Knowledge about Diseases and Vaccination Decision: Implications from Propensity Score Analysis on the U.S. National 2009 H1N1 Flu Survey

Yuxin Yang 1004834673

17/12/2020

Code and data supporting this analysis is available at:

https://github.com/yangyu77/Vaccination_propensity_score_analysis.git (https://github.com/yangyu77/Vaccination_propensity_score_analysis.git)

Abstract

As the COVID-19 pandemic has impacted all nations and populations worldwide, getting vaccines become an urge. However, many potential factors affect the general public's attitudes to vaccination. The present study investigate whether there is a causal inference between individual knowledge about the disease and the vaccination behavior, hoping to provide insights on methods of improving the vaccination rate.

The present study analyzes public-use dataset which obtained from the United States National 2009 H1N1 Flu Survey, because the H1N1 flu is analogous to the COVID-19 pandemic. Propensity score matching and multivariable logistic regression is used to simulate a quasi-experimental design for the obsevational data.

The results confirm the causal inference between individual knowledge and behavior of getting vaccinated, indicating a possible method for improving the will to get vaccine: educating the general public about the disease may increase vaccination rate among society.

Keywords: propensity score analysis; multivariable logistic regression; vaccination; H1N1 flu;

1 Introduction

Moving into 2020, the COVID-19 pandemic has impacted all nations and populations worldwide. Faced with the unstoppable spread of the virus, developing vaccines against this new disease has become the best hope and an urge of many countries (Pouge et al., 2020). However, developing vaccines is quite a different thing from general people accepting the vaccination. More than a few recent studies suggest the public's hesitancy toward vaccination (Chou & Budenz, 2020; Salali & Uysal, 2020). Similarly, the H1N1 influenza pandemic occoured in 2009 reduced public confidence in vaccination. This negative response toward vaccination will cause serious problem to the overall safety of countries under severe situations of the pandemic, as a vaccine could only be useful for the

whole society to back to normal when a sufficiently high percentage of people get vaccinated, yielding widespread immunity (Salali & Uysal, 2020). To possibly reverse this hesitancy toward vaccination, understanding the reasons behind the hesitancy is undoubtedly crucial.

The present study aims to investigate whether the personal knowledge about the diseases have a causal effect on the intent of getting vaccinated. Since the H1N1 influenza pandemic is analogous to the situation today, the H1N1 vaccine program in 2009 can provide a good model. The dataset used for analysis was extracted from the national 2009 H1N1 flu suery on CDC website. A logistic regression model is built to estimate the propensity scores for observations. A *propensity score matching* is used to examine if there is a causal link between individuals' knowledge/understanding about the disease and their intent of getting vaccination. If the causal relationship establishes, then educating the general public about the disease would be an important way to increase vaccination rate among society.

2 Methodology

2.1 Data

Source

The public-use data file used for analysis was obtained from the National 2009 H1N1 Flu Survey (NHFS) implemented by the National Center for Immunization and Respiratory Diseases (NCIRD) and the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC).

Target population

The target population for the NHFS survey was all persons in the Unted States aged 6 months and older.

Frame

The NHFS survey was conducted using a dual-frame sample design, including interviews obtained by both landline telephones and cellular phones from the fifty U.S. states and the District of Columbia.

Sampling methodology and the sample

The NHFS was conducted starting from October 2009 through June 2010. The sample of telephone numbers was obtained with stratification. There are 51 geographic strata for which vaccine coverage levels can be estimated (50 states plus the District of Columbia). To meet the target precisionrequirements, the number of sample telephone number needed in each stratum was pre-calculated by statistical models (see **Figure 1**), and the samples of telephone numbers were drawn independently for each calendar quarter within strata. Additionally, the sample for each state was further divided into random sub-sanples in order to spread the interviews for each sampling area evenly across the entire calendar quarter. Non-working and non-residential landline telephone numbers were eliminated from the sample. There were 396,096 landline telephone numbers and 246,416 cellular phone numbers being released for survey interviewing.
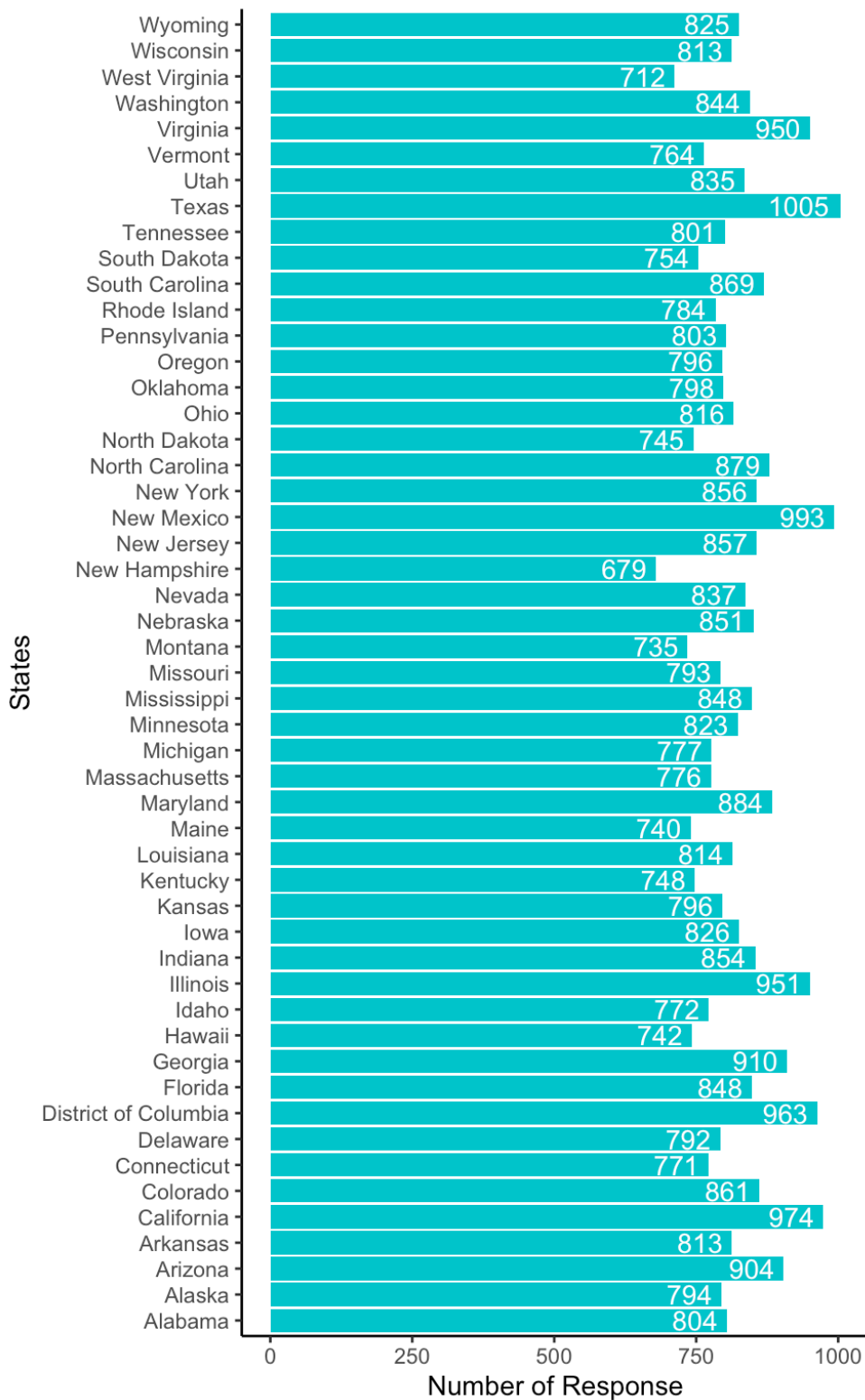
Figure 1. Data Overview: Response count by State

