Exploring the Demographic Aspects of COVID-19 Cases in Toronto: Are the elderly males more susceptible to COVID-19?*

Amie Liu

January 23, 2024

Examining the age and gender patterns of COVID-19 cases in Toronto provides insights into the susceptibility of different demographic groups to the virus. This paper utilizes the dataset from Open Data Toronto, encompassing the period from January 22, 2020, to January 4, 2024. Contrary to the hypothesis that elderly males are more susceptible, the results reveal a pattern, with females aged 20-29 showing the highest susceptibility, exceeding 40,000 cases. This highlights the importance of understanding gender and age-specific patterns for effective public health interventions.

1 Introduction

Global disease pandemics pose significant threats to public health and impact daily life worldwide. In December 2019, the outbreak of Coronavirus Disease 2019 (COVID-19) was identified and is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Dousari, Moghadam, and Satarzadeh 2020). This disease commonly manifests with respiratory symptoms similar to those of pneumonia, colds, or influenza, and some COVID-19 patients may experience severe illness (CDC 2023). Researchers have found that COVID-19 cases tend to be more severe in the elderly, with males experiencing a higher severity rate than females, highlighting the influence of gender and age differences on the severity of COVID-19 (Jin et al. 2020). However, limited information is available regarding these gender and age differences that may influence susceptibility to COVID-19 (Shukla et al. 2022). This leads to a question: "Are the elderly males more susceptible to COVID-19?"

Researchers have collaborated to provide a better understanding of infectious agents and develop techniques to mitigate outbreaks. Public health and social measures (PHSMs) have

*Code and data are available at:https://github.com/AmieLiu/COVID-19_Toronto_Demographic_Aspects_Analysis.git been a significant technique in decreasing the transmission of viruses with pandemic potential (WHO 2021). Understanding the demographic aspects of infection is crucial for exploring how different groups are affected and facilitating the formulation of effective Public Health and Social Measures (PHSMs). Moreover, this understanding can help with the resource allocation for vaccines and predict the demand for medical care across diverse demographic groups.

Given the observation by public health officials that Toronto has a high level of COVID-19 activity (Freeman 2024), I conducted an analysis of the relationship between COVID-19 cases and age groups, as well as genders, in Toronto. My results indicate a pattern of susceptibility in the distribution of COVID-19 cases in Toronto, where the majority of age groups show a higher number of female cases compared to males. Specifically, the group of females aged 20-29 has the highest number of COVID-19 cases—more than 40,000 cases. Contrary to the hypothesis, my data does not show that elderly males are more susceptible to COVID-19. The observed pattern may be influenced by factors beyond age and gender demographics.

This paper is divided into the Data and Limitations sections. Section 2 Data delves into the various sources used for data collection, cleaning, and analyzing processes, leveraging a comprehensive dataset provided by Toronto Public Health on the City of Toronto's Open Data Portal (Health 2024). In addition, it provides a detailed analysis of the findings from the data. Moving forward, Section 3 Limitations reveal potential constraints, including considerations related to occupational exposure, geographical focus, variable measurement, and data accuracy, that might influence the interpretation of the main findings.

2 Data

The data for this paper was obtained from the City of Toronto's Open Data Portal through the 'opendatatoronto' library (Gelfand 2022). The details of the data cleaning process for the dataset 'COVID-19 Cases in Toronto' (Health 2024) are provided in Appendix 4. The data analysis involved using the R programming language (R Core Team 2022) along with the 'tidyverse' (Wickham et al. 2019) 'here' (Müller 2020), 'janitor' (Firke 2023), 'knitr' (Xie 2014) packages for data collection, cleaning, and analysis. Within the 'tidyverse' package (Wickham et al. 2019), I used 'readr' (Wickham, Hester, and Bryan 2024), 'dplyr' (Wickham et al. 2023), 'tidyr' (Wickham, Vaughan, and Girlich 2023), and 'ggplot2' (Wickham 2016) for reading CSV files, manipulating data, tidying data, and creating bar charts.

