Toronto's COVID-19 Impact: Effect of Age and Gender on Outcomes*

Dhruv Gupta

September 26, 2024

The COVID-19 Pandemic caused irreparable damage to the infrastructure of our global systems, highlighting the neccessity to better understand factors that influence health outcomes. This paper uses data from OpenDataToronto to investigate how age and gender affects outcomes of those infected with the virus. We find that in Toronto, age is strongly correlated with mortality and poorer hospitalisation outcomes. Further, we find that males have significantly worse outcomes than females in both mortality rates and hopsitalisation outcomes

Table of contents

*Code and data are available at: https://github.com/dhruv5423/Covid19-R-Project

1 Introduction

The Covid-19 Pandemic has had an unimaginable effect on human lives around the globe. As of April 13, 2024, Worldometer Info estimates that over 704 million people worldwide have contracted the virus, resulting in approximately 7 million deaths (Worldometer 2024). In Canada alone, there have been almost 4.6 million reported cases and more than 38,000 deaths as of July 20, 2024 (Government of Canada 2024).While these statistics paint a harrowing picture of the human toll inflicted, the pandemic has exacerbated existing economic inequalities, destabilized political systems, and put immense pressure on societal infrastructure globally. Looking past the immediate health crisis, lockdowns and restrictions have had immeasurable impacts not only on global supply chains, but also on the mental heatlh of many forced to quarantine or self isolate.

Importantly, Covid-19 has been found to have varying effects across demographics. A 2020 article published in the PLOS Journal found that 'Covid-19 may be associated with worse outcomes in males than in females'. The article found that men are up to 22% more likely to require ICU admission. Lakbar et al. (2020) Moreover, an article published in the Springer Link Journal in 2021 found that older adults, in particular those above the age of 65, face higher mortality rates than their younger counterparts. Weaker immune systems, and the higher likely presence of other conditions can exacerbate the effects of the virus. Hu, Lou, and Meng (2021)

Understanding how demographic variables like age and gender affect outcomes related to contracting viruses is increasingly important in the shaping of future policies and health measures. This paper aims to analyse the differences in outcomes for various age groups and genders among Covid-19 cases in Toronto, in an effort to contribute to deepening our understanding of the risk factors that may impact the lives of those with Covid-19, and possibly in future pandemics as well.

The remainder of this paper is structured as follows. Section 2 discusses data selecting and cleaning techniques, and provides a sample of the raw and cleaned data. Section 3 displays relevant figures found through the analysis - general results, age specific results, and gender specific results. Section 4 discusses these results in detail, the broader significance of the findings, and highlights some weaknesses and next steps that could be taken to improve the report.

2 Data

2.1 A Note on Data Selection and Measurement

Data used in this report was sourced from OpenDataToronto's portal. Gelfand (2022). More specifically, the data set "Covid-19 Cases in Toronto" was used and cleaned for the purposes of this report.

Toronto Public Health (TPH) collects COVID-19 case data in different ways. If an individual tests positive at a healthcare provider or laboratory, it is reported to TPH. Furthermore, in 'high-risk settings', such as hospitals and nursing homes, management is required to report positive cases to TPH. TPH also uses extensive contact tracing methods to monitor potential spreads of COVID-19. City of Toronto (2024b). They released anonymized, person-level data from the start of the pandemic in January 2020 to OpenDataToronto.The data spans from the first reported case on in January of 2020 to February 14th, 2024. In a statement on the website for this dataset, OpenDataToronto states that "As case and outbreak management guidelines changed and COVID-19 specific resources were no longer funded, the level of detail available for cases decreased, and more recent data are less complete and not comparable to previous years. TPH discontinued the production of this report with the final refresh as of February 14, 2024" City of Toronto (2024a).

The data set received a 'Gold' quality score, which takes into account the freshness, usability, metadata, accessibility and completeness of the data. This adds to the credibility of the data. However, there are likely many unreported cases due to the fact that it is not mandatory to report COVID-19 if you test positive using an over-the-counter antigen test. This likely leads to some under-reporting, and hence the data might not be fully representative of the overall effect COVID-19 had in Toronto.

