The Effects of COVID-19 on Older Populations*

Saumya Bakshi

09 February 2022

Abstract

Toronto COVID-19 data was pulled from the City of Toronto Open Portal to analyze the effects of the pandemic on various age groups and find which age groups are most vulnerable to the COVID-19 virus. We noted that although the older populations have a lower number of cases, they have a higher number of deaths, which leads to a higher mortality rate. This information has been used by scientists and policy makers to outline public health guidelines for the city of Toronto.

1 Introduction

The COVID-19 pandemic has impacted every facet of our lives. As we switch our lifestyles and policies to accommodate a changing environment, we can not help but note a skew in the effects of the pandemics on different demographics. Most people that are infected with the virus experience only mild or moderate flu like symptoms. Some might be severely ill and require medical attention; this virus has also proven to be fatal for many. In order to better understand this disproportionate effect of the virus on the general population, we must look at the patient data. It can provide us with an incredible insight into the vulnerable populations and guide public health recommendations for the same.

An older population is more vulnerable to a variety of diseases. In addition to that, since the onset of the pandemic public health guidelines have been employed to advise older populations to be more cautious with the COVID-19 virus. It is thus worth investigating the effects of the COVID-19 virus on various age groups. To localize the study of the effects of this pandemic, we are only investigating the effects of the virus on the city of Toronto. The importance of understanding the impact of the COVID-19 virus on our vulnerable populations in crucial for shaping public healthcare and retirement homes' care facilities and shaping public health guidelines. It is thus important to understand how the virus is impacting us and how the data is being reported. This may also guide our interpretation of the virus and by extension our public health systems.

For this report, I will use the open-access data from the Toronto Public Health Services. I will use this dataset to understand the COVID 19 outbreak in Toronto. More importantly, I will discuss the issues of bias in how the pandemic's data is reported, and the ways in which it might be misleading or misinterpreted. The dataset will be downloaded, processed, and analyzed in `R` (R Core Team 2020) primarily using the `opendatatoronto` (Gelfand 2020), `janitor` (Firke 2021), and `dplyr` (Wickham et al. 2021) packages. Figures and tables will be created with `ggplot2` (Wickham 2016), `kableExtra` (Zhu 2021), `grid` (R Core Team 2021), and `gridExtra` (Auguie 2017). The packages `knitr` (Xie 2014) are used to generate the R markdown report.

2 Data

2.1 Data Source

This report uses the data from Toronto's open data resources updated by the Toronto Public Health Services ("Open Data Dataset," n.d.). This data has been provided by the Provincial Case & Contact Management System (CCM) and contains public demographic and geographic information. It details the severity of all

*Code and data are available at: https://github.com/Saumya510/STA305GIT/tree/main/Paper1

Table 1: COVID-19 Data Subset

id	age_group	classification	outcome
5486561	20 to 29 Years	CONFIRMED	RESOLVED
5486562	19 and younger	CONFIRMED	RESOLVED
5486563	50 to 59 Years	CONFIRMED	RESOLVED
5486564	20 to 29 Years	CONFIRMED	RESOLVED
5486565	30 to 39 Years	CONFIRMED	RESOLVED
5486566	40 to 49 Years	CONFIRMED	RESOLVED

confirmed and probable COVID-19 cases since January 2020. The data is extracted on a weekly basis and is freely available to the general public on Toronto COVID-19 Pandemic Data website or the City of Toronto Open Data Portal. The primary dataset used in this report, was extracted from the Open Data Portal with the opendatatoronto R package (Gelfand 2020) and contains all reported COVID-19 cases in the city of Toronto from January 21, 2020 to February 1, 2022.

2.2 Data Biases

While collecting this patient data, there might be implicit biases that cropped into the data. The neighborhood names are mapped from the patient's residential address and thus might not be an accurate representation of the localisation of the outbreak. Every new record in the system is assigned a unique identifier, which results in duplication of patients that might have contracted COVID 19 twice. During the last surge of the omicron variant of COVID-19, testing in the province was limited to symptomatic and contact traceable people, which results in an under stating of the final results. It is crucial to keep these biases in mind as we analyse the data, primarily because our findings shape the response of the general public to the pandemic.

2.3 Data Structure and Analysis

At the provincial level, a daily report of the COVID patient data is produced by Public Health Ontario. The extracted dataset contains a unique assigned identifier for a patient and includes information such as age group, neighborhood name, gender, reported date, classification and outcome.

In order to better analyse the impact of COVID-19 on various age groups, we limit our imported dataset and remove all rows with any incomplete information. We also use the classification of the case to consider only the confirmed cases of COVID-19. Table 1 is an example of the subsection of the dataset that we are interested in.

To better understand the impact of COVID-19 on different age groups, we look at the total number of confirmed cases, total number of deaths, and the mortality rate in the city of Toronto with respect to various age groups.

