

COVID-19 VACCINATION DISTRIBUTION PRIORITY BY AGE

Deepkumar Pawar



August 20, 2020

CONTENTS

[ABSTRACT 2](#_Toc72084544)

[BACKGROUND 2](#_Toc72084545)

[METHODS 2](#_Toc72084546)

[FINDINGS 2](#_Toc72084547)

[INTERPRETATION 2](#_Toc72084548)

[FUNDING 2](#_Toc72084549)

[INTRODUCTION 2](#_Toc72084550)

[COVID-19 CASE DEMOGRAPHICS DATASET 3](#_Toc72084551)

[METHOD 3](#_Toc72084552)

[RESULTS 4](#_Toc72084553)

[CASE METRIC BY MEDIAN AGE 4](#_Toc72084554)

[CASE METRIC BY SHARE OF POPULATION AGED 65 OLDER 5](#_Toc72084555)

[CASE METRIC BY SHARE OF POPULATION AGED 70 OLDER 6](#_Toc72084556)

[CASE METRIC BY LIFE EXPECTANCY 7](#_Toc72084557)

[SUMMARY 7](#_Toc72084558)

[CANADA INDIVIDUAL CASES AND DEATHS DATASETS 7](#_Toc72084559)

[METHOD 8](#_Toc72084560)

[RESULTS 8](#_Toc72084561)

[NUMBER OF CASES WITH TRAVEL HISTORY VERSUS WITHOUT, BY AGE GROUP IN CANADA 8](#_Toc72084562)

[TOTAL CASES BY AGE GROUP IN CANADA 9](#_Toc72084563)

[NUMBER OF CASES ACQUIRED LOCALLY THROUGH CLOSE CONTACT VERSUS COMMUNITY BY AGE GROUP IN CANADA 9](#_Toc72084564)

[TOTAL CASES ACQUIRED LOCALLY BY AGE GROUP IN CANADA 10](#_Toc72084565)

[NUMBER OF DEATHS BY AGE GROUP IN CANADA 11](#_Toc72084566)

[SUMMARY 11](#_Toc72084567)

[NYC HEALTH DAILY DATA SUMMARY 12](#_Toc72084568)

[METHOD 12](#_Toc72084569)

[RESULTS 12](#_Toc72084570)

[AGE OF COVID-19 DEATHS BY EXISTENCE OF UNDERLYING CONDITIONS 12](#_Toc72084571)

[COVID-19 FATALITY RATE BY AGE 14](#_Toc72084572)

[SUMMARY 14](#_Toc72084573)

[DISCUSSION 15](#_Toc72084574)

[CONCLUSION 15](#_Toc72084575)

[REFERENCES 15](#_Toc72084576)

# ABSTRACT

## BACKGROUND

In 2020, the entire world had been severely disrupted due to the COVID-19 pandemic. Since then it has been imperative to learn as much as possible to help curb the devastating impacts of this virus.

As vaccines are in development around the world, it is becoming increasingly necessary to understand if and how we can determine a priority order for vaccination as well as what this should be.

## METHODS

I analyzed the resources I had available and used them to draw conclusions in favour of or against the arguments I have established.

## FINDINGS

There is evidence to suggest that the younger population requires prioritization. However, there is also evidence to suggest that the older population requires vaccination.

It is also found that using age as the only factor is not ideal to provide a clear prioritization and such a study would benefit from considering other factors such as the presence of underlying health conditions, location, occupation related social and outdoor requirements etc.

## INTERPRETATION

Further research and analysis would be required to determine a clear understanding of vaccine distribution prioritization which should take into account multiple different factors, the importance of these factors as well as a wide range of datasets.

## FUNDING

No funding sources were used for this study.

# INTRODUCTION

The responses to the COVID-19 pandemic have mostly consisted of Non-Pharmaceutical Interventions (NPIs). However, there is ongoing research towards finding a vaccine.

An important aspect of the effectiveness of a vaccine is its distribution. Once developed, it is highly likely that initially, vaccine supplies would be low and infrastructures unable to vaccinate the entire population at once. This is why it is important to determine who should receive this vaccine first and how to prioritize this distribution amongst the population.

This study aims to understand if a vaccine priority can be implemented simply on the basis of age using a statistical analysis and data science approach.

I have established hypotheses for this study in the form of two separate groups of the population that can pose arguments as to why they should be prioritized for vaccination:

* Younger population (roughly 15 to 44 years old)
  + They act as human disease vectors transmitting the COVID-19 virus across the population. They are also likely to be more ignorant of health and safety regulations as compared to older people since they are less likely to be fatally affected by the virus. Vaccinating them could decrease transmissions and spread. In this way, they can also protect the elderly members in their households.
  + The vaccine has a much higher chance of producing an immune response as compared to producing a much weaker immune response in older people.
* Older population (roughly above 65 years old)
  + They are much more fatally vulnerable to the virus and mortality increases in older people.

