

FACTORS AFFECTING INDIVIDUAL DECISION MAKING ON THE COVID-19
VACCINE

Health/Medicine/Science B

Anaswar Jayakumar, Elise Perkins, Jimmy Oribiana, Katie Bews

Northwestern University, Introduction to Data Science

March 7, 2021

1. Introduction

The COVID-19 pandemic has affected the globe economically, personally, and politically. The politicization of the disease has given rise to concerns of the efficacy of the vaccines. Studies have shown that quality of the available evidence regarding the efficacy and effectiveness of vaccinations such as the influenza vaccine remains moderate to poor (Michiels et al. 2011). The aforementioned factors are impacting the decisions of citizens on whether or not to get the COVID-19 vaccine. Over 98 million doses have been distributed in the United States and over 68 million have been administered (Nguyen et al. 2020). This is 6.28% of the US population. This is compared to the UK (1.00%), Israel (35.46%), India (0.11%), Germany (2.30%), France (1.97%), Italy (2.24%), and Spain (2.60%; Our World in Data, n.d.).

The goal of our study is to understand which independent variables including age, education, pre-existing conditions, and vaccine beliefs affect a person's decision to receive/not receive the COVID-19 vaccine. We hypothesize that those who meet the following four conditions are more likely to get the COVID-19 vaccine: People over 65 years of age, those with at least one pre-existing health condition, individuals who have at least a high school education, and those who have received the flu vaccine within the last year. Factors such as experience with prior vaccines, efficacy and safety of prior vaccines, and pre-existing health conditions have been shown to affect a person's decision to take the COVID-19 vaccine. Based on a study performed by the CDC (2020), younger adults, non-Hispanic Black persons and adults with lower levels of education were more likely to

report that they would not receive the COVID-19 vaccine. It was also mentioned that efforts are needed to increase the public's confidence in the safety and efficacy of the COVID-19 vaccines, overall and within specific subpopulations (Nguyen et al. 2020).

2. Sampling Methodology

2.1 Population of Interest/Sampling Frame/Sampling Technique

The population of interest is the entire United States population and our sampling frame includes people with either cell phones or landlines. Our survey will be sent out to a random sample of the United States population, 18 years and older using random digit dialing. We will be using quota sampling for this study. In our survey, we plan to divide the population based on subgroups that were used in a prior CDC study (Nguyen et al. 2020). Dividing the population based on the subgroups used in the CDC study will allow us to gain better insight into the factors that affect a person's decision to receive the COVID-19 vaccine. Since we plan to gather demographic information, information about preexisting health conditions, and information about previous vaccine reception in order to better understand what factors affect a person's decision to receive the COVID-19 vaccine, these subgroups are relevant to our study. We also decided to construct our survey and the questions that we plan to ask based on these subgroups so that our survey would be comprehensive while at the same time minimizing any potential sampling bias and errors. The survey that we plan to use is in Appendix A.

2.2 Sampling Equation

To calculate the sample size needed for a representative sample for proportions, we decided to use the Cochran equation which is the following:

$$n_0 = \frac{Z^2 pq}{e^2}$$

where n_0 is the sample size, Z is the critical value, p is the estimated proportion of an attribute that is present in the population, q is $1-p$, and e is the desired level of precision, also known as the sampling error (Israel 1992). In this study, the following values are used for Z , p , q , and e : 1.96, 49.1, 50.9, and 0.05. We chose 1.96 as Z and 0.05 as e because a 95% confidence interval and 5% precision level are commonly used in statistical studies. We chose 49.1 as p , and consequently 50.9 as q , since 49.1% is the most recently reported COVID-19 vaccine acceptance rate according to an earlier CDC study (Nguyen et al. 2020). We plan to send the sample to at least 3,841 people.

To get the final sample size, the sample size obtained by the Cochran equation was then divided by the response rate which was 0.1 since on average 10% of survey participants would be willing to respond to a survey either through cellphone or landline. The random digit dialing method is commonly used by survey organizations such as the Pew Research Center and works by randomly generating the last two digits of a given telephone number based on area code, telephone exchange, and bank number. A bank is a list of 100 such contiguous phone numbers and the telephone exchanges are selected such that they are proportionally stratified based on county and the telephone exchanges that exist in a given county. In other words, the number of telephone numbers that are randomly sampled within a given county is proportional to the given share of telephone numbers of a particular county within the United States (Pew Research, n.d.).

2.3 Survey Design

Our survey is designed to gather data about study participants' demographic information, pre-existing health conditions, and previous vaccine reception. Demographic variables include age, race, sex, highest level of education achieved, household income,

geographic location, and previous vaccine reception. These questions were modeled after the survey carried out by an earlier study which examined attitudes toward a COVID-19 vaccine (Fisher et al. 2020).

Pre-existing health conditions were defined using the CDC's list of underlying medical conditions that are considered to increase an individual's risk for severe illness. These conditions include cancer, chronic kidney disease, chronic obstructive pulmonary disease, Down syndrome, heart conditions, immunocompromised state from solid organ transplant, obesity, pregnancy, sickle cell disease, smoking, and type 2 diabetes (CDC, n.d.). Investigating these existing health conditions is important because people with comorbidities such as multiple pre-existing health conditions or a combination of pre-existing health conditions and a demographic characteristic such as a certain age are more vulnerable to COVID-19 (Michiels et al. 2011). Lastly, we will collect self-reported prior vaccine reception such as whether a person receives the flu vaccine on a yearly basis.

