Young people should not be blamed in the second wave of COVID-19

YangMing Hu(1003183379)

December 21, 2020

Abstract

This study appplied a GAM poisson model with bayesian analysis to investigate the question of interest that whether there is significant evidence that the mortality rate for people 0-49 years old is much higher in the second wave of COVID-19 began at September. It was found the first wave of covid-19 event mainly affected the old people but not young peple. And it was also found that the second wave of covid-19 event did not mainly affected the young peple as well as old people. So this study suggests we should not to blame young poeple in the COVID-19 event that there is no evidence.

Keywords: covid-19; mortality rate; GAM poisson family;

1 Introduction

In the year 2020, the most important event is the covid-19 event. Lots of peple died due to the cases of covid-19. And according to the USA government official, the first wave of covid-19 occurred in March, 2020. And in this first wave, the covid-19 mainly affects the elderly people much more than younge ones. But in the second wave that considered to began in September, 2020.

The USA government official claimed that the young people who are not want to stay at home but go outside unrestrictl should be response for the wave of covid-19 that covid-10 mainly affects young people including college students than old people. And in this study, the old age group of people refers to be the people with age 70 years old and over, the young age group of people refers to be the people with age under 50 years old which is 0-49 years old. So this study aims to investigate whether the number of deaths of the old age people group is much higher than the historical average for people under 50 years old as well as of the people 70 years old and over.

The followding study is organized as belows that, the methodology would first decribes the data used in the study and the model used in the study in details.Then results would be shown based on the data and the model used. At last, we dicuss the results in the dicussion section. The link to the paper is: https://github.com/Jaydenhu123/STA-304-FINAL.

2 Methodology

2.1 Data

The data source comes from the Weekly number of deaths in Quebec which are taken from the deaths database of the Registre RED maintained by the Institut ISQ. This database is designed to monitor the health status of the population, so it is not a real-time estimate of deaths like world bank data.

When a death occurs, the data would be sent to the ISQ and stored in RED database. And these detahs are mainly come from healthcare facilities such as hospitals and CHSLDs as well as other places like palliative-care homes. Also, there might be sent by Bureau.

And note for this data, no reasons of deaths or gender or other variables recorded, the only recorded information of data used in this study is age grooup, weekly number of deaths and the years. Other information is not used and not interested in this study.

As the data is not collected from a survey or similar approaches, the data is about the population of the deaths in Quebec, so that we discuss the whole population in Quebec or at least the whole frame population as there might be cases not included in the hospitals, CHSLD and palliative-care homes such as accidents but not recorded where there are people died not known to ISQ or cases known to be died but not recorded yet as the database is not a real-time data base, only the data to the year 2017 is complete while the weekly number of weekly deaths disseminated for 2018 to 2020 are not complete yet, however, the results should be very close to the true ones now, this study assumes the data is very close to the true deaths in the years 2018 to 2020 as the deaths in the years before 2018.

2.2 Model

The model used in this study is a GAM poisson model with random time effects. Clearly, the weekly number of deaths have seasonal effects, for example, the number of deaths are always lower in the July than in December might be due to reasons like temperature. Also, as the data is a yearly data started from 2016 to 2020, the weekly number of deaths might be related with lots of factors such as more cars in roads, more developed techniques and so on, thus, the time effect would be also important. At last, poisson family is choosen as the number of deaths must be positive and is a count data which is appropriate for using poisson family. We do not use gamma family as gamma family is for response which is continuous but here, the response is descrete count data, also, we do not use poisson GLM model as we also consider random effects of time.

The model form is as below:

$$Y_i \sim Poisson(\lambda_i)$$

$$log(\lambda_i) = X_i\beta + U_i$$

$$U_i \sim N(0, \sigma_U^2)$$

The prior is,

$$\sigma_U^2 \sim exp(\lambda_1)$$

where, $pr(x > log(1.2)) = 0.5$.

The details of the notations are as follows:

- Y_i is the number of deaths for time t = i.
- λ_i is the mean rate of umber of deaths for time t = i.
- X_i are covariates: the semiannual cycles cycles $cos(2\pi x_i)$, $sin(2\pi x_i)$, and the annual cycles $cos(4\pi x_i)$ and $sin(4\pi x_i)$.
- U_i is random effect of time.

3 Results

Figure 1 shows the Weekly number of deaths for people with 70 years old and over for the 5 years 2016-2020. This time series plot make comparisons the 5 years' weekly deaths numbers for the 70+ years old, obviously, the plot shows in the history the number of deaths in the about 30-th week in a year is lowest while the numbers at the beginning and end of the year has highest number of deaths. And the plot shows a very high peak of deaths due to COVID-19 in the year 2020.