Thus, the study aims to see if the data available supports these arguments.

To do this, I have used three major resources:

* COVID-19 Case Demographics Dataset
* Canada Individual Cases and Death Datasets
* NYC Health Daily Data Summary

For my work, I have used Google Colaboratory, an online runtime that an execute python code that allows me to conduct my analysis.

# COVID-19 CASE DEMOGRAPHICS DATASET

The link to my work using Google Colaboratory is <https://colab.research.google.com/drive/1EKiwXdLCKzcQoP60KxP_L0Sn5jqCAOoJ?usp=sharing>.

This COVID-19 dataset is a collection of the COVID-19 data maintained by Our World in Data. It is updated daily and includes data on confirmed cases, deaths, hospitalizations, testing, and vaccinations as well as other variables of potential interest. (Hasell, et al., 2020)

## METHOD

1. Imported dataset into Google Colab
2. Identified important and relevant parts of dataset and removed other parts
3. Cleaned the data to use
4. Performed Principal Component Analysis on it (Brownlee, 2018)
5. Expressed data as scatter plots
6. Drew conclusions from the plots

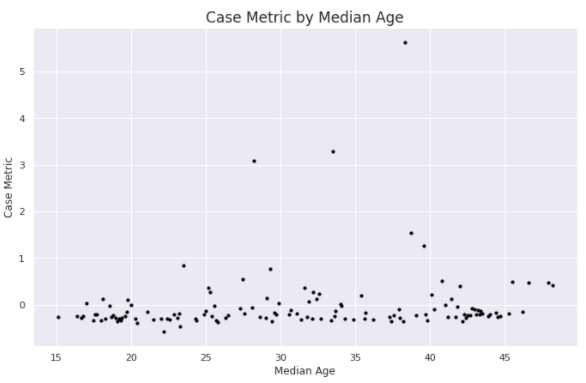
These individual steps have been shown in detail in my work which can be accessed using the link above.

## RESULTS

In this study, I have compared the attributes “Median Age”, ”Share of Population Aged 65 Older”, ” Share of Population Aged 70 Older” and ”Life Expectancy” for each country to a field called “Case Metric” which is a combination of the attributes “Total Cases”, “Days Elapsed since First Case”, “Population” and “Population Density” using Principal Component Analysis for each country. In this way, I was able to compare statistics related to population age against the impact and spread of COVID-19 for different countries.

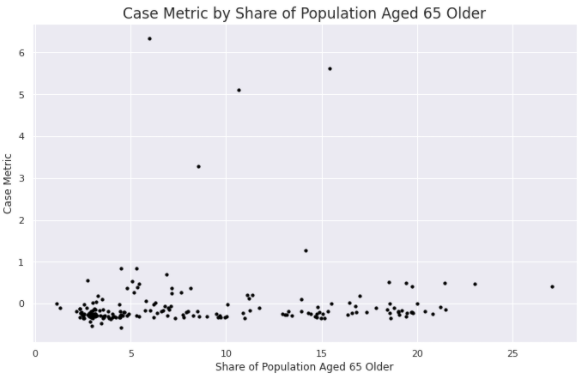
The following are my results:

### CASE METRIC BY MEDIAN AGE



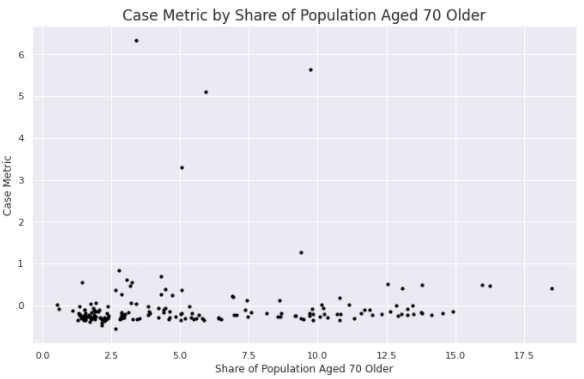
Looking at the Case Metric by Median Age graph, it is seen that the vertical spread at the lower end and higher end of median age is lesser than the vertical spread at the middle. At the middle, where median age is around 25 to 40, the case metric has a much higher variance. This includes what seem to be outliers to the general cluster of points around a case metric of 0. These outliers have a case metric above 2. Clearly, they had a massive surge of cases compared to the others with case metrics near 0. If these points are isolated and examined, they are found to represent the countries India, Brazil and USA with median ages of 28.2, 33.5 and 38.3 respectively. These countries are known to have experienced a very high surge in cases. This suggests that locations with a median age around 25 to 40 have a higher chance to experience a surge in cases. Another trio of interesting points are the ones with median age above 45 and case metric above 0 on the graph. They represent Germany, Italy and Japan which is interesting as they were known to have a surge in cases too in the first few months of the pandemic. The shape of the graph is roughly linear at the ends with it curving upwards in the middle.