3. Possible errors and biases that should be accounted for in our study

One possible error or bias that should be accounted for in our study is researcher bias. Another possible error could be a non-response error. This is when an acquired sample varies from the initial chosen sample. This could arise because the prospective respondents were not contacted or declined to respond and could lead to lack of data rather than incorrect data. While gathering responses, we would need to make sure that the initial respondents are contributing and possibly use follow-up surveys and substitute ways of reaching them if they do not initially respond. People who do not have landlines or cell phones would not be included in our study population due to limitations surrounding use of the Pew sampling method. Pew Research recognizes young adults are more difficult to interview than older

people because of their lifestyles. Therefore, when a household is contacted, they randomly ask for the youngest male or female 18 years of age. This ensures responses from the younger population are captured (Pew Research, n.d.). Other errors that should be accounted for include population specification error, errors in the sampling frame, and selection error.

4. Possible methods of data preparation and analysis

The method of data preparation we are using is a binary yes/no for our COVID-19 survey questions as our dependent variable. For analysis purposes we will treat “undecided” as “no” since these participants have not confirmed they will receive the vaccine. For analysis, we will use analysis of variance, also called ANOVA. ANOVA will let us find out if variations in the mean values amongst multiple groups occur by chance or if they are considerably distinct. Simply put, ANOVA will test the significance of the survey results (Amaresan 2019). In addition, we will look for correlations between acceptance of the COVID-19 vaccine and these types of information as well. We can utilize the data to understand which variables have the highest correlation (positive or negative) to the decision to take or not take the vaccine. Once we understand these variables, we can utilize a classification algorithm like k to the nearest neighbor to help predict the percent of the general population that will take the vaccine. This could be useful for governments/companies to plan for distribution.

5. How our study can be used in further research and by management

The data produced by this survey can be utilized by governments, law makers, suppliers, providers, and advisory committees to plan for and recommend vaccination allocations with the limited supply. It will also provide an understanding of the perception of vaccinations within certain demographics and prioritizes certain groups such as elderly and

frontline healthcare workers. Several studies found similar percentages and trends in vaccination intent and low likelihood of receiving a COVID-19 vaccine among groups disproportionately affected by COVID-19, including Black persons and those with lower educational attainment. Because many of these groups are at increased risk for COVID-19–associated morbidity and mortality It is important to gain insight into the likelihood that people in groups with a high risk for COVID-19–associated morbidity and mortality will receive the vaccine so that efforts can be made to ensure vaccines are reaching these high risk groups (Nguyen et al. 2020). The results can also help either enforce or debunk the perception of vaccines. Some studies have shown that vaccine acceptance is substantially eroded by conspiratorial thinking and is modestly reduced by political conservatism, suggesting that public views of vaccines have been influenced by anti-vaccination misinformation and politicized discourse in recent years (Sarathchandra et al. 2018).

References

- Amaresan, Swetha. 2019. "How to Analyze Survey Results Like a Data Pro." *Hubspot* (blog). August 9, 2019. <https://blog.hubspot.com/service/survey-results>.
- Aprameya, Amrutha. 2016. "Quota Sampling: When To Use It and How To Do It Correctly" *Humans of Data* (blog). April 6, 2016. <https://humansofdata.atlan.com/2016/04/quota-sampling-when-to-use-how-to-do-correctly/>
- Centers for Disease Control and Prevention (CDC). n.d. "People with Certain Medical Conditions." Accessed February 17, 2021. <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html>.
- Cochran, William, G. 1977. *Sampling Techniques (3rd Edition)*. John Wiley & Sons: New Jersey. https://archive.org/details/Cochran1977SamplingTechniques_201703/page/n149/mode/2up
- Danovaro-Holliday, M. Carolina, Emily Dansereau, Dale A. Rhoda, David W. Brown, Felicity T. Cutts, and Marta Gacic-Dobo. 2018. "Collecting and Using Reliable Vaccination Coverage Survey Estimates: Summary and Recommendations from the 'Meeting to Share Lessons Learnt from the Roll-out of the Updated WHO Vaccination Coverage Cluster Survey Reference Manual and to Set an Operational Research Agenda around Vaccination Coverage Surveys', Geneva, 18–21 April 2017." *Vaccine* 36, no. 34: 5150–59. <https://doi.org/10.1016/j.vaccine.2018.07.019>.
- Fisher, Kimberly A., Sarah J. Bloomstone, Jeremy Walder, Sybil Crawford, Hassan Fouayzi, and Kathleen M. Mazor. 2020. "Attitudes Toward a Potential SARS-CoV-2 Vaccine." *Annals of Internal Medicine* 173, no. 12: 964–73. <https://doi.org/10.7326/m20-3569>.
- Israel, Glenn D. 1992. *Determining Sample Size*. Gainesville, Florida: Institute of Food and Agricultural Sciences. <https://www.tarleton.edu/academicassessment/documents/samplesize.pdf>.
- Lewnard, Joseph A., and Sarah Cobey. 2018. "Immune History and Influenza Vaccine Effectiveness." *Vaccines* 6, no. 28. <https://doi.org/10.3390/vaccines6020028>.
- Michiels, Barbara, Frans Govaerts, Roy Remmen, Etienne Vermeire, and Samuel Coenen. 2011. "A Systematic Review of the Evidence on the Effectiveness and Risks of Inactivated Influenza Vaccines in Different Target Groups." *Vaccine* 29, no. 49: 9159–70. <https://doi.org/10.1016/j.vaccine.2011.08.008>.
- Nguyen, Kimberly H., Anup Srivastav, Hilda Razzaghi, Walter Williams, Megan C. Lindley, Cynthia Jorgensen, Neetu Abad, and James A. Singleton. 2021. "COVID-19 Vaccination Intent, Perceptions, and Reasons for Not Vaccinating Among Groups

