Investigation of whether older people result in a higher hospitalized rate when infected by COVID-19*

Tomoya Fujikawa

06 February 2022

Abstract

Age is an important variable in determining whether a symptom of a disease develops to be severe or not in many cases. In this paper, the data used is COVID-19 cases in Toronto, and divides the age group by each 10 years, and indicates whether patients were required to be hospitalized for treatment or not. We visualise the data using tables and graphs, and observe upward trends in the number of hospitalized cases and the percentage of hospitalized people as the age increases. This finding calls for more cautious behaviour by older people who have higher chances of developing a more severe symptom when infected by COVID-19 and possibly an immediate hospitalization to receive a proper care.

1 Introduction

The COVID-19 pandemic has been an ongoing problem since the beginning of year 2020 and has caused a significant change in people's lifestyle. Both schools and works have shifted the media to online in order to minimize the number of infections. Yet, many people are continuously being infected. Even after the vaccines were successfully produced, the new Omicron variant has caused further increase in the number of cases. Among those who were infected, some have been able to treat and recover by themselves at home while others have experienced a more severe damage and were required to be hospitalized for more serious care.

There have been cases where individuals recovered from the virus but were left with prognostic symptoms. Moreover, there seems to be various patterns in how the condition changes. Thus, even if the symptoms are not severe at first, it cannot be underestimated and seeing a doctor is the best option for safety. However, since the number of cases is too large and low severity of symptoms can be cured at home, not everyone can be seen by a doctor. Hence, there is a need of information to allow individuals to make their decisions of whether they should focus on treatment at home or go to a hospital immediately.

For many of the diseases, one variable which tends to be strongly positively related with the development of symptoms is age. It is often the case that the older the patient is, the more severe the symptom is likely going to be for many diseases. It may be true for COVID-19 as well, but there has not been a disclosure of such discovery. Since there have been many cases of young people experiencing a severe symptoms of COVID-19, it is difficult to determine whether seeing the doctor is the best solution or not even for younger individuals.

The aim of this data analysis is to analyze the data of people who have been infected by COVID-19 in Toronto to investigate whether the data shows a strong relationship between the severity of symptoms and age. The conclusion of this analysis can be used by new patients to determine whether they believe hospitalization is necessary for them or not.

---

*Code and data are available at: https://github.com/tomoyasgithub/STA304-P1

The remainder of this paper is: Section 2 introduces the data and visualizes it in various ways. Section 3 explains the results of analysis. Section 4 describes strength and limitations of the data and discusses possible further investigations to improve the analysis. Section 5 summarizes the analysis and states the conclusion.

2 Data

This paper is going to focus on the data about COVID-19 cases in Toronto, acquired from Open Data Toronto Portal (City of Toronto Open Data Portal) and it will be using a statistical programming language, `R` (R Core Team 2021), for analysis. The packages used in this analysis are `tidyverse` (Wickham et al. 2019) and `knitr` (Knitr. R.). The tidyverse package is used to manipulate the data set and knitr package helps to produce tables and graphs for visualization. The original data set acquired from Open Data Toronto contains 277473 observations with 18 variables. In order to analyse and study this massive data, simplification and visualization processes are essential.

The collection of data started in January 2020 when the first case was reported within Toronto. Toronto Public Health are reported with confirmed and probable cases and they keep track of information of all the individuals that apply. The data is being continuously updated on a weekly basis up to present and it is being posted on every Wednesday. There are many variations in how the cases get reported to Toronto Public Health. The source of infection can vary from personal travels to community outbreaks to simply having a close contact with the person who was infected. Specific information about individuals such as age, gender and location of possible infection are also reported. They also keep track of their current status such as currently hospitalized, currently in ICU and currently intubated. Moreover, their records even after the recovery are kept so that the data set can tell if individuals were ever hospitalized, ever in ICU and ever intubated. The data is well managed by the Toronto Public Health. However, note that it is very unlikely that every case in Toronto is being reported because COVID-19 can be cured by treatment at home. Thus, it is only natural to assume that some individuals cure themselves without reporting to Toronto Public Healthcare.

The main focus of this paper, the age of individuals, is being reported in the range of each 10 years, instead of a specific number. The severity of conditions may be interpreted from 3 variables: `Ever Hospitalized`, `Ever in ICU` and `Ever Intubated`. Hospitalization would indicate that there is a need of continuous observation of the conditions of the patient as there is a possibility of sudden change. Experiencing ICU is more serious than simple hospitalization and intubation means the condition is at one of the worst stages it can get.