Figure 2 shows the Weekly number of deaths for people with 0-49 years old and over for the 5 years 2016-2020. This time series plot also make comparisons the 5 years' weekly deaths numbers for the 0-49 years old, obviously, the plot shows the numbers of deaths at the beginning and end of the years are close to the numbers in the middle of the years, and there are no patterns for unusual weekly number of deaths. So the patterns of deaths for 0-49 years old and 70+ years old are indeed different.

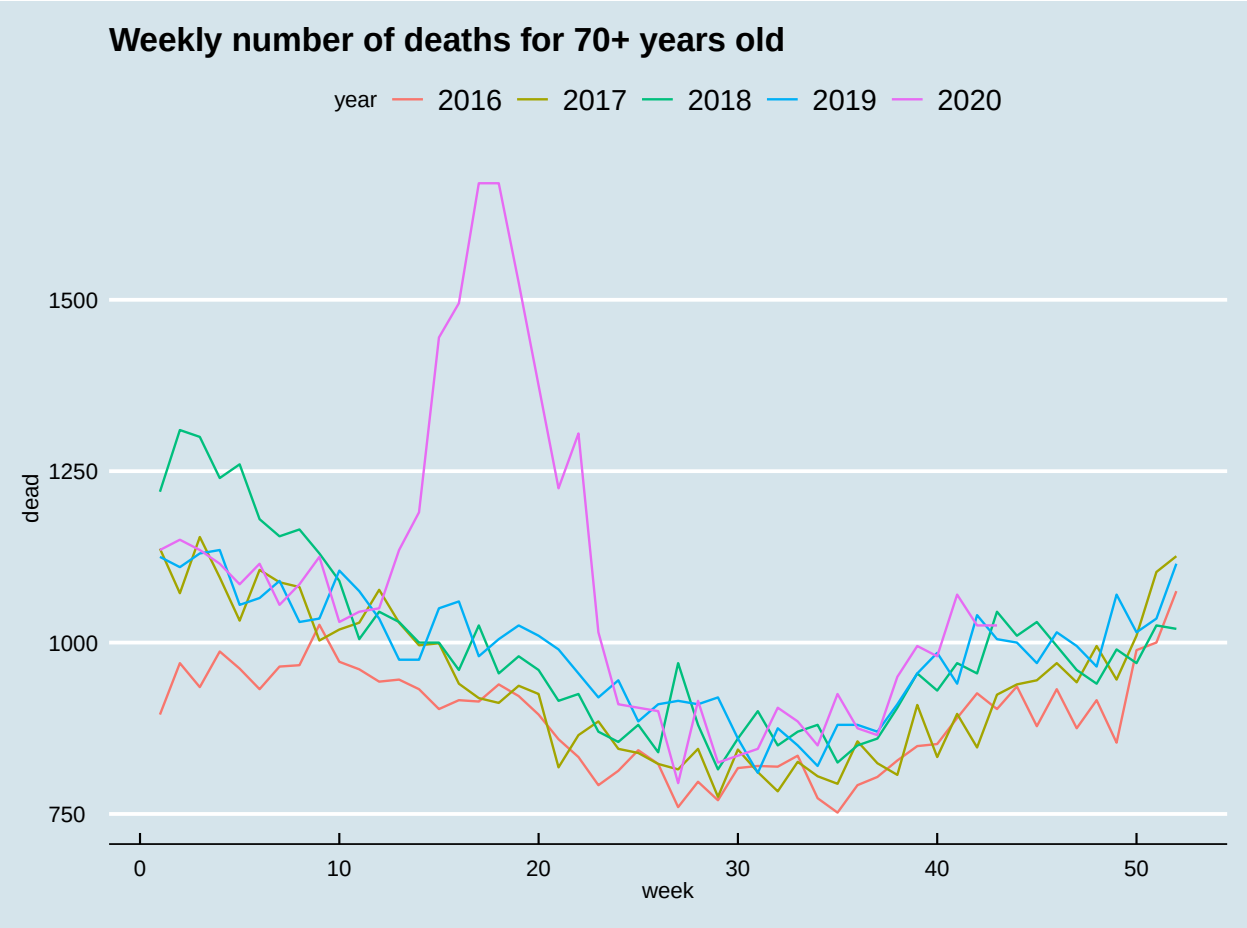

Figure 1: Weekly number of deaths for people with 70 years old and over for the years 2016-2020.

