How did the Number of COVID-19 cases Affect the Distancing Policy of the Canada Government?*

Wong Basil Chi Fung

06 February 2022

Abstract

Two years has passed since we first discovered coronavirus in Toronto. Premier of Ontario, Doug Ford, has declared Toronto as state of emergency serval times and people in Toronto have experienced multiple lockdowns. There have been quite a number of protests during this pademic criticing the decisions of the governemnt. Therefore in this resarch we will obtain a open data from Toronto Open Data Protal to study whether to number of COVID-19 cases throughout the past 2 years is the main indicator for the government to implement the distancing/lockdown policy.

1 Introduction

Over the past two years, the COVID-19 pandemic affected many people around the world and people in Toronto are under lockdown most of the time. There are multiple lockdown from March 2020 to February 2022 (Ranger 2021), with the first from March 2020 to June 2020, second one from December 2020 to March 2021 and the last/most recent one being from December 2021 to February 2022. Compared to other major cities, Toronto has the largest number of days of banning indoor dining, as of May 2021(Levinson-King 2021). With other major cities such as New York, Paris and London have around 250 days of indoor dining restrictions, Toronto has over 360 days in total which means that restaurants in Toronto were closed most of the time during 2020. All these restrictions were implemented by the Canada government and the goal of this research is to find out the key indicators affecting the current policy making process and future plans. Why the government picked the above 3 duration for lockdwon? Is it directly related to the number of COVID-19 cases in those months or are there any other factors affecting them such as fatal rate, source of infection and burden from public health unit and hospitals? In order to find out the correlations, we must analyze the Toronto COVID-19 data set provided by Toronto Public Health (Toronto Public Health 2022). It is very important for us to look into the graphs and numbers reflected from the data set since studying the relationships between variables is the easiest and simplest way to reflect the actual situation of Toronto throughout the pandemic and for us to provide evidence to support/against the government's policies/plans. People in Toronto have been protesting and waiting for the business to reopen, citizens are already really frustrated due to the lockdown and experiencing lockdown fatigue, they are ready to move forward and put this pandemic behind them (D'Mello 2022). Therefore it is important for us to understand the reasons behind all these COVID related policies and this is why this data set is worth studying.

2 Data

2.1 Data Summary

The data for this research is obtained from the portal of Open Data Toronto (Gelfand 2020). The open data portal is an open sources that is entirely free for the general public. It contains a huge variety of data with different categories such as City Government, Locations and Mapping, Community Services, Transportation

*Code and data are available at: https://github.com/basilwongg/Toronto_covid_cases & https://open.toronto.ca.

etc. Out of all the data sets related to COVID-19, I chose the data set documenting all the COVID-19 cases in Toronto, the data is updated on a weekly basis and was first created on January 2020 when the first case was reported (Toronto Public Health 2022). The data set has 18 variables including ID, assigned ID, outbreak associated, age group, neighborhood name, forward station area, source of infection, classification, episode date, reported date, client gender, outcome, currently hospitalized, currently in ICU, currently intubated, ever hospitalized, ever in ICU and ever intubated (Toronto Public Health 2022). We will be focusing on a few of the variables instead of studying all of them, some of the data is relatively hard to study such as neighborhood names since there are 140 different geographically distinct neighborhoods in Toronto (Gelfand 2020) and if we were to plot a graph, it would be almost impossible to present all the columns and thus making it extremely hard to analyze. Therefore we will not be analyzing these types of data, instead we will study those are more trival and well defined.