- Prioritized for Early Vaccination — United States, September and December 2020.” *MMWR. Morbidity and Mortality Weekly Report* 70, no. 6: 217–22.
<https://doi.org/10.15585/mmwr.mm7006e3>.
- Our World in Data. n.d. “Coronavirus (COVID-19) Vaccinations.” Accessed March 1, 2021
<https://ourworldindata.org/covid-vaccinations>.
- Pew Research. n.d. “Our survey methodology in detail.” Accessed March 1, 2021
<https://www.pewresearch.org/methods/u-s-survey-research/our-survey-methodology-in-detail/>.
- QuestionPro*. n.d. “Quota sampling: Definition, types, examples, steps, and more.” *QuestionPro* (blog). Accessed March 1, 2021.
<https://www.questionpro.com/blog/quota-sampling/>.
- Sarathchandra, Dilshani, Mark C. Navin, Mark A. Largent, and Aaron M. McCright. 2018. “A Survey Instrument for Measuring Vaccine Acceptance.” *Preventive Medicine* 109: 1–7. <https://doi.org/10.1016/j.ypmed.2018.01.006>.

This survey is conducted by Northwestern University Students. It will take less than 5 minutes to complete. Your input will help contribute research that aims to get better understanding of which variables impact a person's decision to participate in the COVID 19 vaccination. We thank you for your participation.

* This form will record your name, please fill your name.

1. What is your age?

- ☐ 18-29
- ☐ 30-44
- ☐ 45-59
- ☐ 60-69
- ☐ ≥ 70

2. What is your sex?

- ☐ Male
- ☐ Female
- ☐ Prefer not to answer

3. What is your race?

- ☐ White
- ☐ Black or African American
- ☐ Hispanic or Latino
- ☐ Asian/Pacific Islander
- ☐ Indigenous North Americans
- ☐ Other
- ☐ Prefer not to answer

4. What is the highest level of education you have achieved?

- ☐ No high school diploma
- ☐ High school graduate or equivalent
- ☐ Some college
- ☐ College graduate or above

5. What is your annual household income?

- ☐ <\$30,000
- ☐ \$30,000-<60,000
- ☐ \$60,000-<100,000
- ☐ >=\$100,000

6. What type of community do you live in?

- ☐ Rural area
- ☐ Small city or town
- ☐ Suburb near a large city
- ☐ Large city

7. Do you have any of the following medical conditions? Cancer, chronic kidney disease, chronic obstructive pulmonary disease (COPD), heart conditions (e.g. heart failure, coronary artery disease, or cardiomyopathies), immunocompromised state from solid organ transplant, obesity, pregnancy, sickle cell disease, smoking, type 2 diabetes, or down syndrome

- ☐ No medical conditions
- ☐ 1 medical condition
- ☐ 2 medical conditions
- ☐ ≥ 3 medical conditions

8. How would you rate your overall health?

- ☐ Very good
- ☐ Good
- ☐ Fair
- ☐ Poor

9. What is your political affiliation?

- ☐ Republican
- ☐ Democrat
- ☐ Libertarian
- ☐ Independent
- ☐ Prefer not to answer

10. Where are you most likely to get your news?

- ☐ Print news (Newspapers, magazines)
- ☐ TV network news (i.e. NBC Nightly News)
- ☐ Local TV News
- ☐ Social media
- ☐ Public radio
- ☐ Podcasts
- ☐ Other

11. Have you had either a flu shot or flu spray in the nose within the past year?

- ☐ Yes
- ☐ No
- ☐ I don't remember

12. What is your best guess as to whether you will get the coronavirus within the next 6 months?

- ☐ I don't think I will get the coronavirus
- ☐ I think I will get a mild case of the coronavirus
- ☐ I think I will get seriously ill from the coronavirus
- ☐ I have already had the coronavirus
- ☐ I am self-isolating until I get the vaccine

13. Do you intend to receive the COVID-19 vaccination?

- ☐ Yes
- ☐ No
- ☐ Undecided

14. Please indicate your level of agreement or disagreement with the following statements:

[illegible]

You have completed the survey!

This concludes the survey. We know your time is valuable and we thank you for your participation.