During the survey process, a subset of sampled telephone lines were released each Sunday to the telephone center for dialing. For interviews through landline telephones, one adult was randomly selected as the respondent from all age-eligible (>=18 years) adults in the household. If the household also contained one or more children under 18,

one child was selected at random to take a similar interview as the adult. Cellular telephone numbers were screened to verify that the phone belonged to an adult for personal use, and also that this adult resided in a "cell-phone-only" or "cell-phone- mainly" household.

Key features, strenths and weakness

The NHFS, as a Computer Assisted Telephone Interview survey, collected a large amount of influenza-related data nation-wide. The data includes information about the pH1N1 and seasonal influenza vaccination coverage status, individual knowledge and level of concerns about the H1N1 flu virus, the H1N1 vaccine, etc. The survey also asked whether the respondent had engaged in any of a number of behaviors in response to the potential for an H1N1 flu outbreak, as well as whether they were in any of the selected chronic medical conditions. Demographic information was also requested.

Since the NHFS was a telephone survey, results were weighted to be representative of all persons age 6 months and older. The base sampling weights of adults residing in a household were adjusted according to the number of adults in the household, because they were less likely to be sampled than adults living alone.

Although statistical adjustments were made to account for nonresponse and undercoverage, some bias might remain. Moreover, the data was obtained solely from respondent-reported information. Therefore, the data may reflect recall error from the sampled adult about themselves and the household, and from the adult about the status of the selected child. The NHFS did not collect the nature of the relationship between the selected child and the adult respondent. This may also cause minor effects to the analysis results.

Selected variables

The present study generates a R-format data file `NHFSPUF.RData` from the R program `NHFSPUF.R` provided by the NHFS documentation. The data file `NHFSPUF.RData` is then cleaned and wrangled for analyzing the causal effect of knowledge on the intent to get vaccinated. All missing values, "refuse to answer" values and "unknown" values are excluded from the dataset. All children aged below 18 did not have data about self-reported concerns or opinions, so they were not included in `H1N1_data`. Therefore, a total of 32,071 observations is analyzed.

A list of variables are selected to be included in the prepared dataset `H1N1_data` (See a showcase in Table 1, 2, 3):

- **VACC_H1N1_F**: An indicator for wether the respondent get vaccinated. Value 1 indicates "yes", value 0 indicates "no".

- **VACC_SEAS_COUNT**: The number of seasonal flu vaccinations get by the respondent. The value is allowed to be greater than one only for children under 9 years of age.

- **BEHAVIOR**: This variable indicates the number of actions the respondent took to avoid inffection. It is constructed by summing seven behavioral indicator (value 0 or 1) from the original dataset.

- **CONCERN_H1N1**: This variable indicates the self-reported level of concern about the H1N1 flu. It is constructed by integrating distint levels of concern about H1N1 flu in the original data file. It consists of four levels: "Not at all", "Not very", "Somewhat", "Very".

- **OP_EFFECTIVE_VACC**: Self-reported opinion on the effectiveness of H1N1 vaccine. It consists of four levels: "Very Effective", "Somewhat Effective", "Not Very Effective", and "Not At All Effective".

- **RISK_SICK_WO_VACC**: Percieved risk of getting sick with H1N1 flu without vaccine. It consists of four levels: "Very High", "Somewhat High", "Somewhat Low", "Very Low".

- **WORRY_SICK_W_VACC**: Self-reported extent of worry about getting sick from the H1N1 vaccine. There are four levels: "Very Worried", "Somewhat Worried", "Not Very Worried", and "Not At All Worried".

- **AGE_GROUP**: The age group the respondent belongs to. The variable contains five age groups: "18-34 Years", "35-44 Years", "45-54 Years", "55-64 Years", "65+ Years"

- **SEX_I**: The gender of respondent.

- **EDU**: The adult self-reported education level. There are four levels: "< 12 Years", "12 Years", "Some College", "College Graduate".

- **STATE_NAME**: The self-reported state location. If data is missing or refused to answer, then it equals the sampling state based on telephone exchange.

- **KNOW_H1N1**: The self-reported level of knowledge about H1N1 flu.Value 1 indicates "know a lot", while 0 indicates "knowing little or none". It serves as the treatment in the present study.

