**Abstract** **-** The study conducted inspects the association of social factors on the likelihood of vaccine acceptance. We used the data source “Canadian Perspective Survey Series 3 2020”[1] which was conducted in 10 Canadian provinces. We show that influence of social factors such as – “Concern of infecting others”, “Living with spouse/partner”, “Concern about attending shows movies, sports, festivals” is significant on the parameter “Likelihood of getting vaccine once available”(section). Moreover the factor “Education Level” shows positive relation with the vaccine acceptance amongst the Canadian population. One of our main finding was the effect of social factors on the vaccine acceptance which was not considered in the related studies.

**1. Introduction**

The COVID-19 pandemic might continue to cause massive morbidity and mortality while wreaking havoc on civilizations and economies worldwide. Therefore, vaccine acceptance would play a significant role in defeating the common enemy every Country is in a war with. However, there is an enormous hesitancy amongst the Canadian nationals toward administering the Covid 19 vaccine that is putting the nation at the backdrop.

With Covid becoming a severe life threat to humankind, governments and public health groups must be ready to address vaccination reluctance and boost vaccine literacy so that the public accepts immunization when necessary. In addition, the Canadian government must execute appropriate strategies to enhance trust in and acceptance of the vaccine.

According to the WHO [2], vaccine hesitation is a complicated issue caused by a lack of vaccine uptake despite the vaccine's availability. The causes for the apprehension include everything from lack of confidence in vaccine safety to vaccine adverse effects, personal experiences and beliefs, economic concerns, and even age groups.

We could identify a potential gap in the literature where the experts made predictions about the vaccine hesitancy depending on various factors while completely ignoring each individual's other social factors and daily lives that could have influenced the likelihood of vaccine acceptance. The predictions made in the related studies seem pretty incomplete due to limited factors taken into account.

In our study, we intend to understand the relation of potential social factors that could influence the likelihood of the Covid 19 vaccine amongst the Canadian Population. These findings not only aim to investigate the impact of a few social aspects (such as crowd control during shows, festivals, and sports) that have been overlooked in earlier studies, but also to re-evaluate the findings of some of the previous studies (say, influence of education level on vaccine acceptance). This information should help public health officials focus vaccine promotion messages more effectively.

**2. Literature Review**

Researchers from across the globe have done a notable amount of study on Hesitancy in vaccine administration and the factors influencing it.

**Related Study 1**

In June 2020, Jeffery [3]and his peers performed a country-level analysis of the association of factors - Age, Gender, and Education with the vaccine hesitance. In his study, he randomly selected a sample of 13,426 participants from across 19 countries -Brazil, Canada, China, Ecuador, France, Germany, India, Italy, Mexico, Nigeria, Poland, Russia, Singapore, South Africa, South Korea, Spain, Sweden, United Kingdom, United States. The objective of this research was to analyse the following two statements and draw inferences - 'I will take a COVID-19 vaccine if proven to be safe and effective, and available to me' and 'Once the government has authorised the COVID-19 vaccination as safe and effective, I would follow my employer's advice and acquire one'.

**Related Study 2**

Marzo [4], in October 2020, conducted an internet-based cross-sectional survey on the acceptance of vaccines amongst the Malaysian population and the factors influencing it. The data of 1310 participants aged 18 years or above were collected using the snowball sampling technique. Only 1282 responses were considered for the analysis because of their complete details in the survey form. Following were the set of questions analysed: Age, gender, residence, race, educational level, occupation, number of family members, marital status, and family income are socio-demographic traits. Acceptance, attitude, and vaccination preferences for future COVID-19 immunisation, as well as the importance of identified impact factors on respondents' vaccination decision-making, such as vaccine price, convenience, and physician recommendations. The survey form was triggered to the participants through various social media platforms.

**Related Study 3**

Similar research was performed by Elke [5]that got published on 22nd Sept 2021. He examined the attitude of Austrian Adolescences towards vaccine administration and the association of factors – gender, education status, and migration background with it. A sample of 1442 apprentices and 564 high school students was collated by conducting two cross-sectional surveys. The survey questionnaire constituted the following sections:

Demographic Data (gender, age, migration background)

Would you get vaccinated against the coronavirus if you had the opportunity?

The responses were rated on a 5-point scale ranging from 1 = “no way” to 5 = “definitely”.

**Related Study 4**

Another international survey by Suzanna Awang Bono [6]was published on 17th May 2021 that emphasised on the factors affecting COVID-19 vaccine acceptance among low- and middle-income Countries. A total of 10,183 responses were collected through a cross-sectional online survey. The survey form constituted of the following three sections:

Demographic information (age, gender, country of residence, educational level, studying or working in healthcare, (estimated) age(s) of housemate(s), if any, self-perceived socioeconomic position, and self-perceived area of living)

Eight questions were asked on the participants' health, their knowledge of COVID-19, their adherence to COVID-19 prevention measures, and their fear/worry of contracting COVID-19.

Four questions about participants' readiness to take the COVID-19 vaccination at 90 percent and 95 percent effectiveness levels, as well as their reasons for vaccine apprehension. The responses are collected from social media platforms through a link.

**Related Studies 5**

A quite similar study to our topic was done by Kristyn Frank [7] on 25th August 2020, in which he examined Canadians’ willingness to get a COVID-19 vaccine, group differences and the reasons for vaccine hesitancy. A total of 400 0 respondents took part in the survey. Following were the sections in the survey:

Sociodemographic details (age, sex, and highest level of education)

“When a COVID-19 vaccine becomes available, how likely is it that you will choose to get it?

\*A thorough classification of each of the above related studies along with the objectives, methods, and results is collated in the table 1().