2.1 Covid-19 Cases by Gender

Based on the bar chart of COVID-19 cases in Toronto by gender (see Figure 1), the data reveals that females constitute the largest group, with 219,671 cases, comprising 53.4% of the total cases. Following closely, males represent the second-highest group, with 188,197 cases, accounting for 45.7% of the total cases. Additionally, 3,832 cases fall into the "Other" category, including transgender and not listed gender, making up 0.9% of the total cases.

The gender breakdown of COVID-19 cases highlights a distribution pattern, underscoring the importance of integrating gender-specific considerations into public health interventions. There are no significant differences between female and male cases. However, the higher occurrence of cases among females suggests a need for targeted measurements that consider gender-specific factors. These findings may prompt health officials to communicate and implement outreach strategies, ensuring the effectiveness of preventive measures and vaccination campaigns across diverse gender groups.

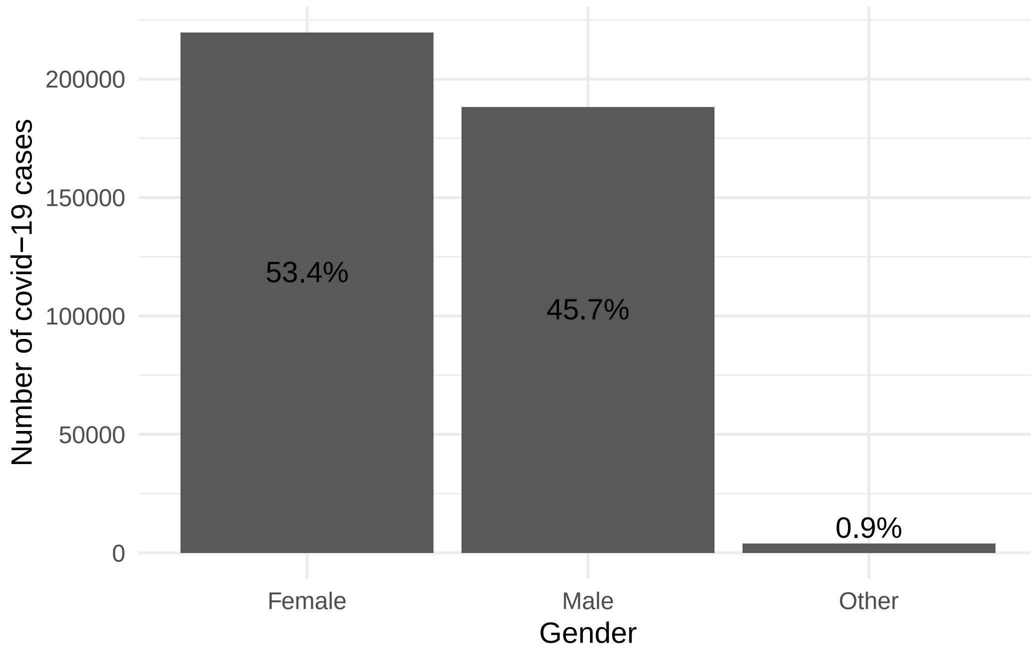

Figure 1: Number of covid-19 cases by gender

2.2 Covid-19 Cases by Age Groups

In Figure 2, the bar chart provides a clear depiction of the distribution of COVID-19 cases across various age groups. Particularly, the age group 20-29 has the highest number of cases, totaling 77,523 instances and representing 18.83% of the overall cases. This is closely followed by the age groups 30-39 and 0-19, with 73,400 (17.83%) and 55,142 (13.39%) cases, respectively. In the middle-age categories, 40-49 and 50-59 years old, there are 13.86% and 12.81% of the total cases, respectively. A gradual decline in COVID-19 cases is observed as the age increases, with the age group 60-69 at 8.62%, 70-79 at 5.73%, 80-89 at 5.51%, and the 90+ category at 3.40%.