### CASE METRIC BY SHARE OF POPULATION AGED 65 OLDER



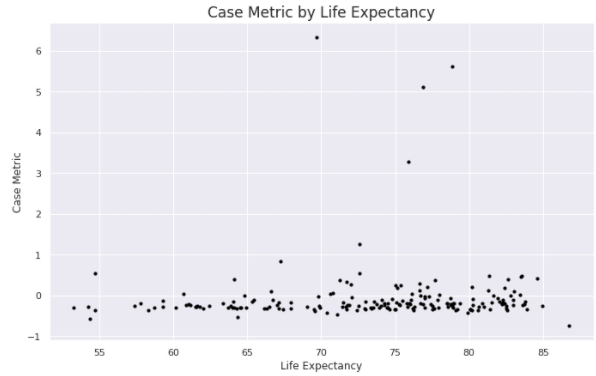
Looking at the Case Metric by Share of Population Aged 65 Above graph, the outliers with a much higher case metric can be seen with values in the range of 5 to 15. A cluster of points can also be seen with share values of around 4. This, however, is a result of more locations having share values in that range rather than a suggestion of an interesting aspect. The vertical spread looks similar throughout the values with a slightly larger vertical spread near the lower end of share values. However, the vertical spreads at different share values cannot be compared because there is a lack of points at the upper end of share values.

### CASE METRIC BY SHARE OF POPULATION AGED 70 OLDER



Looking at the Case Metric by Share of Population Aged 70 Above graph, it is quite similar to the Case Metric by Share of Population Aged 65 Above graph. The values would be shifted slightly to the left considering there is a lesser share of population aged 70 above than there is a share of population aged 65 above.

### CASE METRIC BY LIFE EXPECTANCY



Looking at the Case Metric by Life Expectancy graph, there is are more points with a higher life expectancy. The difference in vertical spread is hard to comment on considering the lesser points with a lower life expectancy.

### SUMMARY

* There is more variation in the spread of cases in locations where the median age ranges from 25 to 40.
* In the range of 25 to 40 for median age, there is a higher chance that the spread of cases becomes uncontrollably high and case numbers increase rapidly.
* The spread of cases in locations with very low or very high median age such as below 20 and above 50 is very low in comparison.

Therefore, this study contributes to the hypothesis that the population aged younger, above 20 and below 45, should be prioritized for vaccination over other age groups.

# CANADA INDIVIDUAL CASES AND DEATHS DATASETS

The link to my work using Google Colaboratory is <https://colab.research.google.com/drive/1hLnxYYbjdMS2qutF1McJcx6spDOBqFsM?usp=sharing>.

This is an individual-level data set of confirmed and presumptive positive cases of COVID-19 in Canada, including demographic characteristics, location, report date, travel history and exposure source. Time series of deaths, recoveries and testing are also recorded. (Berry, Soucy, Tuite, & Fisman, 2020) (Isha Berry, 2020)

## METHOD

1. Imported dataset into Colab
2. Identified important and relevant parts of dataset and removed other parts
3. Cleaned and grouped data to use
4. Expressed data as bar graphs
5. Drew conclusions from the plots

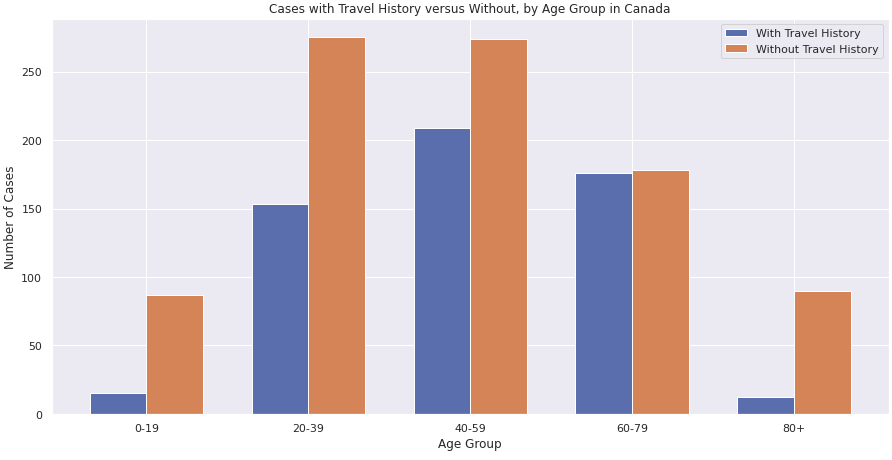
These individual steps have been shown in detail in my work which can be accessed using the link above.