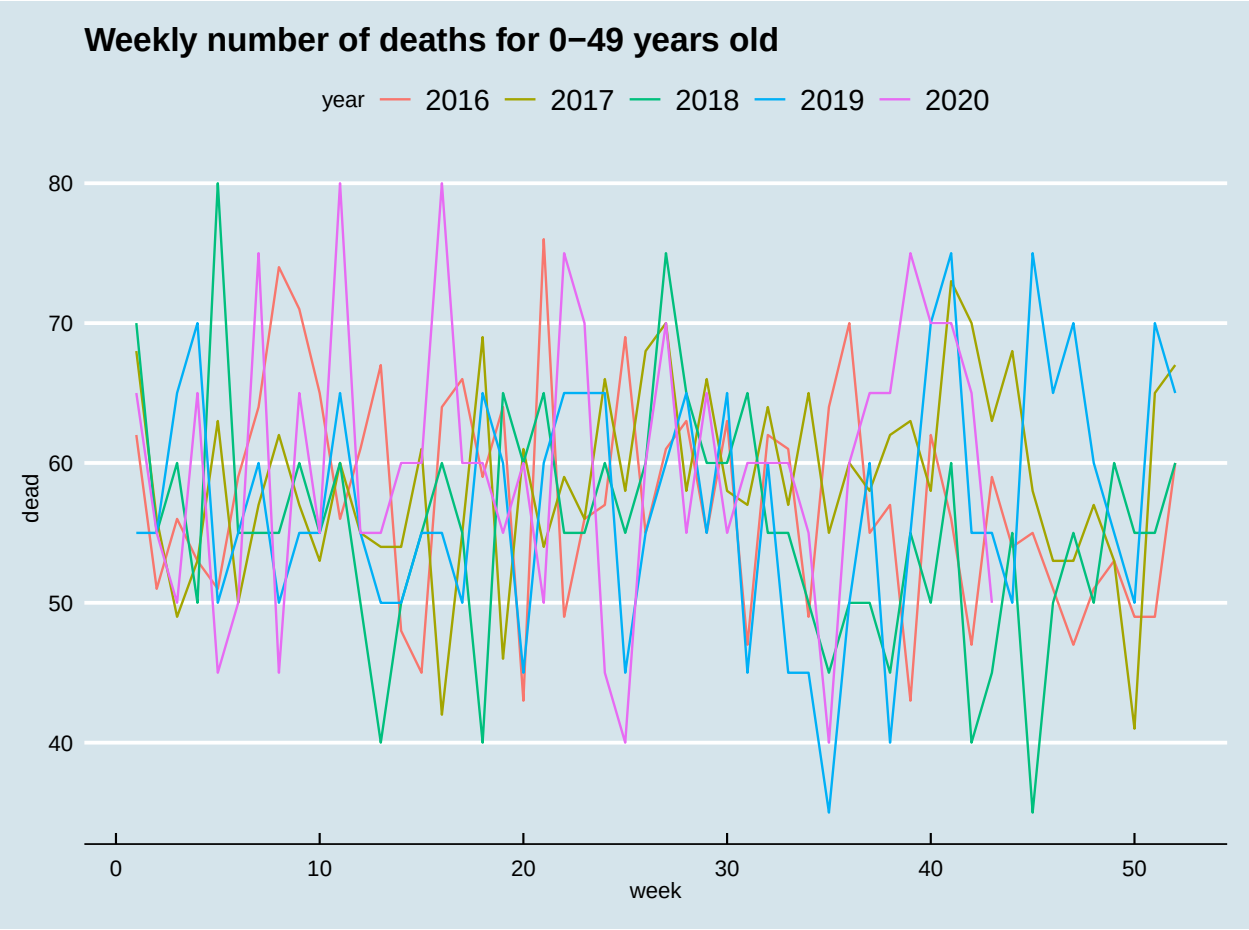

Figure 2: Weekly number of deaths for people with 0-49 years old and over for the years 2016-2020.

