a COVID-19 vaccine once licensed and available. Four months later, the number of U.S. adults willing to receive a SARS-CoV-2 vaccine had sharply declined to as low as 51% [4].

In December 2020, two vaccines against COVID-19 received Emergency Use Authorization (EUA) from the U.S. Food and Drug Administration [5, 6]. The results of phase 3 clinical trials and the subsequent rollout of the Pfizer-BioNTech and NIAID-Moderna vaccines received significant attention in the media. Opinion polls conducted in December 2020 suggested a subsequent increase in public willingness to receive a COVID-19 vaccine, likely due to the widespread availability of data showing the vaccines to be both safe and effective [7].

In this paper, we determine the current degree of COVID-19 vaccine hesitancy in the U.S. and characteristics associated with vaccine hesitancy. This will help public health and community leaders develop effective education and outreach programs to overcome vaccine hesitancy and ensure equitable vaccine distribution and the requisite rapid vaccine uptake.

How We Feel (HWF; http://www.howwefeel.org) is a web and mobile-phone application developed to facilitate the large-scale collection of data about COVID-19 symptoms, SARS-CoV-2 test results, and transmission-mitigating behaviors and sentiments [8]. Users are assigned a randomly generated number that tracks logins from the same device and are otherwise unidentifiable. In December 2020, we fielded a question about users' willingness to receive a COVID-19 vaccine: "If a safe, effective coronavirus vaccine were available, how likely would you be to get yourself vaccinated?" Responses were recorded on a 5-point bipolar Likert

scale, and vaccine hesitancy was defined as a “Very Unlikely,” “Unlikely,” or “Undecided” response.

A total of 34,470 users responded to the vaccine acceptance question. The largest number of respondents came from Connecticut and California with 8,373 and 4,405, respectively (Supp Fig 1a). HWF’s user base is approximately 79% female (Supp Fig 1b) and 83% white (Supp Fig 1c). Users are 18 years of age or older and are equally distributed by age groups (Supp Fig 1d). More than 68% of respondents were non-essential workers, and users cover a diverse range of income groups. All descriptive statistics of the study participants are available in Supp. Tab 1.

In total, 28,687 (83%) were accepting (“Likely” or “Very Likely”) of the vaccine (Fig 1a). After applying a census-based post-stratification weight (see Methods), Vermont (92%) and Washington D.C. (88%) had the highest rates of vaccine acceptance while South Dakota (31%) and Louisiana (23%) had the highest rates of undecided users (Fig 1b). Weighted bar plots of vaccine intent across demographic characteristics revealed that “Undecided” users represented the largest proportion of hesitant users across all demographic groups (Fig 1c, Supp Tab 2). State level hesitancy rates were negatively associated with the average number of users that practiced protective behavior, and were positively associated with cumulative COVID-19 case and death rates by January 10, 2021 (Fig 2a). Unweighted plots are available in Supp Fig 2 and 3.

To assess demographic associations with vaccine hesitancy, we fit a univariate logistic regression with socio-demographic, occupation, preexisting medical conditions, geographical and COVID-19 related predictors (Supp Tab 3) and a multivariable logistic regression model to adjust for potential confounding between the predictors (Fig 2b, Supp Tab 4). We implemented post-stratification weights using census estimates of sex, age, race, and census location (see Methods). People of color reported higher rates of vaccine hesitancy compared to white non-Hispanic users (African American OR, 4.09; CI, 3.57, 4.67; p=3.25e-93). Vaccine hesitancy was more likely among females than males (OR,1.72; CI, 1.55, 1.90; p=7.24e-26); younger users than those over 65 years old; those with three or more preexisting conditions than those with zero (OR, 1.19; CI, 1.06, 1.34; p=0.0045); and parents than non-parents (OR, 1.24; CI, 1.13, 1.36; p=8.15e-6). Individuals that were furloughed or job-seeking were also more vaccine hesitant compared to those working full- or part-time (OR, 1.46; CI, 1.26, 1.68; p=2.89e-7). Respondents from the South (OR, 1.24; CI, 1.04, 1.48; p=0.0157), from less densely populated

areas, or with lower incomes were all more likely to be vaccine hesitant. Users that responded before the Pfizer Emergency Use Authorization (EUA) on December 11, 2020 were more vaccine hesitant than users who responded after the EUA (OR, 1.42; CI, 1.31, 1.54; p=1.80e-17), users who practiced behavior protective against COVID-19 such as mask-wearing or social distancing were less vaccine hesitant (OR, 0.77; CI, 0.70, 0.84; p=1.21e-9), and users that received a COVID test were less vaccine hesitant (OR, 0.80; CI, 0.71, 0.90, p=0.0001).