## RESULTS

In this study, I examined two different individual level datasets. One was related to cases while the other was related to deaths.

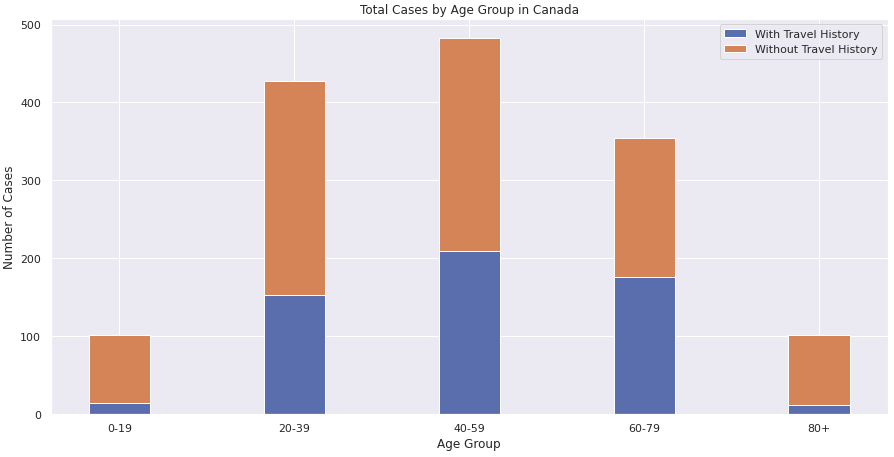
My analysis resulted in several different graphs as shown in the following:

### NUMBER OF CASES WITH TRAVEL HISTORY VERSUS WITHOUT, BY AGE GROUP IN CANADA



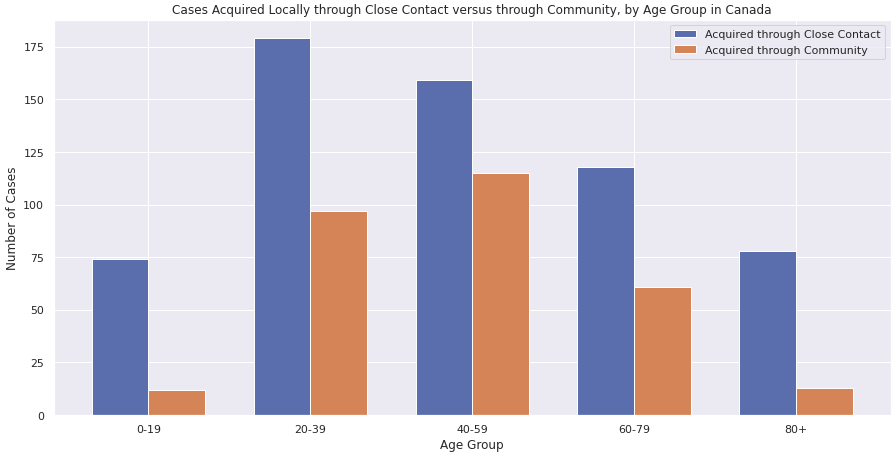
Looking at this graph, it suggests that adults in the age group of 20-59 travel outside the local area and catch COVID-19 much more frequently than persons aged 80 above and 0-19 who do not travel as much. A similar trend shows for cases without travel history.

### TOTAL CASES BY AGE GROUP IN CANADA



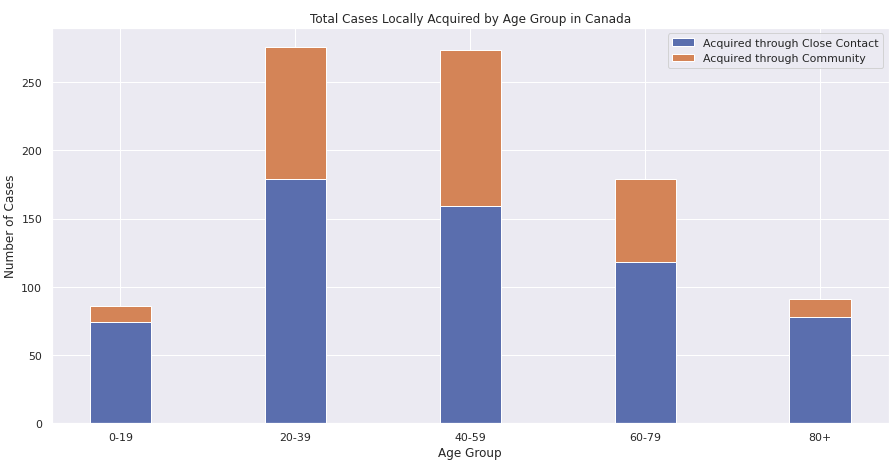
Looking at this graph, it suggests that adults aged 20-79, particularly aged 20-59, are much more susceptible to catching COVID-19 than other age groups.