**Table 1** Related Studies Methods & Results

|  |  |  |
| --- | --- | --- |
| **Related study** | **Methods** | **Conclusions** |
| Hesitant or Not? The Association of Age, Gender, and Education with Potential Acceptance of a COVID-19 Vaccine: A Country-level Analysis. [3] | Descriptive Statistics of demographic variables to inspect each vaccine acceptance question. Univariable & Multivariable logistic regression. | **71.5%** of respondents are very or somewhat likely to accept the vaccine.  Results from recent opinion polling suggests vaccine acceptance ranging from **40%** in the US to **87%** in India.  Likelihood of vaccine acceptance amongst Females was more in countries – France, Germany, Russia, and Sweden as compared to men.  Aged age ≥ 50 in Canada, Poland, Sweden, and UK were more likely to accept vaccine as opposed to those in China.  Education level showed significant relation with the vaccine acceptance with higher level being more likely to accept the vaccine in countries India, US, France, Ecuador whereas Canada, UK, and Spain showed contrasting trends. |
| Factors influencing the acceptability of COVID-19 vaccination: A cross-sectional study from Malaysia. [4] | IBM SPSS version 25 to  analyse data. Descriptive Statistics to outline respondents socio demographic. Chi-square test for analysing importance of link between background characteristics and respondents. Logistic Regression model on  variables with p value <0.05 in bivariate analysis. | **96%** of the participants are likely to accept the vaccine provided the vaccine were successfully developed and approved for future listings.  Factors such as vaccine price **(91%),** vaccine convenience **(95.7%),** & doctor’s recommendation **(97.3%)** too had association with the acceptance rate.  **38.2%** of the participants were willing to get vaccine as soon as it becomes available. However, **61.8%** were reluctant for the immediate administration and wanted to verify the vaccination safety. |
| Education Level and COVID-19 vaccination willingness in adolescents. [5] | IBM SPSS version 26 to analyse data. Independent t tests and chi-square test to analyse sample characteristics differences. Univariate ANOVA to analyse differences in vaccination readiness. | High school students were more likely to accept vaccine **(53%)** as compared to the apprentices **(28%).**  **7%** of students and **22%** of apprentices were reluctant to administer vaccination.  Male adolescents were more likely accept vaccine than female adolescents.  Migration background too had an effect on the vaccine acceptance. |
| Factors Affecting COVID-19 Vaccine acceptance: An international survey among Low- and Middle-Income countries. [6] | IMB SPSS version 27 to analyse data. Descriptive Statistics (Mean & S.D.) for continuous variables. Percentage and Frequency for categorical variables. Chi-square test to analyse vaccine hesitancy reasons. Multiple logistic regression with vaccine acceptance (DV) at 90% & 95% effectiveness. Spearman’s Rank correlation coefficient and Pearson’s correlation coefficient to check multicollinearity. | Acceptance rate for **90%** vaccine effectiveness was **76.4%** whereas for 95% vaccine effectiveness, it was **88.8%.** Vaccine effectiveness led to a significant difference of **12.4%** in vaccine acceptance.  Brazil had maximum likely participants to accept vaccine at **90%** vaccine effectiveness, whereas lowest vaccine acceptance rate was observed in the African countries.  vaccine acceptance was positively associated with other factors such as – Covid-19 knowledge, side effects **(41.2%),** lack of confidence in vaccine effectiveness **(15.1%),** younger age, higher income. |
| Canadian’s willingness to get a COVID-19 vaccine. [7] | IBM SPSS version 25 to analyse data. Descriptive Statistics for IVs. Ordinal Logistics Regression. | **57.5%** of Canadian’s are likely to accept vaccination**, 19%** are somewhat likely, 9% are unlikely, and **9.4%** were unsure once the vaccine becomes available. |

Now to expand the focus on analysing other potential factors affecting the COVID-19 acceptance, we have performed analysis considering other social factors such as – “ “, the hypothesis of which are defined in section ()

**3. Methods**

In this section, we present the methods and procedures involved in our study. We also present our hypotheses that we intend to examine through our model.

**3.1 Procedure**

For our study, we have used the data source “Canadian Perspective Survey Series 3 2020” [1]. It provides information on the recovery of economic and labour activities during COVID-19 and the impacts of COVID-19 on public and social life.  The data was gathered through an online survey conducted in 10 Canadian provinces. The desired target audience was residents of the 10 Canadian provinces aged 15 or older, excluding 2% Canadian population aged 15 or older. The defined audience was approached through an email that spoke about the motive behind this survey and the steps to be followed to participate in the survey. Follow-up reminders via mails and Computer Assisted Telephonic Interview were conducted to encourage people to participate. The survey form had standard response codes to identify all possible outcomes. In addition, proper measures were taken to ensure that the data is clean.

We have used R software to analyse the data. In order to align our data to fit into the models, we made some alternations (section). We selected one dependent variable, four independent variables, and three control variables for this project. The primary variable put under test is “Likelihood of getting vaccine once available,” our dependent variable. We tested the dependency of this variable on independent variables and control variables. The independent variables considered are “Why get tested – concerns of infecting others?”, “Living with spouse or partner,” “​​Health risk concern - Attending shows, festivals, movies, sports,” “Respondents highest level of education ever completed,” and “Rural/Urban Indicator,” “Age Group of Respondent,” “​​Sex of Respondent” being the control variables. We calculated the correlation of the variables (section) to check for multicollinearity amongst the variables.

Multiple linear Regression [8] is used to express the relation between our multiple independent variables and a dependent variable (section), while holding the control variables constant.

*yi​=β0​+β1​xi1​+β2​xi2​+...+βp​xip​+ϵ*

where, for i=n observations:

yi​=dependent variable

xi​=explanatory variables

β0​=y-intercept (constant term)

βp=slope coefficients for each explanatory variable

ϵ=the model’s error term (also known as the residuals)​

Following assumptions must be considered while performing MLR:

Dependent & Independent variables follow a linear relationship [9].

Independent variables must not show multicollinearity.

Residuals to be normally distributed [10] with 0 mean and variance.

Observations (yi​) are selected independently and randomly from the population.