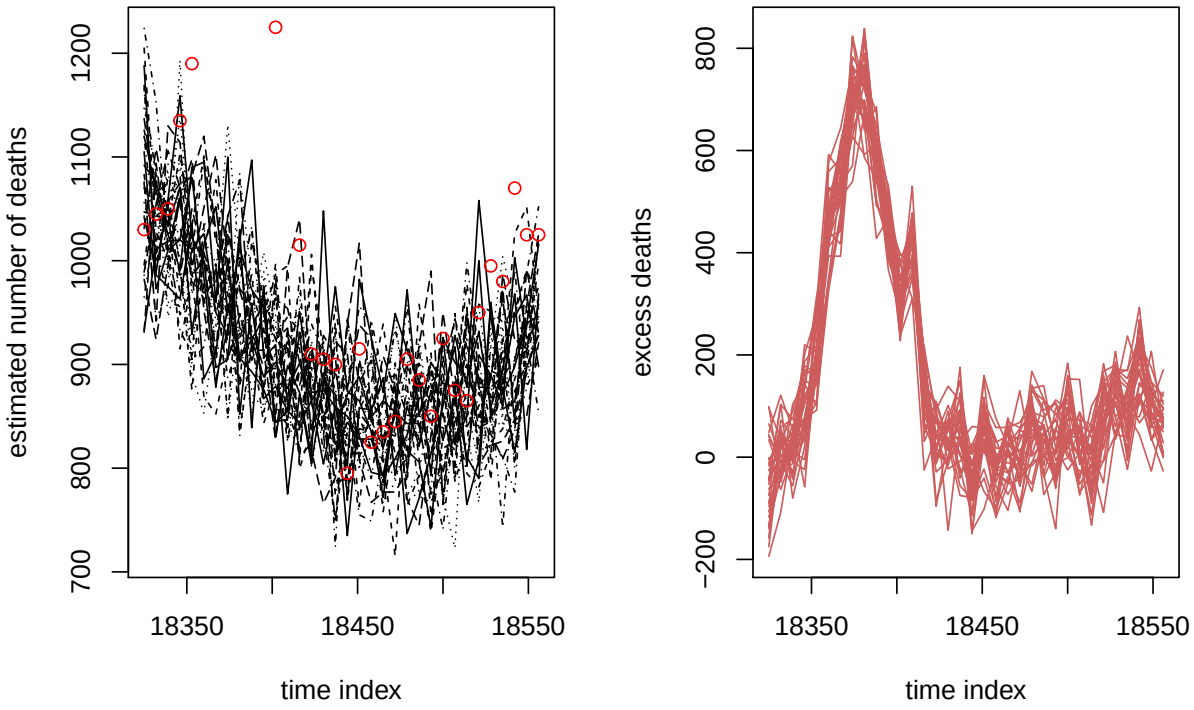

Figure 3: Forecasted number of deaths in the left pannel, related excess deaths in the right pannel, start from the year 2016 to the year 2020, weekly data for peope with 70 years old and over

Figure 3 illustrates the Forecasted number of deaths in the left pannel, related excess deaths in the right pannel, start from the year 2016 to the year 2020, weekly data for peope with 70 years old and over. We can find that the forecasted weekly number of deaths are similar with the weekly number of deaths in the past 5 years in the history and there is a high peak excess number of deaths found in the plot.