Table 1: Variable 1 to 4

	VACC_H1N1_F	VACC_SEAS_COUNT	BEHAVIOR	CONCERN_H1N1
1	0	0	4	Somewhat
3	0	1	4	Somewhat
4	0	0	5	Not very
6	0	1	2	Not very
7	0	0	2	Very
8	0	1	2	Somewhat

Table 2: Variable 5 to 7

	OP_EFFECTIVE_VACC	RISK_SICK_WO_VACC	WORRY_SICK_W_VACC
1	Somewhat Effective	Very Low	Not Very Worried
3	Somewhat Effective	Somewhat High	Somewhat Worried
4	Very Effective	Somewhat Low	Not At All Worried
6	Very Effective	Somewhat Low	Not Very Worried
7	Very Effective	Very High	Very Worried
8	Somewhat Effective	Somewhat Low	Not Very Worried

Table 3: Variable 8 to 12

	AGE_GROUP	SEX_I	EDU	STATE_NAME	KNOW_H1N1
1	65+ Years	Female	Some College	New Mexico	0
3	55-64 Years	Female	12 Years	Idaho	1
4	45-54 Years	Female	College Graduate	Oregon	1
6	55-64 Years	Female	Some College	New York	1
7	45-54 Years	Male	12 Years	Idaho	0
8	65+ Years	Female	College Graduate	Louisiana	0

The figure 2 clearly suggests that respondents' level of knowledge about H1N1 flu is related to the vaccination decision. Respondents with a lot of knowledge about the disease are more likely to get vaccinated. Further analysis is needed to examine the causal inference.
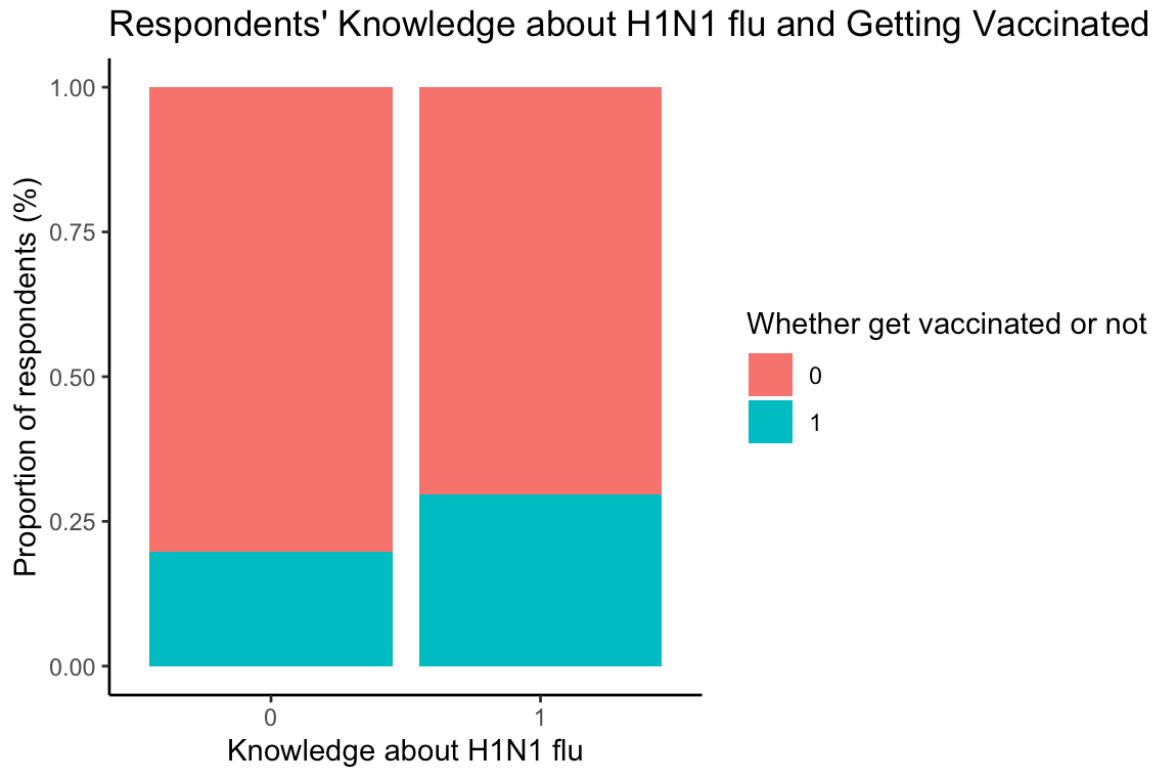
Respondents' Knowledge about H1N1 flu and Getting Vaccinated

Figure 2. Vaccination and knowledge about H1N1 flu

2.2 Propensity Score Analysis

To find whether there is a causal inference between whether having knowledge about the virus and vaccination, the present paper utilized propensity score analysis to simulate a quasi-experiment as random assignment of respondents is infeasible. Propensity scores is a statistical technique useful for evaluating treatment effects when using quasi-experimental or observational data (Rosenbaum & Rubin, 1983). Basically, the propensity score methods improve the comparability of treatment groups on observed characteristics and reduce bias in the effect estimates (Rosenbaum & Rubin, 1983).

There are two assumptions associated with causality that we need to understand before we can use propensity scores. First, the independent variables in the model and the error term should be independent of each other. Ommiting variables is a major problem that violates this assumption by which creating a biased estimate of the effect. Second, the assignment of study participants to treatment conditions is independent of the outcome. This is known as the Ignorable Treatment Assignment Assumption (Olmos & Govindasamy, 2015). To possibly preserve these assumptions, the balance of covariates between treated and the untreated groups need to be examined.

Preliminary Analysis

Before calculating the propensity scores, it is crucial to check the balance of covariates between treated and the untreated groups. A Chi-square test can be used to test for the difference in distribution of categorical variables between two independent groups (Hazra & Gogtay, 2016). The resulting table (p-value = 0) indicates that there is at least one variable creating an imbalance between the two groups.

	chisquare	df	p.value
unstrat	4522.884	72	0

Estimation of Propensity Scores (Model)

The analyses were conducted using RStudio Version 1.2.5042 software. A propensity score for observation was calculated to adjust for potential confounding. Propensity scores were estimated using a multivariable logistic regression model. A logistic regression model is used not only for it allows binary dependent variable but also for it is the default option for propensity scores estimation. The model includes all variables to reduce the chance of hidden bias and avoid imbalance between treated and untreated groups.

