Older Female With High Education Level, High Trust in Public Health Professionals, and Consider COVID-19 as a Serious Threat are Most Likely to Wear a Mask During the Pandemic

Written by: Yuyan Liu
Collaborator: John McAndrews
TA Advisor: David Veitch
Date: April 9, 2021

Abstract

Aiming to identify ways to increase effectiveness of policy published for curbing the spread of Covid-19, this report investigated possible socio-demographic predictors and attitudinal predictors for mask-wearing behaviour among Canadians, and whether people who choose to wore masks also took more other precautions than those who didn't wore masks. The analysis had shown that female tend to have perform mask-wearing than all other genders, have higher income, higher educations, obtain trust toward public health professionals, taken the pandemic as a serious threat are also predictors of higher chance of wearing a mask. Trust in professionals in business discipline and those who related to the government have negative impact on chance of people wearing masks. Lastly, people who performed mask-wearing behaviour are likely to take more other precautions comparing to those who didn't work a mask.

Introduction

Covid-19 had become a global pandemic since March 2020 and had brought dramatic change into people's life. Almost all countries in the world had introduced policies aiming to curb the spread of this diseases as well as helping citizens to adopt the situation, these include asking citizens to stay home, avoid both domestic and international travel, wear masks when they visit public space and many others. Initiated by the Policy, Elections & Representation Lab ("PEARL") at the Munk School of Global Affairs and Public Policy, a series of surveys were conducted regarding public opinions toward politics and public policy during the pandemic. More specifically,with the aim of helping policymakers to publish more effective regulations to stop the pandemic, PEARL are interested in finding predictors that may influence people's mask-wearing decision, which will be helpful for increasing effectiveness of policy published in the future.

There were 18 waves of online surveys conducted from March till August 2020 targeting Canadian citizens that are above 18 years old, some questions had been modified as time progress and some new questions were added. The resulted dataset include 43272 responses and 2903 variables. This report will use these measurements to investigate possible socio-demographic and attitudinal predictors for mask-wearing behaviour. Another question will be discussed about is whether people who wore masks took more or fewer other precautions than those who don't wear masks.

The method section will explain some important data cleaning choices taken before fitting models. The two generalized linear mixed models employed to answer the research question will also be discussed in this section, which include the reason that they got chosen and contexts of the analysis. Answers to the research questions will be concluded in the result section accompanied by relevant statistics, tables and graphs will also be displayed to help communicate the results. Main findings obtained from the analysis will be explained in the context of research questions at the discussion section, any limitations about the statistical model used and suggestions for future analysis will be discussed in this section too. Finally, the appendix section will include the mathematical equation of the two models, as well as additional tables and figures that were relevant to the analysis and results.

Method

Before starting the statistical analysis process, data cleaning was implemented to ensure the accuracy of model build and to remove obstacles during the analysis as much as possible.

After having a glimpse at the dataset, it is noticed that the mask wearing variable, which is the interested response variable, were added at the third wave of surveys. Thus, all observations collected before were not included in the subset of data used for fitting the model. The measuring scale of the variable correspond to respondents' opinion on seriousness of the COVID-19 pandemic originally marked 1 as very serious, and 4 as not serious at all. This scale was flipped in order to match with scales of other attitudinal predictors and to make the resulted numbers more understandable. Moreover, since respondents were allowed to took multiple surveys across different waves, the repeated respid (the unique identification code assigned to each respondents) were removed to satisfy the independent assumption required for most models.

Model 1: Who is most likely to report using a mask?

As wore mask or not is a binary response questions, which means it is not normally distributed, a generalized linear mixed model with binomial family is going to be fitted for finding predictors of mask-wearing behaviour. The generalized linear model was chosen to allow more flexibility for the distribution of the response variable. Moreover, the reason for chosen the mixed model is that individual's mask-wearing decision can be affected by others living in the same neighborhood and regulations in each provinces, meaning it is possible for observations to be correlated within each area. To eliminate this regional effect, province was used as a random intercept in the fitted model.