Breusch-Pagan [11] test, one of the essential steps to check for heteroskedasticity [12] in our model. This test is conducted to inspect the violation of one key assumption [13] in our MLR(section) i.e. Homoscedasticity refers to the fact that the residuals have constant variance at all levels of X. The following hypotheses can be used to run the Breusch-Pagan test:

**Null Hypothesis (H0):** Homoscedasticity is present

**Alternative Hypothesis (HA):** Heteroscedasticity is present

Further, if the p-value of test is less than the significance level (α = .05), we reject our H0 and conclude the presence of heteroscedasticity in our MLR model.

Shapiro-Wilk test [14], is used to examine another key assumption [13] in our MLR model (section), i.e., the random sample drawn follows a normal distribution [10]. It can be performed by using following hypothesis:

**Null Hypothesis (H0):** Population is normally distributed

**Alternative Hypothesis (HA):** Population is not normally distributed

***W=(∑ni=1aix(i))2∑ni=1(xi−¯x)2***

If the p-value value is less than the significance level (α = .05), we reject our H0 and conclude the probability to be not normally distributed.

**3.2 Model**

Ordinal Logistic Regression [15] is a type of regression that uses 'ordered' multiple categories and independent variables to predict the dependent variable.

Let us first develop some notation and study the ideas involved in ordinal logistic regression to understand(link) how to interpret the coefficients. With J categories, let Y be an ordinal outcome. The cumulative probability of Y smaller than or equal to a specific category j=1,.J-1 is then P(Yj). The probability of falling into one of two categories is defined as

*P(Y>J)=0* and dividing by zero is undefined for j=1,..,J-1. You can also put P(Y>j)=1–P(Yj) as an alternative. The log odds is often referred to as the logit.

The ordinal logistic regression model can be defined as

for *j=1,⋯,J−1* and *p* predictors. Due to the parallel lines assumption, the intercepts are different for each category but the slopes are constant across categories, which simplifies the equation above to

**3.3 Hypothesis**

We have tested the following 4 Hypothesis in our model:

**Null Hypothesis: *“****Likelihood of getting vaccine once available”.*

As COVID-19 continues to be a threat to public health, a complete vaccination could decelerate the effect of its various multiple variants. This hypothesis put under test would generate exciting insights and help us understand the influence of factors on the likelihood of vaccine acceptance.

**Hypothesis 1: “***People who are willing to get tested because they are concerned of infecting others are more likely to take the vaccine.”*

With the activities slowly coming into pace, people have started stepping out as a necessity. This further increases the likelihood of infecting others. This hypothesis tells us about the number of people who care for their loved ones and people around them and are likely to take the vaccine.

**Hypothesis 2: “***People who are living with their spouse or partner are more likely to take the vaccine”.*Related articles [16] have shown that vaccinated people living with their unvaccinated partner tend to face a unique set of challenges. This hypothesis helps us analyse the likelihood of vaccine admiration among people living with their unvaccinated spouse/partner, keeping the worry about the challenges aside. Also, to check if both the partners are willing to accept vaccines to overcome the challenges.

**Hypothesis 3: “***People who are very concerned about attending shows, festivals, movies, and sports are more likely to take the vaccine”.*With the lockdown being lifted phase-wise in different provinces across Canada, social activities such as – cinema halls, sports stadiums, cultural events have reopened their doors for the ordinary public. The interest in attending large gathering programs among the ordinary public could be a point of concern if the people willing to attend such programs are unvaccinated. This hypothesis helps us generate interesting insights to analyse the likelihood of vaccine administration to people who are very concerned about being part of such large gatherings.

**Hypothesis 4: “***People with higher levels of education are more likely to take the vaccine”.*

This hypothesis would let us deeply understand the association of education on vaccine acceptance. The results can further be used to create a strategic approach to ensure utmost vaccination acceptance amongst the people of both levels (low & high education levels).

**3.4 Challenging Existing Literature**

The existing studies discuss the influence of limited factors such as - Socio-demographic details, Vaccine confidence, Vaccine effectiveness, side effects, etc., on the likelihood of vaccine acceptance. However, other potential factors that can significantly affect the likelihood of vaccine administration are ignored.

Our study focuses on the association of social factors with the likelihood of vaccine acceptance. The results (section) of our model put the preciseness of the previous studies into question, challenging the limited factors considered in the past.  
Related studies [3, 7] have considered education level as one of the factors influencing the acceptance of vaccines once available. However, the results of two different studies contradict the same geography, i.e. one proved the high education level positively affects vaccine acceptance. In contrast, the other study concluded that the high education level negatively correlates with vaccine acceptance. This contradiction further raises questions on the authenticity of the existing literature.

The results(section) of the association of education level with vaccine acceptance are close to the existing studies. However, no study was done to understand the association of other social factors with vaccine acceptance, making our study distinctive that can be referred to in the future.

**4. Data**

Our data source is The Canadian Perspectives Survey Series (CPSS) Series 3, 2020: Resuming Economic and Social Activities During COVID-19 (CPSS Economic and Social Activity) [1], which is a series of brief online surveys that began in March 2020 and collected data on citizens' knowledge and behaviors in each of Canada's ten provinces. The target population for the Canadian Perspectives Survey Series (CPSS) is residents of the 10 Canadian provinces 15 years of age or older. These individuals were invited to Sign-Up for the CPSS. Those agreeing to join the CPSS were asked to provide an email address. Participants from the Sign-Up that provided valid email addresses formed the probability panel. The participation rate to the panel was approximately 23%. The survey population for all surveys of the CPSS is the probability panel participants. Participants of the panel are 15 years or older as of July 31, 2019. Table 2 provides the appendix for the variables and gives a brief description of each variable. Table 3 lists the variables and gives the number of observations for each category within the variables. Table 4 provides descriptive statistics for all variables, such as the mean, median, min, max and standard deviation.

