Evaluation on factors affecting Americans' participation in contact tracing and quarantine

Yujing Zhang, Yingtong Zhong

University of Pennsylvania, School of Social Policy & Practice

MSSP 897: Applied Linear Modeling

Instructor: Dr. Momin Malik

Policy Research Report 4

December 17, 2021

Contents

1. Introduction

Recently, the new variant of covid-19, Omicron, has triggered new tension around the world. It again aroused our cautiousness and alert in this winter. To better inhibit epidemic, we need the engagement of the whole society. The question is to what extent the public is willing to follow the rules? And how can we improve their cooperation with the government?

The topic of public intension to follow covid policies is not brand new. Last year, a pew research found that most people are willing to cooperate with covid control actions, especially in terms of quarantine (Mcclain & Rainie, 2020). But there are still a large group of people who prefer not to speak of or share their information to control the spread of covid-19. The willingness to follow governmental instructions are affected by some factors. A previous study in Indonesia showed that attitude towards the behavior, subjective norm and perceived behavioral control positively and significantly affect intention to follow "Stay at Home" during the COVID-19 pandemic in this country (Sumaedi et al., 2021). In another research in UK, non-adherence to all social distance rules had a strong association with vulnerability to COVID-19 and control over social distance (Hills & Eraso, 2021).

In this research report, we focus on factors that have influence on people's willingness of contact tracing and quarantine during the pandemic, and will provide suggestions on making more effective covid-19 policies.

2. Data and Method

2.1. Dataset

We use data from Wave 70 of The American Trends Panel (ATP), which is a national,

probability-based online panel of adults living in households in the United States. The topic of this survey is "Religion in public life, social media's role in politics and society, COVID-19 contact tracing". On behalf of the Pew Research Center, Ipsos Public Affairs ("Ipsos") conducted the Wave 70 survey of the panel from July 13, 2020 to July 19, 2020. In total, 10,211 ATP members (both English- and Spanish-language survey-takers) completed the Wave 70 survey. The overall target population for Wave 70 was non-institutionalized persons age 18 and over, living in the US, including Alaska and Hawaii. The sample consisted of 12,981 ATP members that were still active.

From the original 142 variables, we choose 14 of them to create 8 new variables we need in our regression model. There are two dependent variables and 6 independent variables. In this analysis, we separately conducted two regression models to examine factors affecting willingness of both quarantine and contact tracing. We decide to conduct two models because we found that people's attitude towards quarantine has no relationship with their attitude towards contact tracing engagement. Variables are shown below (for each ordinal variable, smaller number means less likely to participate/less agree with the statement/less knowledge):

Dependent Variables

`quarantine`: Willingness to quarantine for at least 14 days if they were told they should do so by a public health official because they had the coronavirus, ranging from 1 to 4.

`con_trac_engage`: Since speak and share are strongly correlated with each other, with a coefficient of 0.797, we combine them into one variable by adding the numbers up. The new variable con_trac_engage is used to evaluate respondents' contact tracing engagement, ranging from 3 to 12.

Independent Variables

`religious`: Whether the respondent attended religious services in person or online in the past month. (0 for no, 1 for yes)

`smedia`: whether the respondent uses social media sites. (0 for no, 1 for yes)

`ctknow`: Respondents' knowledge of contact tracing, ranging from 1 to 4.

`govtrust`: How much the respondents believe that the government will keep personal records safe from hackers or unauthorized users, ranging from 1 to 4.

`horgtrust`: How much the respondent believe that health organizations will keep personal records safe from hackers or unauthorized users, ranging from 1 to 4.

(This is an omitted variable in our regression model. We did not take it into consideration at first because we thought it could be represented by the variable "govtrust", but then we found that the model became better with this variable.)

`party`: Political inclination of the respondent. (0 for Democrat, 1 for Republican)

Table 1. Variables Description

Terms used	Renamed variable in R	Original variable	Scale
Contact willingness	speak	CONTACTHOMEF2_W70+CONTACTPHONEF1_W70	1:Very likely ~ 4:Not at all likely
Information sharing willingness	share	SHAREINFO1_a_W70+SHAREINFO1_b_W70	1:Very likely ~ 8:Not at all likely
Contact tracing engagement	con_trac_engage	speak+share	1:Very likely ~ 12:Not at all likely
Quarantine willingness	quarantine	HEALTHRECF2 + HEALTHRECF1	1:Very likely ~ 8:Not at all likely
Religious engagement	religious	ATTENDMONTH + ATTENDONLINE	0:No ~ 1: Yes
social media usage	smedia	SNSUSE	0:No ~ 1: Yes
Knowledge of contact tracing	ctknow	CTKNOW1	1:A great deal ~ 4:Not at all
Trust in government	govtrust	RECSAFEd	1:Very confident ~ 4:Not at all confident
Trust in health organization	horgtrust	RECSAFEb	1:Very confident ~ 4:Not at all confident
Party affliation	party	PARTY_W70+PARTYLN_W70	0:Democrat ~ 1: Republican

2.2. Method

For this analysis report, we use multiple regression to examine the relationship between people's willingness of participation of covid policy and their social behaviors. Specifically, we are interested in the factors of people's religious participation, social media using, knowledge of contact tracing, trust to public institutions and their political attitude. Although our dependent variables are ordinal, not continuous, the OLS regression model is helpful for finding out the relationship.

We first examine two models based on two different dependents variables and give them a basic interpretation. Then, we use diagnostics for linear regression on the model, in terms of looking for violations of linearity, homoscedasticity, and normality of residuals, and checking

potential outliers. After making appropriate corrections for potential violations, we will compare new model fits to the original one, and draw our conclusion.

3. Descriptive statistics

Limiting the dataset to only non-missing observations across all these variables, we have 9640 observations. We first use 7 variables, the descriptive statistics are given in table 2.

