Reducing the Risk of COVID-19 Infections in Vulnerable ages*

Mohid Sharif

06 February 2022

Abstract

COVID-19 has taken over the world for more than two years now and it is important to study and understand in which circumstances cases rise and fall. I studied and analyzed data obtained about COVID-19 cases from various communities in Toronto specifying age, gender, and source of infection of the patient. I found that younger age groups are more likely to be infected in their household while the seniors are more likely to be infected in community settings. My findings can be used to identify circumstances in which outbreaks are less likely for certain age groups.

1 Introduction

COVID-19 has taken over the world for more than two years now. Thousands of cases have been recorded in this data set and we can use this to our advantage. Analyzing the data set we can compare variables such as age groups and gender to isolate sources of infection which are more likely to infect certain groups and find changes to those circumstances to reduce the likeliness of infection in those groups. Using this technique we can keep making adjustments to what should and should not be allowed during the pandemic to ensure cases don't rise uncontrollably.

I first analysed the data set to understand how cases had been recorded. Every case in the data set records the source of infection, age group, and gender of the patient. These are only a few of the variables but they are also the most important and they are the ones I mainly looked at in this data study. I then created two graphs to compare variables in the data set. The first graph is a comparison of source of infection between age groups, I used these graphs to visualize which age groups were more likely to be infected in certain settings. I found that younger ages (0-19 years) were more likely to be infected in households while seniors are more likely to be infected in community settings or healthcare institutes. The second graph compares the source of infection between gender groups, I used these graphs to analyze which gender groups were more likely to be infected in certain settings. I found that both males and females had almost identical data for how they were being infected, however there were twice as many females being infected in healthcare institutes.

I can use my data to isolate situations in which infections are more likely for certain groups and how we can change that to reduce infections. Firstly younger ages are more likely to get infected in their homes, in this case they are likely infected by a family member. Family members should in this case sanitize regularly while entering and exiting the house and reduce outdoor activity to necessities. Secondly seniors are more likely to be infected in healthcare institutes because of regular visits for their health. For this healthcare institutes should do their best to isolate the non-infected seniors from the infected reducing their risk for infection. Lastly females are more likely then men to be infected in healthcare institutes due to the fact that women dominate nursing, otherwise both males and females are just as likely to be infected in any setting.

The remainder of this paper is: Section 2.

*Code and data are available at: https://github.com/MohidSharif/COVID19_data_paper

2 Data

I obtained my data from the City of Toronto **opendatatoronto** database portal, using the 'opendatatoronto' package (Gelfand 2020) and the statistical programming language **R** (R Core Team 2020). I used the **tidyverse** package for data manipulation (Wickham et al. 2019) and **kableExtra** for table formatting (Zhu 2021). The header includes two lines of code **"usepackage{float}"** which alows the use of float in our R markdown and the line **"floatplacement{figure}{H}"** which keeps the tables and figures locked in the specific place where they are written in R markdown.

This data set records COVID-19 cases across Toronto and records the details of the infected. The data records the **age group, gender, neighborhood, source of infection, and if the infection was outbreak associated or sporadic**. The data classifies cases by age group rather than specifics, age groups are grouped into 10 year age groups (E.g. "20 to 29 Years") except for the first group which is **"19 and younger"** and the last group which is **"90 and older"**. Gender is self-reported and sorted into 7 categories, from which we will look at **males** and **females**.

Source of infection is determined by three factors:

- A public health investigator's assessment of the most likely source of infection
- Being associated with a confirmed COVID-19 outbreak
- Reported risk factors such as contact with a known case or travel

The classifications are very self explanatory categorized as household contact, close contact, community, outbreak associated (this is then further classified into 3 categories, congregate settings, healthcare institutions, and other settings), and travel.

(Table 1) shows the data recorded from the neighborhood of Agincourt North. Most cases in this data set are classified as "No Information" or "Pending" under source of infection, these cases need to be discluded to see cases which give us more valuable information.

Table 1: Outbreaks in Agincourt North

Source of Infection	Age Group	Gender
Travel	30 to 39 Years	FEMALE
Outbreaks, Other Settings	50 to 59 Years	FEMALE
Outbreaks, Congregate Settings	20 to 29 Years	MALE
Community	50 to 59 Years	MALE
Outbreaks, Other Settings	30 to 39 Years	MALE
Outbreaks, Healthcare Institutions	30 to 39 Years	FEMALE
Close Contact	30 to 39 Years	FEMALE
Outbreaks, Healthcare Institutions	19 and younger	MALE
Outbreaks, Healthcare Institutions	50 to 59 Years	FEMALE
Outbreaks, Healthcare Institutions	20 to 29 Years	FEMALE

(Table 1) line 1 can be read as such; a female between the ages of 30 and 39 from Agincourt North likely to be infected due to travel. The table is a good representation of which data is available to us and how we can use it to obtain which sources of infection are most common by counting all cases that fall under that source of infection. I specifically filtered this table to only include cases from Agincourt North to reduce result size and to specify to a single location in Toronto. I also filtered all sources classified as "No Information" or "Pending" since most cases in this dataset as classified as such and this provides us with no information on the case.