The concentration of cases in the age group 20-29 highlights the need for targeted outreach strategies for these younger people. Allocating resources strategically in areas with higher numbers of younger people may be an effective mitigation strategy. Furthermore, the comparatively lower number of cases in older age groups, particularly those aged 60 and above, suggests that these groups may have adopted and implemented effective preventive measures.

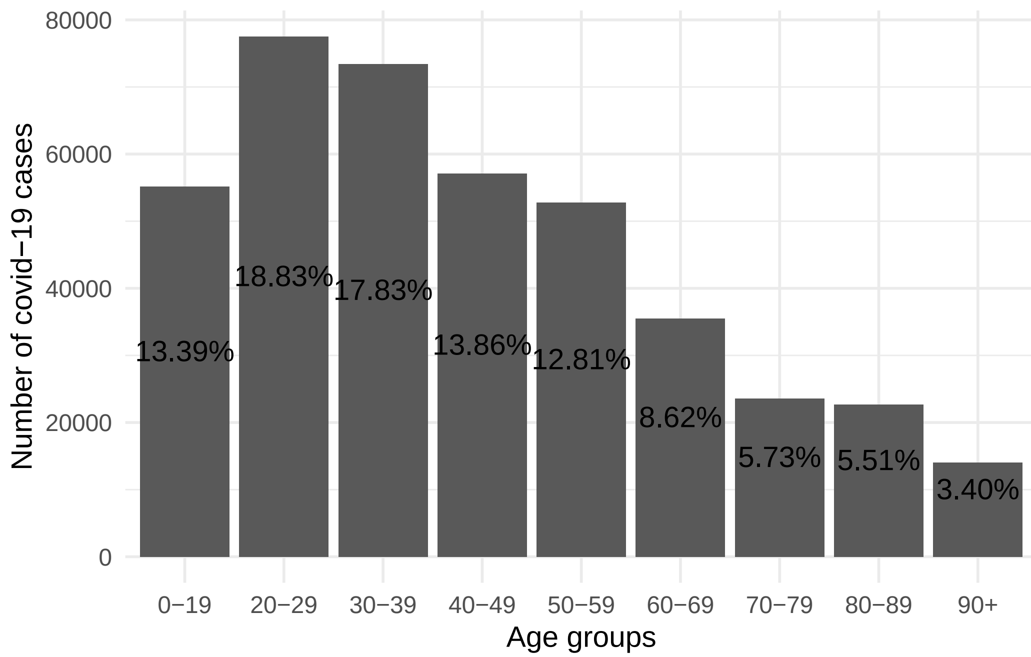

Figure 2: Number of covid-19 cases by age groups

2.3 Distribution of Covid-19 Cases Across Genders and Age Groups

As depicted in Figure 3, the majority of age groups show a higher number of female cases compared to male cases. However, exceptions to this trend are observed in the 0-19 and 70-79 age groups, where male cases exceed females. Moreover, the "Other" category consistently represents the lowest number of cases among all age groups.

The data reveals that the highest incidence of COVID-19 cases is among females aged 20-29, with over 40,000 cases. The second-highest occurs in females aged 30-39, while the third-highest is observed among males in the 20-29 age group. These findings emphasize the significance of understanding age and gender-specific patterns. This paper seeks to address the question of whether elderly males are more susceptible to COVID-19; however, the data does not support this hypothesis, indicating that patterns of susceptibility may include influences beyond age and gender demographics.