Nominal logistic regression (see Methods) evaluated whether vaccine hesitancy was driven by "Undecided" vs. "Unlikely/Very Unlikely" responses (Supp Tab 5), and was also conducted with a weighted analysis (Supp Tab 6). Hesitancy in healthcare workers, those aged 55-65, Asian users, and those in locations with a median income between $70,000 and $100,000 was driven by the "Undecided" group, whereas hesitancy in the unemployed, those with 3+ preexisting conditions, and southern users was driven by the "Unlikely" group. We performed the weighted multivariable and nominal regression analyses with a less restrictive threshold for the trimming weights (Supp Tab 7) and found similar results.

Nationwide, we found lower rates of vaccine hesitancy than previously estimated, potentially related to confidence in the COVID-19 vaccines now available. Increased hesitancy was associated with minority race/ethnicity, living in less densely populated regions, and being a healthcare worker. A large proportion of these populations were undecided about COVID-19 vaccination, suggesting that targeted outreach may improve vaccine uptake.

Black respondents had the highest rates of COVID-19 vaccine hesitancy relative to other racial and ethnic groups, consistent with other surveys [4, 9-12]. The history of racist practices within the U.S. healthcare system and research community, such as during the Tuskegee Syphilis Study [13], and racial health disparities that persist today likely contribute to our findings. Dispelling concerns within the Black community requires extensive, sustained, structured outreach and will be critical to efforts to contain and eliminate COVID-19. The National Institutes of Health's Community Engagement Alliance (CEAL) provides a model for such outreach, targeting populations that have been hardest hit by the COVID-19 pandemic [14].

Education and outreach efforts must target several additional populations. This includes healthcare workers, rural residents, and young adults. Because large proportions of these populations were undecided about COVID-19 vaccination, outreach to these groups must also

provide reliable vaccine information tailored to the needs of each community, and different outreach strategies may be needed to address the concerns of those who were undecided and those who were unlikely.

Vaccine hesitancy in healthcare workers warrants particular attention because it raises the possibility of sustained COVID-19 vulnerability in an essential worker population critical to caring for the ill and elderly. Previous work has found that U.S. nurses had the highest degree of COVID-19 vaccine hesitancy among healthcare workers [15]. Future versions of the HWF application will provide more insight into the roles of healthcare workers who express hesitancy.

Clarifying regional foci of vaccine hesitancy will be critical in federal resource allocation to combat vaccine hesitancy. We identified the greatest level of COVID-19 vaccine hesitancy in the South followed by the Midwest, while a contemporaneously conducted CDC-sponsored survey found that hesitancy was most prevalent in the Northeast, followed by the South [12]. However, a recent release of the U.S. state and county-level vaccination rates and allocated dose usage data by the CDC showed that Southern states have lower vaccination rates and lower allocated dose usages [16].

A study limitation is that our sample may not be generalizable to the broader American public. How We Feel users are self-selecting and more likely to have a baseline level of concern about COVID-19. The user base is inherently skewed by a large proportion of users residing in Connecticut and California and by regional age discrepancies. Census-adjusted, weighted analysis help correct the sampling bias but may not completely remove the potential for bias, and interpretation of our findings should note this.

Further work is needed to better understand how vaccine uptake relates to vaccine hesitancy in the U.S. and to understand how knowledge, attitudes, and behaviors surrounding COVID-19 vaccines change over time. As COVID-19 vaccines become more widely available in the U.S., our findings reveal the urgent need to address vaccine hesitancy in tandem with vaccine distribution.