Table 2. Descriptive Statistics of All Variables

Variable	Type	Min	Max	Mean	Median	S.D.	NA
con_trac_engage	Ordinal	3	12	5.7	5	2.46	105
quarantine	Ordinal	1	4	2.15	2	1.17	31
religious	Dummy	0	1	0.41	0	0.49	35
smedia	Dummy	0	1	0.76	1	0.42	16
ctknow	Ordinal	1	4	1.96	2	0.93	80
govtrust	Ordinal	1	4	2.59	2	0.88	28
party	Dummy	0	1	0.42	0	0.49	340

Our two dependent variables are ordinal variables. Notice that a smaller number means more willingness to participate. With their mean and median, we can find that in general, the public is willing to participate in contact tracing if necessary, but are not that willing to participate in quarantine when they are required.

Among all of the 6 independent variables, we have 3 dummy variables and 2 ordinals. For dummy variables, we can see that 41% respondents had participated religious services within

a month when they answered the question; 76% of them had the experience using social medias. As for their party affiliation, there are 42% consider themselves more as a republican and 58% are more as a democrat.

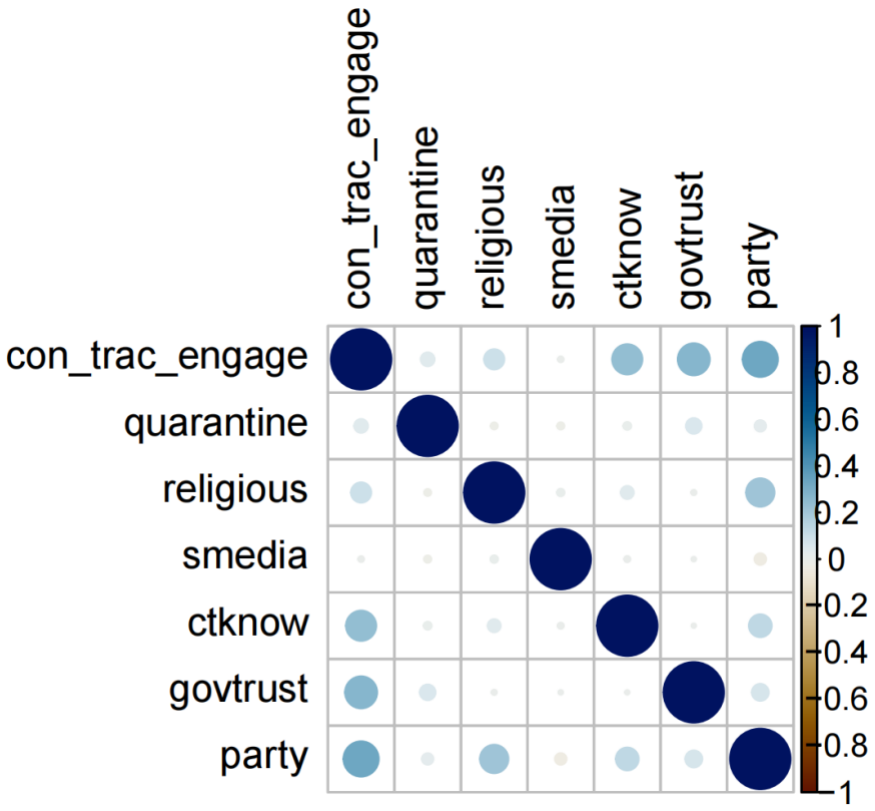
From the mean and median of `ctknow`, which are 1.96 and 2, we know that majority Americans at least have somewhat heard about "contact tracing". When looking at `horgtrust` and `govtrust`, we find that their mean values are larger than 2 while medians are both 2. This means on the one hand, most people are somewhat believing in government can protect their personal information, while on the other hand, number of people who are not at all confident with the government is larger than those who are very confident.

Next, we examine correlations between variables. The correlation matrix is shown in Table 3 and Figure 1.

Table 3. Correlation Matrix

	con_trac_engage	quarantine	religious	smedia	ctknow	govtrust	party
con_trac_engage	1						
quarantine	0.0488	1					
religious	0.1077	-0.0115	1				
smedia	0.0077	-0.0125	0.0135	1			
ctknow	0.2457	0.015	0.0435	0.0088	1		
govtrust	0.2763	0.0638	0.0063	0.0047	0.0045	1	
party	0.3346	0.033	0.216	-0.0337	0.1384	0.075	1

Figure 1. Correlogram



Fortunately, there do not appear to be any potential collinearity issues. There is a large correlation between contact tracing engagement and independent variables such as `ctknow`, `govtrust` and `party`. However, we don't find any strong correlation between another dependent variable, `quarantine` with other variables. We need to further examine it in the regression model to see if there is a relationship exist.

4. Basic Regression Model

In this part, we consider two regression models. One with dependent variable of `quarantine` and the other with dependent variable of `con_trac_engage`. The results are shown in Table 4.

quarantine~religious+smedia+ctknow+govtrust+party (lm_quarantine)

con_trac_engage~religious+smedia+ctknow+govtrust+party (lm_con_trac_engage)