The general form of the multivariable logistic regression model is:

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon_i$$

The multivariable logistic regression model specifically for the analysis is:

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_{vacc\_seas\_count} + \beta_2 X_{BEHAVIOR} + \beta_3 X_{CONCERN\_H1N1Notvery} + \beta_4 X_{CONCERN\_H1N1Somewhat} +$$

$$\beta_5 X_{CONCERN\_H1N1Very} + \beta_6 X_{OP\_EFFECTIVE\_VACCNotVeryEffective} + \beta_7 X_{OP\_EFFECTIVE\_VACCSomewhatEffective} +$$

$$\beta_8 X_{OP\_EFFECTIVE\_VACCVeryEffective} + \beta_9 X_{RISK\_SICK\_WO\_VACCSomewhatLow} + \beta_{10} X_{RISK\_SICK\_WO\_VACCVeryHigh} +$$

$$\beta_{11}X_{RISK\_SICK\_WO\_VACCVeryLow} + \beta_{12}X_{WORRY\_SICK\_W\_VACCNotVeryWorried} +$$

$$\beta_{13}X_{WORRY\_SICK\_W\_VACCSomewhatWorried} + \beta_{14}X_{WORRY\_SICK\_W\_VACCVeryWorried} + \beta_{15}X_{AGE\_GROUP35-44Years} +$$

$$\beta_{16}X_{AGE\_GROUP45-54Years} + \beta_{17}X_{AGE\_GROUP55-64Years} + \beta_{18}X_{AGE\_GROUP65+Years} + \beta_{19}X_{SEX\_IMale} +$$

$$\beta_{20}X_{EDU12Years} + \beta_{21}X_{EDUCollegeGraduate} + \beta_{22}X_{EDUSomeCollege} + \beta_{23}X_{STATE\_NAMEAlaska} +$$

$$\beta_{24}X_{STATE\_NAMEArizona} + \beta_{25}X_{STATE\_NAMEArkansas} + \ldots + \beta_{71}X_{STATE\_NAMEWestVirginia} +$$

$$\beta_{72}X_{STATE\_NAMEWisconsin} + \beta_{73}X_{STATE\_NAMEWyoming}$$

The vaariable p in the log odds $log\left(\frac{p}{1-p}\right)$ represents the best guess at the probability of respondent having a lot of knowledge about the H1N1 flu, regardless of whether the respondents have knowledge or not.

The values of all β coefficients are listed in the "Estimate" column.

```
## 
## Call:
## glm(formula = KNOW_H1N1 ~ VACC_SEAS_COUNT + BEHAVIOR + CONCERN_H1N1 +
##     OP_EFFECTIVE_VACC + RISK_SICK_WO_VACC + WORRY_SICK_W_VACC +
##     AGE_GROUP + SEX_I + EDU + STATE_NAME, family = binomial,
##     data = H1N1_data)
## 
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8350  -0.9610  -0.6626   1.1454   2.3644
## 
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       -1.871154   0.117990 -15.859  < 2e-16 ***
## VACC_SEAS_COUNT                    0.349258   0.023379  14.939  < 2e-16 ***
## BEHAVIOR                           0.062385   0.009650   6.465 1.01e-10 ***
## CONCERN_H1N1Not very              -0.129784   0.039489  -3.287 0.001014 **
## CONCERN_H1N1Somewhat               0.084822   0.040232   2.108 0.035002 *
## CONCERN_H1N1Very                   0.190061   0.046673   4.072 4.66e-05 ***
## OP_EFFECTIVE_VACCNot Very Effective -0.146973  0.067370  -2.182 0.029141 *
## OP_EFFECTIVE_VACCSomewhat Effective -0.232652  0.059379  -3.918 8.93e-05 ***
## OP_EFFECTIVE_VACCVery Effective     0.072565   0.061013   1.189 0.234310
## RISK_SICK_WO_VACCSomewhat Low      -0.213033   0.029093  -7.322 2.44e-13 ***
## RISK_SICK_WO_VACCVery High          0.214360   0.046318   4.628 3.69e-06 ***
## RISK_SICK_WO_VACCVery Low          -0.250360   0.033762  -7.416 1.21e-13 ***
## WORRY_SICK_W_VACCNot Very Worried  -0.297755   0.027117 -10.980  < 2e-16 ***
## WORRY_SICK_W_VACCSomewhat Worried  -0.343615   0.031350 -10.960  < 2e-16 ***
## WORRY_SICK_W_VACCVery Worried      -0.300307   0.045684  -6.574 4.91e-11 ***
## AGE_GROUP35-44 Years               0.310239   0.036468   8.507  < 2e-16 ***
## AGE_GROUP45-54 Years               0.413444   0.033778  12.240  < 2e-16 ***
## AGE_GROUP55-64 Years               0.314620   0.033719   9.331  < 2e-16 ***
## AGE_GROUP65+ Years                -0.109446   0.035144  -3.114 0.001844 **
```

```
## SEX_IMale                       -0.281262  0.022253 -12.640  < 2e-16 ***
## EDU12 Years                      0.589841  0.052239  11.291  < 2e-16 ***
## EDUCollege Graduate              1.687724  0.049803  33.888  < 2e-16 ***
## EDUSome College                  1.140691  0.050622  22.534  < 2e-16 ***
## STATE_NAMEAlaska                 0.262577  0.110956   2.366 0.017958 *
## STATE_NAMEArizona                0.414455  0.106942   3.875 0.000106 ***
## STATE_NAMEArkansas              -0.130058  0.114519  -1.136 0.256085
## STATE_NAMECalifornia             0.229030  0.105807   2.165 0.030419 *
## STATE_NAMEColorado               0.383517  0.107857   3.556 0.000377 ***
## STATE_NAMEConnecticut            0.281463  0.111133   2.533 0.011320 *
## STATE_NAMEDelaware               0.218055  0.111151   1.962 0.049787 *
## STATE_NAMEDistrict of Columbia   0.265533  0.105026   2.528 0.011463 *
## STATE_NAMEFlorida                0.427325  0.108510   3.938 8.21e-05 ***
## STATE_NAMEGeorgia                0.016295  0.109326   0.149 0.881514
## STATE_NAMEHawaii                 0.014475  0.113843   0.127 0.898822
## STATE_NAMEIdaho                  0.295595  0.112145   2.636 0.008393 **
## STATE_NAMEIllinois               0.117044  0.107019   1.094 0.274094
## STATE_NAMEIndiana                0.126822  0.110832   1.144 0.252509
## STATE_NAMEIowa                   0.217942  0.110692   1.969 0.048965 *
## STATE_NAMEKansas                 0.166750  0.111293   1.498 0.134054
## STATE_NAMEKentucky               0.053597  0.114853   0.467 0.640742
## STATE_NAMELouisiana             -0.180499  0.114010  -1.583 0.113380
## STATE_NAMEMaine                  0.514472  0.111786   4.602 4.18e-06 ***
## STATE_NAMEMaryland               0.348942  0.107442   3.248 0.001163 **
## STATE_NAMEMassachusetts          0.453304  0.110461   4.104 4.07e-05 ***
## STATE_NAMEMichigan               0.190998  0.112020   1.705 0.088186 .
## STATE_NAMEMinnesota              0.364961  0.109465   3.334 0.000856 ***
## STATE_NAMEMississippi           -0.269169  0.114392  -2.353 0.018621 *
## STATE_NAMEMissouri               0.194008  0.112137   1.730 0.083614 .
## STATE_NAMEMontana                0.321746  0.113006   2.847 0.004411 **
## STATE_NAMENebraska               0.173807  0.109805   1.583 0.113450
## STATE_NAMENevada                 0.248579  0.110133   2.257 0.024003 *
## STATE_NAMENew Hampshire          0.501218  0.113870   4.402 1.07e-05 ***
## STATE_NAMENew Jersey             0.188901  0.109079   1.732 0.083313 .
## STATE_NAMENew Mexico             0.219564  0.105585   2.079 0.037572 *
## STATE_NAMENew York               0.308298  0.108354   2.845 0.004437 **
## STATE_NAMENorth Carolina         0.082597  0.109684   0.753 0.451422
## STATE_NAMENorth Dakota           0.009101  0.115027   0.079 0.936934
## STATE_NAMEOhio                   0.242652  0.110577   2.194 0.028205 *
## STATE_NAMEOklahoma               0.128638  0.112237   1.146 0.251740
## STATE_NAMEOregon                 0.353370  0.110405   3.201 0.001371 **
## STATE_NAMEPennsylvania           0.378414  0.110256   3.432 0.000599 ***
## STATE_NAMERhode Island           0.431497  0.110656   3.899 9.64e-05 ***
## STATE_NAMESouth Carolina        -0.017162  0.110447  -0.155 0.876517
## STATE_NAMESouth Dakota           0.329105  0.112023   2.938 0.003305 **
## STATE_NAMETennessee              0.247798  0.111198   2.228 0.025851 *
## STATE_NAMETexas                  0.055763  0.106265   0.525 0.599753
## STATE_NAMEUtah                   0.310527  0.109179   2.844 0.004452 **
## STATE_NAMEVermont                0.631012  0.110749   5.698 1.21e-08 ***
```

```
## STATE_NAMEVirginia                       0.223211   0.106029    2.105 0.035274 *
## STATE_NAMEWashington                     0.313213   0.109019    2.873 0.004066 **
## STATE_NAMEWest Virginia                  0.257692   0.115341    2.234 0.025472 *
## STATE_NAMEWisconsin                      0.452558   0.110218    4.106 4.03e-05 ***
## STATE_NAMEWyoming                        0.346583   0.109650    3.161 0.001573 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 56045  on 42178  degrees of freedom
## Residual deviance: 51260  on 42106  degrees of freedom
## AIC: 51406
##
## Number of Fisher Scoring iterations: 4
```