2.2 R Packages

As we want to focus on the relationship between the number of COVID-19 cases and the government distancing policy, we will investigate our data set based on the age group, source of infection, date of infection, outcome and ever hospitalized columns. We will be using R (R Core Team 2020) throughout the whole research with the aid of different packages provided by R. The package **opendatatoronto** (Gelfand 2020) is used to extract data from the open data Toronto portal which is the main package we will be using for our research, the raw data set is retrieved by obtaining the specific package from the open data portal. **dplyr** (Wickham et al. 2021) is used for data manipulation focusing on data frames, **tidyverse** (Wickham et al. 2019) is used for data exploration and visualization, **janitor** (Firke 2021), **tidyr** (Wickham 2021) and **lubridate** (Grolemund and Wickham 2011) are used for cleaning the data set so the developer can analyze the data in a more organized way while **ggplot2** (Wickham 2016), **gridExtra** (Auguie 2017) and **scales** (Wickham and Seidel 2020) are the most important packages because they are used to generate graphs and table in order for us to present our findings. We will be using bar charts, histograms and tables to deliver our findings, based on our data characteristic, different types of graphs/tables will come in handy.

2.3 Source of Infection

In order to investigate the relationship, first we need to understand how COVID-19 is spread. According to the US. Department of Health & Human Services (US. Department of Health & Human Services 2021), COVID-19 is spread along the below three main paths:

1. Staying in the same area and breathing in air with an infected person who's exhaling droplets and particles that contain virus
2. Have close contact with the droplets and particles that contain virus with our nose, mouth and eyes
3. Contacting our nose, mouth and eyes with our body part containing virus on them

Therefore it is pretty obvious that the interaction between people is the main channel for the virus to spread, however out of all the social activities, which of them is the most common channel for the virus to spread and turn into a pandemic. The infection sources are divided into 9 fields namely travel, 3 types of outbreak under different circumstances, household contact, close contact and community with the remaining cases as no information and pending (Gelfand 2020). The the outbreak of the virus, it is mainly divided into three subtypes: congregate setting, healthcare institutions and others. Household contact and close contact are cases where the person acquired COVID-19 through family members, roommates, close friends or even co-worker, however infection acquired from school is classified as "other setting - outbreak" (Gelfand 2020).

From (Figure 1) we can see that there are around 120000 cases classified as no information. The number of no information cases are way more than other sources, this may cause a bias in the government policy making process as they are unable to determine the source of the COVID-19 virus. The cause of the no information category might due to the huge medical burden in Toronto Public Health Unit, therefore the government can only decide their policies depending on the information they have.

Neglecting the no information bar, the second largest group is community infection with around 63000 cases followed by household contact of around 34000 cases. The rest of the categories are all below 20000 with