Table 4. Original Models

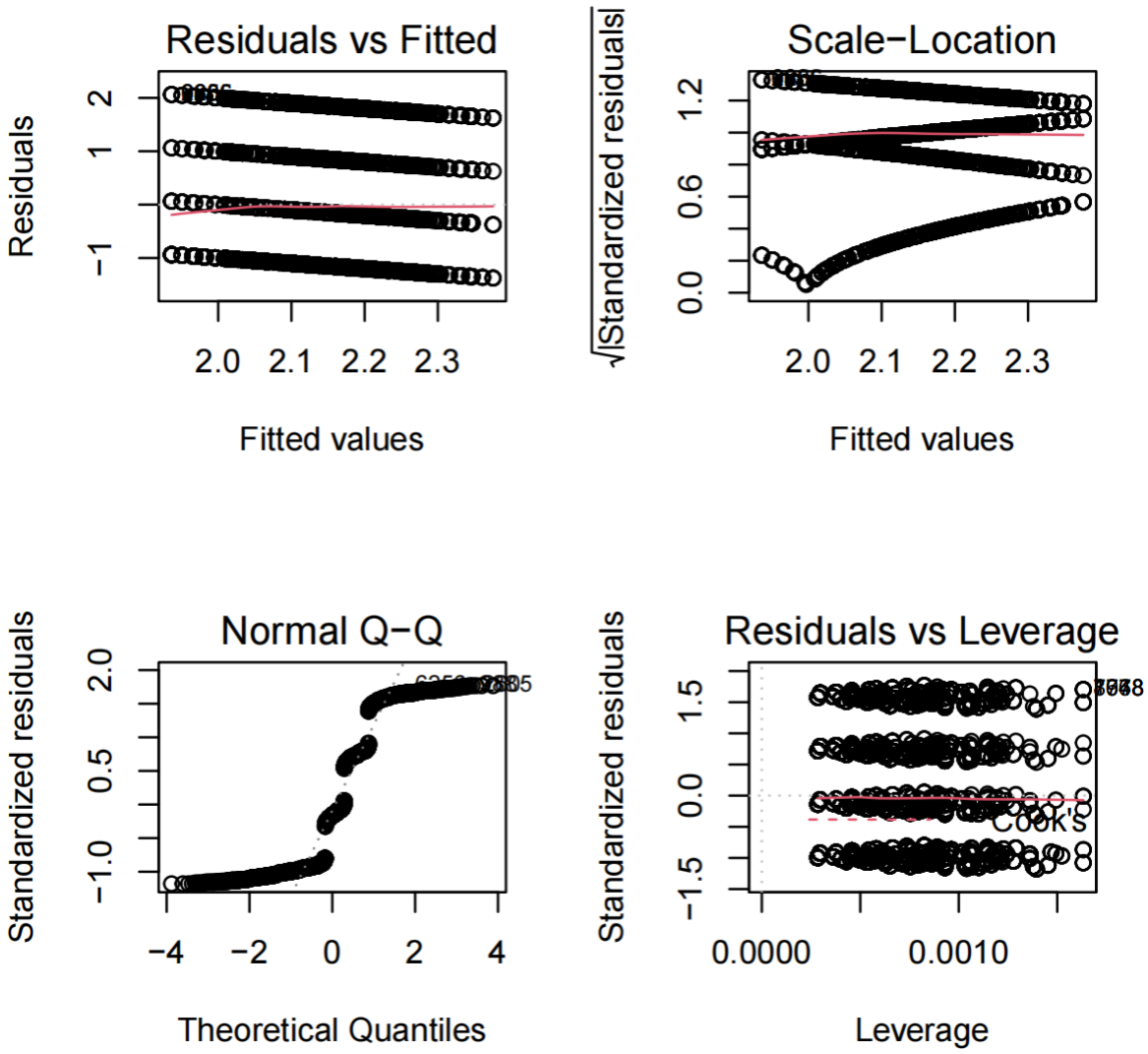
	Dependent variable:	
	quarantine	con_trac_engage
	(1)	(2)
religious	-0.045*	0.183***
	(0.025)	(0.046)
smedia	-0.032	0.079
	(0.028)	(0.052)
ctknow	0.014	0.538***
	(0.013)	(0.024)
govtrust	0.082***	0.712***
	(0.014)	(0.025)
party	0.072***	1.394***
	(0.025)	(0.047)
Constant	1.916***	2.054***
	(0.050)	(0.093)
Observations	9,640	9,640
R^2	0.005	0.218
Adjusted R^2	0.005	0.218
Residual Std. Error (df = 9634)	1.168	2.172
F Statistic (df = 5; 9634)	10.597***	537.266***
Note:		*p**p***p<0.01

For both two variables, we interpret p-values at the 0.05 level, where any p-values above the critical value of 0.05 indicate insufficient evidence to reject a null hypothesis that there is no relationship between a given independent variable and the response, and p-values below the critical value of 0.05 indicate sufficient evidence to reject a null hypothesis. Furthermore, we interpret R^2 as a measure of how much variance in the response is explained by the given combination of the independent variables. And note that a negative coefficient in the table represents more likely to participate, because in our dataset, larger numbers mean not to do or participate something.

The first model examines the relationship between willingness to quarantine for at least 14 day and a set of independent variables. We can see that R^2 value is 0.005, which means those independent variables can only explain 0.5% of the variance in this model. And the plots of this model, such as residuals vs. fitted value, Q-Q plot and residuals vs. leverage (shown below), also prove that the model does not fit very well. So, overall it's not that meaningful on showing a relationship between willingness to quarantine and the factors we chose as independent variables, and we will not continue working with it in the following analysis.

However, there are still three coefficients significant, which tell religious activity, knowledge of contact tracing and party affiliation are related to the willingness to quarantine. It somehow tells us people's participation can be affected by these factors, although the model itself is bad. More specifically, controlling other variables, those who had religious activities within a moth are more likely to participate in required quarantine; those who are confident in government on protecting their information are more willing to quarantine; and republicans are less likely to participate than democrats.

Figure 2. lm_quarantine Plots



The second model about factors on public's willingness to participate in contact tracing activities, which we will use for further analysis in this report, performs much better than the first one. This model can explain 21.8% of the total variance of the dependent variable, and all coefficients except the one of social medial using are significant.

This model tells us that religious activity, knowledge of contact tracing, confidence to the government and party affiliation are significantly associated with one's willingness of participating

in contact tracing during the pandemic. The constant tells us that with an average knowledge of contact tracing and average confidence in the government, a democrat who didn't attended religious activities and never uses social media will give a score of 2.05 on his willingness to participate in contact tracing. This score does not have practical meaning because the scale starts at 3, but we can just regard it as a very strong willingness. When controlling other variables, a person who attended religious service or who is self-identified as a republican is less likely to cooperate in contact tracing; and with more knowledge about contact tracing and more confidence towards government, one will be more likely to participate.

In our following analysis, we will drop lm_quarantine and use lm_con_trac_engage as our Model 1.

5. Assumptions Diagnose of Original Regression Model

In this section, we will perform diagnostics of the second model. Linearity, homoscedasticity and normality of residuals will be diagnosed to check whether violations exist.

