COVID-19 Cases in Toronto*

Fatcors of mortality rate

Ruibo Sun

5 Feb 2023

This paper examines the mortality rate difference between male and female COVID-19 patients in Canada from 2020 to 2023, taking into account the impact of new virus variants. Using data from Ontario's national Case & Contact Management System, the study will present findings using data visualization tools and update previous findings that showed higher mortality rates in male patients. The study will also examine other variables that may impact mortality rate and address the knowledge gap between current COVID-19 variants and previous studies. The authors will explain the limitations of the data, including the possibility of missing cases due to a lack of testing or symptoms.

Introduction

The ongoing COVID-19 pandemic has profoundly impacted the entire world, and Toronto Public Health is no exception. The virus was first declared a Public Health Emergency of International Concern by the World Health Organization on January 30, 2020, after 171 reported fatalities. However, by the end of 2020, the number of reported fatalities had risen to 1,813,188. According to preliminary estimates, at least 3 million deaths worldwide in 2020 can be attributed to COVID-19, a significantly higher number than the official estimate. The most recent COVID-19 deaths reported to WHO now exceed 3.3 million globally. (Organization (2022)) In response to the COVID-19 pandemic, Toronto Public Health has compiled and analysed data on confirmed and probable COVID-19 cases. This data set includes important demographic, geographic, and severity information for all cases reported and managed by Toronto Public Health since the first case was reported in January 2020. The data set includes both sporadic cases in the community and outbreak-associated cases. However, it is important to note that this data is subject to change and continuous improvement. (Health Data Research Network (2022))The current research is mostly published in 2020 or 2021, and

*Code and data are available at: https://github.com/ruibosun/paper_1/tree/main

these results might not be representative of the current situation. Hence, this paper aims to examine the mortality rate difference between male and female patients in Canada from 2020 to 2023, taking into account the impact of the new variants of the virus on the outcome. From the first outbreak till now, seven variants of the virus have been identified by WHO, including Epsilon (B.1.427 and B.1.429), Zeta (P.2), Eta (B.1.525) and Iota (B.1.526), Theta (P.3), Kappa (B.1.617.1), and Lambda (C.37). (Siddiqui, Alhamdi, and Alghamdi (2022)) The study will update the previous findings, which showed that male patients tend to have a higher mortality rate than female patients, based on data from July 2020. The research aims to determine if the new variants of the virus have affected the mortality rate for different groups and how it has impacted male and female patients differently. (Hoffmann and Wolf (2020)) Some study also suggests that mortality rates across the board were significantly higher for the immediately younger age group. Patients between the ages of 60 and 69 had the highest mortality risk compared to those between the ages of 50 and 59. Furthermore, Patients over the age of 80 had the highest mortality rate. (Bonanad et al. (2020)) The study will also examine other potential variables and their characteristics that may contribute to reducing the chance of getting COVID-19 and possibly reducing the mortality rate. This study aims to address the knowledge gap between the current COVID-19 variant and previous studies. To do this, the authors will present findings from Ontario's national Case & Contact Management System using a variety of data visualization tools, including plots and tables (CCM). The data will be presented clearly and succinctly with these visuals, making it simpler to comprehend the COVID-19 pandemic's current state of affairs and emerging trends in Ontario. The bias and limitation of the data will also be explained in the Section of Limitation. This following analysis is processed in R (R Core Team (2022)) with packages of tidyverse(Wickham et al. (2019)), dplyr(Wickham et al. (2022)). The tables is constructed via gt(Iannone et al. (2022)) and scales(Hadley Wickham (2022)). The package inside of tidyverse helps to create the plots via ggplot2(Wickham et al. (2019)). This paper is knitted as PDF file by the packages of R markdown. ("R Markdown - Dynamic Documents for r" (n.d.))

Data

Collection

The raw data is extracted from Ontario's provincial Case & Contact Management System(CCM). The CCM system serves as a data repository. Public health departments are required by law to gather COVID-19 data and submit it to the Ministry of Health for regional, national, and local surveillance. This data in the spreadsheet may be subject to change and updates due to ongoing public health investigations and quality control efforts. The data will be refreshed and replaced every week and will be extracted and posted every week. It is possible that the numbers may differ from those presented in other sources because the data is being extracted at different times and from various sources. Additionally, as some

COVID-19 cases go unreported due to a lack of symptoms or testing, the numbers in CCM might not accurately represent all cases, which will be discussed in Section of Limitation.