Gender, age and education level of respondents were selected as possible socio-demographic predictors. Seriousness of the pandemic, and 12 variables measuring trust level toward professionals in different disciplines were selected as possible attitudinal predictors. Before actually fitting a model, factor analysis was used to categorize the 12 trust variables into 3 groups: trust level toward professionals related to public health, toward people related to business, and toward government (The specific numbers was shown in Table 1 at the Appendix). Factor analysis is a technique that group variables based on their correlation with each other, new variables then can be created by combining values of variables that are grouped together. This was employed to reduce the large number of variables used in a model, which may be problematic as the model got too complicated.

Model 2: Do people who wear masks take more or fewer other precautions than people who don't wear masks?

The response variable for this part is the amount of precautions a respondent had took, it is a non-negative numeric variable and a poisson family will be used for the model. Similar as above, people's choice of taking precautions are related with the choice of those around them and the regional regulations, thus the correlation within each region need to be taken into consideration. As a result, a generalized linear mixed model using province as a random effect was fitted to solve the research question.

A new variable was created as a sum of all the precautions took by each respondents other than wearing a mask, and it will act as the dependent variable for the model. The main predictor this report will focus on is the mask wearing variable, it was used as a categorical predictor in this model and values were converted to Yes or No. Other predictors chosen to provide additional information include gender, income, and education level.

Result

Model 1

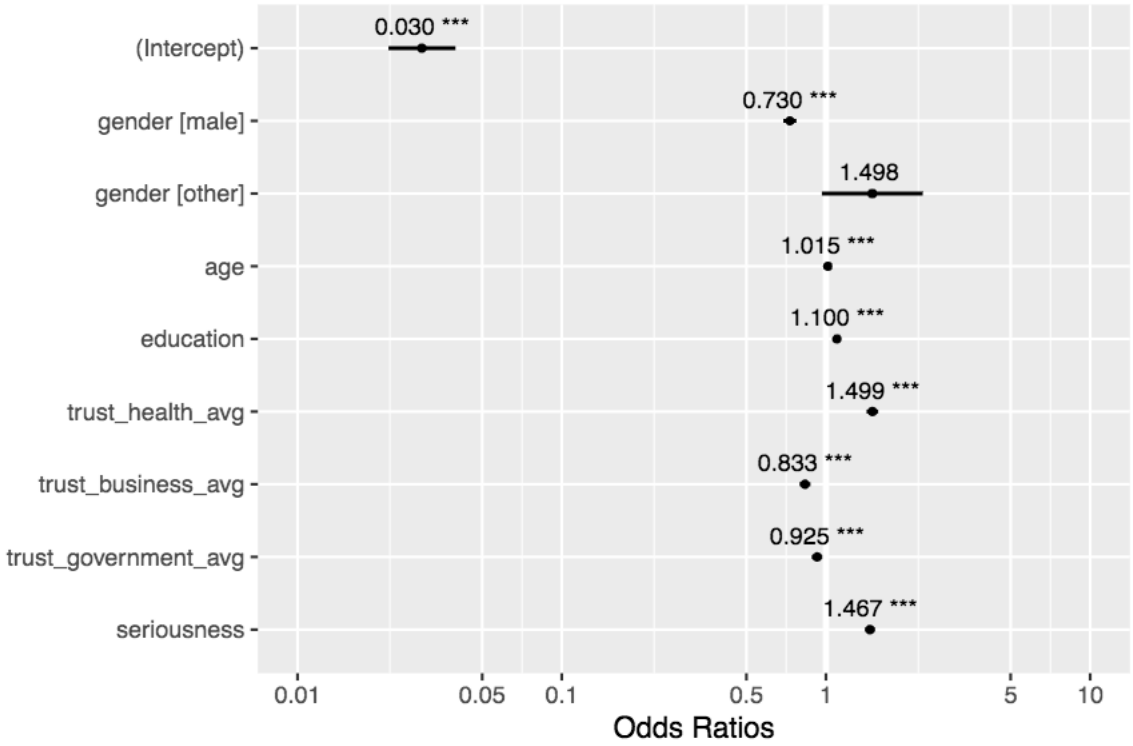

Figure 1: GLMM Estimation and Confidence Interval Regaridng the Mask-wearing Behaviour
Points respresent point estimate of parameters, lines aross them represent their 95% confidence interval, stars located after the numbers indicate the statistical significance of result