This project uses R Core Team (2023) to simulate, download, clean, plot, and analyse the data provided by (**opendatatoronto-set?**).

2.2 Raw Data

In it's original form, the dataset contains more than 414,000 entries regarding information on cases of Covid-19 in Toronto. Table 1 and Table 2 below are samples of the first three rows of the raw data, separated into two tables for readability.

Table 1: COVID-19 Case Raw Data (Part 1)

X_id	Outbreak.Associated	Age.Group	Neighbourhood.Name	FSA	Source.of.Infection	Classification
1	NO	50 to 59 Years	Willowdale East	M2N	Travel	CONFIRMED
2	NO	50 to 59 Years	Willowdale East	M2N	Travel	CONFIRMED
3	NO	20 to 29 Years	Parkwoods-Donalda	M3A	Travel	CONFIRMED

Table 2: COVID-19 Case Raw Data (Part 2)

Episode.Date	Reported.Date	Client.Gender	Outcome	Ever.Hospitalized	Ever.in.ICU	Ever.Intubated
2020-01-22	2020-01-23	FEMALE	RESOLVED	No	No	No
2020-01-21	2020-01-23	MALE	RESOLVED	Yes	No	No
2020-02-05	2020-02-21	FEMALE	RESOLVED	No	No	No

(Table Separated using Tips from (Stack Overflow 2015) and LLMs)

2.3 Data Cleaning

There were a variety of issues and fixes needed to be done to the raw data in order to get it ready for analysis. In order to download, clean and run tests on the data, the package Tidyverse, Wickham et al. (2019) was used.

When cleaning the data, the first step was to select appropriate variables of interest to analyse. The raw data had many different variables to pick from, but this report importantly selected the variables 'Age Group', 'Client Gender', 'Reported Date', and 'Outcome'. This was done to run some basic analysis for the differences in outcomes for age groups and genders. Additionally, the variables 'Ever Hospitalized', 'Ever In Icu', and 'Ever Intubated' were selected to provide further depth to analysis regarding the severity of cases across age and gender.

Next, the Janitor package as part of Tidyverse was used to clean up the names of the variables, converting them to snake lowercase format; for example, 'Age.Group' was cleaned to 'age_group'. Another key step in the cleaning process was to arrange the data in chronological order, and then match Case ID's to the chronologically adjusted dates. This was done in order to have a clearer picture on the time trends of mortality and case progression of COVID-19 in Toronto. Further, the variable 'Client Gender' took on many different values in the raw data, so for simplicity, genders other than Male and Female were combined to take the value 'Other'. Finally, NA values were removed. After running tests, there were 422 NA values associated with the raw data, which were dropped.

Table 3 is a sample of the first 3 rows of the new cleaned data table.

Table 3: Sample of COVID-19 Case Cleaned Data

x_id	age_group	client_gender	reported_date	ever_in_icu	ever_hospitalized	ever_intubated	outcome
1	50 to 59 Years	FEMALE	2020-01-23	No	No	No	RESOLVED
2	50 to 59 Years	MALE	2020-01-23	No	Yes	No	RESOLVED
3	19 and younger	MALE	2020-02-04	No	No	No	RESOLVED

3 Results

3.1 General Results

Figure 1 displays the progression of COVID-19 cases in Toronto from January 2020 to March 2024. Figure 2 importantly highlights mortality rates by level of hospitalisation. That is, the mortality rates for those hospitalised, put in the ICU, and/or intubated.