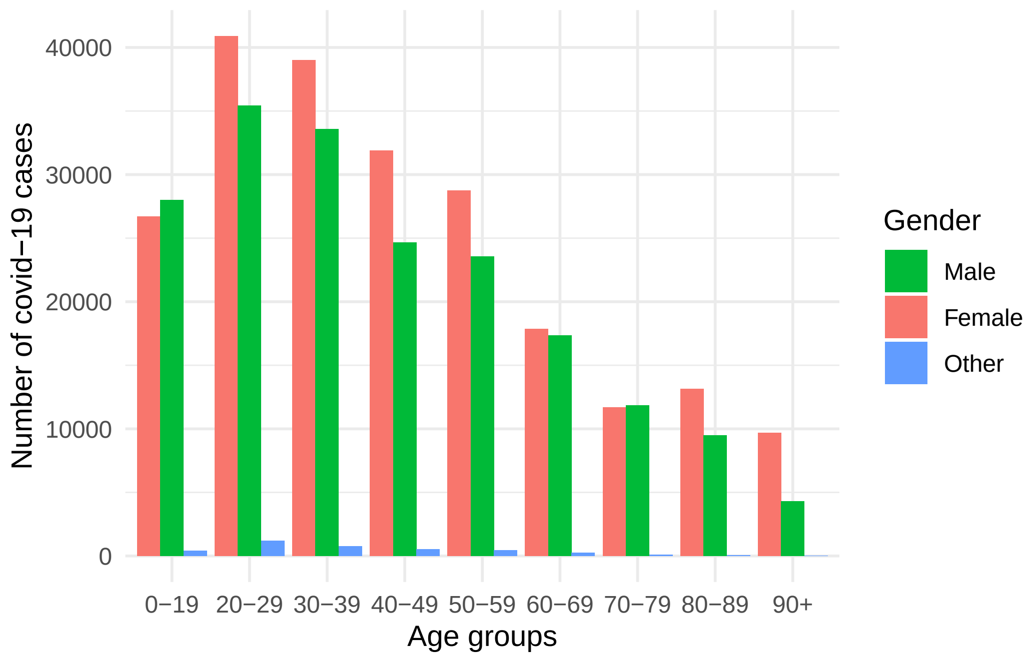

Figure 3: Distribution of covid-19 cases across genders and age groups

3 Limitations

While this paper provides valuable insights into the age and gender patterns of COVID-19 cases in Toronto, several limitations should be taken into consideration, including occupational exposure, geographical focus, variable measurement, and data accuracy.

3.1 Occupational Exposure

Occupational exposure is a key factor that may influence the demographic distribution of COVID-19 cases. Jobs involving close interaction with the public, such as those in healthcare and social assistance, could pose elevated risks of exposure. It is important to note that within these sectors, there is a higher prevalence of female workers compared to male workers and those aged 25-54 compared to other age groups (Bank 2023).

3.2 Geographical focus

The focus on Toronto in this paper may limit the generalizability of research findings to other areas or populations, given that diverse areas may exhibit distinct patterns. Therefore, the results presented here may not be universally applicable. The scope of this paper is specifically targeted at Toronto, and other areas are beyond the intended scope of this paper.

3.3 Variable Measurement

The comparison of COVID-19 cases by gender and age group in Toronto raises considerations about the representativeness of the observed patterns with the population. The resident population data indicates variations in the proportions of males and females, with females slightly outnumbering males (Department 2024), and the age distribution, particularly the largest population (20-29 age group) (Department 2023), should be noticed. This difference in the underlying population composition may have implications for the interpretation of COVID-19 cases. Although the population data aligns with the observed COVID-19 case data, it is important to know that the general population's proportions of gender and age groups should be taken into account. This limitation highlights the importance of considering demographic factors, such as comparing the resident population to the COVID-19 case data, to ensure a more comprehensive pattern of my findings.

3.4 Data Accuracy

The data used for this analysis may not fully capture all COVID-19 cases within the population of Toronto. The City of Toronto highlights a significant change in testing guidance by the Ministry of Health starting in 2021 (Health 2021), leading to a limitation in the available data, as only specific subsets of the population are being tested. Certain data may no longer accurately represent the broader population (Health 2021). Moreover, certain people infected with COVID-19 may not report to the Ministry of Health, and those who show no symptoms could be overlooked in the dataset. This limitation leads to the challenge of accurately assessing the data on COVID-19 cases.

Future research could address these limitations by exploring other demographic variables and considering a wider geographical scope. This would provide a more comprehensive understanding of the factors that influence the patterns of COVID-19 cases in diverse populations.