### NUMBER OF CASES ACQUIRED LOCALLY THROUGH CLOSE CONTACT VERSUS COMMUNITY BY AGE GROUP IN CANADA



My comments on this graph have been combined with those on the following graph.

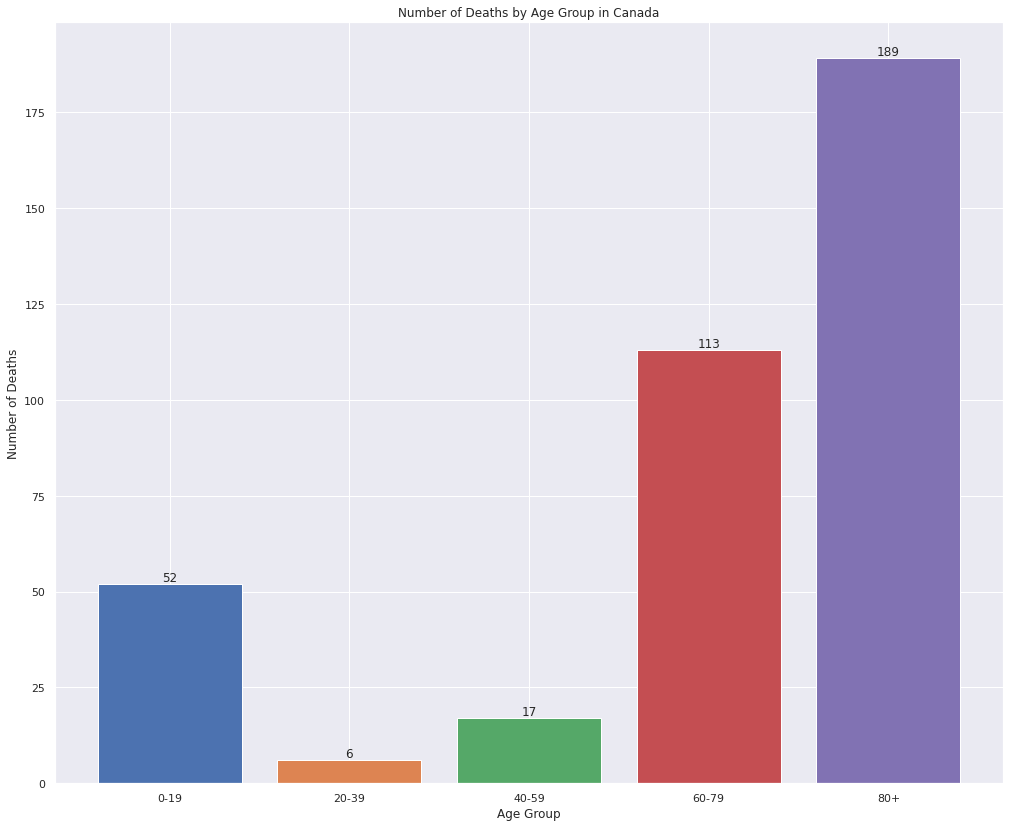
### TOTAL CASES ACQUIRED LOCALLY BY AGE GROUP IN CANADA



It appears that “Close Contact” is the main way that COVID-19 is acquired locally, compared to “Community”. “Close Contact” is when COVID-19 is transmitted through close contact with an infected person. “Community” is when the causes of getting infected are unknown.

Also, if we consider figures for “Close Contact”, we can see that it is highest in age group 20-59, particularly 20-39. Even for “Community”, it follows a similar trend except the highest is the age group 40-59.

### NUMBER OF DEATHS BY AGE GROUP IN CANADA



Looking at this graph, it suggest that people aged 60 and over have a much higher chance of death due to COVID-19 than people aged 20 to 59. Children and teenagers however seem to have a significantly high death rate. This would require separate investigation.

