

Age-Driven Disparities in COVID-19 Fatality Rates – A Toronto Perspective*

Kaavya Kalani

January 25, 2024

This paper presents an analysis of COVID-19 cases in Toronto, emphasizing the disproportionate impact on different age groups. Leveraging data from the City of Toronto's Open Data Library, my analysis sheds light on the severity of outcomes, with a particular focus on the vulnerability of older individuals. This analysis includes data preprocessing details, variable descriptions, and visualizations. Findings underscore the need for targeted interventions for the elderly population in pandemic response strategies.

Table of contents

1	Introduction	2
2	Data	2
2.1	About the dataset	3
2.2	Preprocessing and data cleaning	3
2.3	Variables of interest	3
3	Results	4
4	Conclusion	8
5	Appendix	9
5.1	Glimpse of the dataset	9
5.2	Data cleaning	9
	References	10

*Code and data are available at: https://github.com/kaavyakalani26/covid19_toronto.git

1 Introduction

The COVID-19 or the coronavirus pandemic is a global pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Wikipedia 2024). The novel virus was first identified in Wuhan, China in December 2019, and spread worldwide in early 2020. COVID-19 was declared a pandemic by the World Health Organization on March 11, 2020 (WHO 2020). As of the present date, the pandemic has resulted in over 700 million reported cases and approximately 7 million deaths worldwide (worldometers 2024).

This unprecedented health crisis has affected individuals across diverse demographics, transcending geographical boundaries and health conditions. Notably, studies have consistently suggested that older individuals face a heightened susceptibility and increased risk of severe outcomes compared to their younger counterparts (Disease Control and Prevention 2023).

In this paper, I try to uncover these trends. I use the data of all the COVID-19 cases reported to Toronto Public Health since the first case which was reported in January 2020 until December 2022. My focus is on investigating age disparities to understand if older individuals exhibit a higher susceptibility to contracting the virus. Additionally, I analyze mortality rates across different age groups, with a specific emphasis on the elderly population to ascertain if age plays a significant role in COVID-19 severity. Further, I explore the correlation between age and the need for hospitalization or ICU admission, shedding light on healthcare utilization patterns. This study aims to contribute insights into the nuanced impact of COVID-19 on diverse age demographics, particularly emphasizing the vulnerability of older individuals. These findings intend to inform targeted public health interventions and strategies.

The paper is further organised into three sections: Data, Results, and Conclusion. In the Data section, I discuss how the dataset to be used for the analysis was obtained, pre-processed and cleaned. I will also explain the variables of interest in the dataset for my analysis. The Results section will then highlight and discuss the trends and associations found during the analysis. Lastly, the Conclusion section will summarize the major takeaways from my analysis.

2 Data

The data used for analysis in this paper was obtained from the City of Toronto’s Open Data Library using the library `opendatatoronto` (Gelfand 2022). The dataset is called “COVID-19 Cases in Toronto” (Data 2024). Data was cleaned and analysed using the statistical programming software R (R Core Team 2023) along with the help `tidyverse` (Wickham et al. 2019), `knitr` (Xie 2014), `ggplot2` (Wickham 2016), `here` (Müller 2020), `arrow` (Richardson et al. 2023).

2.1 About the dataset

The “COVID-19 Cases in Toronto” dataset obtained for this analysis contains demographic, geographic, and severity information for all confirmed and probable cases reported to and managed by Toronto Public Health since the first case was reported in January 2020. The data are extracted from the provincial Case & Contact Management System (CCM). (Data 2024) The dataset is overwritten bi-weekly and the version I extracted was last refreshed on January 17, 2024.

2.2 Preprocessing and data cleaning

The data was preprocessed and cleaned to contain only the entries that were essential for my analysis. This included me keeping entries that were confirmed cases which ended up with a known outcome i.e either fatal or resolved. This excluded probable or active cases. Another decision taken was to use records till 31 December, 2022 due to the multiple changes in criteria and updating methods throughout 2023. These decisions helped me refine my dataset and get precise results for my analysis.