The dependent variable choses here is HR\_10, which corresponds to the question - " When a COVID-19 vaccine becomes available, how likely is it that you will choose to get it?". In our data, the dependent variable is coded as - ”1 = Very Likely”, “2 = Somewhat Likely”, “3 = Somewhat Unlikely”, “4 = Unlikely”​. We recode these values in increasing order such that our new values become -– “1 = Unlikely”, “2 = Somewhat Unlikely”, “3 = Somewhat Likely”, “4 = Very Likely”​. The independent variable HR\_05C and FCR\_05 correspond to the following questions - "If testing were widely available to all Canadians, why would you go to get tested for the COVID-19 virus? - If I did not have any symptoms but had concerns of infecting others" and " Are you currently living with a spouse or partner?" respectively. They were originally coded as – “1 = Yes”, “2 = No”​. We have recoded it into - “1 = Yes”, “0 = No”. The other independent variables are PTC\_05E, which asks the question - "How concerned are you about the health risks of resuming the following activities as the COVID-19 safety measures are relaxed? - Attending shows, festivals, movies or sporting events", and variable PEDUC\_LC​ - that corresponds to "Respondents highest level of education ever completed".

We have excluded all values that were skipped for each variable in our model.

The control variables include Rural/Urban Indicator coded as "1=Rural", 2="Urban", Sex coded as "1=Male", "2=Female" and Age Group of Respondent coded as "1=15 to 24 years old","2=25 to 34 years old", "3=35 to 44 years old", "4=45 to 54 years old", "5=55 to 64 years old", "6=65 to 74 years old", "7=75 years and older".

**Table 2** Appendix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable​** | **Description​** | **Category​** | **Type** | **Source** |
| HR\_10​ | Likelihood of getting vaccine once available​ | Dependent ​ | Discrete/Ordinal | cpss-5311-E-2020-ec-social |
| HR\_05C​ | Why get tested – concerns of infecting others?​ | Independent ​ | Discrete/Binary | cpss-5311-E-2020-ec-social |
| FCR\_05 | Living with spouse or partner​ | Independent ​ | Discrete/Binary | cpss-5311-E-2020-ec-social |
| PTC\_05E​ | Health risk concern - Attending shows, festivals, movies, sports ​ | Independent ​ | Discrete/Ordinal | cpss-5311-E-2020-ec-social |
| PEDUC\_LC​ | Respondents highest level of education ever completed ​ | Independent ​ | Discrete/Ordinal | cpss-5311-E-2020-ec-social |
| RURURB​ | Rural/Urban Indicator​ | Control ​ | Discrete/Binary | cpss-5311-E-2020-ec-social |
| AGEGRP​ | Age Group of Respondent​ | Control ​ | Discrete | cpss-5311-E-2020-ec-social |
| SEX​ | Sex of Respondent​ | Control ​ | Discrete/Binary | cpss-5311-E-2020-ec-social |

**Table 3** Number of observations

|  |  |
| --- | --- |
| **Variable​** | **Count​** |
| HR\_10​ (Likelihood of getting vaccine once available)  Very likely | ​  2597 |
| Somewhat likely | 723 |
| Somewhat unlikely | 176 |
| Very unlikely | 348 |
| HR\_05C​ (Why get tested – concerns of infecting others?​) | ​ |
| Yes | 1413 |
| No | 2784 |
| FCR\_05 (Living with spouse or partner) | ​ |
| Yes | 2746 |
| No | 1459 |
| PTC\_05E​ (Health risk concern - Attending shows, festivals, movies, sports) | ​ |
| Not at all concerned | 227 |
| Somewhat concerned | 1102 |
| Very concerned | 2590 |
| PEDUC\_LC​ (Respondents highest level of education ever completed) | ​ |
| Less than high school diploma or its equivalent | 239 |
| High school diploma or a high school equivalency certificate | 795 |
| Trade certificate or diploma | 364 |
| College/CEGEP/other non-university certificate or diploma | 1002 |
| University certificate or diploma below the bachelor's level | 153 |
| Bachelor's degree (e.g.- B.A., B.Sc., LL.B.) | 1040 |
| University certificate, diploma, degree above the BA level | 616 |
| RURURB​ (Rural/Urban Indicator) | ​ |
| Rural | 869 |
| Urban | 3340 |
| AGEGRP​ (Age Group of Respondent) | ​ |
| 15 to 24 years old | 203 |
| 25 to 34 years old | 543 |
| 35 to 44 years old | 711 |
| 45 to 54 years old | 678 |
| 55 to 64 years old | 924 |
| 65 to 74 years old | 845 |
| 75 years and older | 305 |
| SEX​ (Sex of Respondent) | ​ |
| Male | 1930 |
| Female | 2279 |

**Table 4** Descriptive Statistics

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Min** | **1st Qu.** | **Median** | **Mean** | **SD** |  | **3rd Qu** | **Max** |
| HR\_10 | 1.000 | 3.000 | 4.000 | 3.456 | 0.939 |  | 4.000 | 4.000 |
| HR\_05C | 0.0000 | 0.0000 | 0.0000 | 0.3533 | 0.478 |  | 1.0000 | 1.0000 |
| FCR\_05 | 0.0000 | 0.0000 | 1.0000 | 0.6596 | 0.473 |  | 1.0000 | 1.0000 |
| PTC\_05E | 1.000 | 2.000 | 3.000 | 2.592 | 0.613 |  | 3.000 | 3.000 |
| PEDUC\_LC | 1.000 | 3.000 | 4.000 | 4.421 | 1.883 |  | 6.000 | 7.000 |
| RURURB | 1.0 | 2.0 | 2.0 | 1.8 | 0.4 |  | 2.0 | 2.0 |
| SEX | 1.000 | 1.000 | 2.000 | 1.537 | 1.64 |  | 2.000 | 2.000 |
| AGEGRP | 1.000 | 3.000 | 4.000 | 4.221 | 0.498 |  | 6.000 | 7.000 |

Bivariate Analysis of independent variables against our dependent variable is shown in Plot 1, 2, 3 and 4. Bivariate Analysis of control variables against our dependent variable is shown in Plot 5, 6 and 7.