### SUMMARY

* Adults in the age group of 20 to 79 travel outside the local area and catch COVID-19 much more frequently than persons aged 80 above and 0 to 19 who do not travel as much.
* Adults aged 20 to 59 are the main spreaders of COVID-19 for locally acquired cases.
* Adults aged 60 and over, particularly 80 above, have a much higher chance of death due to COVID-19 than people aged 20 to 59.

Therefore, the first two points here contribute to the hypothesis that the younger population aged 20 to 59 should be vaccinated first in an effort to curb their significantly major transmission of the virus.

The third point contributes to the hypothesis that the older population aged 60 above should be prioritized for vaccination on account of their higher risk of death.

# NYC HEALTH DAILY DATA SUMMARY

This is not a dataset, rather it is a report that presents statistics found in a tabular format for New York City.

This data is taken from New York City Health (NYC Health, 2020) (NYC Health, 2020).

## METHOD

1. Identified tables containing relevant data
2. Drew conclusions from the relevant tables

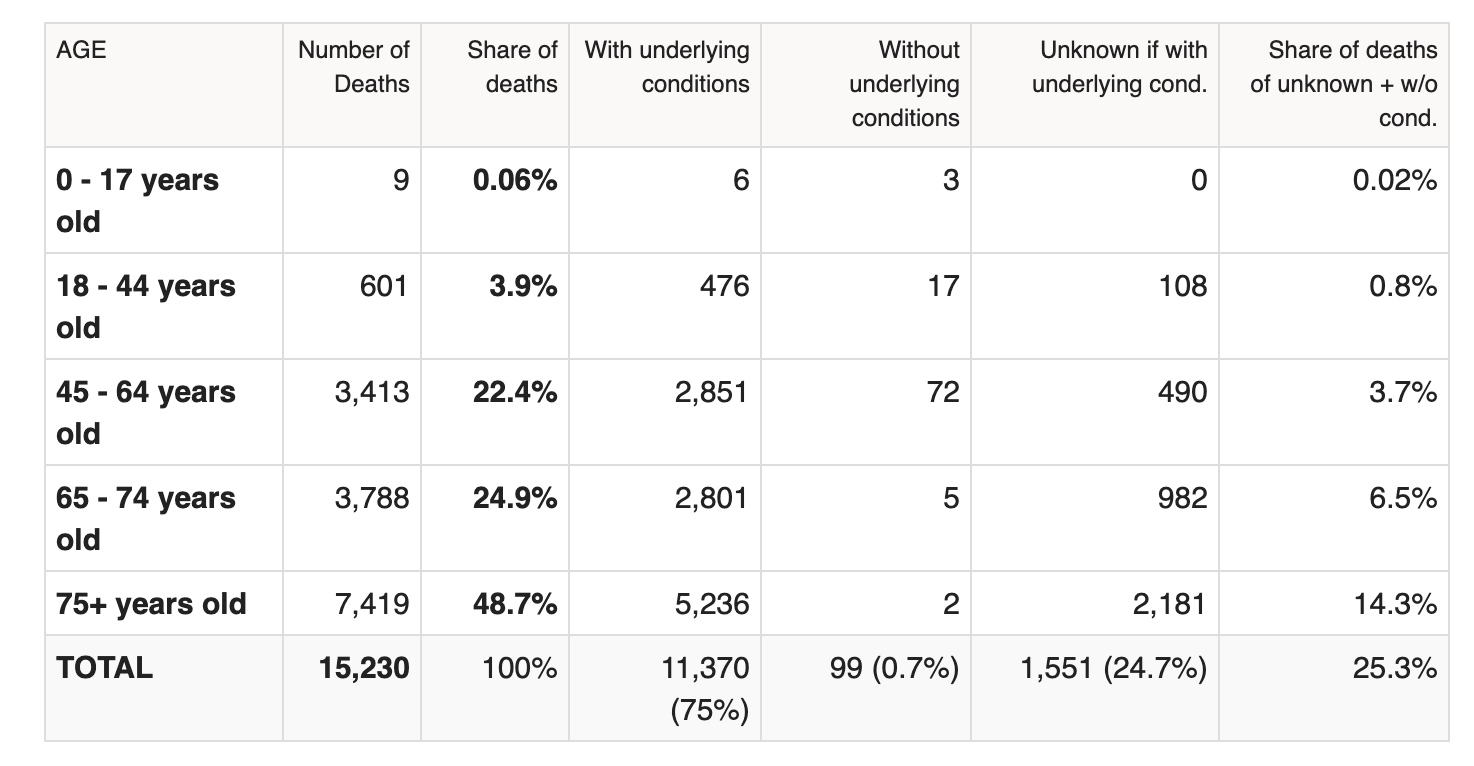
## RESULTS

The relevant tables here were the ones related to the age of a person.

