Statistical Modeling for Relative and Excess Mortality

with Applications to COVID-19 Pandemic

Jeffrey Barlow

Relative Mortality

Under the relative mortality model, assume we are given independent and non-negative

random variables $X_1, \ldots, X_n$, with hazard rate function $\mu_i(t)\alpha(t,\theta)$, where $\mu_i(t)$ is known. Then the

intensity process is given by $\lambda_i(t,\theta) = \alpha(t,\theta)\mu_i(t)1(X_i \geq t)$, where $i = 1, \ldots, n$.

The Nelson-Aalen style estimator for the relative mortality model is given by

$$\hat{A}(t) = \int_0^t \frac{J(u)}{\sum_{i=1}^n \mu_i(u) Y_i(u)} dN(u)$$

where $Y_i(t) = 1(X_i \geq t)$ and $Y(t)$ is the sum of all Y_i's. $J(u)$ is an indicator that the denominator is

positive.

The relative mortality for a time interval $t \in [t_{m-1}, t_m)$ is estimated by

$$\hat{\alpha}(t_m) = \frac{\hat{A}(t_m) - \hat{A}(t_{m-1})}{t_m - t_{m-1}} = \frac{1}{\hat{\mu}(t_m)} \frac{N(t_m) - N(t_{m-1})}{(t_m - t_{m-1}) Y(t_{m-1})}$$

where $N(t_m) - N(t_{m-1})$ is the number of deaths in the interval $[t_{m-1}, t_m)$.

The variance of the relative mortality is estimated by

$$v\hat{a}r(\hat{\alpha}(t_m)) = \frac{1}{\hat{\mu}^2(t_m)} \frac{N(t_m) - N(t_{m-1})}{(t_m - t_{m-1})^2 Y^2(t_{m-1})} = \frac{\hat{\alpha}^2(t_m)}{N(t_m) - N(t_{m-1})}$$

which leads us to the 95% confidence intervals given by

$$\hat{\alpha}(t_m) \pm 1.96 \sqrt{v\hat{a}r(\alpha(t_m))}$$

Excess Mortality

Under the excess mortality model, assume the hazard rate function is of the form $\mu_i(t) + \alpha(t,\theta)$, where $\mu_i(t)$ is known. Then the intensity process is given by $\lambda_i(t,\theta) = (\alpha(t,\theta) + \mu_i(t))1(X_i \geq t)$. The Nelson-Aalen style estimator for the excess hazard model is given by

$$\hat{A}(t) = \int_0^t \frac{J(u)}{\sum_{i=1}^n Y_i(u)} dN(u) - \int_0^t \sum_{i=1}^n \mu_i(u) \frac{Y_i(u)}{Y(u)} du$$

Note that this is the simply the ordinary Nelson-Aalen estimator minus an integrated average population hazard.

The excess mortality for a time interval $t \in [t_{m-1}, t_m)$ is estimated by

$$\hat{\alpha}(t_m) = \frac{\hat{A}(t_m) - \hat{A}(t_{m-1})}{t_m - t_{m-1}} = \frac{N(t_m) - N(t_{m-1})}{(t_m - t_{m-1}) Y(t_{m-1})} - \hat{\mu}(t_m)$$

The variance of the relative mortality is estimated by

$$\hat{var}(\hat{\alpha}(t_m)) = \frac{N(t_m) - N(t_{m-1})}{(t_m - t_{m-1})^2 Y^2(t_{m-1})}$$

which leads us to the 95% confidence intervals given by

$$\hat{\alpha}(t_m) \pm 1.96 \sqrt{\hat{var}(\alpha(t_m))}$$

Objectives

The objective is to determine how the chances of dying in a given week changed as a result of the coronavirus pandemic for different age groups and different ethnicities. We will use the models for relative mortality and excess mortality.

About the Data

The data for this project came from the CDC. There were two data sets: one listed the number of deaths per week from each of six age groups from 2015 to 2021; the other data set

listed the number of deaths per week for six ethnicities over the same time period. Data from recent weeks was noted as incomplete, since many jurisdictions are slow to report the number of deaths to the US Department of Health. However, the CDC has provided estimates in lieu of actual numbers. The data was downloaded on 5 April 2021 and can be found at this address: https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm.

The age groups given in the CDC data are under 25 years, 25-44 years, 45-64 years, 65-74 years, 75-84 years, and 85 years and older. The ethnicities in the CDC data are Hispanic, Non-Hispanic American Indian or Alaska Native, Non-Hispanic Asian, Non-Hispanic Black, Non-Hispanic White, and Other.