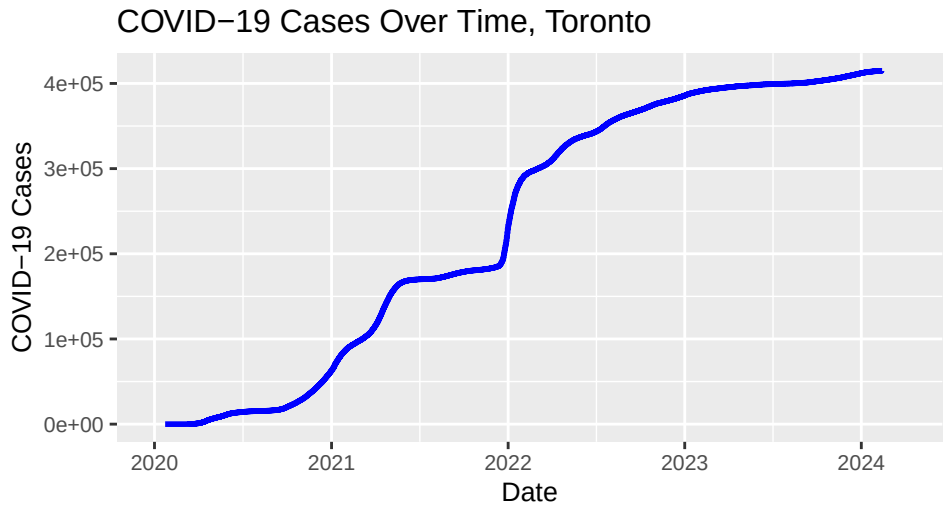

Figure 1: COVID-19 Case Progression Over Time, Toronto

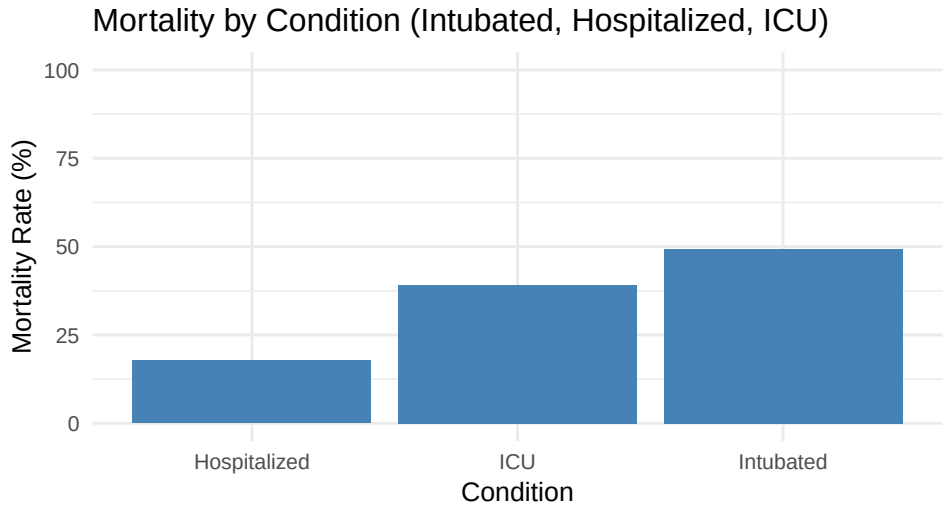

Figure 2: Level of Hospitalisation vs Mortality Rates

3.2 Age vs Outcomes

Figure 3 shows the differences in mortality rates for different age groups at the time of contracting COVID-19. Figure 4 shows the hospitalisation, ICU, or intubation rates for different age groups.