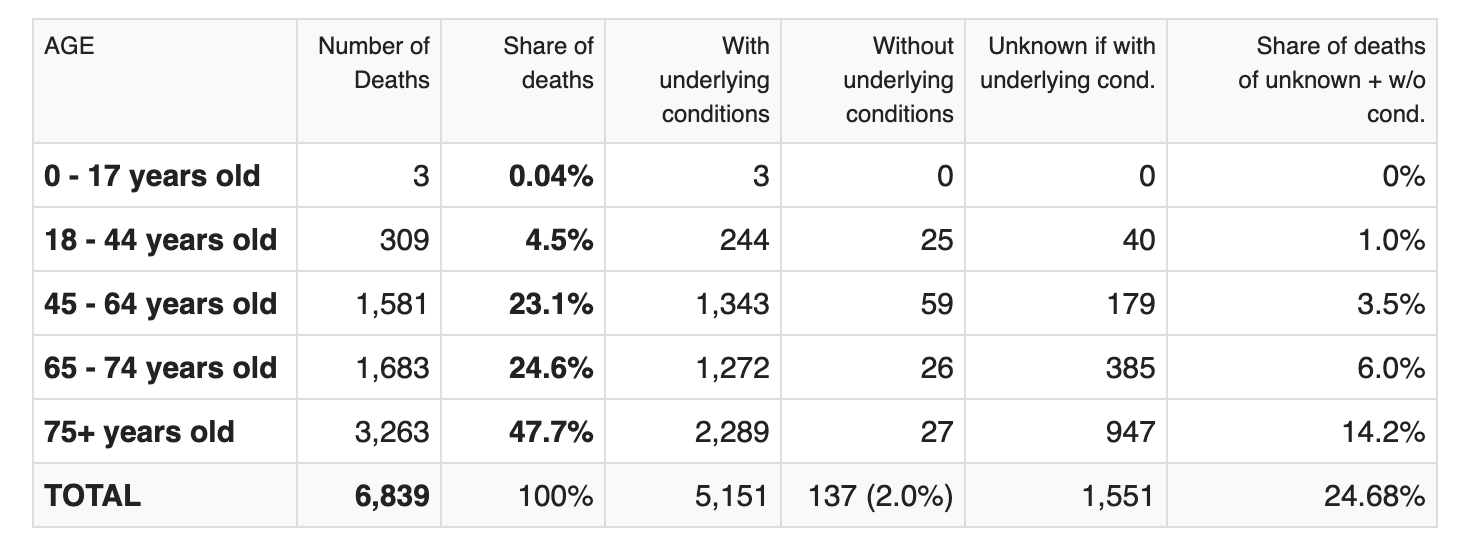
They are as following:

### AGE OF COVID-19 DEATHS BY EXISTENCE OF UNDERLYING CONDITIONS

May 13th, 2020

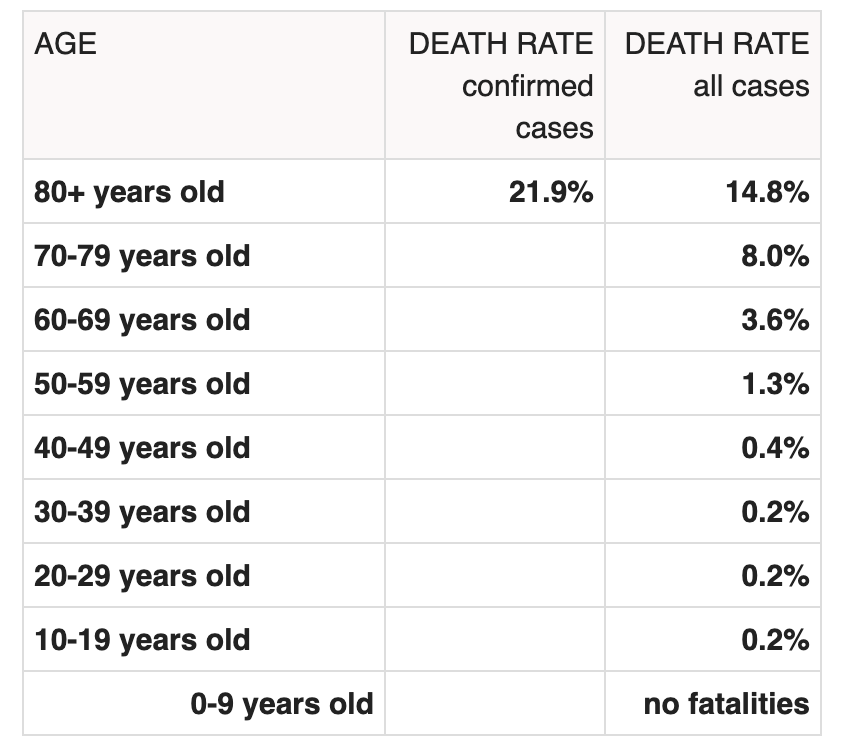


April 14th, 2020



Looking at these two tables, it contains two columns for deaths of people “With underlying conditions” and “Without underlying conditions”. It can be seen that cases that have underlying conditions have a much higher risk of death as age increases. This risk seems to be similar from 45 years to 74, but almost doubles for ages 75 older. The risk seems much lower for ages 44 below in comparison to 44 above. However, it is still higher than for the cases without underlying conditions. For cases without underlying conditions, the risk seems to be quite low from 18 to 75 above years, with no one younger than 18 reported dead. The deaths for 45 to 64 years is about twice as much for years 65 to 74 and 75 above in the April 14 table and the difference is much higher in the May 13 table. Thus, for a person without underlying conditions, there is a much higher risk of death at ages 45 to 64 years than any other. Also, the deaths for 0 to 17 years and 75 above is almost the same. It suggests that underlying conditions are the main factor in death and them being more common and deadly as age increases shows in these statistics.

### COVID-19 FATALITY RATE BY AGE



Here, the fatality rate or the death rate is the percentage ratio of the number of deaths in that age group to the total number of cases. It denotes the probability of death and the percentages do not have to add up to a 100 because we relate the number of deaths to the number of cases and not all cases die.

It can be seen that the death rate remains the same (a low 0.2%) for 10 to 39 years of age after which it starts growing slightly at first but then increases rapidly to give a death rate of 14.8%. The death rate for confirmed cases at 80 above years is higher at 21.9% (a confirmed case is one which has been tested and resulted in a positive, not all cases are tested and confirmed). Children below the age of 10 years have shown no fatalities.

### SUMMARY

This study, rather than clearly support one of the established hypotheses, makes it apparent that there are other factors apart from age that are as significant or more when it comes to determining a priority for vaccine distribution. The key factor considered here being the presences of underlying conditions as that seems to have a drastic effect on the mortality of COVID-19 cases.

Therefore, it would be highly beneficial to consider the effect of this factor and other possible factors when discussing vaccine distribution priority. Such investigation of other factors however, is beyond the scope of this paper.

# DISCUSSION

As has been made clear by this study, the determination of a vaccine priority order is a very vast area of research.

It is unclear which age group should be prioritized for vaccination as there is evidence that supports the vaccination of both the younger population as well as the older population. There is also undeniable evidence that age must not be the sole factor to consider, rather it would be more plausible to consider a variety of different factors in conjunction with another such as age and underlying conditions. Other possible factors might include location, current NPIs, occupation, mental health, particular vaccine properties, etc.

Hence, further research must be conducted on the topics mentioned above and I hope that this study can provide information as well as a guide towards this further research.

# CONCLUSION

In conclusion, it is possible to determine a priority order for vaccine distribution. However, it is not plausible to use age as the sole determining factor. Other factors such as underlying conditions and possibly location, current NPIs, occupation, mental health, particular vaccine properties, etc. may play a major role in determining the most beneficial prioritization of a future COVID-19 vaccine.

# REFERENCES

Berry, I., Soucy, J.-P. R., Tuite, A., & Fisman, D. (2020, April 14). Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. *CMAJ, 192*(15). doi:https://doi.org/10.1503/cmaj.75262

Brownlee, J. (2018, March 2). *How to Calculate Principal Component Analysis (PCA) from Scratch in Python*. Retrieved August 6, 2020, from Machine Learning Mastery: https://machinelearningmastery.com/calculate-principal-component-analysis-scratch-python/

Gellin. (2020).

Hasell, J., Mathieu, E., Beltekian, D., Appel, C., Gavrilov, D., Giattino, C., . . . Roser, M. (2020). A cross-country database of COVID-19 testing. *Scientific Data, 7*(345). doi:https://doi.org/10.1038/s41597-020-00688-8

Isha Berry, J.-P. R. (2020, April 14). *Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada*. Retrieved August 11, 2020, from cmaj: https://www.cmaj.ca/content/192/15/E420

NYC Health. (2020, May 12). *Daily Data Summary*. Retrieved August 10, 2020, from NYC Government: https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-daily-data-summary-deaths-05132020-1.pdf

NYC Health. (2020, April 14). *Daily Data Summary*. Retrieved August 10, 2020, from NYC Government: https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-daily-data-summary-deaths-04152020-1.pdf