**Plot 2**

**Plot 1**

Chart, scatter chart

Description automatically generated 

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

**Plot 4**

**Plot 3**

**Plot 6**

**Plot 5**

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

**Plot 7**

Chart, scatter chart

Description automatically generated

**5. Results**

In two parts below, we provide the findings of our investigation. For the complete sample of Canadian population, we perform the linear regression followed by ordinal logistic regression in the first section. In the second section, the outcomes for the four hypotheses are reported.

**5.1 Linear Regression model results**

The outcomes of our hypotheses tests using multiple linear regression model are presented in this section. As we know that, several striking trends appeared from univariable and multivariable logistic regression: women in France, Germany, Russia, and Sweden were considerably more likely to accept a vaccine than men in these countries. In Canada, Poland, France, Germany, Sweden, and the United Kingdom, older (greater than 50) respondents were much more favorable to vaccination than younger respondents, although the opposite tendency was observed in China [3]. The majority of respondents in our data are living in the urban area. We have used Sex, Age and Rural/urban indicator as control in our model.

Table 5 and 6 contain the results for our multiple linear regression model.

**Table 5** Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 1.651855 | 0.106272 | 15.544 | < 2e-16 \*\*\* |
| HR\_05C | 0.301936 | 0.030067 | 10.042 | < 2e-16 \*\*\* |
| FCR\_05 | 0.084619 | 0.030387 | 2.785 | 0.005386 \*\* |
| PTC\_05E | 0.472721 | 0.023853 | 19.818 | < 2e-16 \*\*\* |
| PEDUC\_LC | 0.025711 | 0.007697 | 3.341 | 0.000845 \*\*\* |
| RURURB | 0.095843 | 0.036172 | 2.65 | 0.008094 \*\* |
| AGEGRP | 0.050158 | 0.008823 | 5.685 | 1.41e-08 \*\*\* |
| SEX | -0.053515 | 0.02884 | -1.856 | 0.063590 |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8572 on 3623 degrees of freedom

Multiple R-squared: 0.1582

Adjusted R-squared: 0.1565

F-statistic: 97.23 on 7 and 3623 DF

p-value: < 2.2e-16

AIC: 9195.708

**Table 6** Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | 1Q | Median | 3Q | Max |
| -2.9496 | -0.3269 | 0.252 | 0.504 | 1.8107 |

However, we need to ensure if the assumptions of Linear regression model hold good. We proceed to test for homoscedasticity and normality. Plots 1,2 and 3 for residuals show there is some heteroscedasticity in our model. To confirm this hypothesis, we use Breusch-Pagan test. BP Test gives following values as shown in Table 7. The null hypothesis is that the residual variance remains constant [12]. We can reject the null hypothesis and infer that the assumption of homoscedasticity is violated because the p-value is less than 0.05. To test for normality assumption, Shapiro-Wilks test is performed. Shapiro's test's null hypothesis is that the population is normally distributed [14]. We can't assume normality because the p-value is less than 0.05, as shown in Table 8.

**Chart

Description automatically generated Chart, line chart

Description automatically generated Chart, histogram

Description automatically generated**

**1. Residual Plot**

**3. Density Plot for Residuals**

**2. Normal Q-Q Plot**

**Table 7** Breusch-Pagan Test

|  |  |  |
| --- | --- | --- |
| BP | df | p-value |
| 266.5 | 7 | < 2.2e-16 |

**Table 8** Shapiro-Wilks Test

|  |  |
| --- | --- |
| Value | p-value |
| 0.6204 | < 2.2e-16 |

Due to heteroscedasticity, the result of linear regression is not valid. Also, the assumption of normality is violated. Since our dependent variable is categorical, we proceed with Ordinal Logistic Regression.

**5.2 Ordinal Logistic Regression model results**

In this section we present the results of our Ordinal Logistic Regressiontests. One of the basic assumptions of ordinal logistic (and ordinal probit) regression is that each pair of outcome groups has the same relationship. To put it another way, ordinal logistic regression assumes that the coefficients describing the relationship between, say, the lowest and all higher categories of the response variable are the same as those describing the relationship between the next lowest and all higher categories, and so on. As defined earlier [15], in R (polr) the ordinal logistic regression model is parameterized as

logit (P (Y≤ j))=βj0–η1x1–⋯–ηpxp

For Hypothesis 1, the results for coefficients are present in table 9. We can say that for a one unit increase in HR\_05C, we would expect a 0.92 increase in the expected value of HR\_10 in the log odds scale, given that all the other variables in the model are held constant. As the p-value is much less than 0.05, the results are significant, and we cannot reject the hypothesis. Interpreting the Odds Ratio for the same tells us that for respondents who do think of concern of infecting others as a reason to get tested, the odds of being more likely (i.e., very, or somewhat likely versus unlikely) to accept the vaccine is 2.5 times that of respondents who do not think of concern of infecting others as a reason to get tested, holding constant all other variables. The regression model can be written as follows:

logit (P (Y ≤ 1)) = -0.5355 - (0.92) \*HR\_05C - (0.376) \*RURURB - (0.2) \*AGEGRP - (0.033)\*SEX

logit (P (Y ≤ 2)) = -0.056 - (0.92) \*HR\_05C - (0.376) \*RURURB - (0.2) \*AGEGRP - (0.033) \*SEX

logit (P (Y ≤ 3)) = 1.116 - (0.92) \*HR\_05C - (0.376) \*RURURB - (0.2) \*AGEGRP - (0.033) \*SEX