2.3 Variables of interest

I have 6 variables I will be using in my analysis from the original dataset. My variables of interest are :

- **ID**
 - Numerical variable which is a unique row identifier for the database.
- **Age Group**
 - Categorical variable for the age at the time of illness (in years) split into 9 groups: 19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90+.
- **Reported Date**
 - The date on which the case was reported to Toronto Public Health.
- **Outcome**
 - Categorical variable for the outcome of the case, with two possible values:
 - * **Fatal:**
 - Any case that has died and has been marked as Outcome equals ‘Fatal’ and Type of Death does not equal ‘Disease of Public Health Significance was unrelated to cause of death’ in the provincial reporting system (CCM). (Data 2024)

* **Resolved:**

- A case outcome description in CCM of 'Recovered' OR Case outcome description is equal to 'Fatal' AND Type of Death is equal to 'Disease of Public Health Significance was unrelated to cause of death' OR Today's date is more than 14 days from the episode date AND the case is not currently hospitalized/intubated/in ICU AND Case outcome description is not equal to 'Fatal' where Type of Death is not equal to 'Disease of Public Health Significance was unrelated to cause of death'. (Data 2024)

- **Ever Hospitalized**

- Categorical variable taking the value of either “yes” or “no” for whether they were hospitalized related to their COVID-19 infection.

- **Ever in ICU**

- Categorical variable taking the value of either “yes” or “no” for whether they were admitted to the intensive care unit (ICU) related to their COVID-19 infection.

I will now analyse the dataset to uncover trends and draw conclusions.

3 Results

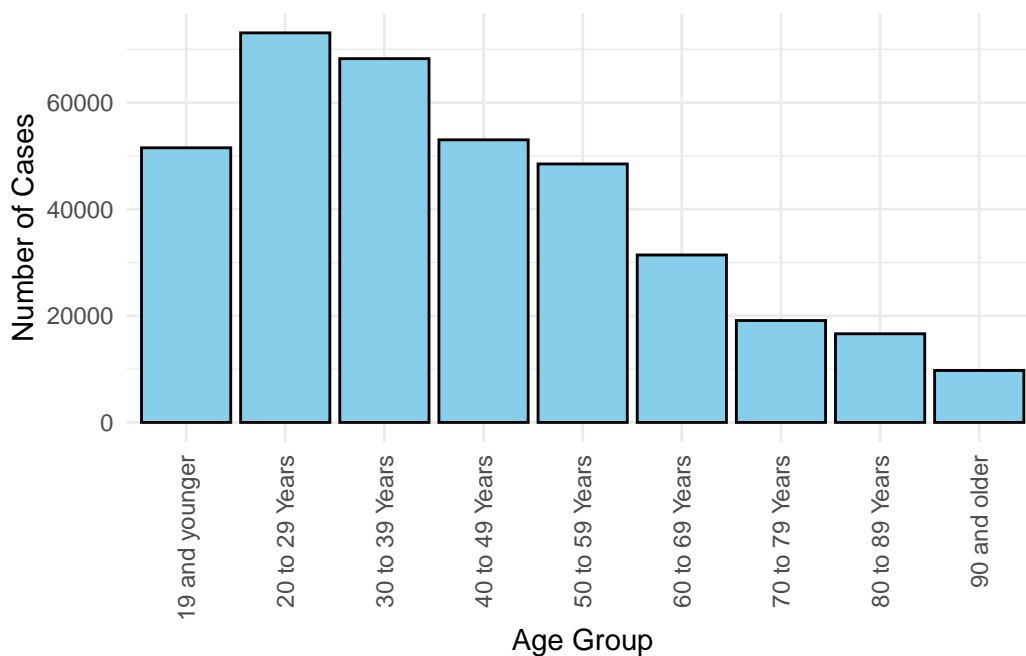


Figure 1: Number of confirm COVID-19 cases for each age group

The graph depicted in Figure 1 illustrates the distribution of confirmed COVID-19 cases across different age groups, offering a snapshot of the affected population. Notably, a discernible trend emerges, with the most of cases observed in the 20 to 29 age group. This observation aligns with the demographic composition of a metropolitan city, where the working-age population tends to be more predominant.

The prominence of the 20 to 29 age group can be attributed to the higher interaction levels among the working-class individuals within this range. Their daily activities, social interactions, and workplace engagements may contribute to an elevated risk of exposure to the virus. However, as I traverse older age groups, the number of cases progressively diminishes. This decline can be attributed to factors such as a natural decrease in population numbers due to aging and, unfortunately, mortality rates.