5.1. Linearity

Figure 3. Correlations Plots

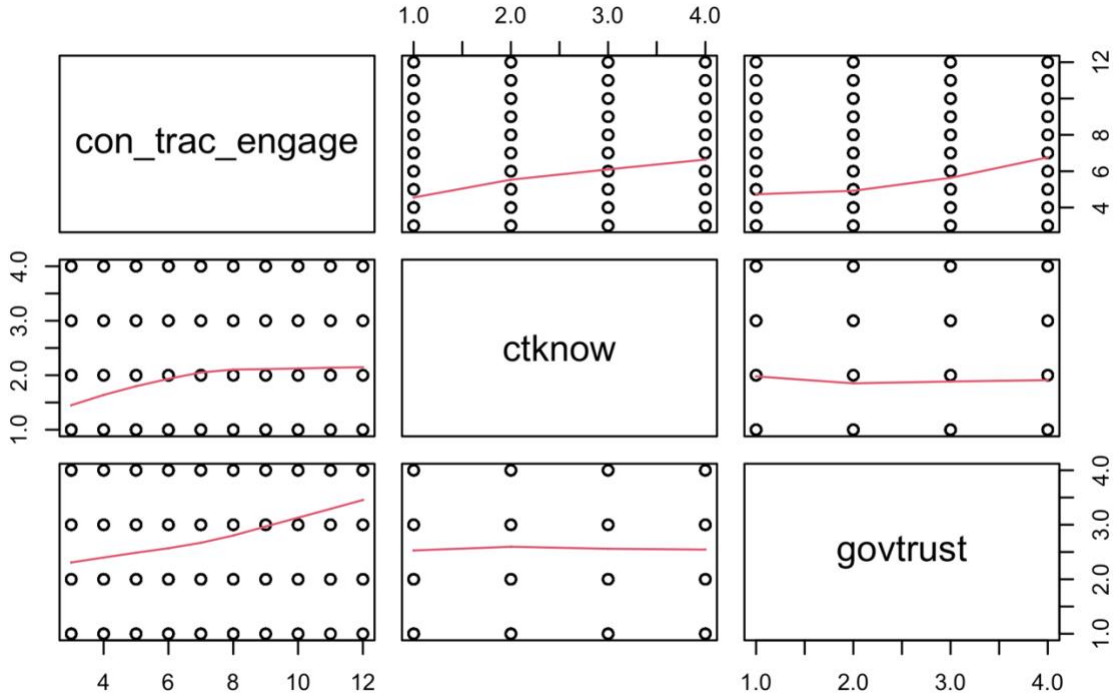
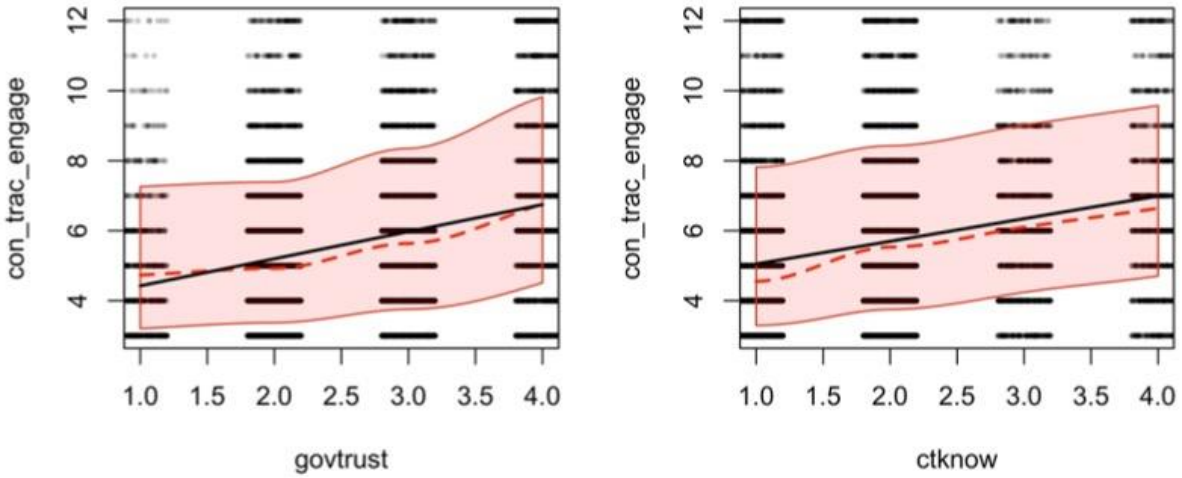


Figure 4. Independent Variables Correlations Plots



To test whether the relationship between independent variables (IVs) and Dependent Variable (DV) is linear, a scatter plot matrix of each IV against the DV and other IVs is produced. We only examine linearity of continuous variables since dummy variables meet the assumption of linearity by definition.

As we can see from the scatter plot matrix in figure 3, knowledge of contact tracing (`ctknow`) and trust in governments (`govtrust`) are linearly associated with contact tracing engagement (`con_trac_engage`). Though the scatter points sparse untypically due to small variance of variable range.

Bivariate plots in the figure 4 also show that both `ctknow` and `govtrust` have slight nonlinearities, but a line of best fit (in black line) still falls within a lowess confidence interval (in red line and area). Thus, it can be reaffirmed that linearity assumption is met.

5.2. Homoscedasticity

Next, we evaluate the assumption of homoscedasticity by examining whether variance of the model residuals is constant for each value of IVs, we plot the model residual values against the fitted values for the model.

As we can see from standard plot in figure 5 we might see evidence of heteroskedasticity when looking at the residuals spreading unevenly in two tails. To get more information we plot a lowess confidence interval. We can see from figure 6 that variance seems to slightly over the range of fitted values. However, it is still a rather appropriate model to interpret because the LOWESS fit line is rather flat around the zero line.