References

1. Huang, C., et al., *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.* Lancet, 2020. **395**(10223): p. 497-506.

2. Corey, L., et al., *A strategic approach to COVID-19 vaccine R&D.* Science, 2020. **368**(6494): p. 948-950.
3. Randolph, H.E., and Barreiro, L.B. *Herd Immunity: Understanding COVID-19.* Immunity, 2020. **52**(5): p. 737-741.
4. Funk, C. and Tyson, A. *Growing Share of Americans Say They Plan To Get a COVID-19 Vaccine – or Already Have*. 2021: Pew Research Center.
5. Baden, L.R., et al., *Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine.* N Engl J Med, 2020.
6. Polack, F.P., et al., *Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine.* N Engl J Med, 2020. **383**(27): p. 2603-2615.
7. Hamel, L., Kirzinger, A., Muñana, C., Brodie, M., *KFF COVID-19 Vaccine Monitor: December 2020*, in *KFF COVID-19 Vaccine Monitor*. 2020, Kaiser Family Foundation.
8. Allen, W.E., et al., *Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing.* Nat Hum Behav, 2020. **4**(9): p. 972-982.
9. Hamel, L., Sparks, G., Brodie, M., *KFF COVID-19 Vaccine Monitor: February 2021*. 2021, Kaiser Family Foundation.
10. File, T., and Mohanty, A. *Around Half of Unvaccinated Americans Indicate They Will "Definitely" Get COVID-19 Vaccine*. 2021.
11. Ruiz, J.B., and Bell, R.A. *Predictors of intention to vaccinate against COVID-19: Results of a nationwide survey.* Vaccine, 2021. **39**(7): p. 1080-1086.
12. Nguyen, K.H., et al., *COVID-19 Vaccination Intent, Perceptions, and Reasons for Not Vaccinating Among Groups Prioritized for Early Vaccination - United States, September and December 2020.* MMWR Morb Mortal Wkly Rep, 2021. **70**(6): p. 217-222.
13. Brandt, A.M., *Racism and research: the case of the Tuskegee Syphilis Study.* Hastings Cent Rep, 1978. **8**(6): p. 21-9.
14. NIH. *NIH Community Engagement Alliance (CEAL)*. Accessed Mar. 15, 2021; Available from: https://covid19community.nih.gov/.
15. Shaw, J., et al. *Assessment of U.S. health care personnel (HCP) attitudes towards COVID-19 vaccination in a large university health care system.* Clin Infect Dis, 2021.
16. See How the Vaccine Rollout Is Going in Your County and State. Accessed March 27. Available from *https://www.nytimes.com/interactive/2020/us/covid-19-vaccine-doses.html*
17. Altieri, N., et al. *Curating a COVID-19 Data Repository and Forecasting County-Level Death Counts in the United States.* Harvard Data Science Review., 2021(Special Issue 1).
18. *US Coronavirus Cases & Deaths by State*. Accessed Jan. 10, 2021; Available from: https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/.

Figures:

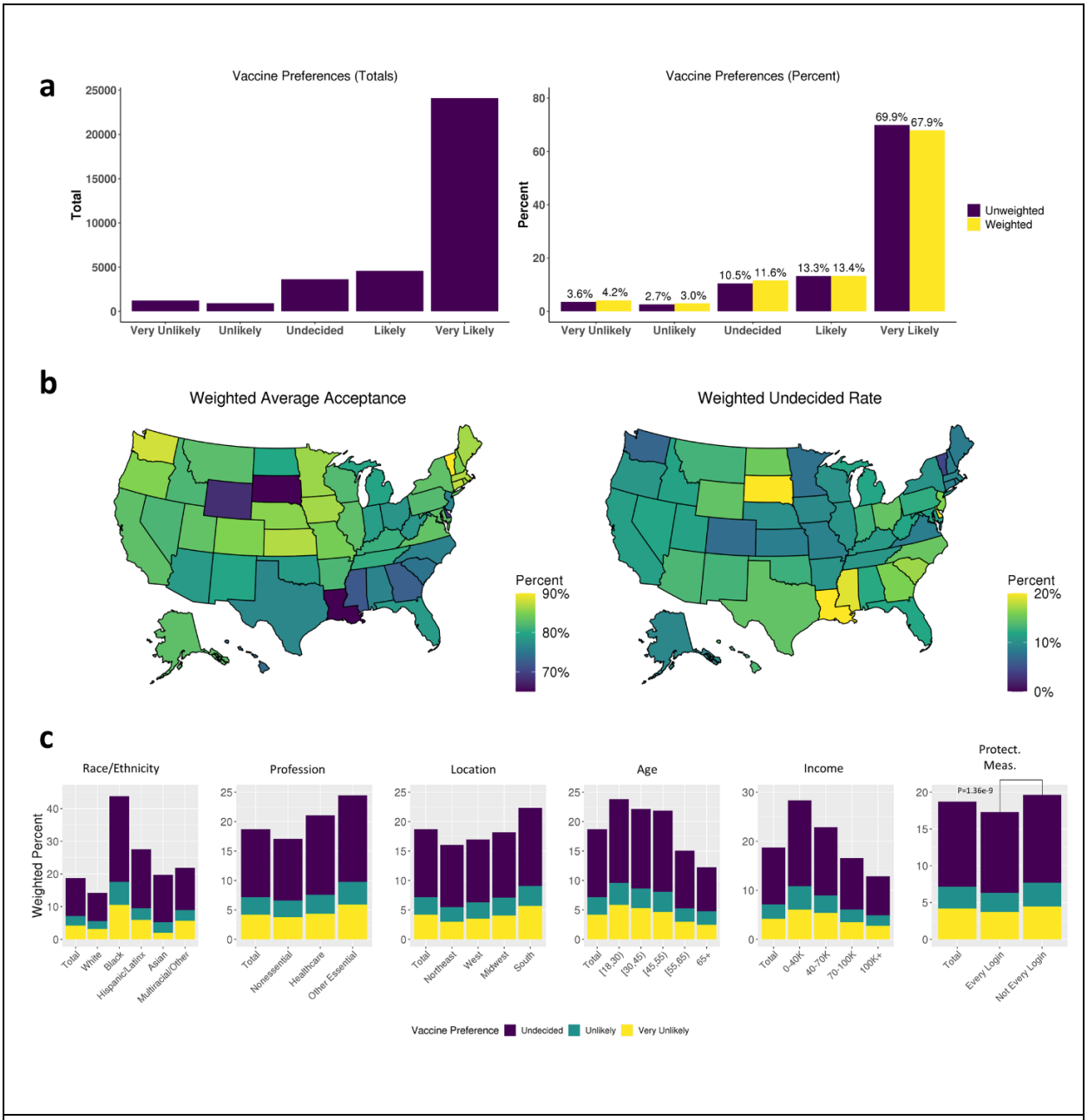

Figure 1, Hesitancy rates: (a) (Left) Number of responses and (Right) unweighted and weighted percentages. **(b)** Weighted average acceptance and undecided rates by state. **(c)** Weighted percentages of hesitant responses by race/ethnicity, profession, location, age, income, and use of protective measures.