It also becomes evident that relying solely on raw case numbers to comprehend the true impact of COVID-19 across different age groups can be misleading. Given the substantial disparity in population sizes between older and younger demographics, using raw numbers may not accurately capture the relative severity of the virus’s effects. Thus, my analysis will use proportions to gain a more accurate understanding of the impact of COVID-19 on diverse age groups.

Table 1: Statistics for the outcome for confirmed COVID-19 cases.

outcome	count	proportion
fatal	4721	0.0127153
resolved	366564	0.9872847

The statistics presented in Table 1 offer an overview of the outcomes for confirmed COVID-19 cases in Toronto. Notably, the majority of cases exhibit a ‘resolved’ outcome, providing a glimpse of the resilience and recovery of a substantial portion of the affected population. However, a closer examination reveals 1.27% of the cases culminate in a fatal outcome.

While this percentage might appear modest at first glance, its significance becomes apparent when placed in perspective. Considered in absolute terms, it translates to 1 in every 100 individuals who contracted COVID-19 in Toronto meeting a fatal outcome. This stark reality underscores the gravity of the pandemic, aligning with its classification as one of the most lethal global health crises in recent history. (Maya Prabhu 2021)

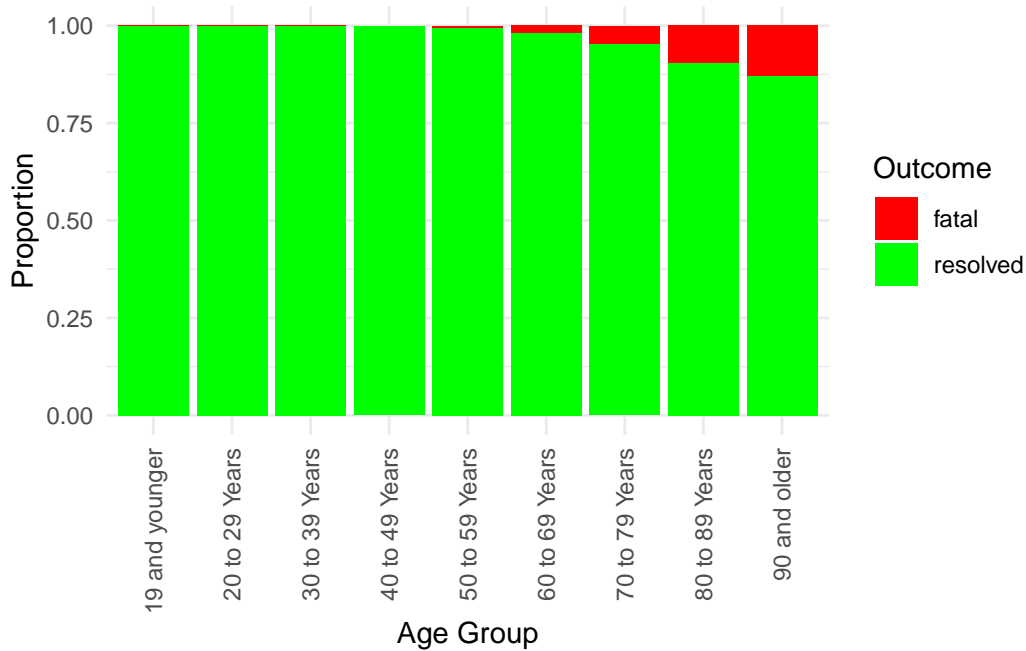


Figure 2: Outcome proportions for confirmed COVID-19 cases

Table 1 highlighted how 1 in every 100 people ended with a fatal outcome. The insights from Figure 2 underscore a critical aspect of COVID-19 outcomes, particularly when viewed through the lens of different age groups. Notably, the graph reveals a trend in fatality rates, with individuals aged 50 and below exhibiting relatively negligible fatality rates. However, as we progress beyond the age of 50, an exponential increase in fatalities becomes evident. Upon noticing the last group aged 90 and older, a striking 15% of patients are noted to experience a fatal outcome.

This escalation in fatality rates among the elderly imparts the perspective – older individuals are significantly more endangered when confronted with COVID-19. These findings align with broader epidemiological observations, emphasizing that older age is a significant risk factor for adverse outcomes associated with COVID-19.