Cleaning:

The raw data was very organized and all values were correctly stored in each cell. However, in the Age Group variable, some cells contain an empty quotation (i.e., " "). Since the sample size is large and these rows are only a tiny proportion, these missing values were removed directly. Moreover, the raw data from Open Toronto Portal contains 32,000 observations and 15 variables until February 1st, 2023. These variables are _id, Assigned_ID, Outbreak Associated, Age Group, Neighbourhood Name, Source of Infection, Classification, Episode Date, Reported Date, Client Gender, Ever Hospitalized, Ever in ICU and Ever Intubated. For simplicity purposes, only four variables are begging selected: (1) Age Group, (2) Client Gender, (3) Source of Infection and (4) Outcome. The variable of These is different categories used to classify the source of infection of COVID-19 cases.

Analysis

Based on the demonstration of Figure 1, it is clear that most patients are from the age range of 20 to 29, followed by the age group of 30 to 39 years old. However, the third largest group jumps to the group of 40 to 49 years. In contrast to the senior group, specifically patients who are older than 60 years, their bars are much lower than the young group. Generally, the younger patient takes a large portion of the data and this portion decrease as age increase. This indicates that most of the people who were infected by COVID-19 and recorded in this data are relatively young. Nevertheless, this data might be deeply influenced by survival bias. According to the current study and analysis, the senior group has the highest mortality rate in this pandemic crisis. (Bonanad et al. (2020)) It is possible that some senior patients were infected but deceased before getting to the hospital. In that case, they, as part of the dataset, are not recorded in the current data. Hence, this explains why the proportion of senior patients is significantly lower.

Based on the demonstration of Figure 1, it is clear that most patients are from the age range of 20 to 29, followed by the age group of 30 to 39. However, the third largest group jumps to the group of 40 to 49 years. In contrast to the senior group, specifically patients who are older than 60 years, their bars are much lower than the young group. Generally, the younger patient takes a large portion of the data and this portion decrease as age increase. This indicates that most people infected by COVID-19 and recorded in this data are relatively young. Nevertheless, this data might be deeply influenced by survival bias. According to the current study and analysis, the senior group has the highest mortality rate in this pandemic crisis. (Bonanad et al. (2020)) It is possible that some senior patients were infected but deceased before getting to the hospital. In that case, they, as part of the dataset, are not recorded in the current data. Hence, this explains why the proportion of senior patients is

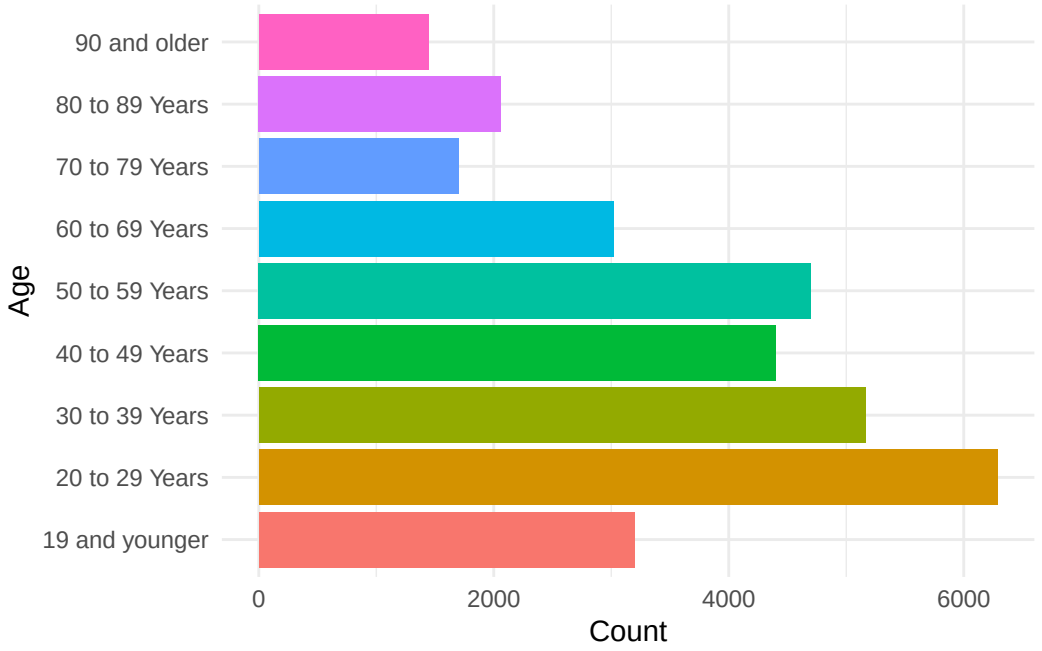

Figure 1: Age Group

significantly lower.

The data also collected the source of infection for the patients on the record. This variable includes several values.