Figure 5. Residuals vs Fitted, model 1

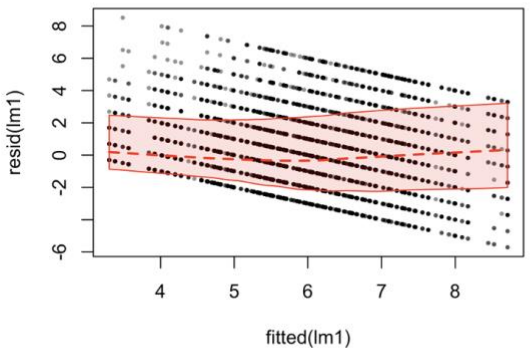
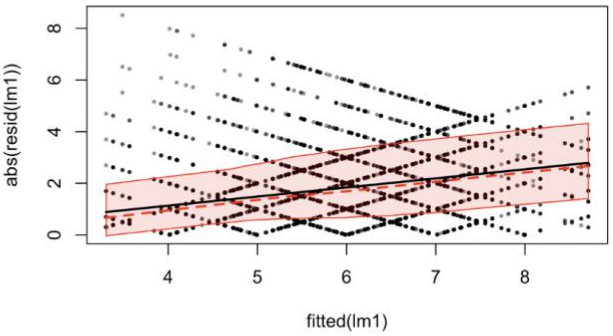


Figure 6. Abs value of Residuals vs Fitted, model 1



5.3. Normality of Residuals

From the histogram of the residuals in figure 7, the distribution of residuals of the original model is roughly a standard normal distribution with mean zero. For a more precise look, we can refer to Q-Q plot in figure 8. Specifically, the distribution is basically symmetrical. However, the right and left tails seem to deviate off the dark dash line with some potential outliers on the right tail. We will examine what to address in the next steps.

Figure 7.Histogram of model 1 residuals

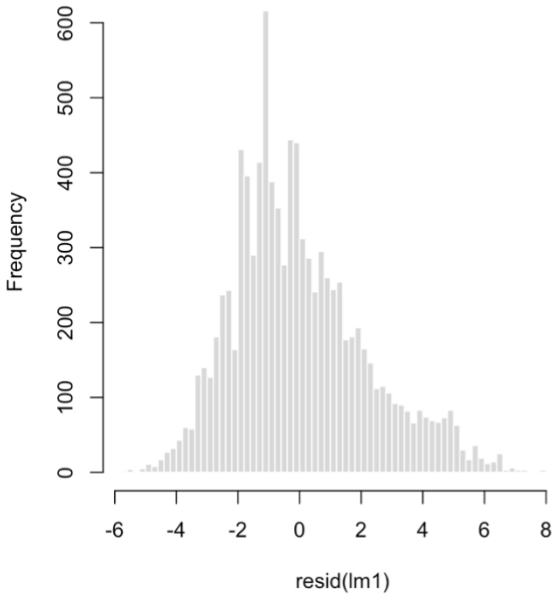
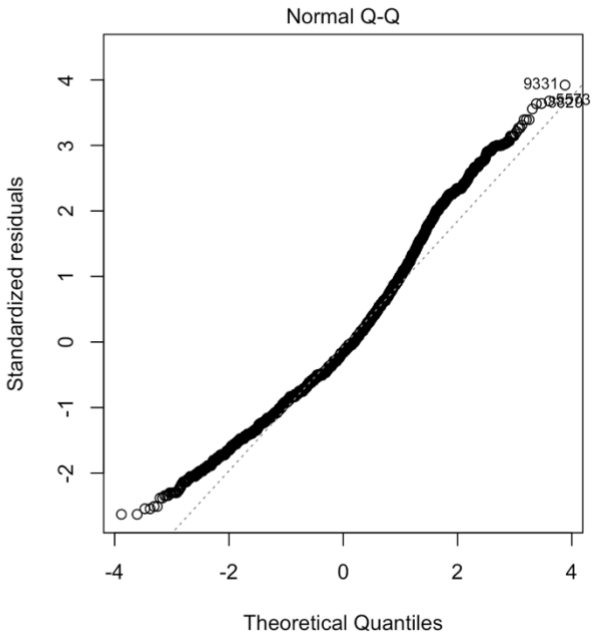


Figure 8.Q-Q Plot of model 1 residuals



6. Model iteration

6.1. Added Variable as Omitted Relevant Variable

To improve our model performance, we consider adding trust in health organization (`horgtrust`) variable into our model. On the one hand, we would like to assess whether trust in health organization affects contact tracing engagement.

Correlation analysis

We start by testing the `horgtrust` with other independent variables and the results are presented in Table 5 It is obvious that `horgtrust` is correlated with other independent variables in the model.

Table 5. horgtrust Correlation analysis

Other IVs	Pearson Coefficient	P value
religious	0.075	p<0.01
party	0.284	p<0.01
govtrust	0.581	p<0.01
smedia	-0.042	p<0.01
ctknow	0.091	p<0.01

Regression analysis

Then we test the correlation between `horgtrust` and the dependent variable. As presented in Table 6, `horgtrust` is statistically significant with $p<0.05$. This indicates `horgtrust` is a determinant of the dependent variable `con_trac_engage`.

Based on the correlation and regression analysis, `horgtrust` meets the definition of omitted variable and it should be included in our model.

Table 6.Regression Results of Omitted Relevant Variable

	Dependent variable:
	con_trac_engage
horgtrust	1.198***
	(0.027)
Constant	2.794***
	(0.067)
Observations	9,640
R^2	0.175
Adjusted R^2	0.175
Residual Std. Error	2.231 (df = 9638)
F Statistic	2,039.800*** (df = 1; 9638)
Note:	*p<0.1; **p<0.05; ***p<0.01

6.2. Outliers Investigation

As we diagnosed in the previous Q-Q plot and residuals vs. fitted plot, we spot some potential outliers. After adding the omitted variable, we investigate the outlier of model X. To take a closer look at specific case with high leverage and influence, we have a set of outliers to investigate: the observations in rows 2454, 7939 and 9359. In the table 7, we can see that those cases are not as extreme outliers as the variance is little in our dataset. Moreover, as we can see from figure 10, none of these have a particularly large value for Cook's distance, and so it may not be necessary to remove them as they are still valid samples.