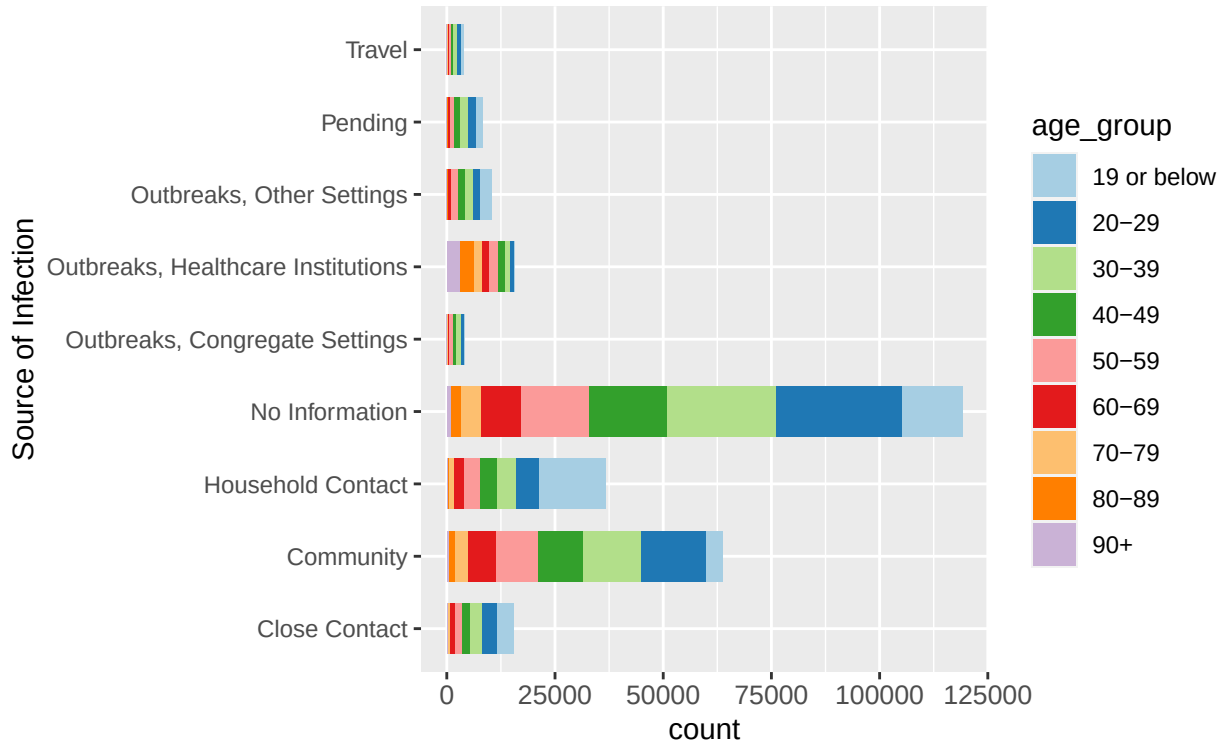

Figure 1: Source of Infections with Age Distribution

travel being the shortest column. The result fits the policy made by the Canada government on the March 20 stating that the Canada USA border was closed to all non-essential travelers (Ranger 2021), ever since the border is shutdown in March, the number of traveling infection has turned from the largest category to the smallest category (Toronto Public Health 2022) therefore it reflected the border shutdown policy is indeed effective. With household contact being the third largest group of infection source, this result did not surprise me because since 2020, Toronto has gone through lockdown three times and hence if someone is infected before lockdown, his/her family will then has a really high chance of getting infected as well. then we proceed with community infection, our second largest source of infection. This is the largest known group of infection in the data set and is the main factor affecting the government's policy. Below we will investigate more on community infection and see how it truly affect the social distancing and lockdown policy in the past 2 years.

According to (Figure 2), a obvious pattern can be easily spotted, apart from the less than or equal to 19 age group, from 20-29 years old onward, the number of community infection is inversely proportional to the increase in age group. With school being one of the largest locations for social students to gather, 8In 2020 March, Ontario Premier Doug Ford announces that all school should be closed and students should not be returning to school for in person learning (Ranger 2021). March 17,2020 Ford declared a state of emergency, ordered some business to be closed including but not only daycares, bars and restaurants, theatres and private school (Ranger 2021). Most of these are places where teenagers usually go and will have social gathering, therefore to avoid these gatherings, the government decided to shutdown those locations.

The group of age with 90+ seems to have the smallest amount of COVID-19 cases, however there might be a bias in the data set affecting the policy making process because number of year who are 90 years old or above are way less than the other age group, so the number of community cases for are hence relatively lower. In March 22, 2020 there was an outbreak in a nursing home in Bobcaygeon (kawarthaNOW 2020), 14 staffs and 3 residents are confirmed to be positive with 50+ more people living there with COVID-19 symptoms. This showed that the outbreak among elderly is also severe in the early stage of COVID-19 outbreak, however throughout the past 2 years, there are not much COVID-19 related policies for the elderly and that might due to the bias showed in the figure. Next we will be looking at the fatal rate of COVID-19.