Based on the fitted multivariable logistic regression model, we are able to calculate the propensity score for each obsevation. Although some of the covariates do not show significance in predicting treatment, these covariats are still included in the propensity score calculation, because the estimating model should be as inclusive as possible to reduce the chance of hidden bias and avoid imbalance between treated and untreated groups.

The propensity score `.fitted` for each observation is obtained by substituting the β coefficients with according values and computing the sum of intercept β_0 and all βX terms. Every propensity score is pridicted and placed into the dataset for each respondent.

For example, a repondent would have a propensity score of $-1.871154 + 0.349258 X_{VACC\_SEAS\_COUNT} + \$$
0.062385X_{BEHAVIOR} + 0.084822X_{CONCERN_H1N1Somwhat} +
(-0.232652)X_{OP_EFFECTIVE_VACCSomewhat Effective} +

$$(-0.250360)X_{RISK\_SICK\_WO\_VACCVeryLow} + (-0.297755)X_{WORRY\_SICK\_W\_VACCNotVeryWorried} +$$

(-0.109446) X_{AGE_GROUP65+ Years} + 0 X_{SEX_IFemale} + 1.140691 X_{EDUSome College} +\$\$ 0.219564 X_{STATE_NAMENew Mexico}\$

Propensity Score Matching

We uses the matching function in arm package, which finds the closest of the ones that were not treated, to match with each one that was treated. Since there are in total 16055 respondent assigned in the treated group (KNOW_H1N1 = 1), the matched dataset `H1N1_data_matched` has 32110 observations.