The table 1 shows the first ten rows of the data set, which only shows two important variables for this paper: age group of the individuals and if they were ever hospitalized or not. This process was done by first removing the empty observations in order to make the analysis as possible using `na.omit` function. Then, using the `select` function, the two variables, `Age Group` and `Ever Hospitalized` were pulled out. The variable, `Age Group` was altered slightly by using the `mutate` and `case_when` functions(dplyr). These functions were used to change the names of the age groups, for example "19 and younger" to "~19" and "20 to 29 Years" to "20~29." The intention of displaying this extracted data set is simply to show how the data was collected to keep track of all COVID-19 cases reported to Toronto Public Health. Simplifications as such allows for a clear visualization of the data set, hence, less confusion to the reader. However, simply looking at this table does not allow any conclusion to be made. Even if the table is extended to show every observation, it is difficult to observe any correlation between age and whether individuals were hospitalized. Thus, it needs to be cleaned in order to further analyse whether the age of infected individuals affect the level of symptoms shown when infected by COVID-19.

Table 1: First ten rows of the simplified data set

Age Group	Hospitalized
50~59	No
50~59	Yes
20~29	No
60~69	No
60~69	No
50~59	No
80~89	No
60~69	Yes
50~59	No
60~69	No

Table 2: Number of hospitalized and infected people in each age group

Age Groups	Hospitalized	Total
19 and younger	275	41599
20 to 29 Years	459	57571
30 to 39 Years	745	50774
40 to 49 Years	1090	39179
50 to 59 Years	1885	36055
60 to 69 Years	2471	22164
70 to 79 Years	2553	10966
80 to 89 Years	2541	8168
90 and older	1159	4486

The table 2 now shows three variables: the age group, the number of people who were ever hospitalized, and the total number of infected people in each age group. The cleaning of data set allows for a better visualization and easier interpretation of the data set. The number of total patients can easily be compared to the total number of patients who were ever hospitalized in each group. Since the purpose of this investigation is to examine the relationship between the age and the necessity of external help in recovering from the infection, it is necessary to compare the total number of patients to those who required external help in order to analyse the probability of hospitalization for each age group.

Figure 1:Distribution of COVID−19 cases in Toronto by age groups
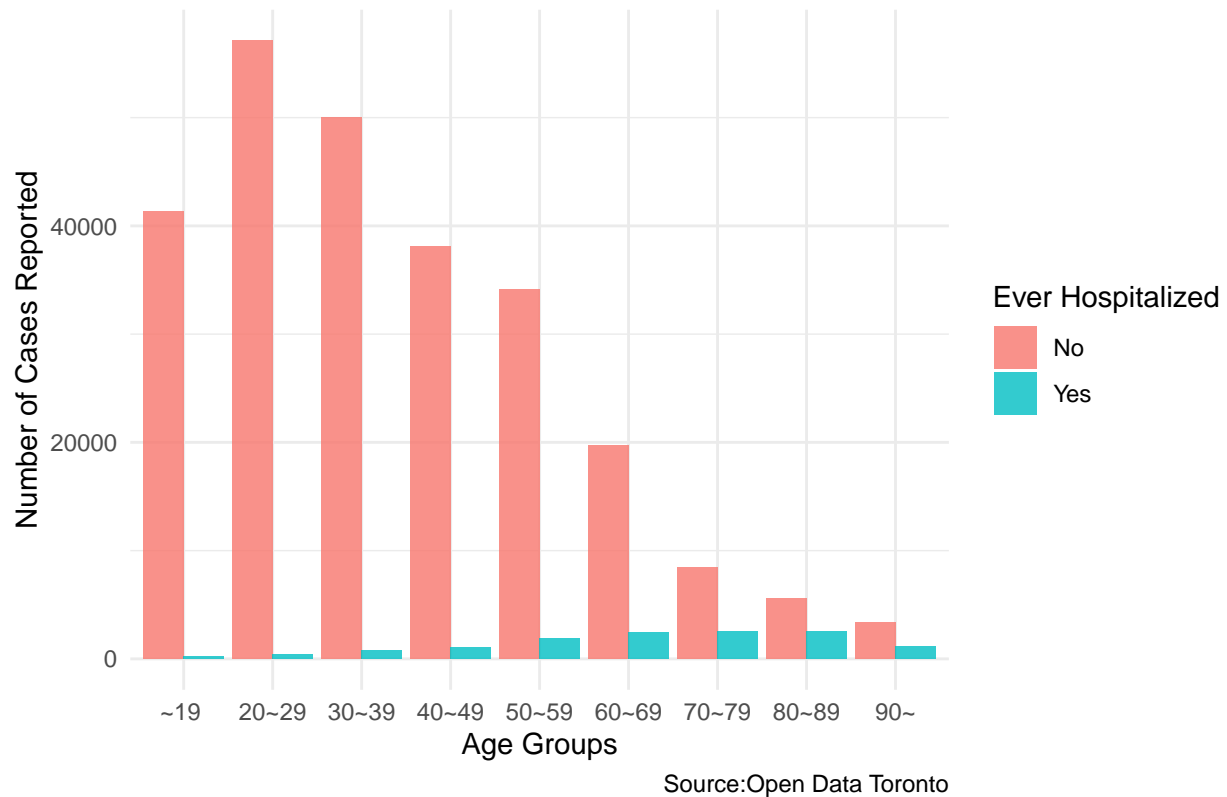