Figure 9.Cook's distance, model 2

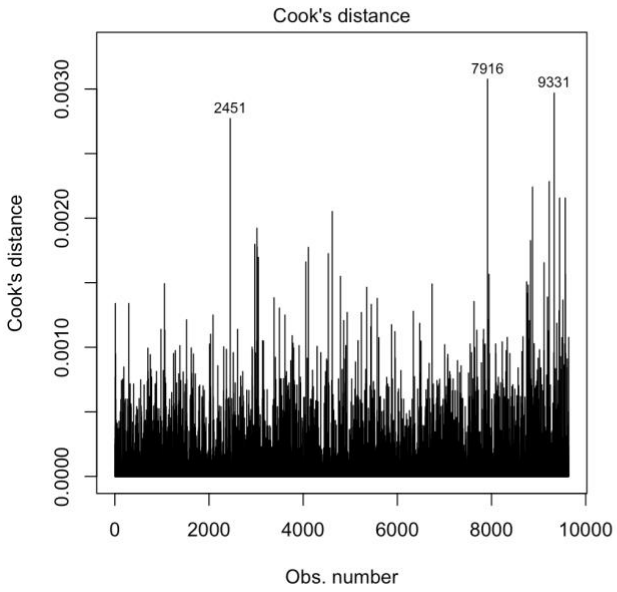


Figure 10.Residuals vs Leverage, model 2

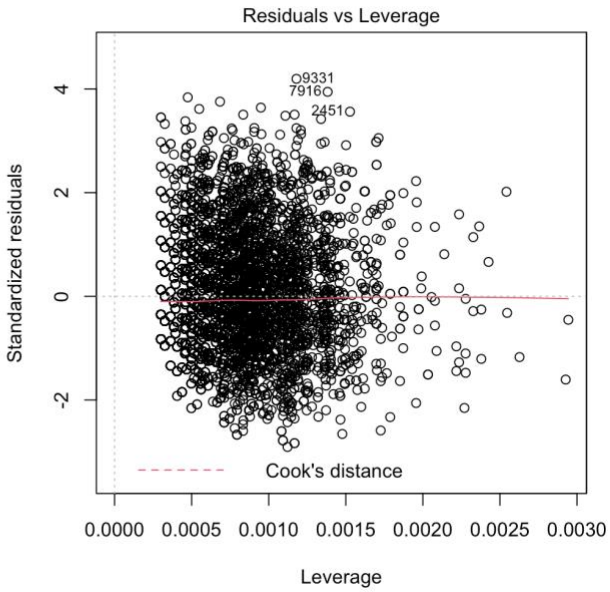


Table 7.Outliers Observations

Variable	Row		
	2451	7916	9331
speak	4	4	4
share	8	8	8
con_trac_engage	12	12	12
quarantine	1	1	4
religious	0	1	1
smedia	0	0	0
ctknow	4	1	1
horgtrust	1	1	1
govtrust	1	3	1
party	0	0	0

6.3. Interaction effect

As we are interested in whether party affiliation (`party`) interacts with religious

engagement (`religious`) in affecting contact tracing engagement (`con_trac_engage`), we will include interaction effect of `religious` and `party` in our model.

Before incorporating interaction effects into a multiple regression model, we will center continuous variable `ctknow` into `cctknow`, `horgtrust` into `chorgtrust`, `govtrust` into `cgovtrust` to deal with problems of multicollinearity when adding interaction effects. In addition, centering variables also make it easier to interpret the regression results.

7. Model interpretation

7.1. Diagnosis improvement

We can see that diagnosis of model 3 improves a little bit compared to model 1, especially the Q-Q plots and residuals vs leverage plots in that the distribution of residuals is slightly more normal and the standardized residuals are estimated as closer to zero across the range of leverage.

Unfortunately, we were not able to meet the perfect homoscedasticity assumption, nor the long tails that made the distribution of residuals a little bit wider than a normal distribution. Yet, model 3 is still a plausible, explainable model to interpret after detail diagnosis and iteration.