For Hypothesis 2, the results for coefficients are present in table 10. We can say that for a one unit increase in FCR\_05, we would say that for a one unit increase in FCR\_05, we would expect a 0.22 increase in the expected value of FCR\_05 in the log odds scale, given that all the other variables in the model are held constant. As the p-value is much less than 0.05, the results are significant, and we cannot reject the hypothesis. This also means that for respondents who live with their spouse/partner, the odds of being more likely (i.e., very, or somewhat likely versus unlikely) to accept the vaccine is 1.25 times that of respondents who do not live with their spouse/partner, holding constant all other variables. The regression model can be written as follows:

logit (P (Y ≤ 1)) = -0.6776 - (0.22) \* FCR\_05 - (0.40) \*RURURB - (0.18) \*AGEGRP - (0.04) \*SEX

logit (P (Y ≤ 2)) = -0.2070 - (0.22) \* FCR\_05 - (0.40) \*RURURB - (0.18) \*AGEGRP - (0.04) \*SEX

logit (P (Y ≤ 3)) = 0.9365 - (0.22) \* FCR\_05 - (0.40) \*RURURB - (0.18) \*AGEGRP - (0.04) \*SEX

For Hypothesis 3, the results for coefficients are present in table 11. We can say that for a one unit increase in PTC\_05E, we would say that for a one unit increase in PTC\_05E, we would expect a 1.028 increase in the expected value of PTC\_05E in the log odds scale, given that all the other variables in the model are held constant. Interpreting the Odds Ratio for the same tells us that for respondents who do think of concern of infecting others as a reason to get tested, the odds of being more likely (i.e., very, or somewhat likely versus unlikely) to accept the vaccine is 2.5 times that of respondents who do not think of concern of infecting others as a reason to get tested, holding constant all other variables. As the p-value is much less than 0.05, the results are significant, and we cannot reject the hypothesis. For respondents who feel attending shows, festivals, movies, sports could be a heath concern, the odds of being more likely (i.e., very, or somewhat likely versus unlikely) to accept the vaccine is almost 2.8 times that of respondents who do not feel that resuming activities like attending shows, festivals, movies, sports could be a heath concern, holding constant all other variables. The regression model can be written as follows:

logit (P (Y ≤ 1)) = 1.066- (1.028) \* PTC\_05E - (0.286) \*RURURB - (0.137) \*AGEGRP - (-0.101) \*SEX

logit (P (Y ≤ 2)) = 1.578 - (1.028) \* PTC\_05E - (0.286) \*RURURB - (0.137) \*AGEGRP - (-0.101) \*SEX

logit (P (Y ≤ 3)) = 2.815 - (1.028) \* PTC\_05E - (0.286) \*RURURB - (0.137) \*AGEGRP - (-0.101) \*SEX

For Hypothesis 4, the results for coefficients are present in table 12. We can say that for a one unit increase in PEDUC\_LC, we would say that for a one unit increase in PEDUC\_LC, we would expect a 0.109 increase in the expected value of PEDUC\_LC in the log odds scale, given that all the other variables in the model are held constant. As the p-value is much less than 0.05, the results are significant, and we cannot reject the hypothesis. Interpreting the Odds Ratio for the same tells us that for respondents who have a higher level of education ever completed, the odds of being more likely (i.e., very, or somewhat likely versus unlikely) to accept the vaccine is 1.07 times that of respondents with one lesser level of education, holding constant all other variables. The regression model can be written as follows:

logit (P (Y ≤ 1)) = -0.507- (1.028) \* PEDUC\_LC - (0.322) \*RURURB - (0.184) \*AGEGRP - (0.029) \*SEX

logit (P (Y ≤ 2)) = -0.0354 - (1.028) \* PEDUC\_LC - (0.322) \*RURURB - (0.184) \*AGEGRP - (0.029) \*SEX

logit (P (Y ≤ 3)) = 1.114 - (1.028) \* PEDUC\_LC - (0.322) \*RURURB - (0.184) \*AGEGRP - (0.029) \*SEX

Results of consolidated logistic regression containing all our independent variables together with control variables are shown in table 13. The analysis confirms our previous results with significant results. The lower AIC value 6253.163 compared to previous model shows this is a better model.

**Table 9** H1 :People who are willing to get tested because they are concerned of infecting others are more likely to take the vaccine

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Value | Std. Error | t value | p value |
| HR\_05C | 0.91886574 | 0.08054809 | 11.407666 | 3.83E-30 |
| RURURB | 0.37636281 | 0.08634001 | 4.3590776 | 1.31E-05 |
| AGEGRP | 0.20380301 | 0.02173411 | 9.3771029 | 6.78E-21 |
| SEX | 0.03327259 | 0.07153014 | 0.4651548 | 6.42E-01 |
| 1|2 | -0.5355165 | 0.22546056 | -2.3752113 | 1.75E-02 |
| 2|3 | -0.0564413 | 0.22383433 | -0.2521567 | 8.01E-01 |
| 3|4 | 1.11681604 | 0.22414844 | 4.982484 | 6.28E-07 |

Intercepts:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Value | Std. Error | t value |
| 1|2 | -0.5355 | 0.2255 | -2.3752 |
| 2|3 | -0.0564 | 0.2238 | -0.2522 |
| 3|4 | 1.1168 | 0.2241 | 4.9825 |

Residual Deviance: 6540.622

AIC: 6554.622

|  |  |  |  |
| --- | --- | --- | --- |
|  | OddsRatio | 2.5% | 97.5% |

HR\_05C 2.506446 2.1429700 2.938873

RURURB 1.456976 1.2291357 1.724363

AGEGRP 1.226057 1.1750451 1.279563

SEX 1.033832 0.8985235 1.189382

**Table 10** H2  :People who are living with their spouse or partner are more likely to take the vaccine

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Value | Std. Error | t value | p value |
| FCR\_05 | 0.22032172 | 0.07387193 | 2.982482 | 2.86E-03 |
| RURURB | 0.40276414 | 0.08544478 | 4.7137362 | 2.43E-06 |
| AGEGRP | 0.18017364 | 0.02138356 | 8.4258012 | 3.58E-17 |
| SEX | 0.04721444 | 0.07084088 | 0.6664858 | 5.05E-01 |
| 1|2 | -0.6776071 | 0.22844511 | -2.9661701 | 3.02E-03 |
| 2|3 | -0.207044 | 0.22672203 | -0.9132065 | 3.61E-01 |
| 3|4 | 0.9365209 | 0.22663193 | 4.1323431 | 3.59E-05 |