Explanation of Calculations

I compared the weekly number of deaths from March 2019 to March 2020 and from March 2020 to March 2021 for each of the age groups and ethnicities. I estimated the number of people in these groups using population data from the US Census population clock (https://www.census.gov/popclock/). I recorded the total number of people in each group at the beginning of the March 2019 and March 2020. I then subtracted the number of deaths for each week in order to get an estimate for the number of people in that group at each week of that time period. I divided the number of deaths for each week by the number of people in that group on that week to get a weekly death rate.

I compared the weekly death rates for these groups in two ways. First, I divided the weekly death rates during the pandemic by the weekly death rates from before in order to estimate the relative mortality. Second, I subtracted the weekly death rates for before and during the pandemic to get an estimate for the excess death rate. The following charts show the results

of these calculations. To estimate the cumulative relative and excess mortality, I took the integral.
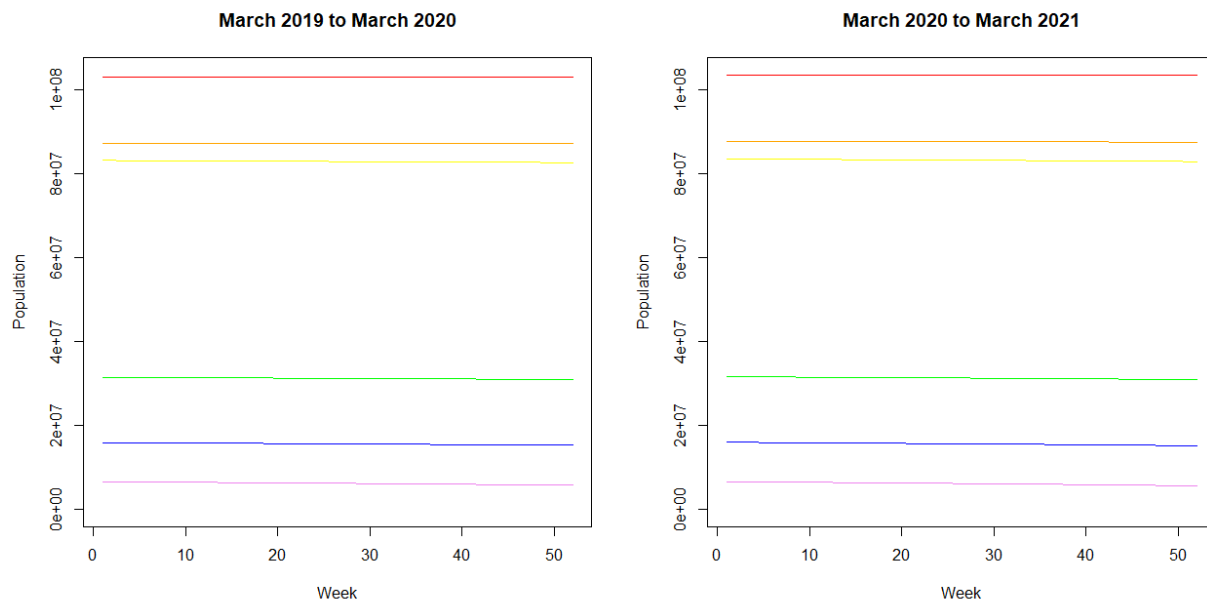
Charts