Figure 11. Residuals vs Fitted, model 1

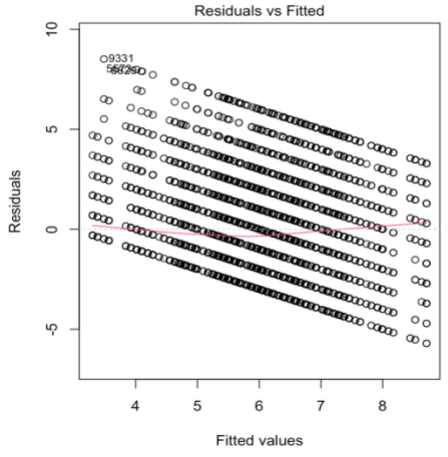


Figure 12. Residuals vs Fitted, model 3

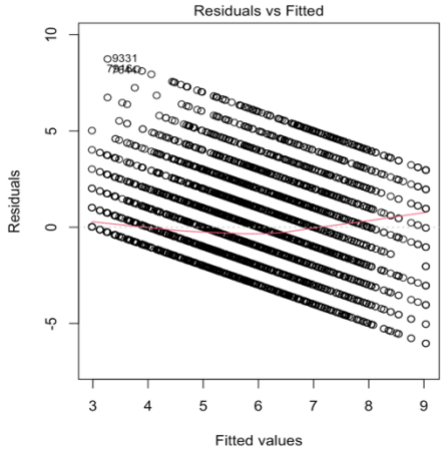


Figure 13. Q-Q Plot of residuals, model 1

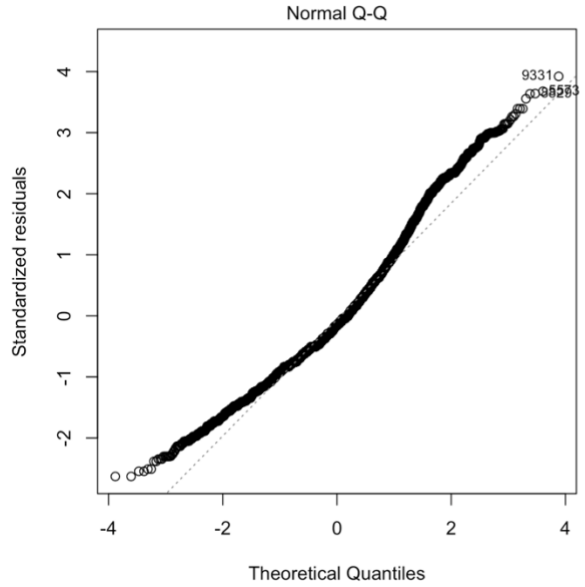


Figure 14. Q-Q Plot of residuals, model 3

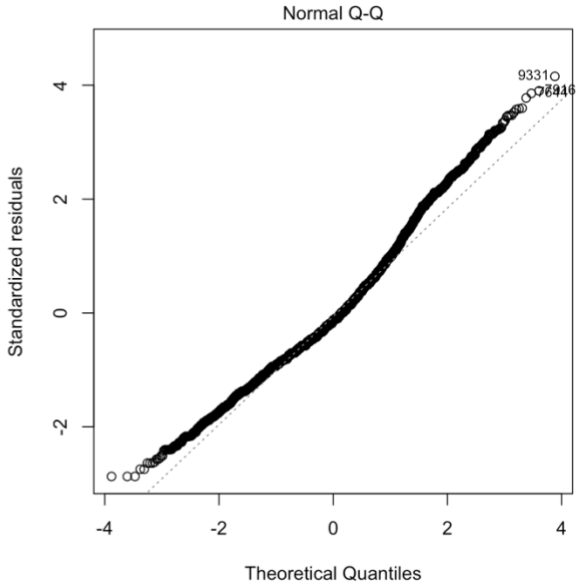


Figure 15. Residuals vs Leverage, model 1

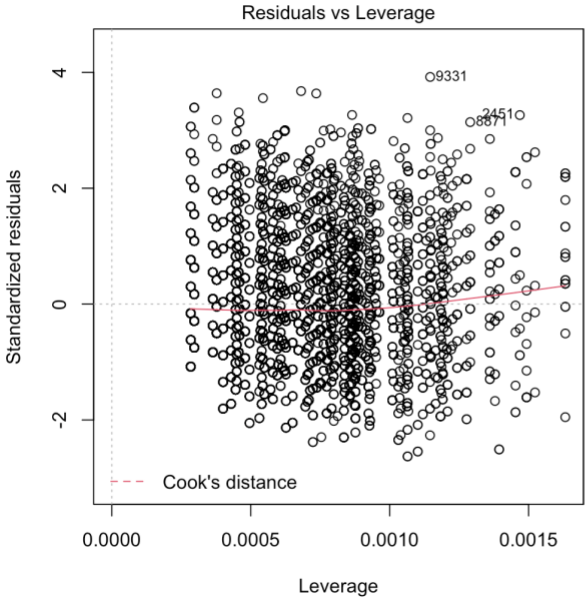
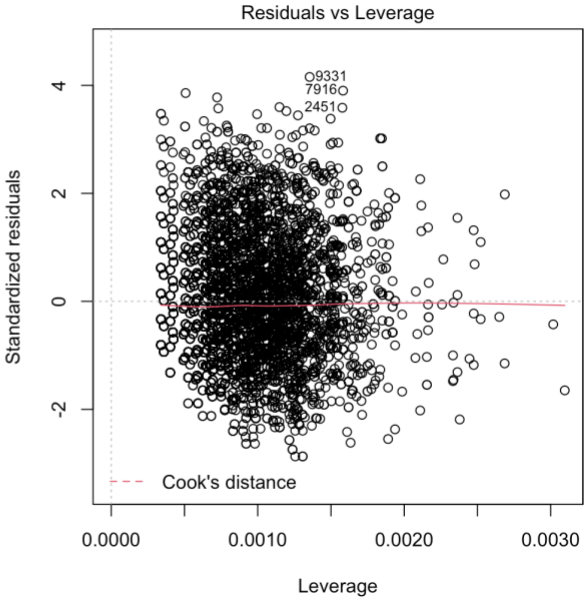


Figure 16. Residuals vs Leverage, model 3



7.2. Model Performance and interpretations

Table 8.Regression Results, Model 1, 2, 3

	Dependent variable:		
	con_trac_engage		
	(1)	(2)	(3)
religious	0.183***	0.154***	0.282***
	(0.046)	(0.045)	(0.061)
party:religious			-0.282***
			(0.090)
smedia	0.079	0.142***	0.144***
	(0.052)	(0.051)	(0.051)
ctknow	0.538***	0.496***	
	(0.024)	(0.023)	
govtrust	0.712***	0.264***	
	(0.025)	(0.030)	
horgtrust		0.813***	
		(0.032)	
cctknow			0.494***
			(0.023)
chorgtrust			0.810***
			(0.032)
cgovtrust			0.267***
			(0.030)
party	1.394***	1.071***	1.194***
	(0.047)	(0.047)	(0.061)
Constant	2.054***	1.451***	4.999***
	(0.093)	(0.093)	(0.052)
Observations	9,640	9,640	9,640
R^2	0.218	0.266	0.267
Adjusted R^2	0.218	0.266	0.266
Residual Std. Error	2.172 (df = 9634)	2.104 (df = 9633)	2.103 (df = 9632)
F Statistic	537.266*** (df = 5; 9634)	582.623*** (df = 6; 9633)	501.255*** (df = 7; 9632)
Note:			*p<0.1; **p<0.05; ***p<0.01

Model Performance

In table 8, we can combine the original regression model 1, model 2 with omitted relevant variable `horgtrust` added and model 3 with interaction effects of `party` and `religious`.

As we can see, the R^2 and AIC are 0.218 and 42320.5 for model 1, 0.266 and 41708 for model 2, 0.267 and 41701 for model 3. It is apparent that model 2 and 3 perform better than model 1 considering R^2 and AIC. Thus, we are confident in interpreting model 2 as main effects model and model 3 as the interaction model.