Intercepts:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Value | Std. Error | t value |
| 1|2 | -0.6776 | 0.2284 | -2.9662 |
| 2|3 | -0.2070 | 0.2267 | -0.9132 |
| 3|4 | 0.9365 | 0.2266 | 4.1323 |

Residual Deviance: 6672.449

AIC: 6686.449

|  |  |  |  |
| --- | --- | --- | --- |
|  | OddsRatio | 2.5% | 97.5% |

FCR\_05 1.246478 1.0780819 1.440236

RURURB 1.495954 1.2642483 1.767411

AGEGRP 1.197425 1.1483821 1.248810

SEX 1.048347 0.9123751 1.204458

**Table 11** H3  :People who are very concerned about attending shows, festivals, movies, and sports are more likely to take the vaccine

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Value | Std. Error | t value | p value |
| PTC\_05E | 1.0282755 | 0.05767268 | 17.829509 | 4.17E-71 |
| RURURB | 0.2868861 | 0.08757247 | 3.275985 | 1.05E-03 |
| AGEGRP | 0.1371886 | 0.02211625 | 6.20307 | 5.54E-10 |
| SEX | -0.1018075 | 0.07301445 | -1.394348 | 1.63E-01 |
| 1|2 | 1.0659926 | 0.24749474 | 4.307132 | 1.65E-05 |
| 2|3 | 1.5787405 | 0.2474944 | 6.378894 | 1.78E-10 |
| 3|4 | 2.8149773 | 0.25105518 | 11.212584 | 3.54E-29 |

Intercepts:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Value | Std. Error | t value |
| 1|2 | 1.0660 | 0.2475 | 4.3071 |
| 2|3 | 1.5787 | 0.2475 | 6.3789 |
| 3|4 | 2.8150 | 0.2511 | 11.2126 |

Residual Deviance: 6362.546

AIC: 6376.546

|  |  |  |  |
| --- | --- | --- | --- |
|  | OddsRatio | 2.5% | 97.5% |

PTC\_05E 2.7962396 2.4979109 3.131690

RURURB 1.3322725 1.1211094 1.580431

AGEGRP 1.1470445 1.0984307 1.197928

SEX 0.9032034 0.7826105 1.041997

**Table 12** H4  :People with higher levels of education are more likely to take the vaccine

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Value | Std. Error | t value | p value |

PEDUC\_LC 0.10930398 0.01881502 5.8094004 6.27E-09

RURURB 0.32279965 0.08627367 3.7415777 1.83E-04

AGEGRP 0.18429758 0.02132249 8.6433426 5.46E-18

SEX 0.02912098 0.07083558 0.4111066 6.81E-01

Intercepts:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Value | Std. Error | t value |
| 1|2 | -0.5071 | 0.2295 | -2.2098 |
| 2|3 | -0.0354 | 0.2278 | -0.1553 |
| 3|4 | 1.1144 | 0.2279 | 4.8895 |

Residual Deviance: 6647.418

AIC: 6661.418

|  |  |  |  |
| --- | --- | --- | --- |
|  | OddsRatio | 2.5% | 97.5% |

PEDUC\_LC 1.115501 1.0751510 1.157460

RURURB 1.380989 1.1651444 1.634163

AGEGRP 1.202374 1.1532689 1.253824

SEX 1.029549 0.8960128 1.182833

**Table 13**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Value | Std. Error | t value | p value |

HR\_05C 0.82098091 0.08283658 9.910849 3.734577e-23

FCR\_05 0.17273456 0.07698530 2.243734 2.484950e-02

PTC\_05E 0.96119204 0.05821805 16.510207 3.098174e-61

PEDUC\_LC 0.07599575 0.01969599 3.858437 1.141144e-04

RURURB 0.23912283 0.08943113 2.673821 7.499247e-03

AGEGRP 0.14939990 0.02237590 6.676823 2.441776e-11

SEX -0.09230138 0.07400059 -1.247306 2.122853e-01

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1|2 | 1.53299034 | 0.25979778 | 5.900706 | 3.619493e-09 |
| 2|3 | 2.05768072 | 0.26009835 | 7.911164 | 2.549930e-15 |
| 3|4 | 3.33026278 | 0.26439396 | 12.595835 | 2.225905e-36 |

Intercepts:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Value | Std. Error | t value |
| 1|2 | 1.5330 | 0.2598 | 5.9007 |
| 2|3 | 2.0577 | 0.2601 | 7.9112 |
| 3|4 | 3.3303 | 0.2644 | 12.5958 |

Residual Deviance: 6233.163 AIC: 6253.163

|  |  |  |  |
| --- | --- | --- | --- |
|  | OddsRatio | 2.5% | 97.5% |

HR\_05C 2.2727281 1.9343137 2.676635

FCR\_05 1.1885506 1.0216920 1.381671

PTC\_05E 2.6148116 2.3332871 2.931566

PEDUC\_LC 1.0789580 1.0381072 1.121448

RURURB 1.2701345 1.0649386 1.512224

AGEGRP 1.1611372 1.1113863 1.213292

SEX 0.9118303 0.7885632 1.053990

**6. Implications and Conclusion**

In this section we will attempt to convey to the reader how the results can help the experts in getting a better understanding about various factors that can possibly affect vaccine acceptance. These results not only intend to explore the effect of few social factors (such as concern of large gatherings during shows, festivals, and sports) that have been neglected in previous studies but also reassess the results of some of the previous studies (say, influence of education level on vaccine acceptance). This data should aid public health officials in better targeting vaccine promotion messaging.