Two back to back histograms are plotted to determine the match between groups. Figure 3 and 4 together show an improvement in the match between the two distributions of propensity scores after the match.
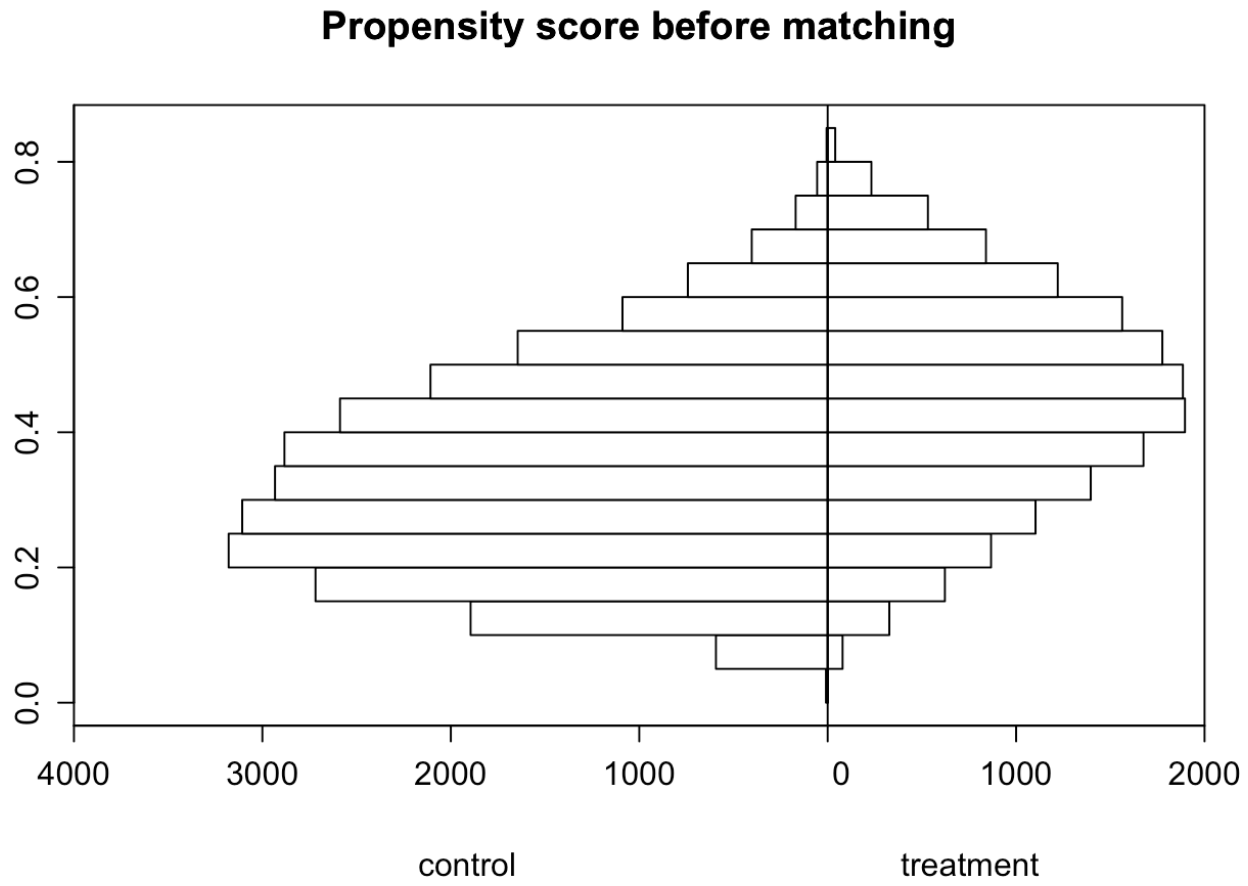
Propensity score before matching

Figure 3. Propensity score before matching
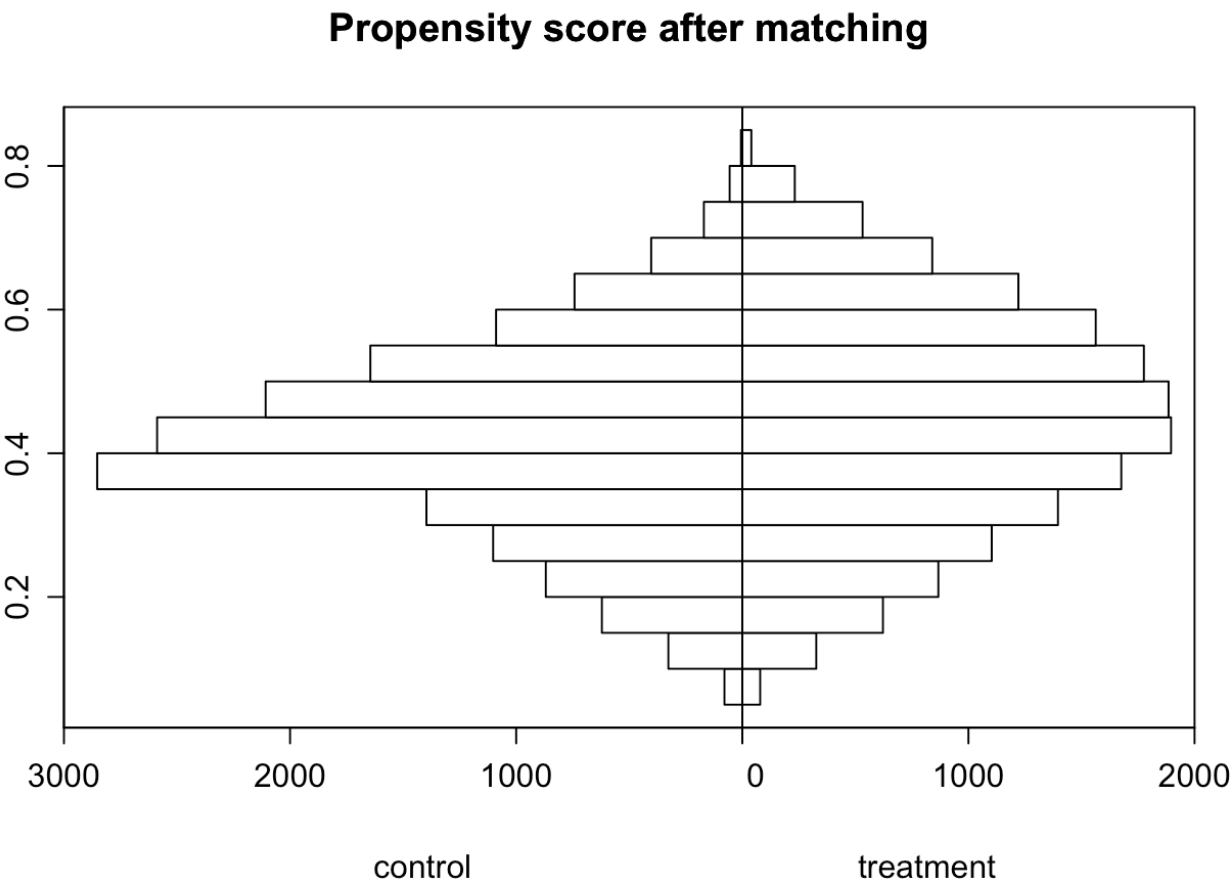
Propensity score after matching

Figure 4. Propensity score after matching

3 Results

The present study utilizes a propensity score matching with multivariable logistic regression model to simulate a quasi-experimental design for obsevational data. Results show a significant relationship between having knowledge about the H1N1 flu and get vaccination in the context of quasi-experimental design. Therefore, it is conclude that the knowledge about diseases has a causal effect on the vaccination behavior.