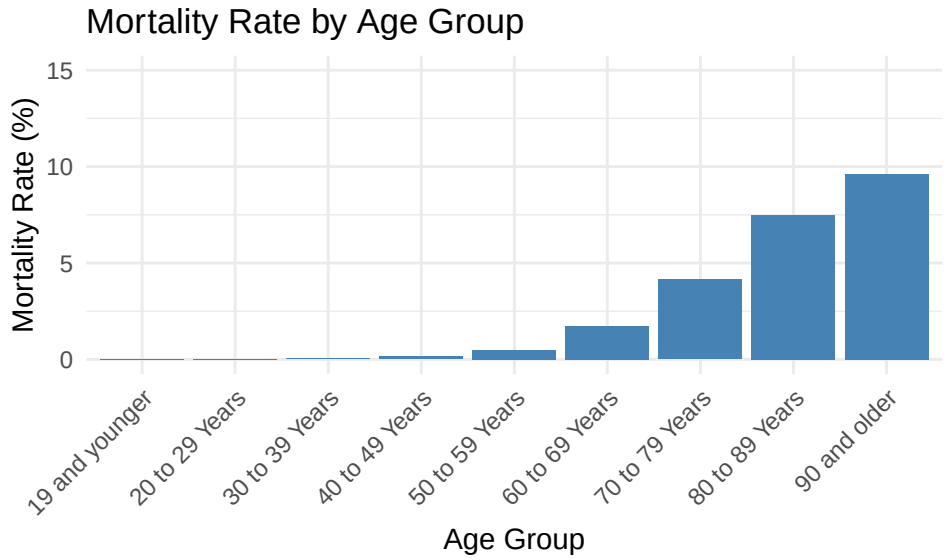

Figure 3: Age Group vs Mortality Rates

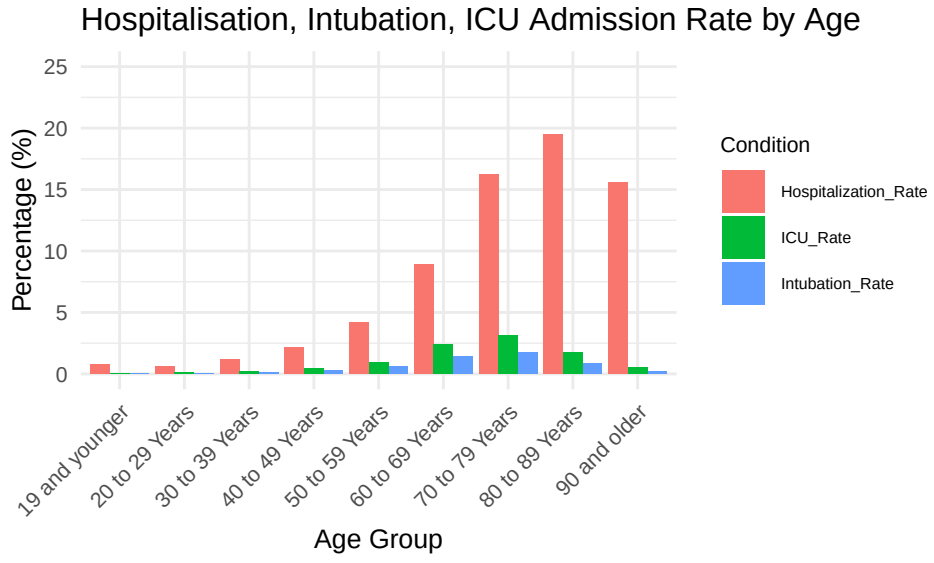

Figure 4: Hospitalisation, Intubation, ICU Admission Rate by Age

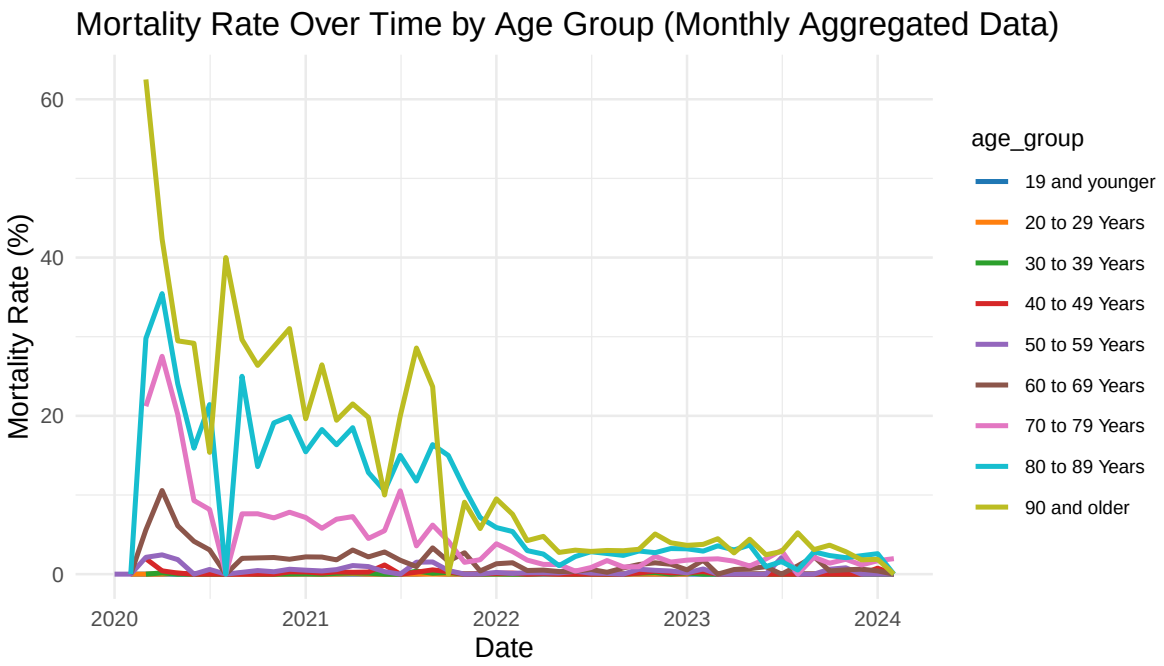

Figure 5: Mortality Rate Over Time by Age Group

3.3 Gender vs Outcomes

Figure 6 graphs COVID-19 mortality rates by gender. Figure 7 graphs hospitalisation, ICU and/or incubation rates by gender. That is, how likely a person who has contracted COVID-19 is likely to be in any of these conditions, by gender.