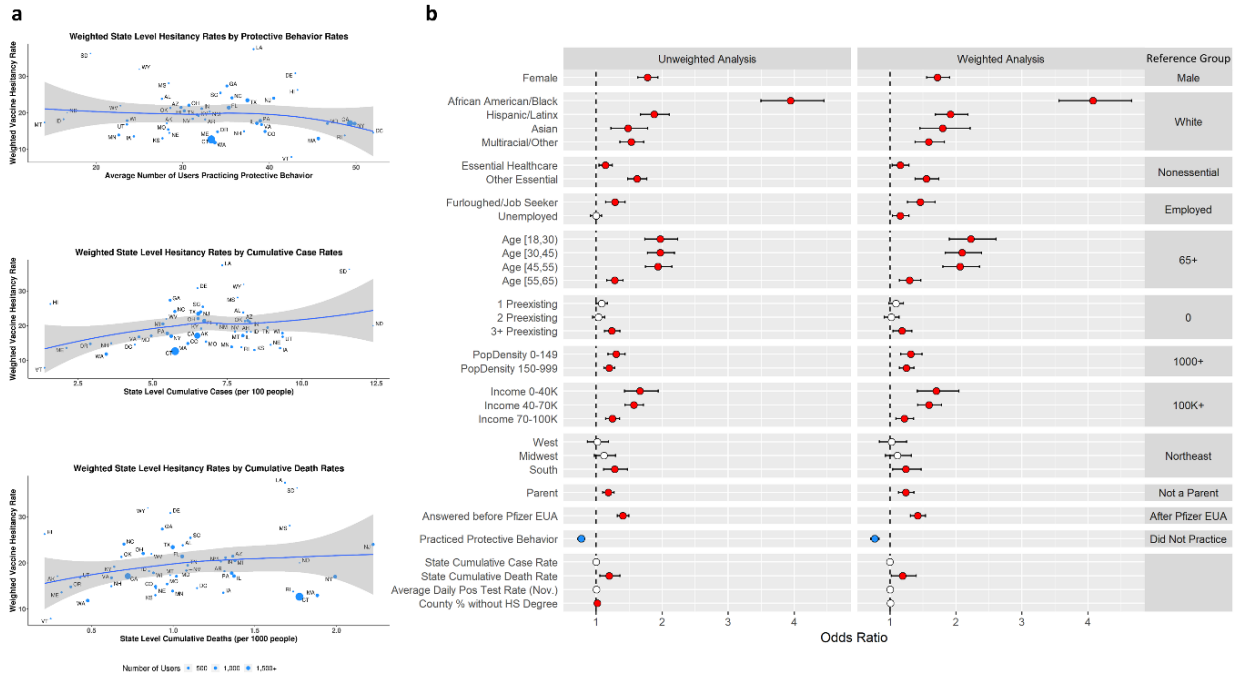

Figure 2, Logistic regression-based association analysis results: (a) State level weighted hesitancy rates by average number of users practicing protective behavior, and cumulative case (/100 individuals) and death (/1000 individuals) rates. **(b)** Forest plots for (Left) unweighted and (Right) weighted multivariable logistic regression analyses with 95% confidence intervals. Non-significant variables at the 0.05 level (white), significant positive associations (red), and significant negative associations (blue).

Online Methods:

Ethics statement:

The HWF application was approved as exempt by the Ethical & Independent Review Services LLP IRB (Study ID 20049–01). The analysis of HWF data was also approved as exempt by Harvard University Longwood Medical Area Institutional Review Board (IRB) (Protocol no. IRB20- 0514) and the Broad Institute of MIT and Harvard IRB (Protocol no. EX-1653). Informed consent was obtained from all users and the data were collected in de-identified form.

Open-source software: We used the following open-source software in the analysis.

- R: http://www.r-project.org
- Tidyverse: http://www.tidyverse.org
- Data.table: https://CRAN.R-project.org/package=data.table
- nnet https://CRAN.R-project.org/package=nnet
- censusapi https://CRAN.R-project.org/package=censusapi
- survey https://CRAN.R-project.org/package=survey
- ggplot2 https://CRAN.R-project.org/package=ggplot2
- cowplot https://CRAN.R-project.org/package=cowplot

Data Collection:

Data were collected from the How We Feel web and mobile application on vaccine hesitancy between December 4th, 2020 and February 9th 2021. Users were asked "If a safe, effective coronavirus vaccine were available, how likely would you be to get yourself vaccinated?" Responses were given on a bipolar 5-point Likert scale from "Very Unlikely" to "Very Likely", with "Undecided" being the middle value. Users were asked the vaccine hesitancy question at regular intervals and the most recent response was used for the analysis. Users also self-reported race/ethnicity, sex, age, occupation, and preexisting conditions. Users who identified as "other" in the gender response were dropped due to small sample size. Median neighborhood income was obtained from the user's zip code by using the American Community Survey 5-year average results from 2018. Population density was calculated at the county level for each user based on data from the Yu Group at University of California at Berkeley [17]. State level case and death rates were obtained from USAFACTS [18].

Race/ethnicity was defined using distinct groups corresponding to “white,” “Black/African-American,” “Hispanic/Latino,” and “Asian” if the user only selected that respective racial group. Users which answered more than one race or ethnicity or selected an option other than the ones listed above were placed in a “multiracial/other” category.

During each login, users reported whether they left their home and for what reason. If they left home, they were then asked what types of protective measurements they used while away (mask, social distancing, cloth mask, and/or avoiding public transportation). We defined “protective behavior” to be if a user either stayed home or wore a mask when outside the home. If the user said that they did not wear a mask outside the home but engaged only in outdoor exercise and maintained physical distance from others, then they were also considered to be practicing protective behavior. We then created a variable that was coded as “1” if they practiced protective behavior during all logins and a “0” if they failed to be protective during at least one login.

Modeling:

Users were considered to be vaccine hesitant if they responded “Very Unlikely,” “Unlikely,” or “Undecided” to the vaccine question. Using vaccine hesitancy as the outcome, a logistic regression was fit using several demographic variables as predictors to identify characteristics of users that were more or less vaccine hesitant. Both a univariate (Supp Tab 3) and a multivariable model (Fig 2, Supp Tab 4) were performed to adjust for potential confounding. Only responses from users residing within the United States were used in the modelling. Corresponding odds ratios and 95% confidence intervals are provided, and statistical significance was assessed at the 0.05 level. Analyses were conducted using R (v 3.5.1).