	(1)
(Intercept)	-2.539 ***
	(0.213)
VACC_SEAS_COUNT	1.879 ***
	(0.037)
BEHAVIOR	-0.048 ***

	(0.014)
CONCERN_H1N1Not very	-0.179 **
	(0.063)
CONCERN_H1N1Somewhat	-0.334 ***
	(0.063)
CONCERN_H1N1Very	-0.447 ***
	(0.070)
OP_EFFECTIVE_VACCNot Very Effective	-0.124
	(0.171)
OP_EFFECTIVE_VACCSomewhat Effective	0.848 ***
	(0.145)
OP_EFFECTIVE_VACCVery Effective	1.666 ***
	(0.145)
RISK_SICK_WO_VACCSomewhat Low	-0.990 ***
	(0.037)
RISK_SICK_WO_VACCVery High	0.473 ***
	(0.054)
RISK_SICK_WO_VACCVery Low	-1.559 ***
	(0.050)
WORRY_SICK_W_VACCNot Very Worried	-0.284 ***
	(0.038)
WORRY_SICK_W_VACCSomewhat Worried	-0.081
	(0.042)
WORRY_SICK_W_VACCVery Worried	0.045
	(0.063)

AGE_GROUP35-44 Years	-0.149 **
	(0.054)
AGE_GROUP45-54 Years	-0.257 ***
	(0.050)
AGE_GROUP55-64 Years	-0.075
	(0.049)
AGE_GROUP65+ Years	-0.299 ***
	(0.050)
SEX_IMale	0.104 **
	(0.032)
EDU12 Years	0.306 ***
	(0.089)
EDUCollege Graduate	0.353 ***
	(0.085)
EDUSome College	0.312 ***
	(0.086)
STATE_NAMEAlaska	0.198
	(0.160)
STATE_NAMEArizona	0.034
	(0.152)
STATE_NAMEArkansas	0.109
	(0.164)
STATE_NAMECalifornia	0.200
	(0.152)
STATE_NAMEColorado	0.000

	(0.152)
STATE_NAMEConnecticut	-0.137
	(0.160)
STATE_NAMEDelaware	-0.086
	(0.160)
STATE_NAMEDistrict of Columbia	-0.193
	(0.150)
STATE_NAMEFlorida	-0.016
	(0.158)
STATE_NAMEGeorgia	-0.155
	(0.161)
STATE_NAMEHawaii	0.313 *
	(0.159)
STATE_NAMEIdaho	-0.058
	(0.166)
STATE_NAMEIllinois	-0.022
	(0.155)
STATE_NAMEIndiana	0.077
	(0.160)
STATE_NAMEIowa	0.122
	(0.154)
STATE_NAMEKansas	0.149
	(0.157)
STATE_NAMEKentucky	-0.181

	(0.168)
STATE_NAMELouisiana	-0.315
	(0.171)
STATE_NAMEMaine	0.251
	(0.155)
STATE_NAMEMaryland	0.056
	(0.151)
STATE_NAMEMassachusetts	0.217
	(0.151)
STATE_NAMEMichigan	-0.211
	(0.164)
STATE_NAMEMinnesota	0.021
	(0.153)
STATE_NAMEMississippi	-0.613 ***
	(0.183)
STATE_NAMEMissouri	-0.150
	(0.163)
STATE_NAMEMontana	-0.060
	(0.163)
STATE_NAMENebraska	0.210
	(0.153)
STATE_NAMENevada	0.201
	(0.157)
STATE_NAMENew Hampshire	0.182
	(0.159)

STATE_NAMENew Jersey	-0.194
	(0.161)
STATE_NAMENew Mexico	0.044
	(0.152)
STATE_NAMENew York	-0.230
	(0.158)
STATE_NAMENorth Carolina	0.199
	(0.156)
STATE_NAMENorth Dakota	0.405 *
	(0.162)
STATE_NAMEOhio	-0.175
	(0.159)
STATE_NAMEOklahoma	-0.021
	(0.162)
STATE_NAMEOregon	0.056
	(0.157)
STATE_NAMEPennsylvania	-0.211
	(0.161)
STATE_NAMERhode Island	0.405 **
	(0.151)
STATE_NAMESouth Carolina	-0.231
	(0.163)
STATE_NAMESouth Dakota	0.396 **
	(0.154)
STATE_NAMETennessee	-0.157

	(0.160)
STATE_NAMETexas	-0.184
	(0.156)
STATE_NAMEUtah	0.053
	(0.153)
STATE_NAMEVermont	0.064
	(0.154)
STATE_NAMEVirginia	0.117
	(0.149)
STATE_NAMEWashington	0.224
	(0.154)
STATE_NAMEWest Virginia	0.027
	(0.165)
STATE_NAMEWisconsin	0.217
	(0.156)
STATE_NAMEWyoming	0.261
	(0.154)
KNOW_H1N1	0.162 ***
	(0.030)
N	32110
logLik	-13837.026
AIC	27822.052