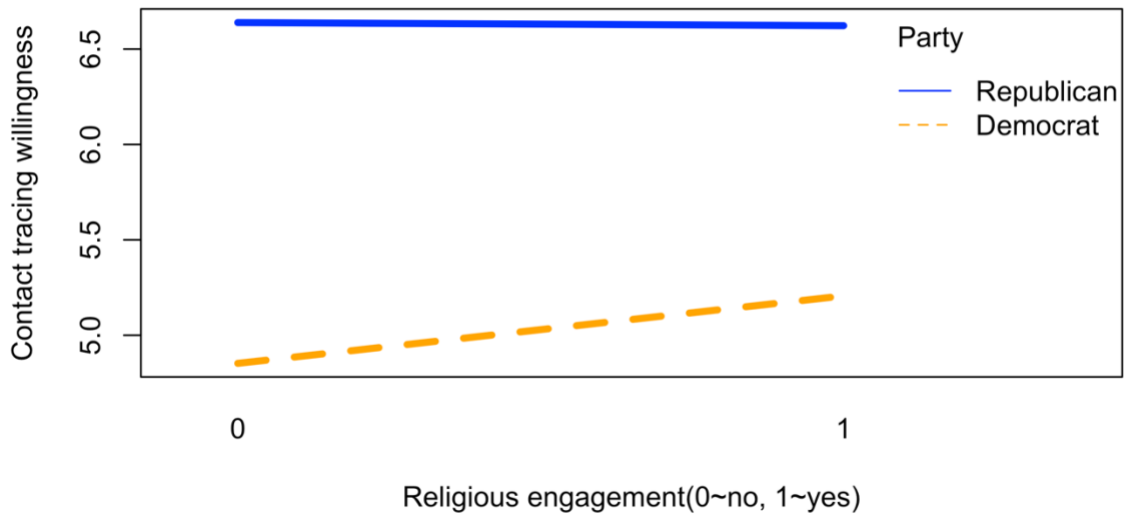
Model interpretations

Both model 2 and 3 are statistically significant with p-value less than .05 for the F-test. The coefficients of all the independents variables are statistically significant with $p<0.05$. Interpreting model 2, holding other X variables constant: people who use social media is 0.142 unit less likely to engage in covid contact tracing programs than those who do not use; the likelihood of contact tracing engagement willingness increases by 0.496 unit if people have more knowledge of Covid contact tracing programs; people trust public health organizations and federal government are 0.813 and 0.264 unit more likely to engage in covid contact tracing programs respectively than those who do not.

In the interaction model (model 3), we find that the p-value for the `party*religious` interaction term is significant ($p<0.05$), meaning that there is evidence for `party` being moderated by `religious` (i.e., the impact of religious engagement on people's contact tracing engagement depends on their party affiliation). Interaction effects plot are showed in figure X.

Since we have centered all of our continuous variables, the intercept term in model 3 tells us the expected likelihood of contact tracing engagement score (1 denotes very likely and 12 very unlikely) for the "average" person (i.e., a person who is self-identified democrat, a person who DOES NOT engage in religious activities, a person who DOES NOT use social media, a person who has medium knowledge of contact tracing programs and a person who has average trust in public health organizations and federal government.)

Fugure 17. Interaction between Party and Religious



To interpret the interaction effect of party affiliation and religious engagement, on average, people's willingness in engaging contact tracing programs is 0.282 unit higher for those who do not actively engage in religious practice than who DO. However, this disparity is not uniform across party affiliation. A self-identified republican who DOES NOT engage in religious practice is 1.194 less likely to engage in contact tracing program than democrat who DOES NOT engage in religious practice. A self-identified republican who DOES engage in religious practice is 0.282 more likely to engage in contact tracing program than democrat who DOES NOT engage in

religious practice.

8. Conclusion

We started by examining initial models of quarantine and contact tracing, but only retained the second one because the former model lacks credibility in interpretations. Interestingly to say, people's attitudes towards Covid corporation especially quarantine are quite random, with only some clues indicating that people who attended religious services within a month, trust federal government, and is self-identified as democrat might more likely to comply with quarantine requirements.

In terms of 'speak' and 'share' contact tracing information sharing and engagement, the model seems to work well in outcomes delivery. After detailed diagnosis and model iteration, we conclude from our statistically significant model that religious engagement, social media usage, knowledge of contact tracing programs, party affiliation and trust in health organization and federal government are factors affecting people's Covid contact tracing engagement. Based on the interaction effects evaluations, the effect of party affiliation on contact tracing engagement is clearly moderated by religious engagement. Replicants who are less active in religious engagement in person or online show high willingness to engage in contact tracing programs. People who have high trust in public health organizations shows the highest correlation to contact tracing engagement.

Those findings inspire government and policy makers in dealing with Covid public health crisis. Increasing popular trust in people and advocating and educating certain groups of people may help improve public's willingness to comply with Covid contact racing engagement and requirements, compliance that is one of the most significant factors in battle with Covid.

9. Limitation

There are three main limitations in our analysis report:

1) Dataset and variable choosing

Our dataset does not include demographic data, which may be influential on the final results. Generally, respondents' age, race and gender, or maybe where they live in should be considered in the regression model, but we missed all the information.

Within the available data, there may have other omitted variables we didn't notice, just like `horgtrust`. Since our final regression model is still not perfect, adding other factors into the model may help us improve it.

2) Low variance

All our variables are dummy or ordinal, and ordinal variables usually ranges from 1 to 4, which results in a very small variance. This small-scale stops us from diving deeper into the subtle differences in the willingness to participate among the huge number of respondents.

3) Regression method

OLS may not be the best choice for dealing with dataset in ours type, with only dummy or ordinal variables. Although the result is readable and interpretable, there may be other analysis methods that will perform better with the data.

Reference

Hills, S., & Eraso, Y. (2021). Factors associated with non-adherence to social distancing rules during the COVID-19 pandemic: A logistic regression analysis. *BMC Public Health*, *21*(1), 352. https://doi.org/10.1186/s12889-021-10379-7

Mcclain, C., & Rainie, L. (2020, October 30). The Challenges of Contact Tracing as U.S. Battles COVID-19. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/2020/10/30/the-challenges-of-contact-tracing-as-u-s-battles-covid-19/

Sumaedi, S., Bakti, I. G. M. Y., Rakhmawati, T., Widianti, T., Astrini, N. J., Damayanti, S., Massijaya, M. A., & Jati, R. K. (2021). Factors influencing intention to follow the "stay at home" policy during the COVID-19 pandemic. *International Journal of Health Governance*, *26*(1), 13–27. https://doi.org/10.1108/IJHG-05-2020-0046