As shown in the figure 1 above, a 18 years old female had not completed any schooling, don't have any trust in any groups and don't think the pandemic is a serious threat will have about 3% chance to wore a mask. For socio-demographic predictors, keeping all other variables constant, male are about 27% less likely to choose wore mask and those marked their gender as other are 49% more likely to wore mask comparing to female, but the difference for other gender is not statistically significant with a p-value greater than 0.05. Moreover, people are more likely to have wore masks when they are more educated, and the chance for them wearing a mask increase as age increase, with the rate of about 1.5% each year of age. Regarding attitudinal predictors, people are more likely to choose wore masks when they have more trust in public health professionals and think the pandemic is a serious situation, less likely to wore masks when they are more trusted in government and people related to business and government. Table 2 in the Appendix section provide more statistic including parameter estimation, standard error and p-values of the result.

In general, older female with high education level, trust a lot toward public health professionals and don't believe in the two other groups, and consider COVID-19 as a serious threat will be most likely to report using a mask.

Model 2

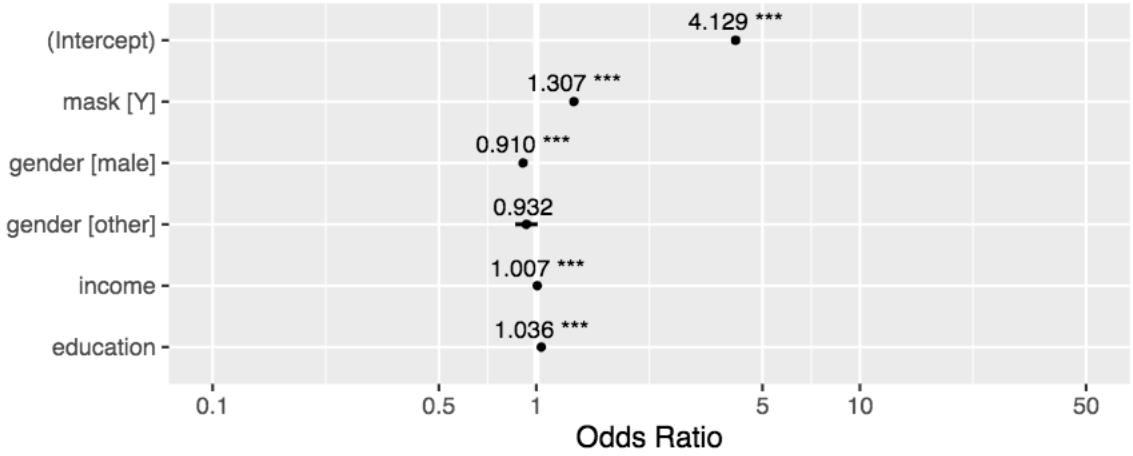

Figure 2: GLMM Estimation and Confidence Interval About Number of Precautions Took.
Points in the figure respresent point estimate, lines aross them represent the 95% confidence interval, the stars located after the numbers indicate the statistical significance of result

As shown in Figure 2 above, female respondents that are 18 years old, don't have any income, had not completed any schooling, and didn't wore masks (referred as the baseline group) tend to take about 4 other precautions. A respondent with similar condition except wearing a mask will tend to take 1.3 times the number of precautions taken by those who didn't wear one, which will be about 6 precautions other than mask-wearing. Some other results include that male respondent tend to take 0.91 times the number of precautions taken by the baseline group, which is also approximately 4; those marked themselves as other genders take 0.93 times the number of precautions taken by female, but there are lack of statistical evidence to support this answer under a 5% significance level. Having higher income and higher education can also increase the chance of people taking more precautions,but the effect won't make much difference since their estimation are about 1 times and 1.04 times the number of precautions taken by baseline group respectively.

Table 3 in the Appendix section had provide more statistic including parameter estimation, standard error and p-values of the result.

