Which groups are at highest risk for low uptake of flu vaccine?

Chris Okitondo and Irene Cavros

4/7/23

# 1. Summary/Abstract

Influenza, or flu, is an extremely contagious respiratory illness caused by a virus. It can cause a range of symptoms, including fever, cough, sore throat, body aches, headache, chills, and fatigue (CDC, Key Facts about Seasonal Flu Vaccine, 2022). It can be particularly dangerous for certain groups, such as older adults, young children, and individuals with underlying health issues, often resulting in severe illness or even death  Although influenza vaccinations are widely accessible, many people in the United States do not receive them as recommended. This study examines influenza vaccination uptake among US adults from 2012-2021. Using a subset of the National Health Interview Survey (NHIS) data set, researchers provide important insights into which groups are at highest risk for under-coverage of flu vaccine in order to better inform targeted public health interventions for improving uptake.

# 2. Introduction

### 2.0.1 General Background Information

Influenza, commonly known as the flu, is a contagious respiratory illness which comes from influenza viruses. It can cause mild to severe illness, and in some cases, it can lead to hospitalization or even death. Symptoms of the flu include fever, cough, sore throat, body aches, headache, chills, and fatigue (CDC, About Flu, 2022). It can spread easily from person to person through droplets that are released when infected individuals cough, sneeze, or talk.

Vaccines are a crucial tool in preventing the spread of infectious diseases, including the flu. On an individual level, the best way to protect against the flu is by getting vaccinated each year. At the population level, high vaccine coverage is a powerful strategy for reducing spread of the virus (CDC, Key Facts about Seasonal Flu Vaccine, 2022).

The U.S. Centers for Disease Control and Prevention (CDC) recommends that everyone six months of age and older receive a flu vaccine each year, with rare exceptions (CDC, Key Facts about Seasonal Flu Vaccine, 2022). Influenza is a constantly evolving virus that mutates rapidly, and uncontrolled spread gives rise to many different strains, so new vaccines are developed each year to protect against the most common strains of the virus that are expected to circulate during the upcoming flu season. It is especially important for individuals who are at higher risk of severe illness or complications from the flu, such as young children, older adults, pregnant women, and individuals with underlying medical conditions.

In terms of the availability and cost, there have been significant changes for flu vaccines over the past 50 years (CDC, Influenza Historic Timeline, 2019). In the 1960s, flu vaccines were not widely available and were primarily given to people considered to be at high risk for complications. The cost of the vaccine was relatively high and was not covered by most insurance plans. In the 1970s and 1980s, flu vaccines became more widely available and were recommended for more people, including children and healthy adults. The cost of the vaccine began to decrease as more people received it, and insurance coverage for the vaccine became more common. In the 1990s and 2000s, the availability of flu vaccines continued to increase, and the cost of the vaccine became more affordable for a larger proportion of the U.S. population.

Today, flu vaccines are frequently available in doctor’s offices, pharmacies, community clinics, workplace clinics, school-based clinics, and local health departments, among others. They are often covered by insurance plans, which has increased their accessibility. Many public health clinics and community organizations also offer flu vaccines at low or no cost, in order to better target historically unreached populations.

Despite the availability and recommendation of influenza vaccines, many people in the United States still do not receive them as recommended. In fact, the CDC reported that only 49.2% of adults aged 18 years and older received a flu vaccine during the 2020-2021 flu season (CDC, Inequities in Flu Vaccine Uptake 2022). This low uptake of influenza vaccine highlights the need for understanding the factors that contribute to under-coverage of the vaccine in certain populations.

Racial/ethnic minorities have a long history of reduced vaccination rates relative to white Americans (Khan 2018). Since 2010, flu vaccination coverage has been consistently lower among Black, Hispanic, and AI/AN adults. During the 2021–2022 season, flu vaccination coverage was 54% among White adults, 42% among Black adults, 38% among Hispanic adults, and 41% among AI/AN adults (CDC, Inequities in Flu Vaccine Uptake 2022). There are many reasons for these inequities, including lack of access to health care and insurance, missed opportunities to vaccinate, and misinformation and distrust.