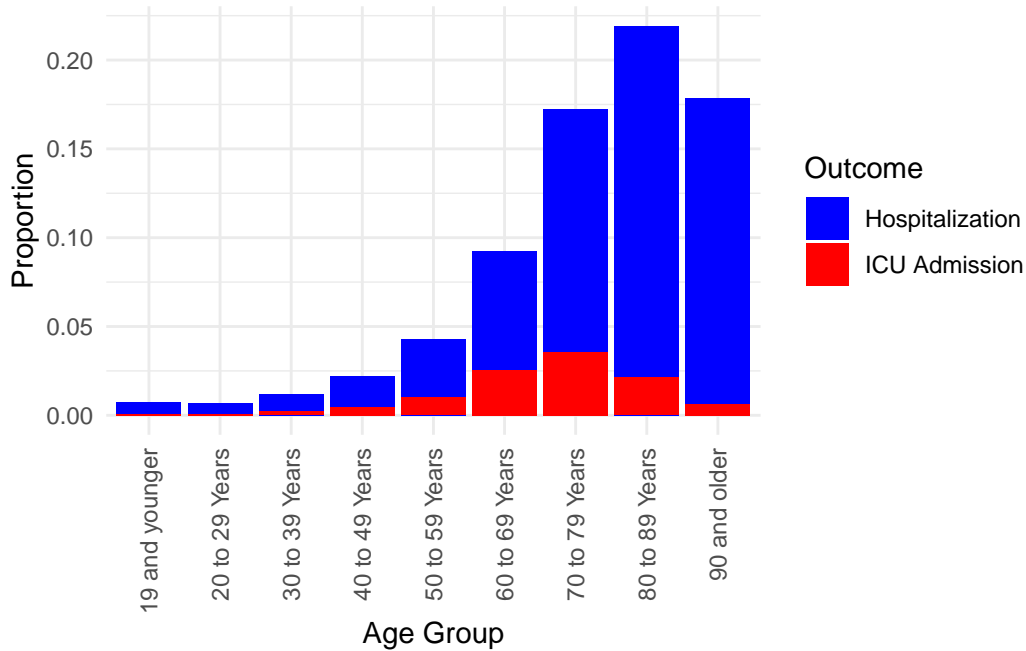


Figure 3: Proportion of people who were hospitalised or admitted to the ICU while having COVID-19

The insights offered by Figure 3 provide a view of the proportional distribution of individuals who required hospitalization or admission to the ICU due to COVID-19 across different age groups. The pattern reveals an exponential increase in these proportions with increasing age, reaching a peak for individuals between 80 to 89 years old where approximately 22% of individuals required hospitalization or ICU admission. This observation further reinforces the narrative that COVID-19 posed heightened risks for older individuals.

The visualization presented in Figure 4 delves into the proportional dynamics of deaths versus resolved cases after hospitalization across different age groups. The pattern highlights a significant escalation in the proportion of fatalities with increasing age. This observation can be connected with the earlier findings from Figure 3, which underscored the substantial increase in hospitalization proportions with advancing age.

Connecting these insights offers the perspective that as age increases, not only do more individuals require hospitalization, but a considerable number among them succumb to fatal outcomes. This interconnected analysis serves to solidify the overarching conclusion: older individuals are more vulnerable and endangered by COVID-19.