Appendix

4 Data Cleaning

The dataset 'COVID-19 Cases in Toronto' (Health 2024) was utilized and obtained from 'opendatatoronto' library (Gelfand 2022). It comprises 412,122 records and 15 variables, including assigned ID, age group, neighborhood name, episode date, client gender, and more within the COVID-19 cases in Toronto. The data cover the period from January 22, 2020, to January 4, 2024, and are structured in CSV file format.

I conducted a data-cleaning procedure, focusing on variable selection, simplifying variable names, and handling missing values. The primary variables selected for analysis were chosen based on their relevance to my research question, relating to age and gender. If one categorical value was missing, then the entire row was removed Therefore, the cleaned COVID-19 cases dataset now includes three variables: case ID, gender, and age group (see Table 1). The total number of rows has been reduced from 412,122 to 411,700.

Table 1: First six rows of COVID-19 cases by age groups and gender

Case ID	Age group	Gender
1	50-59	Female
2	50-59	Male
3	20-29	Female
4	60-69	Female
5	60-69	Male
6	50-59	Male

References

Bank, Job. 2023. “Health Care and Social Assistance: Ontario 2022-2024.” https://www.jobbank.gc.ca/trend-analysis/job-market-reports/ontario/sectoral-profile-health-care.

CDC. 2023. “About COVID-19.” https://www.cdc.gov/coronavirus/2019-ncov/your-health/about-covid-19.html.

Department, Statista Research. 2023. “Population of Toronto, Ontario in Canada in 2022, by Age.” https://www.statista.com/statistics/1317461/population-toronto-canada-age/.

———. 2024. “Resident Population in Metropolitan Areas in Canada as of July 2022, by Gender.” https://www.statista.com/statistics/446726/canada-resident-population-by-metropolitan-area-and-gender/.

Dousari, Amin Sadeghi, Majid Taati Moghadam, and Naghmeh Satarzadeh. 2020. “COVID-19 (Coronavirus Disease 2019): A New Coronavirus Disease.” *Infection and Drug Resistance* 13: 2819–28. https://doi.org/10.2147/IDR.S259279.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Freeman, Joshua. 2024. “Influenza Activity in Toronto Higher Than Pre-Pandemic Norm as Some Hospitals Warn That They’re over Capacity.” https://www.cp24.com/news/influenza-activity-in-toronto-higher-than-pre-pandemic-norm-as-some-hospitals-warn-that-they-re-over-capacity-1.6726562.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Health, Toronto Public. 2021. “COVID-19: Archived Dashboards.” https://www.toronto.ca/community-people/health-wellness-care/health-programs-advice/respiratory-viruses/covid-19/covid-19-pandemic-data/covid-19-archived-dashboards/.

———. 2024. “COVID-19 Cases in Toronto.” https://open.toronto.ca/dataset/covid-19-cases-in-toronto/.

Jin, Jian-Min, Peng Bai, Wei He, Fei Wu, Xiao-Fang Liu, De-Min Han, Shi Liu, and Jin-Kui Yang. 2020. “Gender Differences in Patients with COVID-19: Focus on Severity and Mortality.” *Frontiers in Public Health* 8 (152). https://doi.org/10.3389/fpubh.2020.00152.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

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

Shukla, Durgesh, Sumit Singh Bhadoria, Manoj Bansal, and Richa Changulani. 2022. “Evolution of the Pandemic: Analysis of Demographic Characteristics of COVID-19-Infected Patients During Its Two Waves in Gwalior District of Central India. Journal of Family Medicine and Primary Care.” *Journal of Family Medicine and Primary Care* 11 (4): 1314–21. https://doi.org/10.4103/jfmpc.jfmpc_1189_21.

WHO. 2021. “WHO Global Technical Consultation on Public Health and Social Measures During Health Emergencies.” https://www.who.int/news/item/05-09-2021-who-global-technical-consultation-on-public-health-and-social-measures-during-health-emergencies.

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

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2023. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

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