Racial and ethnic minorities have a longstanding history of lower vaccination rates compared to white Americans. Since 2010, Black, Hispanic, and American Indian/Alaska Native adults have consistently had lower flu vaccination rates than White adults. In the 2021-2022 flu season, vaccination coverage for White adults was 54%, while it was 42% for Black adults, 38% for Hispanic adults, and 41% for AI/AN adults (CDC, Inequities in Flu Vaccine Uptake 2022).  The reasons for these disparities are multifactorial, and may include limited access to healthcare and insurance, missed opportunities for vaccination, and misinformation and mistrust.

In prior studies, lower income levels were associated with decreased odds of influenza vaccine receipt (Gaskin 2022, Vashist 2022), and having health insurance coverage has also been shown to be associated with increased influenza vaccine coverage (Cambou 2021, Vashist 2022). Influenza vaccination rates were also found to be lower in rural than in urban areas among adults of all age groups and both genders (Jain 2022). However, there are still many variables of interest from the NHIS dataset that have yet to be evaluated in detail. Additionally, the periods of study from all prior research all end prior to the onset of the COVID-19 pandemic, which could have altered or heightened disparities even more.

# 3. Methods

### 3.0.1 Description of data and data source

The data source used for this project is the National Health Interview Survey (NHIS). It is a harmonized set of data beginning in 1963 with information on general health status, acute and chronic illness, functional limitations, access to care, insurance coverage, and health behaviors for the U.S. population. On average, the survey covers 100,000 persons in 45,000 households each year. It is conducted in-person and by telephone, and its sample is designed to be representative of the civilian United States population (NHIS, About).

#### 3.0.1.1 Study population

Individuals were eligible for the study if they were interviewed as part of the NHIS household survey between the years of 2012 and 2021. Table 1 below provides a summary of demographic characteristics of our study population.

### 3.0.2 Statistical analysis

The raw data for the period of interest directly extracted from the NHIS IPUMS survey initially had 792,916 observations. NHIS Researchers then selected their predictors of interest to model the association between the outcome and variables. Variables assessed are as follows:

1. Demographics: age, sex, race, ethnicity, veteran status

2. Socioeconomic status: education level, total combined family income, employment status, paid sick leave as part of job benefits

3. General health: perceived health status, categorical BMI

4. Conditions: ever told had asthma, ever told had cancer, ever told had coronary heart disease

5. Health behaviors: alcohol drinking status, cigarette smoking status

6. Access to care: has usual place for medical care, medical care delayed due to cost (past 12 months)

7. Health insurance: health insurance coverage status

8. Mental health: anxiety level, depression level.

After processing the raw data and removing all observations with missing data for any of the predictors of interest, the remaining 106,980 individuals were included in the analysis. Because of the binary outcome, logistic regression was utilized to model the association between the outcome and variables of interest. Researchers utilized machine learning to split data and create a workflow. Data was split randomly where 75% of included participants made up the training dataset and 25% made up the testing dataset.

# 4. Results

### 4.0.1 Exploratory/Descriptive analysis

|  |
| --- |
|  |

\*\* Table 1: Demographic characteristics of the study population \*\*

### 4.0.2 Loading Table 1 from the Exploratory Analysis

result\_table1 <- readRDS(here("results", "table1\_demographics.rds"))  
result\_table1