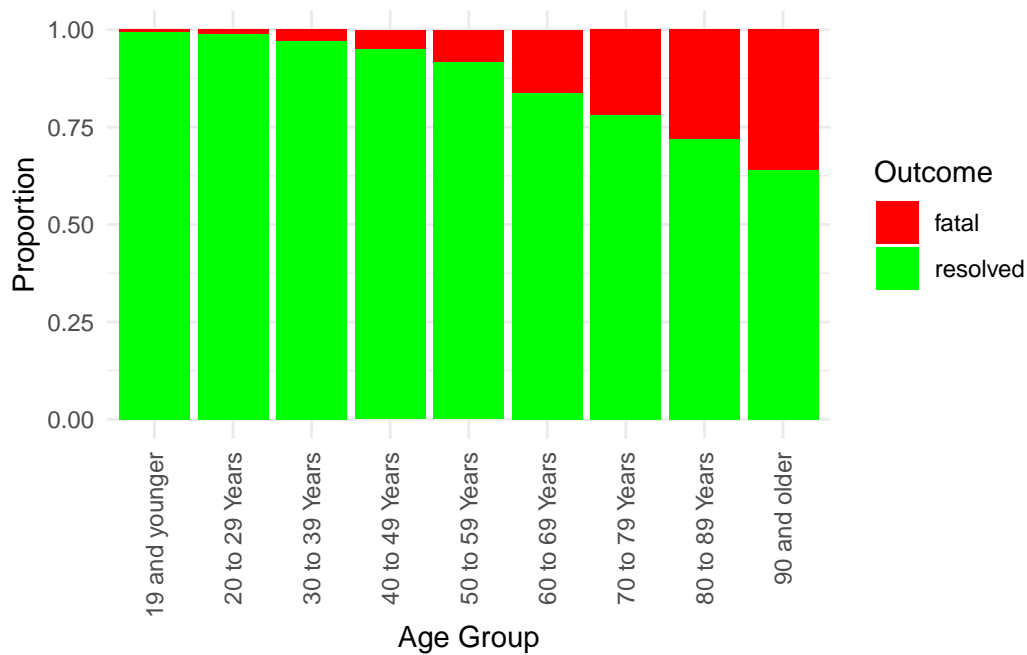


Figure 4: Outcome for those were hospitalised having COVID-19

4 Conclusion

In conclusion, the comprehensive analysis underscores a visible pattern revealing that older individuals exhibit a heightened vulnerability to a fatal outcome upon contracting COVID-19. This assertion is substantiated by the increase in mortality rates observed in higher age categories. Moreover, the findings illuminate a parallel trend, indicating an elevated likelihood of ICU admission as age progresses. The cumulative evidence strongly suggests that advancing age is a critical determinant in the severity of COVID-19 outcomes, emphasizing the imperative for targeted public health measures and specialized care strategies for the elderly population.

5 Appendix

5.1 Glimpse of the dataset

Table 2: Sample of cleaned Covid-19 data

ID	Age Group	Reported Date	Outcome	Ever Hospitalized	Ever in ICU
1	50 to 59 Years	2020-01-23	resolved	no	no
2	50 to 59 Years	2020-01-23	resolved	yes	no
3	20 to 29 Years	2020-02-21	resolved	no	no
4	60 to 69 Years	2020-02-25	resolved	no	no
5	60 to 69 Years	2020-02-26	resolved	no	no
6	50 to 59 Years	2020-02-27	resolved	no	no

Table 2 provides a glimpse of the cleaned COVID-19 data. The data was used to uncover trends and make associations.

5.2 Data cleaning

The data cleaning steps consisted of multiple filtering decision. Those are:

- Keeping only confirmed cases. The original dataset included probable cases also but for my analysis, I wanted to find trend among people who were confirmed to have COVID so for precision in trend, I omitted the probable cases.
- Keeping records of cases till Dec 31, 2022. This was because the dataset page documents how criteria and updating methods were changed multiple time throughout 2023. To ensure coherent and similarly-collected data for precision of result, this call was taken.
- Cases which are still active were removed as analysing the outcome of the case is one of my trends to observe which having active cases would not help.
- The last minor cleaning included removing empty age categories as that is a variable of interest and changing the datatype for the Reported date variable.

References

- Data, Toronto Open. 2024. “COVID-19 Cases in Toronto.” <https://open.toronto.ca/dataset/covid-19-cases-in-toronto/>.
- Disease Control, Centers for, and Prevention. 2023. “COVID-19 Risks and Information for Older Adults.” <https://www.cdc.gov/aging/covid19/index.html>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://sharlagelfand.github.io/opendatatoronto/>.
- Maya Prabhu, Jessica Gergen. 2021. “History’s Seven Deadliest Plagues.” <https://www.gavi.org/vaccineswork/historys-seven-deadliest-plagues>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://here.r-lib.org/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoş Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2023. *Arrow: Integration to ‘Apache’ ‘Arrow’*. <https://github.com/apache/arrow/>.
- WHO. 2020. “WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020.” <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>.
- 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>.
- Wikipedia. 2024. “COVID-19 Pandemic.” https://en.wikipedia.org/wiki/COVID-19_pandemic.
- worldometers. 2024. “Coronavirus Cases.” <https://www.worldometers.info/coronavirus/>.
- 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>.