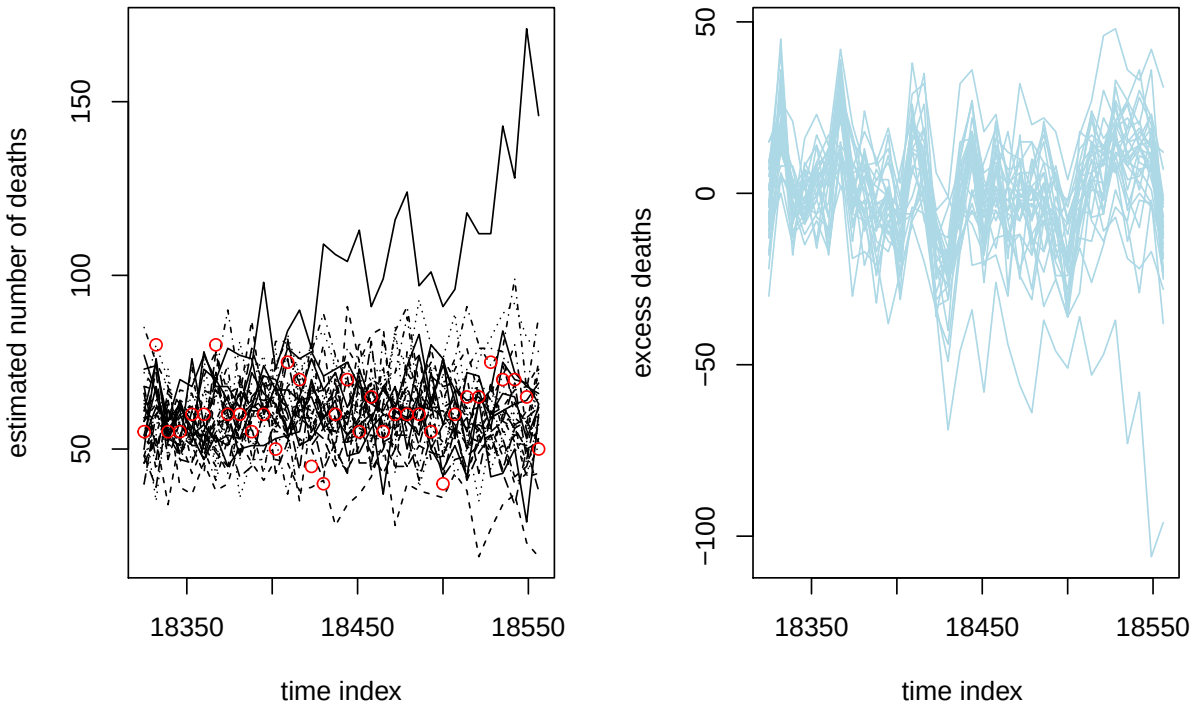

Figure 4: Forecasted number of deaths in the left pannel, related excess deaths in the right pannel, start from the year 2016 to the year 2020, weekly data for peope with 0-49 years old

Figure 4 illustrates the Forecasted number of deaths in the left pannel, related excess deaths in the right pannel, start from the year 2016 to the year 2020, weekly data for peope with 0-49 years old. We can find that the forecasted weekly number of deaths are also similar with the weekly number of deaths in the past 5 years in the history and there is no clear pattern of excess number of deaths found in the plot for this age group.

4 Discussion

First, to answer the question of interest that in the year 2020, the first wave of the COVID-19 epidemic which is mainly consider in the March, April and May mainly afffected the old people rather than younger ones. To investigate the question of interest, in figure 1 we compare the past 5 years in history with the year 2020

by the weekly number of deaths, clearly, we can find in the year 2020, there is an unusual pattern between about 12th week and 20th week which is about in the months March, April and May, it means there is a peak in this period. So that it indicates there is indeed a high number of death in this period in the year 2020 compared to the same period in the 5 years in history. And for the young people aged with 0-49 years old, figure 2 shows there is no such unusual pattern found across the weeks in the year 2020 compared with the past 5 years in the same periods, the weekly number of deaths are just mixed up for this age group. So combined with the two times series plots, it seems that in the first wave of the COVID-19 epidemic, it indeed affeced old people more than younger ones.

This is also verified by the GAM model with poisson family that we estimate the predicted weekly number of deaths, so we have both predicted weekly number of deaths for age group with 70 years old and over as well as for age group 0-49 years old, figures 3 and 4 show the predicted numbers by the models. Clearly, we can find in the left pannels that if there are no COVID-19 events, the predicted patterns of weekly number of deaths would be very similar with the past 5 years for both age groups. However, due to COVID-19, we compute the excess deaths for both age groups which can be consideres as the deaths caused by COVID-19, the results are shown in the right pannels that we found there is a very large excess deaths for old people with age 70 years old and over in the March, April and May but it is not true for younger age group 0-49 years old. Thus, in the first wave, we can conclude COVID-19 epidemic mainly affected old people.

Then, in the second wave, which began in September, as it was claimed that there are lots of young people acting irresponsibly, the young poeple should be the main groups that COVID-19 affected. However, figure 3 and figure 4 do not show patterns that there are high peaks for the young age group, it is just like the old people. This means that the young people is not appear to show more weekly number of deaths in the second wave compare with old people.

At last, besides our findings, we should also note that the data is not complete for the years 2018 to 2020, so the results might be slightly different when the whole population of the number of deaths for the years 2018 to 2020 are given. But what ever, we should not blame young people in the second wave as there is no evidence shows the young group mainly affected in this wave.

5 References

1. Data for 2020 are updated every two weeks, if the database completeness rate permits. Please note that the next release is scheduled for January 14, 2021
2. D. Kourounis, A. Fuchs, and O. Schenk, Towards the next generation of multiperiod optimal power flow solvers, IEEE Transactions on Power Systems, vol. PP, no. 99, pp. 1-10, 2018.
3. Erich Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https:

//CRAN.R-project.org/package=RColorBrewer

4. Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.

5. Hadley Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

6. Hadley Wickham, Romain Fran<U+00E7>ois, Lionel Henry and Kirill Müller (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.3. https://CRAN.R-project.org/package=dplyr

7. Jeffrey B. Arnold (2019). ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. R package version 4.2.0. https://CRAN.R-project.org/package=ggthemes

8. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.