|  | no | yes | Overall |
| --- | --- | --- | --- |
|  | (N=58425) | (N=48555) | (N=106980) |
| **factor(sex)** | | | |
| male | 29881 (51.1%) | 20902 (43.0%) | 50783 (47.5%) |
| female | 28544 (48.9%) | 27653 (57.0%) | 56197 (52.5%) |
| **Age** | | | |
| Mean (SD) | 45.0 (16.1) | 54.6 (17.8) | 49.3 (17.6) |
| Median [Min, Max] | 44.0 [18.0, 85.0] | 57.0 [18.0, 85.0] | 49.0 [18.0, 85.0] |
| **factor(race)** | | | |
| white | 46388 (79.4%) | 40081 (82.5%) | 86469 (80.8%) |
| black | 8356 (14.3%) | 5090 (10.5%) | 13446 (12.6%) |
| native american | 643 (1.1%) | 496 (1.0%) | 1139 (1.1%) |
| asian | 3038 (5.2%) | 2888 (5.9%) | 5926 (5.5%) |
| **factor(hispanic)** | | | |
| non-hispanic | 49333 (84.4%) | 43787 (90.2%) | 93120 (87.0%) |
| hispanic | 9092 (15.6%) | 4768 (9.8%) | 13860 (13.0%) |
| **factor(education)** | | | |
| less\_than\_hs | 6784 (11.6%) | 4528 (9.3%) | 11312 (10.6%) |
| high\_school | 15161 (25.9%) | 10228 (21.1%) | 25389 (23.7%) |
| some\_college | 19099 (32.7%) | 14086 (29.0%) | 33185 (31.0%) |
| bachelor | 11690 (20.0%) | 11268 (23.2%) | 22958 (21.5%) |
| graduate | 5691 (9.7%) | 8445 (17.4%) | 14136 (13.2%) |
| **factor(fam\_income)** | | | |
| < $50,000 | 29164 (49.9%) | 20515 (42.3%) | 49679 (46.4%) |
| $50,000-$99,999 | 17097 (29.3%) | 14222 (29.3%) | 31319 (29.3%) |
| $100,000+ | 12164 (20.8%) | 13818 (28.5%) | 25982 (24.3%) |
| **factor(employ\_status)** | | | |
| unemployed | 2484 (4.3%) | 931 (1.9%) | 3415 (3.2%) |
| employed | 42869 (73.4%) | 29153 (60.0%) | 72022 (67.3%) |
| retired | 13072 (22.4%) | 18471 (38.0%) | 31543 (29.5%) |
| **factor(paid\_sick)** | | | |
| no | 27943 (47.8%) | 16920 (34.8%) | 44863 (41.9%) |
| yes | 30482 (52.2%) | 31635 (65.2%) | 62117 (58.1%) |
| **factor(health\_ins\_status)** | | | |
| no | 9935 (17.0%) | 1869 (3.8%) | 11804 (11.0%) |
| yes | 48490 (83.0%) | 46686 (96.2%) | 95176 (89.0%) |
| **factor(med\_access)** | | | |
| no | 11046 (18.9%) | 2547 (5.2%) | 13593 (12.7%) |
| yes | 47379 (81.1%) | 46008 (94.8%) | 93387 (87.3%) |
| **factor(med\_care\_delayed\_cost)** | | | |
| no | 51052 (87.4%) | 45059 (92.8%) | 96111 (89.8%) |
| yes | 7373 (12.6%) | 3496 (7.2%) | 10869 (10.2%) |
| **factor(health\_status)** | | | |
| excellent | 16670 (28.5%) | 11618 (23.9%) | 28288 (26.4%) |
| very\_good | 20113 (34.4%) | 16656 (34.3%) | 36769 (34.4%) |
| good | 15133 (25.9%) | 13376 (27.5%) | 28509 (26.6%) |
| fair | 5199 (8.9%) | 5261 (10.8%) | 10460 (9.8%) |
| poor | 1310 (2.2%) | 1644 (3.4%) | 2954 (2.8%) |
| **factor(bmi)** | | | |
| underweight | 964 (1.7%) | 633 (1.3%) | 1597 (1.5%) |
| average | 19469 (33.3%) | 15396 (31.7%) | 34865 (32.6%) |
| overweight | 20276 (34.7%) | 17002 (35.0%) | 37278 (34.8%) |
| obese | 17716 (30.3%) | 15524 (32.0%) | 33240 (31.1%) |
| **factor(asthma)** | | | |
| no | 51449 (88.1%) | 41304 (85.1%) | 92753 (86.7%) |
| yes | 6976 (11.9%) | 7251 (14.9%) | 14227 (13.3%) |
| **factor(cancer)** | | | |
| no | 54623 (93.5%) | 41766 (86.0%) | 96389 (90.1%) |
| yes | 3802 (6.5%) | 6789 (14.0%) | 10591 (9.9%) |
| **factor(chd)** | | | |
| no | 56727 (97.1%) | 45045 (92.8%) | 101772 (95.1%) |
| yes | 1698 (2.9%) | 3510 (7.2%) | 5208 (4.9%) |
| **factor(alcohol\_status)** | | | |
| nondrinker | 9219 (15.8%) | 7708 (15.9%) | 16927 (15.8%) |
| former\_drinker | 7791 (13.3%) | 8159 (16.8%) | 15950 (14.9%) |
| drinker | 41415 (70.9%) | 32688 (67.3%) | 74103 (69.3%) |
| **factor(smoking\_status)** | | | |
| nonsmoker | 34750 (59.5%) | 29338 (60.4%) | 64088 (59.9%) |
| former\_smoker | 12027 (20.6%) | 13685 (28.2%) | 25712 (24.0%) |
| smoker | 11648 (19.9%) | 5532 (11.4%) | 17180 (16.1%) |