From here I further cleaned the data for graphing. I cleaned the data by removing unnecessary or irrelevant information from columns such as “No Information” or “Pending” from source of infection. I wanted my graph to show how susceptible ages were being infected, I learnt that vulnerable ages range from the ages of 0-19 and 70+ (Davies et al. 2020). I then filtered the ages that I needed to analyze and graphed the count data in a bar graph.

(Figure 1) shows the comparison of cases between vulnerable age groups for each source of infection.

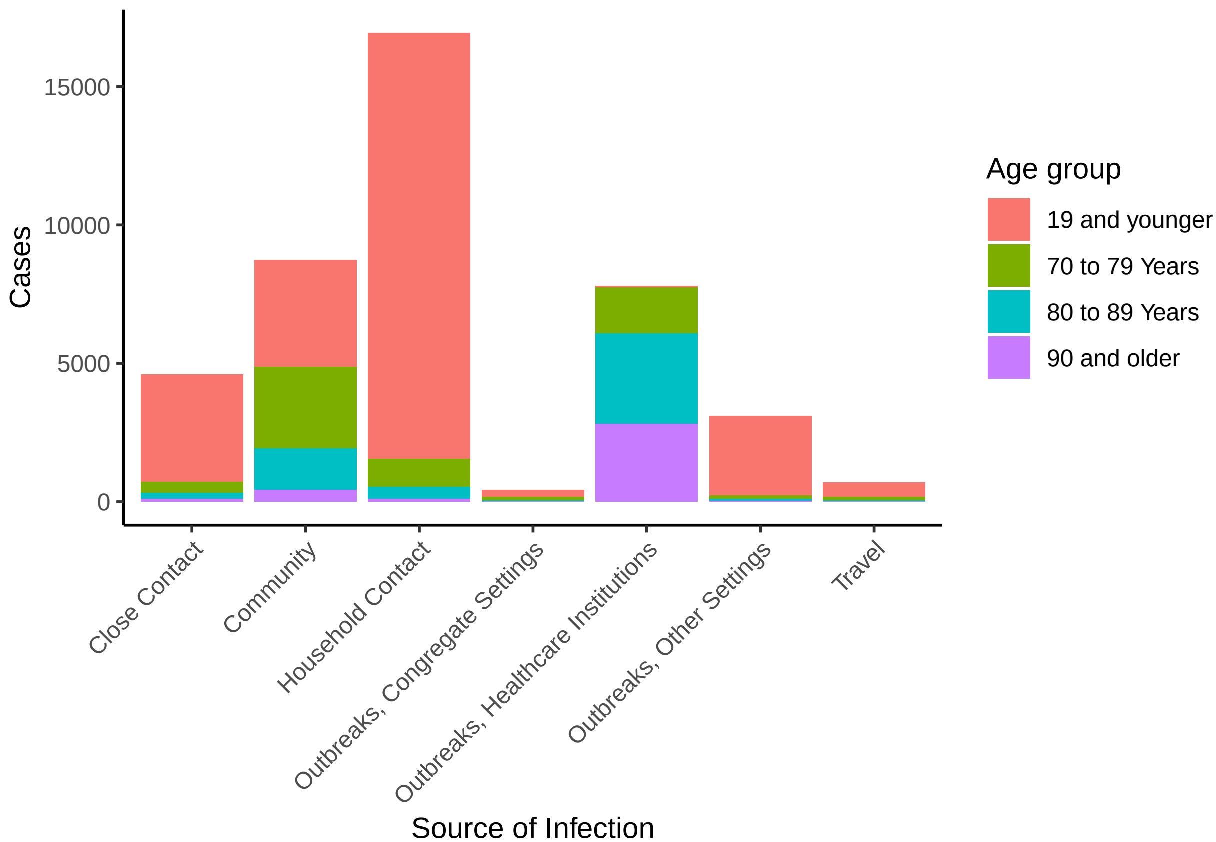

Figure 1: Source of Infections by Age group

The graph shows a great visual comparison of the amount of infections per age group by each source of infection. The biggest observation here is the difference in cases between the seniors and children/youth in the Household Contact category. We can see here that children/youth are much more likely to be infected in their household. Seniors are most likely to be infected in healthcare institutes likely due to their frequent visits to the hospital due to old age. The next most likely source of infection for them is community settings since seniors are often involved in community events and gatherings. Both children/youth and seniors have very low cases in the travel and congregate settings categories.