Using the same covariates as in the logistic regression, a nominal logistic regression was fit to assess if results from the logistic regression were driven by individuals being more likely to be in the “Undecided” or “Unlikely” groups. The 5-point Likert scale was reduced to a 3-level bipolar variable for modelling purposes by combining “Very Unlikely” with “Unlikely” and “Very Likely” with “Likely”.

Weighted Analysis:

To adjust our analyses to a user base that matches the major U.S. census demographics, we implemented a weighted analysis using post-stratification weights. Using the census population estimates of sex, race, age, and census location, a population-based joint

distribution was obtained. A user base distribution was also calculated using the same breakdown, and the two proportions were then matched per user. The post-stratification weight was then calculated by dividing the census proportion by the sample proportion plus 1e-4 to avoid issues with smaller user base probabilities. To avoid over or underweighting individuals, the post-stratification weights were trimmed to be between 0.3 and 3 using the *trimWeights* function in the *survey* R package. The weighted analysis was then conducted using the *svyglm* function. Two separate weighted logistic regressions were conducted. One compared the "Undecided" group vs. the "Likely" group, while the other compared the "Unlikely" group vs. the "Likely" group. To assess the choice of the weight trimming bounds, sensitivity analyses were conducted for both of the above weighted analyses (Supp Tab 7/8) using a threshold of 0.1 and 5. Supplementary Figure 4 provides the distribution of the post-stratification weights.

IPW Analysis:

To formally assess if there was a difference in vaccine hesitancy between those that received a prior positive COVID test and those that received a negative test, we need to adjust for the demographic biases associated with receiving a COVID test. We first fit a weighted logistic regression to model the probability of receiving a test using all individuals and all demographic features that have been reported in previous analyses while applying the same weighted procedure as above. The coefficients, 95% confidence intervals, and p-values for this analysis are available in Supplementary Table 9. The fitted probabilities were then used as inverse probability weights (IPWs) in a weighted logistic regression model for vaccine hesitancy only including individuals which had received a COVID test. The same predictors for previous weighted models were used and a new variable designating if a user tested positive or negative was included. To avoid extreme high or low weights, the fitted probabilities were trimmed to be between 0.1 and 0.9 or 0.05 and 0.95. The results of both of these models are available in Supplementary Table 10.

Acknowledgements:

S.D.M. is supported by the United States National Institutes of Health [grant T32ES007069] and a grant from the Partners in Health during preparation and writing of this manuscript. E.A.H. is supported by the United States National Institutes of Health [grant T32A1007524] during preparation and writing of this manuscript. X.L. is supported by a grant from the Partners in Health. D.S. is supported by United States National Institutes of Health [grant T32GM135117]. F.Z. is supported by the Howard Hughes Medical Institute, the McGovern Foundation, and J.

and P. Poitras and the Poitras Center. The How We Feel Project is a non-profit corporation. The How We Feel Project thanks many operational volunteers and the HWF participants who took our survey and allowed us to share our analysis. Funding and in-kind donations for the How We Feel Project came from B. and D. Silbermann, F. Zhang and Y. Shi, L. Harp McGovern, D. Cheng, A. Azhir, K.H. Yoon and the Bill & Melinda Gates Foundation.

Data Availability:

This work used data from the How We Feel project (http://www.howwefeel.org/). The data are not publicly available, but researchers can apply to use the resource. Researchers with an appropriate IRB approval and data security approval to perform research involving human subjects using the HowWeFeel data can apply to obtain access to data used in the analysis.

Code Availability:

The analysis code developed for this paper can be found online at https://github.com/mccabes292/HWF_VaccineHes_Paper.

Contributions:

E.A.H. and S.D.M. initiated the project. S.D.M. led data analysis and figure production. S.D.M., D. S. and S. S. cleaned the data. E.A.H., and A.S. contributed to analysis. R. M. provided feedbacks on analysis. S.D.M. and E.A.H. wrote the manuscript with M.T. and X.L. D. C., W. A., R.P., B.S., F. Z., X. L. designed and implemented the How We Feel application. E.A.H., L.D.C., and J.R.C., M. T., designed and pilot-tested the vaccine hesitancy question. M.T. and X.L. supervised all aspects of the work.