**6.1 Implications to practice**

Looking at the results, we find they are significant for all four hypotheses, and they cannot be rejected. From the results we can see how people who would prefer to get tested due concerns of infecting others as well as individuals who live with their spouse - are more likely to accept the vaccine. This information can be used in advertising strategies to emphasize more on the message how important immunization is, not just for the health of the individual, but also for the health of their loved ones and people around them. Likewise, people who are concerned about resuming large gatherings during festivals, sports, movies, shows etc. are more likely that they would go for the covid 19 vaccine once available. This information is useful in understanding the fact that many people who are still cautious when it comes to participating in large scale events prefer vaccination as one of the safety measures. Targeted advertising and messages coming from famous sports and movie personalities can help promote the importance of vaccination during their shows and events. In terms of education level, people with a higher level of education are somewhat more likely to take the vaccine. Previous research has shown that education level was associated with lower vaccination acceptance in some countries, while opposite trends were seen in others [3, 7]. Based on the results from our analysis, we find that more efforts might be needed to improve willingness towards vaccination among people with a relatively lower education level.

**6.2 Limitations**

Our results may be useful to experts and public health officials in better targeting vaccine promotion messaging. The study does, however, have certain limitations. For instance, most of the responses of individuals are self-reported. Due to this, intentional or unintentional errors are quite possible. For more reliable information, data collection via anonymous submissions could be a better alternative in future. Another limitation associated with our current study is that although we see slightly more vaccine acceptance in more educated people, however it is not clear what could be the cause of this behavior. For instance, is it because people with higher education background tend to have better trust on the government and health system? Or is it the lack of effective communication from the government that the people with lower level of education seem to be hesitant? Further research can be conducted in future to understand this behavior.

**6.3 Conclusion**

The objective of our study was to investigate if certain social factors could possibly affect vaccine acceptance among the people of Canada. We began with simple linear regression model to test our hypotheses. After having few of the linear regression assumptions violated, we finally proceeded with ordinal logistic regression to get more concrete results. In all cases, we saw our hypotheses turning up with statistically significant results. In the first hypothesis, we demonstrated that a person who is willing to get tested because they are afraid of infecting others is almost 2.5 times more likely to receive the vaccine. In the second hypothesis, we show that living with a spouse or partner increases individual's chances of accepting the vaccine by 1.25 times. According to the third hypothesis, those who are concerned about resuming involvement in activities such as movies, sports, and festivals have a 2.8 times higher chance of accepting the vaccine than others. In the final hypothesis, we show that those with a greater level of finished education are 1.12 times more likely to get vaccinated than those with a lower level of education. With the available results, professionals and public health officials may find the current study valuable in better tailoring vaccine promotion messaging.

**References**

[1] “Canadian Perspective Survey Series 3, 2020: Resuming Economic and Social Activities During COVID-19 Study Documentation,” 2021.

[2] E. Dubé, C. Laberge, M. Guay, P. Bramadat, R. Roy, and J. Bettinger, “Vaccine hesitancy: An overview,” *Human Vaccines and Immunotherapeutics*, vol. 9, no. 8, pp. 1763–1773, Aug. 2013, doi: 10.4161/HV.24657.

[3] J. v. Lazarus *et al.*, “Hesitant or Not? The Association of Age, Gender, and Education with Potential Acceptance of a COVID-19 Vaccine: A Country-level Analysis,” *https://doi.org/10.1080/10810730.2020.1868630*, vol. 25, no. 10, pp. 799–807, 2021, doi: 10.1080/10810730.2020.1868630.

[4] R. R. Marzo *et al.*, “Factors influencing the acceptability of COVID-19 vaccination: A cross-sectional study from Malaysia,” *Vacunas*, Sep. 2021, doi: 10.1016/J.VACUN.2021.07.007.

[5] E. Humer, A. Jesser, P. L. Plener, T. Probst, and C. Pieh, “Education level and COVID-19 vaccination willingness in adolescents,” *European Child & Adolescent Psychiatry*, vol. 1, p. 1, 2021, doi: 10.1007/S00787-021-01878-4.

[6] S. A. Bono *et al.*, “Factors Affecting COVID-19 Vaccine Acceptance: An International Survey among Low- and Middle-Income Countries,” *Vaccines*, vol. 9, no. 5, p. 515, May 2021, doi: 10.3390/VACCINES9050515.

[7] “Canadians’ willingness to get a COVID-19 vaccine: Group differences and reasons for vaccine hesitancy.” https://www150.statcan.gc.ca/n1/pub/45-28-0001/2020001/article/00073-eng.htm (accessed Dec. 06, 2021).

[8] “Multiple Linear Regression (MLR) Definition.” https://www.investopedia.com/terms/m/mlr.asp (accessed Dec. 06, 2021).

[9] “Linear Relationship Definition.” https://www.investopedia.com/terms/l/linearrelationship.asp (accessed Dec. 06, 2021).

[10] “Normal Distribution Definition.” https://www.investopedia.com/terms/n/normaldistribution.asp (accessed Dec. 06, 2021).

[11] “The Breusch-Pagan Test: Definition & Example.” https://www.statology.org/breusch-pagan-test/ (accessed Dec. 06, 2021).

[12] “Heteroskedasticity - Overview, Causes and Real-World Example.” https://corporatefinanceinstitute.com/resources/knowledge/other/heteroskedasticity/ (accessed Dec. 06, 2021).

[13] “The Four Assumptions of Linear Regression - Statology.” https://www.statology.org/linear-regression-assumptions/ (accessed Dec. 06, 2021).

[14] “Shapiro-Wilk Test: What it is and How to Run it - Statistics How To.” https://www.statisticshowto.com/shapiro-wilk-test/ (accessed Dec. 06, 2021).

[15] “Ordinal Logistic Regression | R Data Analysis Examples.” https://stats.idre.ucla.edu/R/Dae/Ordinal-logistic-regression/ (accessed Dec. 06, 2021).

[16] “What Happens When One Partner Is Vaccinated and the Other Isn’t? | Vogue.” https://www.vogue.com/article/what-happens-when-one-partner-is-vaccinated-and-the-other-isnt (accessed Dec. 06, 2021).

[17] " Enhancing public trust in COVID-19 vaccination: The role of governments."

https://www.oecd.org/coronavirus/policy-responses/enhancing-public-trust-in-covid-19-vaccination-the-role-of-governments-eae0ec5a/