Figure 1 is a bar graph showing the distribution of COVID-19 cases in Toronto divided by age groups and the colors show the difference in whether an individual was ever hospitalized or not. This bar graph allows for clear visualization of trend in the increase and decrease in the number of cases for each age group.

3 Results

Both visualizations, the graph and the table 2, show clear trends that there is a decreasing pattern in the total number of cases as the age increases, other than the 19 and younger age group, but an increasing pattern in the number of people ever hospitalized. These patterns indicate that as the age increases, more percentage of people among those who were infected have hospitalized for treatment. Thus, it is more likely that the older the patient is, the more severe the symptoms of COVID-19 tends to be.

4 Discussion

The data shows a very clear trend and the conclusion that there is a positive relationship between age and the level of symptoms seems very accurate. However, there may be a few variables we may need to consider before completely agreeing to what the data shows.

The data was collected through all the reported cases to Toronto Health Public. Thus, the data is very reliable. All the information recorded in the data set must be credible. However, there are countless cases of COVID-19 which are not reported to Toronto Health Public. It is reasonable to assume that there are more cases within younger age groups which were not reported. This is because if they live with their parents, they can be treated at home. Also, for people who has jobs, if they get infected, it is necessary to report to the company which then would be reported to Toronto Health Public. The government of Canada has disclosed

the information about the rights and responsibilities of employees. The employees have the responsibility to follow the management's directions in reporting to work and workplace health procedures in the context of COVID-19 (Government of Canada). Therefore, it is possible that there are more cases within younger age groups and possibly more people who required hospitalization if they had been in different circumstances such as living alone or working at a company.

Another point to discuss is the definition of hospitalization. This is because there are different levels of treatment patients receive in the hospital(COVID-19 Information at UHN). Moreover, in the data set, there are variables such as `Ever in ICU` and `Ever Intubated` which may be better variables to determine the level of symptoms. However, only the variable, `Ever Hospitalized` was chosen to be discussed because the purpose of this study is to determine if infected individuals require external assistance in recovery or not. A further collection of data on the levels of treatment each individual received helps to make a stronger understanding of the correlation between the age and severity of conditions.

Last point to discuss is the possibility of misjudging symptoms. COVID-19 symptoms vary from: fever, cough, headache, shortness of breath, runny nose, physical dysfunction, nausea, diarrhea, muscle aches, loss of taste or smell to eye pain(COVID-19 Information at UHN). It is very probable that people assume they are not infected by COVID-19 virus, but instead simply feel sick or suspect to have other diseases. Especially fever, cough, headache and runny nose can occur from simply catching a cold, thus, people may not report it to the workplace or the Toronto Public Health. Also, since the COVID-19 testing is not cheap, it is possible that people try to recover without testing. Such possibilities definitely affects the result of the analysis.

5   Conclusion

Although the data set has high credibility as it records the cases reported to Toronto Health Public and is being updated weekly, there are some confounding variables to consider when reaching the conclusion. Furthermore, there is a limit to measure the severity of conditions with the available variables in the data set. A more detailed collection of data such as indicating the levels of symptoms numerically may help to understand the correlation between age and severity of conditions. In conclusion, the result observed from this data analysis can be used as a factor for individuals from each groups to decide whether they seek for external help if they get infected in the future.

6 References

A general vectorised if - case_when. - case_when • dplyr. (n.d.). https://dplyr.tidyverse.org/reference/case_when.html

Canada, P. H. A. of. (2021, December 22). Government of Canada. Canada.ca. https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection.html?utm_campaign=not-applicable&utm_medium=vanity-url&utm_source=canada-ca_coronavirus

COVID-19 Information at UHN. Covid-19 information at uhn. (n.d.). https://www.uhn.ca/Covid19#

Knitr. R. (n.d.). https://www.r-project.org/nosvn/pandoc/knitr.html

Open data dataset. City of Toronto Open Data Portal. (n.d.). https://open.toronto.ca/dataset/covid-19-cases-in-toronto/

Secretariat, T. B. of C. (2021, November 10). Government of Canada. Canada.ca. https://www.canada.ca/en/government/publicservice/covid-19/rights-responsibilities.html

The R project for statistical computing. R. (n.d.). https://www.r-project.org/

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