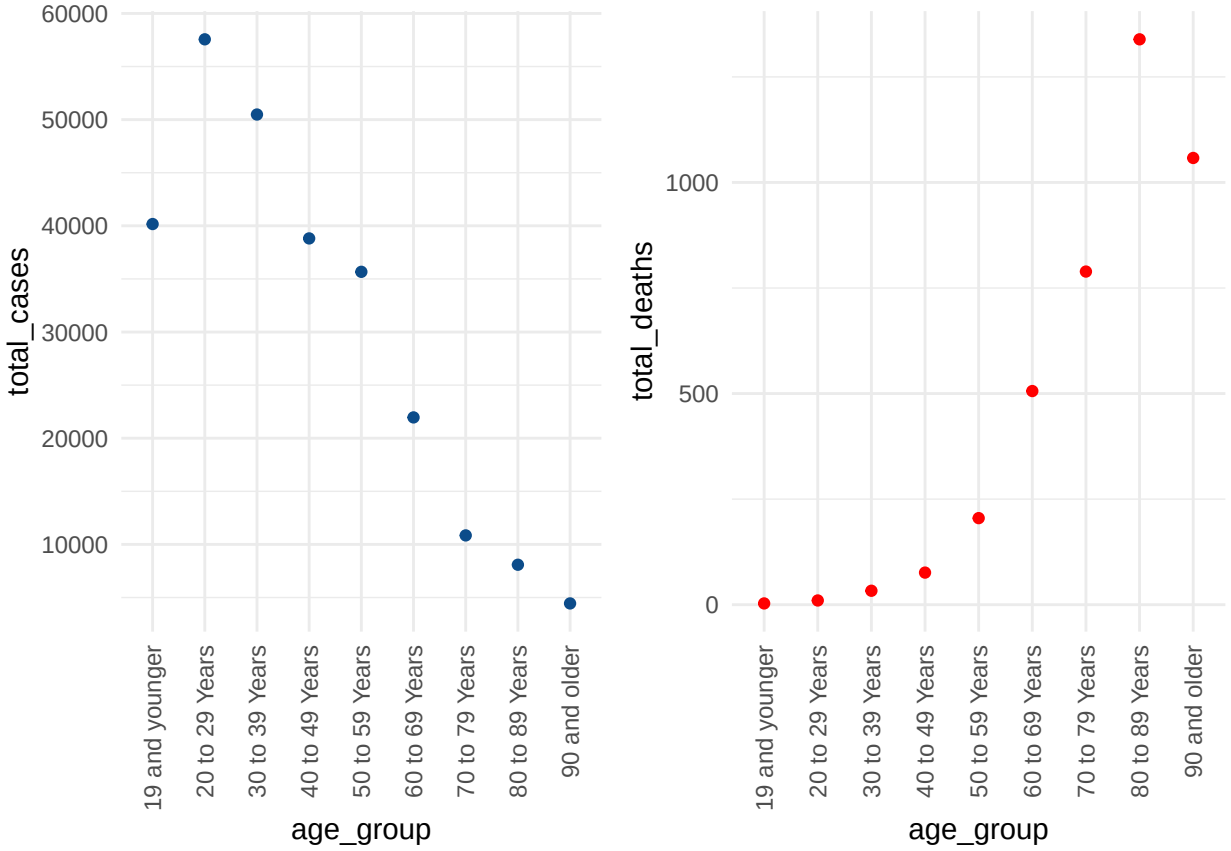
3 Results

Table 2 shows that the highest number of cases was observed in the age groups of 20 - 29 years followed by 30 to 39 years. However, the highest number of deaths was in the age groups of 89 to 90 years followed by 90 and older. The mortality rate (calculated by total_deaths/total_cases) was the highest for the age groups of 90 and older followed by 80 to 89 years. It can also be noted that the mortality rate is monotonically increasing, which points at a positive correlation between age and mortality rate.

The plot on the left shows the number of total cases for each age group. We note that the infection is highest for younger age groups (20 - 29 years and 30 to 39 years) and lowest for older age groups (80 to 89 years and 90 and older). In contrast, the plot on the right details that the number of deaths is the highest for older age groups (80 to 89 years and 90 and older) and lowest for younger age groups (20 - 29 years and 30 to 39 years).

Table 2: COVID 19 Cases by Age Group

age_group	total_cases	total_deaths	mortality_rate
19 and younger	40167	3	0.0000747
20 to 29 Years	57562	10	0.0001737
30 to 39 Years	50482	33	0.0006537
40 to 49 Years	38815	76	0.0019580
50 to 59 Years	35667	205	0.0057476
60 to 69 Years	21959	506	0.0230429
70 to 79 Years	10849	789	0.0727256
80 to 89 Years	8085	1339	0.1656153
90 and older	4443	1058	0.2381274



4 Discussion

In this paper, we study the effect of COVID-19 on various age groups in the City of Toronto. We noted a monotonic increase in the mortality rate with respect to age group. This has guided our public health measures in the city of Toronto. Our analysis corroborates the guidelines for older populations and retirement facilities.

We also noted an increased number of COVID-19 infections in the younger populations. This could be attributed to the work conducted by them as front line workers. However, this would require an enhanced socio-economic analysis of the people that were infected. It would also be worthwhile to study the variation in the impact of COVID-19 in its three different waves.

References

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

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.

"Open Data Dataset." n.d. *City of Toronto Open Data Portal*. https://open.toronto.ca/dataset/covid-19-cases-in-toronto/.

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

———. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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

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

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.