\*\*

### 4.0.3 Full analysis

##### 4.0.3.0.1 Model 1: Focusing on only demogrphics characteristics using the flu vaccine uptake as our outcome

term estimate std.error statistic  
1 (Intercept) -2.412331604 0.0334164527 -72.1899366  
2 age 0.034347387 0.0003974164 86.4266897  
3 sexfemale 0.355237904 0.0130976192 27.1223264  
4 raceblack -0.240890761 0.0204033476 -11.8064332  
5 racenative american 0.231326583 0.0632861504 3.6552481  
6 raceasian 0.114989390 0.0284889803 4.0362761  
7 hispanichispanic -0.177889489 0.0209758170 -8.4806942  
8 educationhigh\_school 0.004245455 0.0249208828 0.1703573  
9 educationsome\_college 0.185041675 0.0244858985 7.5570710  
10 educationbachelor 0.424188493 0.0265939096 15.9505879  
11 educationgraduate 0.701967529 0.0294263747 23.8550462  
12 fam\_income$50,000-$99,999 0.132219857 0.0159985840 8.2644724  
13 fam\_income$100,000+ 0.353068166 0.0179804150 19.6362635  
 p.value  
1 0.000000e+00  
2 0.000000e+00  
3 5.370848e-162  
4 3.615769e-32  
5 2.569334e-04  
6 5.430631e-05  
7 2.238543e-17  
8 8.647291e-01  
9 4.122468e-14  
10 2.822273e-57  
11 8.976145e-126  
12 1.403139e-16  
13 7.577252e-86

### 4.0.4 Model 2: Using all our predictors of interest in the model with flu vaccine uptake as our outcome

term estimate std.error statistic  
1 (Intercept) -2.239921505 0.0379425281 -59.034588  
2 age 0.030696805 0.0004333642 70.833735  
3 sexfemale 0.347985133 0.0133861311 25.995945  
4 raceblack -0.236115703 0.0206769405 -11.419277  
5 racenative american 0.250033086 0.0637233155 3.923730  
6 raceasian 0.132068451 0.0288590192 4.576332  
7 hispanichispanic -0.192490023 0.0212801177 -9.045534  
8 educationhigh\_school 0.004827906 0.0252706377 0.191048  
9 educationsome\_college 0.158552543 0.0249416764 6.356932  
10 educationbachelor 0.378764844 0.0272631661 13.892915  
11 educationgraduate 0.647484649 0.0301351349 21.486038  
12 fam\_income$50,000-$99,999 0.130690475 0.0162269769 8.053902  
13 fam\_income$100,000+ 0.348323226 0.0183432206 18.989208  
14 asthmayes 0.334923861 0.0193313991 17.325381  
15 canceryes 0.279409838 0.0230819284 12.105134  
16 chdyes 0.492680238 0.0324541351 15.180815  
17 smoking\_statusformer\_smoker 0.056881954 0.0165999208 3.426640  
18 smoking\_statussmoker -0.401446680 0.0199201482 -20.152796  
19 alcohol\_statusformer\_drinker 0.057943801 0.0243219525 2.382366  
20 alcohol\_statusdrinker -0.022146202 0.0195529294 -1.132628  
 p.value  
1 0.000000e+00  
2 0.000000e+00  
3 5.503490e-149  
4 3.350107e-30  
5 8.718844e-05  
6 4.731995e-06  
7 1.489357e-19  
8 8.484880e-01  
9 2.058229e-10  
10 6.993054e-44  
11 2.103115e-102  
12 8.019570e-16  
13 2.094648e-80  
14 3.026826e-67  
15 9.920715e-34  
16 4.738537e-52  
17 6.110985e-04  
18 2.543362e-90  
19 1.720178e-02  
20 2.573704e-01

### 4.0.5 Model Evaluation

#### 4.0.5.1 ROC curve from the full model

|  |
| --- |
|  |

#### 4.0.5.2 ROC curve from the reduced model

|  |
| --- |
|  |

# 5. Discussion

### 5.0.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

\*\*To be completed (PART 5)

### 5.0.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

\*\*To be completed (PART 5)

### 5.0.3 Conclusions

*What are the main take-home messages?*

\*\*To be completed (PART 5)

# 6. References

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

\*\*To be completed (PART 5)