*** p < 0.001; ** p < 0.01; * p < 0.05.

4 Discussion

Levels of Education and % of Respondents Knowledgeable About H1N1 flu
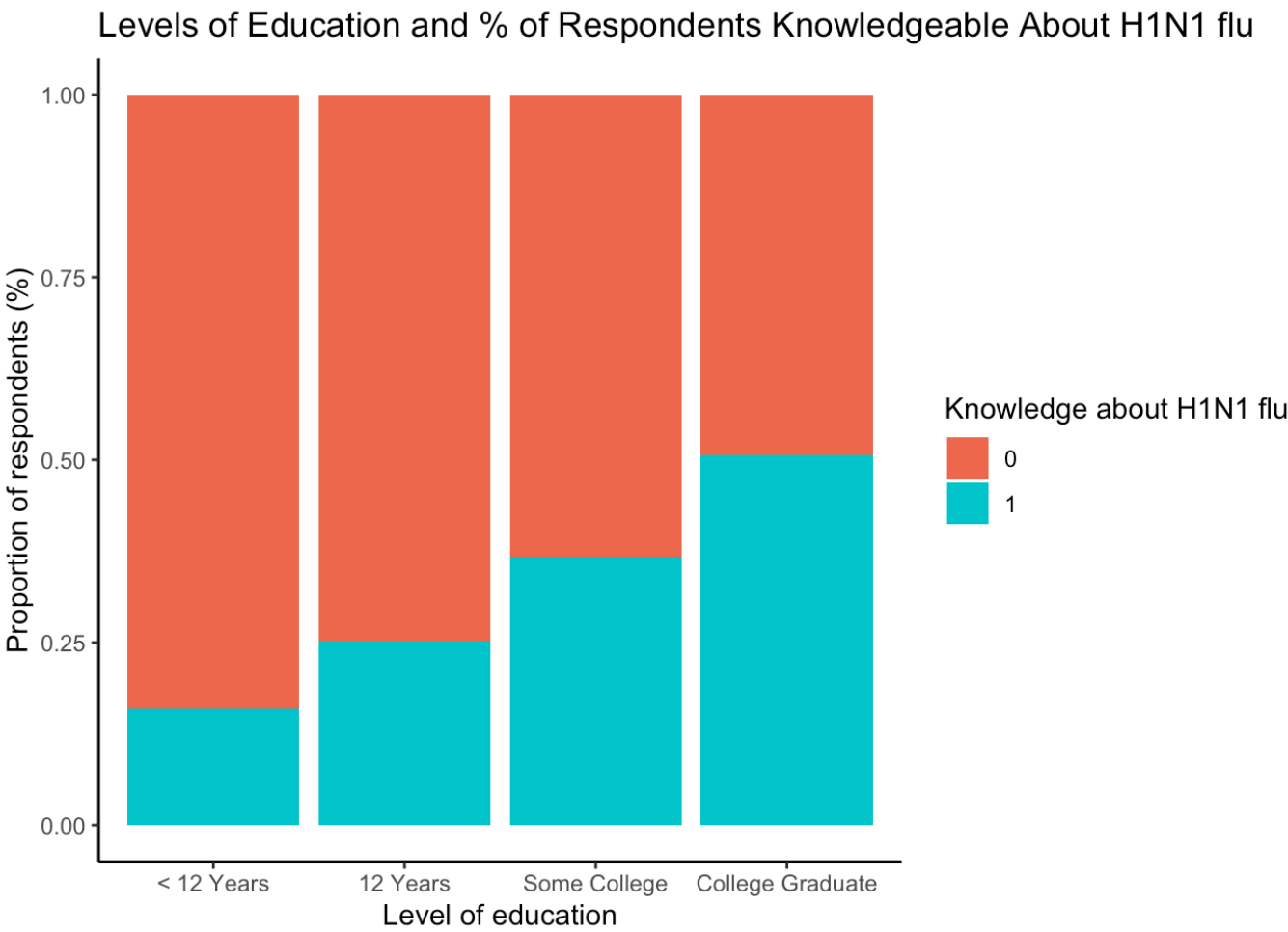
Data Overview: Levels of education and knowledge about H1N1 flu

References

Andrew Gelman and Yu-Sung Su (2020). arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. R package version 1.11-2. https://CRAN.R-project.org/package=arm (https://CRAN.R-project.org/package=arm)

Austin, P. C., & Merlo, J. (2017). Intermediate and advanced topics in multilevel logistic regression analysis. Statistics in medicine, 36(20), 3257–3277. https://doi.org/10.1002/sim.7336 (https://doi.org/10.1002/sim.7336)

Ben B. Hansen and Jake Bowers (2008). Covariate balance in simple, stratified and clustered comparative studies. Statistical Science. 23(2):219–236.

Chou, W. S., & Budenz, A. (2020). Considering emotion in covid-19 vaccine communication: Addressing vaccine hesitancy and fostering vaccine confidence. Health Communication, doi:http://dx.doi.org/10.1080/10410236.2020.1838096 (doi:http://dx.doi.org/10.1080/10410236.2020.1838096)

David Robinson, Alex Hayes and Simon Couch (2020). broom: Convert Statistical Objects into Tidy Tibbles. R package version 0.7.0. https://CRAN.R-project.org/package=broom (https://CRAN.R-project.org/package=broom)

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01 (doi:10.18637/jss.v067.i01).

Frank E Harrell Jr, with contributions from Charles Dupont and many others. (2020). Hmisc: Harrell Miscellaneous. R package version 4.4-1. https://CRAN.R-project.org/package=Hmisc (https://CRAN.R-project.org/package=Hmisc)

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. https://CRAN.R-project.org/package=dplyr (https://CRAN.R-project.org/package=dplyr)

Hao Zhu (2020). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.1. https://CRAN.R-project.org/package=kableExtra (https://CRAN.R-project.org/package=kableExtra)

Hazra, A., & Gogtay, N. (2016). Biostatistics series module 4: Comparing groups - categorical variables. Indian Journal of Dermatology, 61(4), 385-392. doi:10.4103/0019-5154.185700 (doi:10.4103/0019-5154.185700)

Jake Bowers, Mark Fredrickson, and Ben Hansen (2019). RItools:Randomization Inference Tools. R package version 0.1-17.

Olmos, A., & Govindasamy, P. (2015). Propensity Scores: A Practical Introduction Using R. Journal Of MultiDisciplinary Evaluation, 11(25), 68-88. Retrieved from https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/431 (https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/431)

Pogue, K., Jensen, J., Stancil, C., Ferguson, D., Hughes, S., Mello, E., . . . Poole, B. D. (2020). Influences on Attitudes regarding Potential COVID-19 Vaccination in the United States. doi:10.20944/preprints202009.0338.v1 (doi:10.20944/preprints202009.0338.v1)

Rosenbaum, P., & Rubin, D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 70(1), 41-55. doi:10.2307/2335942 (doi:10.2307/2335942)

Salali, G. D., & Uysal, M. S. (2020). COVID-19 vaccine hesitancy is associated with beliefs on the origin of the novel coronavirus in the UK and Turkey. Psychological Medicine, 1-3. doi:10.1017/s0033291720004067 (doi:10.1017/s0033291720004067)

U.S. Department of Health and Human Services (DHHS). National Center for Health Statistics. The National 2009 H1N1 Flu Survey, Hyattsville, MD: Centers for Disease Control and Prevention, 2012. Information about the NHFS is located at http://www.cdc.gov/nchs/nis/about_nis.htm#h1n1 (http://www.cdc.gov/nchs/nis/about_nis.htm#h1n1).

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686 (https://doi.org/10.21105/joss.01686)