Discussion

In general, older female with high education level, trust a lot toward public health professionals, don't believe in business professionals or the government, and consider COVID-19 as a serious threat will be most likely to report using a mask. Moreover, people who performed mask-wearing behaviour are likely to take more other precautions comparing to those who didn't work a mask; gender, income and education don't seem to have a huge effect on people's choice to took more precautions or not.

There are several limitations of the result obtained, a major one is that observations with repeated respid (the identification code for each respondents) were removed to obtain a valid model, the model will be more complete if all observations were included. Add respid as a random intercept might is the alternative way to keep the independent assumption of the model, but the lme4 package employed for this report wasn't able to handle it. Fitting a Bayesian model from inla package with appropriate priors and other settings might solve this problem and this is the potential next step of analysis for the first research question.

Another limitation of this analysis is that we have not specify priors for the random effect in both models and were using the default prior set by the package. If more time was allowed, the models could be benefitted and arrived at a more accurate result from using priors that are specifically chosen based on the data. More specifically, this could be solved by taking more time to test different priors by checking prior and posterior distribution of the data until finding the more suitable one.

Appendix

Mathmatical Expresson of Models

The equation for the first model is as the following:

$$Y_i \sim Binomial(1, \mu_i)$$

$$log(\frac{\mu_i}{1-\mu_i}) = \beta X + Province_i$$

Y_i is the dichotomous response variable representing whether the i^{th} respondent had wore a mask or not, with 1 meaning a positive response and 0 the opposite. μ_i is the probability of i^{th} respondent choose to wear a mask based on all the predictors. X are the predictors as described in the method section and β are their estimated coefficient. $Province_i$ is the fitted random effect.

The equation for the second model is as the following:

$$Y_i \sim Poisson(\mu_i)$$

$$log(\mu_i) = \beta X + Province_i$$

Y_i is the number of precautions the i^{th} respondent had took and μ_i is the average number of precautions took among all respondents. X are the predictors as described in the method section and β are their estimated coefficient. $Province_i$ is the fitted random effect.

Relevant Tables

Table 1: Result of Factor Analysis

	Factor 1	Factor 2	Factor 3
trust_experts	0.593	NA	0.218
trust_economists	0.362	0.638	0.155
trust_scientists	0.789	0.159	0.159
trust_doctors	0.668	0.160	0.259
trust_legal	0.301	0.549	0.252
trust_uni_professors	0.565	0.315	0.198
trust_federalgov	0.350	0.266	0.592
trust_provincial	0.230	0.321	0.498
trust_publichealth	0.640	0.157	0.484
trust_border	0.312	0.399	0.394
trust_business	NA	0.626	0.280
trust_financial	0.297	0.709	0.138

Table 2: Parameter Estimation of GLMM Regaridng the Mask-wearing Behaviour (Model 1)

	Estimate	Std. Error	z value	Pr(>|z|)	OddsRatio
(Intercept)	-3.5207	0.1489	-23.6499	0.0000	0.0296
gendermale	-0.3146	0.0290	-10.8428	0.0000	0.7301
genderother	0.4038	0.2242	1.8010	0.0717	1.4975
age	0.0146	0.0009	16.7300	0.0000	1.0147
education	0.0951	0.0074	12.7890	0.0000	1.0997
trust_health_avg	0.4047	0.0257	15.7524	0.0000	1.4989
trust_business_avg	-0.1821	0.0239	-7.6233	0.0000	0.8335
trust_government_avg	-0.0775	0.0232	-3.3447	0.0008	0.9254
seriousness	0.3835	0.0218	17.6016	0.0000	1.4674

Table 3: Parameter Estimation of GLMM About Number of Precautions Took (Model 2)

	Estimate	Std. Error	z value	Pr(>|z|)	OddsRatio
(Intercept)	1.4181	0.0135	104.8083	0.0000	4.1294
maskY	0.2677	0.0055	48.7694	0.0000	1.3069
gendermale	-0.0939	0.0054	-17.2423	0.0000	0.9103
genderother	-0.0702	0.0406	-1.7284	0.0839	0.9322
income	0.0072	0.0017	4.1467	0.0000	1.0072
education	0.0356	0.0015	24.3105	0.0000	1.0362