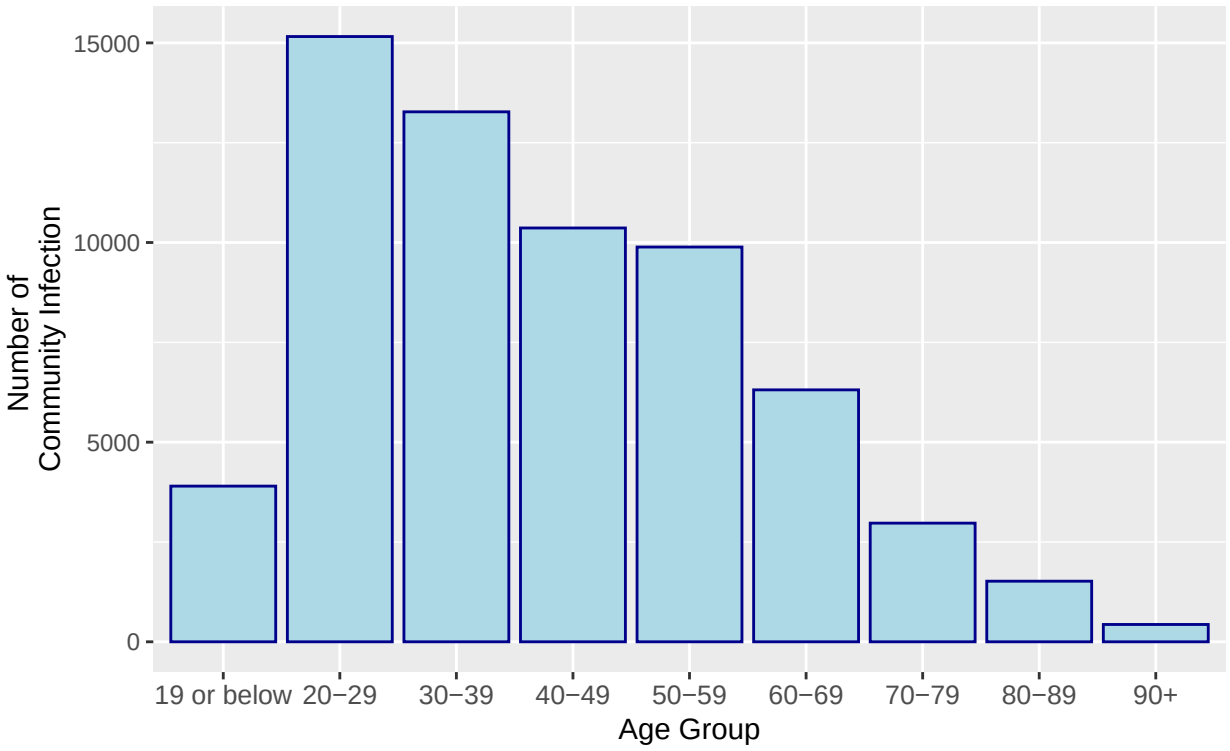

Figure 2: Age Distribution

2.4 Fatal Rate

The reason why COVID-19 was causing such a panic around the world is that COVID-19 virus can cause severely respiratory illness, post COVID conditions and even death. Before the first COVID-19 vaccine was invented in December 11,2020 by Pifzer, the first company to receive approval from Food and Drug Administration (Solis-Moreira 2021), there is no other way for people around the world to fight against this virus, the only thing we can do is to maintain social distancing and keep our personal hygiene such as washing our hands, wearing a mask and avoid poorly ventilated spaces and crowds (US. Department of Health & Human Services 2021). The outcome of COVID-19 are divided into 3 different types: fatal, resolved and active. We are only interested in those cases with outcome "fatal" in this part and will look into the data of the fatal rate due to COVID-19 in Toronto to see how this data affected or biased the government policy making.

From Table 1, we can see that number of deaths in Toronto for people aged 19 or below is 3 while the number of deaths for people aged 90+ is 1065. The huge difference in the number is a strong evidence showing that younger people generally have a better immune system to defend COVID-19. If we take a closer look at the column for fatal rate we will be surprised by how the fatal rate of age group 90+ is 23.5% while age group 19 or below only has a fatal rate of 0.007%. Comparing with other age groups, people who are below 60 years old have a fatal rate of less than 1% while people who are 60-69 has fatal rate of 2.5%, 7.2% for 70-79 age group and 16.5% for 80-89 age group. We can say that there is an exponential growth on the fatal rate, therefore elderly will be at a really high risk if they are tested positive for COVID-19. From Figure 1 and we have mentioned earlier, household contact is one of the major infections sources. If one of the family members has been infected, he/she may easily pass on the virus to the elderly at home causing severe illness with a 23.5% death rate. Therefore the distancing/lockdown measures are not set up to prevent the elderly to go out, in fact they are implemented to avoid their family members spreading the virus to them after being infected in the community. Apart from the distancing/lockdown policies, there are also a few hotline for COVID-19 support specifically for elderly (City of Toronto 2022).