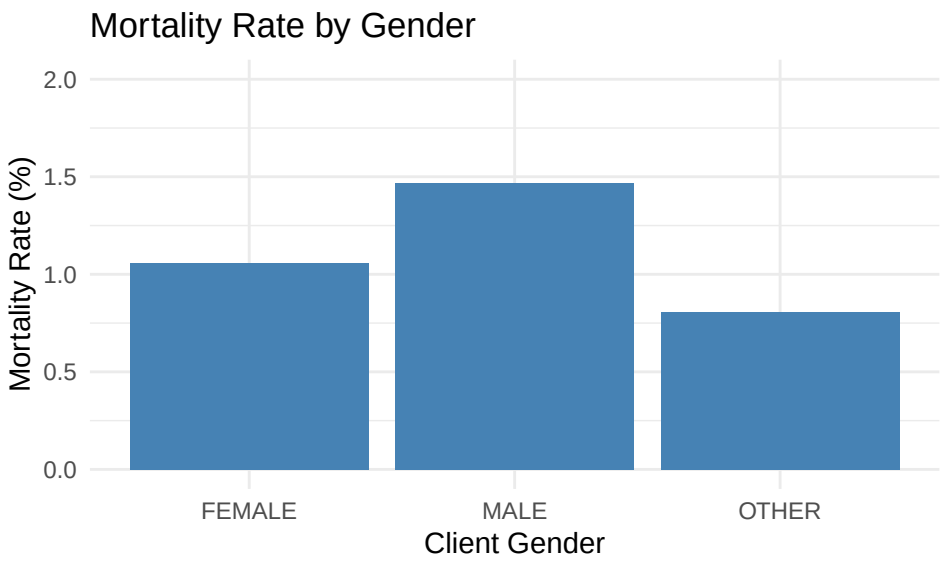

Figure 6: COVID-19 Mortality by Gender

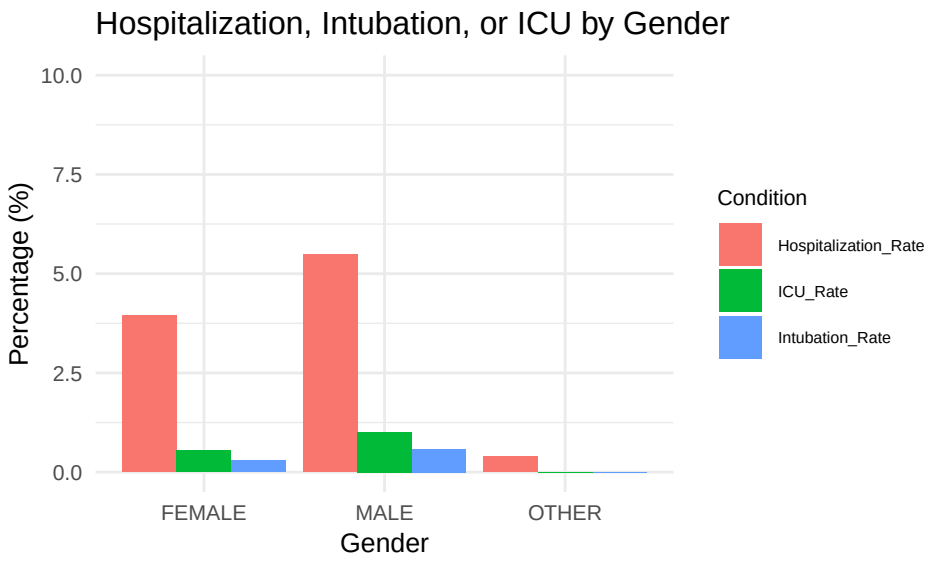

Figure 7: Hospitalisation, ICU, Intubation Rate by Gender

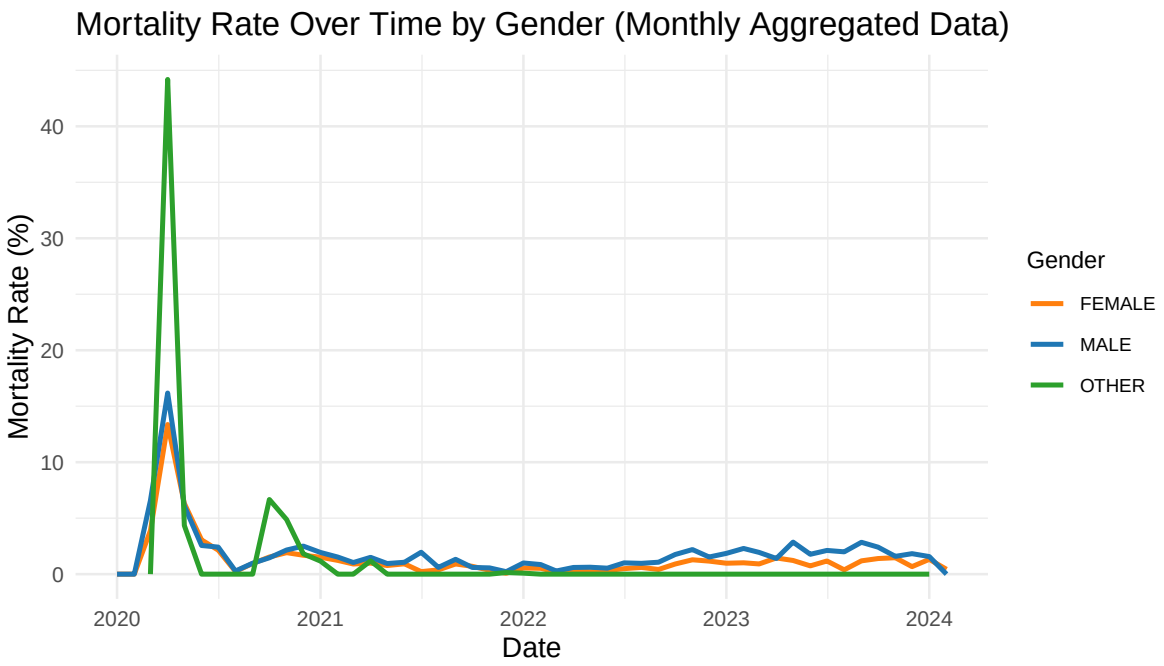

Figure 8: Mortality Rate Over Time by Gender

4 Discussion

4.1 Effect of Age on Outcomes

Understanding the effect age has on COVID-19 is extremely important both for potential victims and those in charge of policy to put in place correct and appropriate health measures for different ages.