Our biggest concern from seeing this comparison should be to reduce the risk of infection in household settings for children/youth and the risk of infection in healthcare institutes. We should try our best to sanitize before and after entering the household and to limit our outdoor activities to only necessities. For healthcare institutes we should try to keep the infected away from the other patients in the facility as much as possible. Including sanitation stations in every corridor and room, and sanitizing as often as necessary is a great way to prevent the spread of the virus.

Now lets compare the source of infections between men and women. I used the same filters from (Figure 1) for the source of infections and changed the age filters to gender filters. I then filtered out any gender classifications other than male and female since I was only interested in the two. The I proceeded to graph the count of cases per each source of infection and graph them in a bar graph.

2) shows the comparison of cases between males and females for each source of infection.

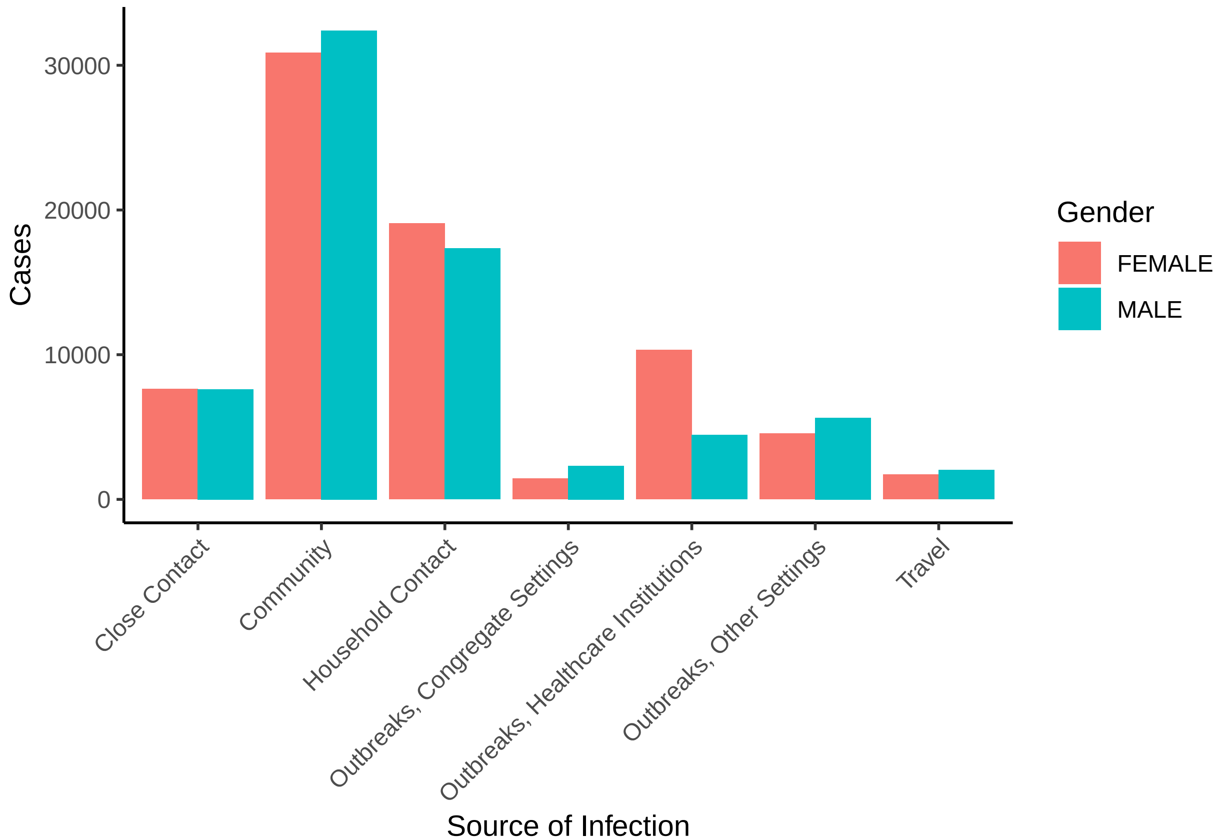

Figure 2: Source of Infections by Gender

Observing the graph we can see that both males and females are very similar in the amount of cases in each source of infection. The most important information we can gather from this graph is that there are twice as many cases for women in healthcare institutes compared to men. This is most likely due to the fact that nursing is female dominated and having to take care of infected patients they have a very high chance of catching the virus. Other than this we can see that both males and females are mainly getting infected in community and household settings. Therefore we should encourage social distancing, regular sanitization and limit outdoor activity to only necessary activities.

References

Davies, Nicholas G., Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, and Rosalind M. Eggo. 2020. "Age-Dependent Effects in the Transmission and Control of COVID-19 Epidemics." *Nature News*. Nature Publishing Group. https://www.nature.com/articles/s41591-020-0962-9#:~:text=We%20estimate%20that%20susceptibility%20to,people%20aged%20over%2070%20years.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*.

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

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.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*.