Table 1: Toronto Covid-19 Fatal Rate (by Age Group)

Age Range	Num of Death(s)	Num of cases	Fatal Rate
19 or below	3	42,283	0.007%
20-29	11	59,451	0.019%
30-39	34	52,344	0.065%
40-49	79	40,026	0.197%
50-59	209	36,701	0.569%
60-69	508	22,562	2.252%
70-79	803	11,129	7.215%
80-89	1,357	8,244	16.460%
90+	1,065	4,537	23.474%

2.5 Public Health Unit and Hospitals Burden

In our last session, we will examine the correlations between the total number of COVID-19 cases and hospitalized cases. The burden of the public health unit and hospitals have been greatly increased since pandemic, therefore we would like to study the statistics of the hospitalized people and cross check it with the dates of lockdown so that we can determine if there is a direct relationship between them. Does the number of hospitalized cases lead to the lockdown policy in order to relieve the public health unit burden? Does the timeline of the COVID-19 policy (Ranger 2021) match the general trend of the graphs in Figure 3?

Figure 3: Number of COVID-19 Cases Per Month

From Figure 3, on the right side, the graph with red bars is reflect the total number of infected people in the past 2 years, we cannot notice any cases in 2020 January and February because there are only a few cases in these two months thus the number is too small and cannot be shown on the graph. There is a significant increase in March 2020 and a drop in May 2020, starting September 2020 there is a increasing trend with a

drop in February and March 2021 and a peak in April 2021 of around 35,000 cases. The number of cases remain below 5,000 from June to November 2021, however in December 2021 there is a sudden raise and the cases in that month exceed 50,000. The graph on the left in Figure 3 is the number of hospitalized people per month. We can notice that the graph has a very similar pattern with the graph on the right, the peak basically overlapped with each other and it precisely reflected the period of lockdown. The major peaks are April 2020, January 2021, April 2021 and December 2021, all of these dates the government had implemented enhanced distancing policies with April 2021 Toronto was declared as state of emergency in order to control the spread of COVID-19 and relieve the burden of public health unit and health care related occupations.

3 References

Auguie, Baptiste. 2017. *GridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra.

City of Toronto. 2022. "COVID-19: Community Supports." https://www.toronto.ca/community-people/health-wellness-care/covid-19-wellness-during-the-pandemic/covid-19-seniors-vulnerable-people/.

D'Mello, Collin. 2022. "'Frustration Is Real' Ontario Politicians Agree on Eliminating Covid-19 Lockdown." https://toronto.ctvnews.ca/frustration-is-real-ontario-politicians-agree-on-eliminating-covid-19-lockdowns-1.5761605.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

kawarthaNOW. 2020. "COVID-19 Outbreak at Pinecrest Nursing Home in Bobcaygeon the Largest in Ontario." https://kawarthanow.com/2020/03/27/covid-19-outbreak-at-pinecrest-nursing-home-in-bobcaygeon-the-largest-in-ontario/.

Levinson-King, Robin. 2021. "Toronto Lockdown - One of the World's Longest?" https://www.bbc.com/news/world-us-canada-57079577.

Ranger, Michael. 2021. "Timeline: A Year of Pandemic Life in Toronto." https://toronto.citynews.ca/2021/03/11/timeline-a-year-of-pandemic-life/.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Solis-Moreira, Jocelyn. 2021. "COVID-19 Vaccine: How Was It Developed so Fast?" https://www.medicalnewstoday.com/articles/how-did-we-develop-a-covid-19-vaccine-so-quickly.

Toronto Public Health. 2022. "COVID-19 Cases in Toronto." https://open.toronto.ca/dataset/covid-19-cases-in-toronto/.

US. Department of Health & Human Services. 2021. "Centers for Disease Control and Prevention." https://www.cdc.gov/coronavirus/2019-ncov/faq.html.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2021. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Dana Seidel. 2020. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.