- "Household contact" refers to cases acquired from a confirmed or probable case within a household setting.
- "Close contact with a case" refers to cases acquired from close contact with a confirmed or probable case.
- "Outbreaks" refers to cases linked to known confirmed outbreaks in different settings like congregate settings, healthcare institutions, and other settings.
- "Travel" refers to cases that traveled outside of Ontario prior to their symptoms or test date.
- "Community" refers to cases that did not travel, were not a close contact, and were not part of a known outbreak.

- "No information" refers to cases with no information on the source of infection.

The left plot Figure 2 visualizes the bar plots of the source of infection, the age range and the outcome of COVID-19 from the patients. Patients who are unfortunately deceased due to COVID-19 are primarily beyond 70 to 79 years. Patients who are 80 to 89 years have twice as much as death compared to the age group of 70 to 79. However, it is noticeable that the group of 90 and older has a lower death than those aged 80 to 89 years. Furthermore, the majority of patients in these age groups are infected by the outbreaks at Healthcare institutions, followed by the source of community.

The right plot is for patients who recovered from COVID-19. As the plot shows, most of the patients who recovered are from the age group of 20 to 29, they are relatively young among all the patients, so they have a high recovery rate and low mortality rate. In addition, those patients are infected from different sources, and the infection source is relatively evenly distributed among other age groups. In contrast the fatal group, senior patients have a lower recovery rate and higher mortality rate than the younger groups.

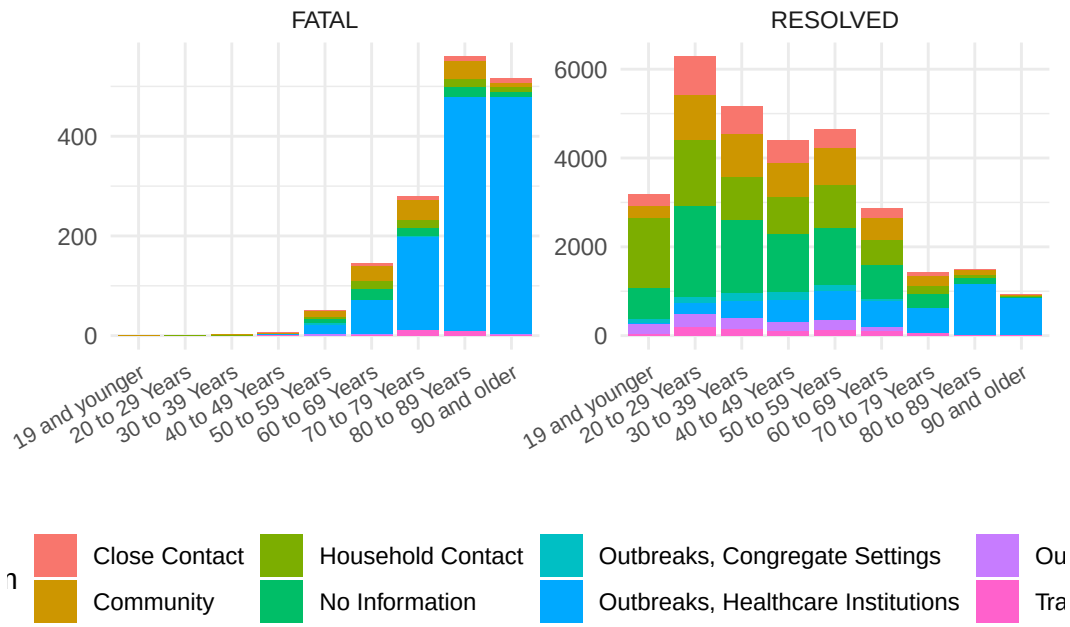

Figure 2: Gender Group

The table Figure 3 shows the summaries of the different age groups and their mortality rate to verify the result via numeric values. There were 805 female fatalities (2.52%) and 15,790 resolved cases (49.40%) among female patients. Similarly, there were 735 male fatalities (2.30%) and 14,486 resolved cases (45.32%) among male patients. For non-binary and other gender

categories, there were only 1 and 7 resolved cases, respectively, with a low mortality rate. There were 23 fatalities (0.07%) and 112 resolved cases (0.35%) among patients of unknown gender. Interestingly, the existing research suggests that male patients have a higher mortality rate than female patients. (Hoffmann and Wolf (2020)) The mortality rate measures the number of deaths in a population or a specific group. In this case, the data shows that 2.52% of female patients had fatal outcomes, while 2.30% of male patients had fatal outcomes. The higher mortality rate for female patients suggests that, in this study, female patients had a higher risk of death than male patients. However, it's important to consider other factors contributing to this difference, such as age, underlying medical conditions, access to healthcare, and treatment.