Looking at Figure 3, we see a clear message - age group and mortality are positively correlated. More specifically, mortality past the ages of 50-59 shoots from under half a percent to 1.72% for 60-69 years of age. This only worsens as individuals get older. For the age groups 70-79 and 80-89, the mortality rate if one contracts COVID-19 shoots to 4.14% and 7.50% respectively. Moreover, those who contract COVID-19 at ages of 90 and above have a mortality rate of almost 10% (9.62%).

In contrast, those younger than 50 at the time of contraction have mortality rates under 0.5%. In fact, those 19 and younger have a mortality rate of only 0.007%. Clearly, age and mortalty are very strongly correlated.

Perhaps of more relevance are chances of hospitalisation, admission to the ICU, and intubation. In reference to Figure 4, we see that the older population is much more susceptible to having more severe cases of COVID-19. For all ages above 70, the chance of hospitalisation is above

15%, peaking at 19.46% for 80-89 year olds. Compared to the younger population, this is a stark difference. Those under 30 have hospitalisation rates of under 30%. This increases with age: hospital admission rates for those with COVID-19 is 1.15% for 30-39 year olds, 2.13% for 40-49 year olds, 4.15% for 50-59 year olds, and 8.91% for 60-69 year olds.

discussion about weaker immune systems with age (bring in some literature review)

4.2 Effect of Gender on Outcomes

Similar to age, understanding differences in outcomes for genders matter greatly to virologists and as a matter of public safety. Looking at Figure 6, we find that males that contract COVID-19 have a much higher mortality rate when compared to females (1.47% vs 1.06%). According to this data, males who contract COVID-19 are almost 1.4 times more likely to die than females.

Similarly, looking at Figure 7, we see that males are more likely to be hospitalised, put in the ICU, and intubated.

discussion about how Covid-19 affects the male body vs female; lit and scientific review; is this a worldwide trend?

4.3 Broader Discussion and Takeaways

Reference the importance of Figure 2 - those intubated have almost a 50% chance of death. (what?!)

4.4 Weaknesses and next steps

Measurement weaknesses (stopped collecting data), might be important to look at selection biases (are there more male cases than females? what about Other genders? are they affect results?

5 LLM Disclosure

ChatGPT Data Analyst was used to generate code and help fix bugs for this assignment. A full LLM Disclosure can be found on the GitHub Repository under "Other - LLM - usage.txt"

References

City of Toronto. 2024a. "COVID-19 Cases in Toronto." https://open.toronto.ca/dataset/covid-19-cases-in-toronto/.

———. 2024b. "COVID-19: What to Do If You Have COVID-19." 2024. https://www.toronto.ca/community-people/health-wellness-care/health-programs-advice/respiratory-viruses/covid-19/covid-19-what-you-should-do/covid-19-what-to-do-if-you-have-covid-19/.

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

Government of Canada. 2024. "COVID-19: Current Situation in Canada." https://health-infobase.canada.ca/covid-19/current-situation.html.

Hu, D., X. Lou, and N. et al. Meng. 2021. "Influence of Age and Gender on the Epidemic of COVID-19." *Wien Klin Wochenschr* 133: 321–30. https://doi.org/10.1007/s00508-021-01816-z.

Lakbar, I., D. Luque-Paz, J. L. Mege, S. Einav, and M. Leone. 2020. "COVID-19 Gender Susceptibility and Outcomes: A Systematic Review." *PLOS ONE* 15 (11): e0241827. https://doi.org/10.1371/journal.pone.0241827.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Stack Overflow. 2015. "How to Hold Figure Position with Figure Caption in PDF Output of Knitr." https://stackoverflow.com/questions/29696172/how-to-hold-figure-position-with-figure-caption-in-pdf-output-of-knitr. https://stackoverflow.com/questions/29696172/how-to-hold-figure-position-with-figure-caption-in-pdf-output-of-knitr.

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.

Worldometer. 2024. *COVID-19 Coronavirus Pandemic*. https://www.worldometers.info/coronavirus/.