Gender	Outcome	Number	Mortality Rate
FEMALE	FATAL	805	2.5184%
FEMALE	RESOLVED	15790	49.3978%
MALE	FATAL	735	2.2994%
MALE	RESOLVED	14486	45.3183%
NON-BINARY	RESOLVED	1	0.0031%
OTHER	RESOLVED	7	0.0219%
TRANSGENDER	RESOLVED	6	0.0188%
UNKNOWN	FATAL	23	0.0720%
UNKNOWN	RESOLVED	112	0.3504%

Figure 3: Mortality Rate

The recovery rate, on the other hand, measures the number of patients who recover from a disease or condition. In this case, the data shows that 49.40% of female patients had resolved outcomes, while 45.32% of male patients had resolved outcomes. The higher recovery rate for female patients suggests that, in this study, female patients had a better chance of recovery than male patients.

Limitation

Although the data is collected by Ontario's provincial Case & Contact Management System, there are some issues. First, the reported number of COVID-19 cases in CCM likely falls short of the true extent of the disease as individuals who are asymptomatic or have mild symptoms may not seek testing or medical attention, leading to their cases going unrecorded. Factors like limited testing availability, access to healthcare, and concerns about stigma or consequences can also play a role in under reporting. Second, this data is collected on a weekly basis; new data will be uploaded every Tuesday. Hence, this current paper relies on the data before February 1st, 2023, which might need to be more representative. Third, this is related to the

first bias, yet there is another survival bias in this data as well. Survival bias is an error in statistical analysis where only successful outcomes or data points are considered and failures are ignored. This can lead to a biased view and incorrect conclusions, especially in situations where the success rate is being studied, such as evaluating a new medical treatment. This can result in overestimating the results.(Worklife (2020)) In this case, senior patients are the group with the highest mortality rate, it is likely that some of the older patients pass away before their data can be collected by Ontario's provincial Case & Contact Management System. Additionally, patients from other age groups may also suffer from this bias.

Conclusion

In light of the COVID-19 pandemic, the current study evaluated the mortality rate disparity between male and female patients in Canada from 2020 to 2023. The majority of COVID-19 cases in Canada are reported in the age groups of 20 to 29, 30 to 39, and 40 to 49, according to data from Ontario's Case & Contact Management System. The study also discovered that the mortality rate was impacted by the virus' new variants differently in male and female patients. The study also looked at how other potential factors, like the infection source, might affect how COVID-19 cases turned out. Due to ongoing public health investigations and quality control efforts, it is crucial to be aware that the data may be subject to updates and changes. Moreover, some COVID-19 cases might not be reported, which could affect the accuracy of the study's numerical results. The study highlights the need for ongoing monitoring and analysis and offers insightful information about the situation as it stands and new trends in the COVID-19 pandemic in Canada. Further research can investigate a more reliable data source and avoid having biases to conduct accurate results.

References

Bonanad, Clara, Sergio García-Blas, Francisco Tarazona-Santabalbina, Juan Sanchis, Vicente Bertomeu-González, Lorenzo Fácila, Albert Ariza, Julio Núñez, and Alberto Cordero. 2020. "The Effect of Age on Mortality in Patients with COVID-19: A Meta-Analysis with 611,583 Subjects." *Journal of the American Medical Directors Association*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7247470/.

Hadley Wickham, Dana Seidel. 2022. "How Survivorship Bias Can Cause You to Make Mistakes." https://cran.r-project.org/web/packages/scales/index.html.

Health Data Research Network, Canada. 2022. "Inventory/4763." *HDRN Canada - Inventory/4763*. https://www.hdrn.ca/en/inventory/4763/.

Hoffmann, Christian, and Eva Wolf. 2020. "Older Age Groups and Country-Specific Case Fatality Rates of COVID-19 in Europe, USA, and Canada - Infection." *SpringerLink*. https://link.springer.com/article/10.1007/s15010-020-01538-w#Sec1.

Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, and JooYoung Seo. 2022. *Gt: Easily Create Presentation-Ready Display Tables*. https://CRAN.R-project.org/package=gt.

Organization, World Health. 2022. "COVID-19 Dashboard." *World Health Organization*. https://covid19.who.int/table.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

"R Markdown - Dynamic Documents for r." n.d. https://rmarkdown.rstudio.com.

Siddiqui, Sazada, Heba Waheeb Saeed Alhamdi, and Huda Ahmed Alghamdi. 2022. "Recent Chronology of COVID-19 Pandemic." *Frontiers*. https://www.frontiersin.org/articles/10.3389/fpubh.2022.778037/full.

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. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Worklife, BBC. 2020. "How 'Survivorship Bias' Can Cause You to Make Mistakes." *BBC Worklife*. https://www.bbc.com/worklife/article/20200827-how-survivorship-bias-can-cause-you-to-make-mistakes.