This first chart shows the number of deaths for each age group for one year periods before and during the pandemic:
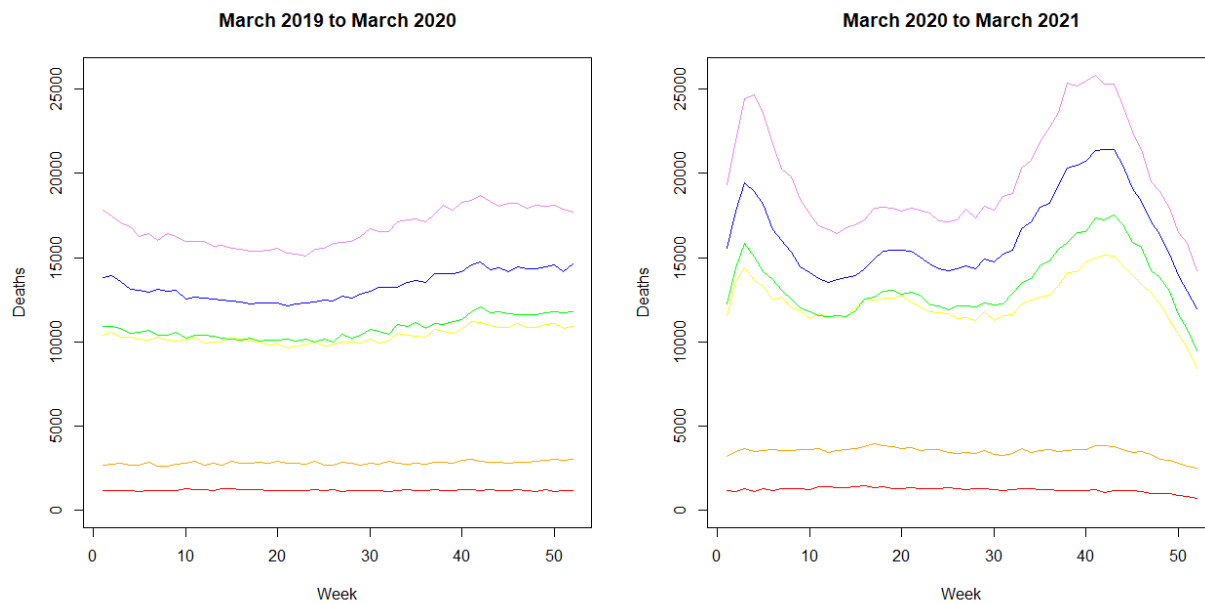
Color code:

- Under 25 years
- 25-44 years
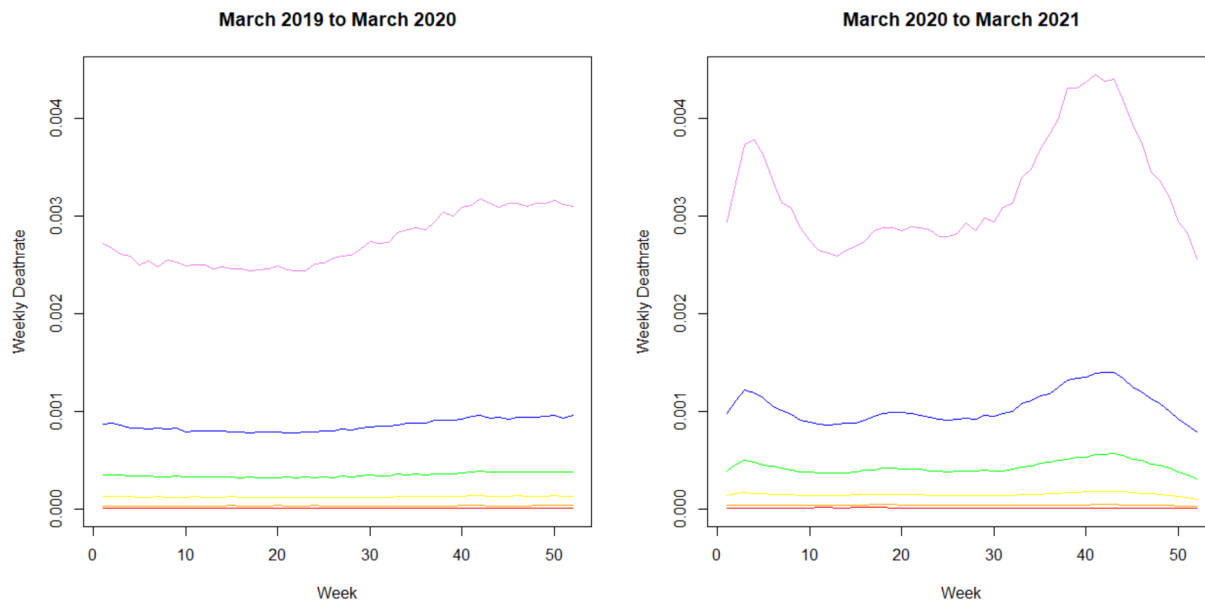- 45-64 years

- 65-74 years
- 75-84 years
- 85 years and older

This next chart shows how the population of the six age groups changed over time:

March 2019 to March 2020

March 2020 to March 2021

Color code:

- Under 25 years
- 25-44 years
- 45-64 years

- 65-74 years
- 75-84 years
- 85 years and older

This next chart shows the weekly death rate for the same time periods.

March 2019 to March 2020

March 2020 to March 2021

Color code:

- Under 25 years
- 25-44 years
- 45-64 years

- 65-74 years
- 75-84 years
- 85 years and older

These next six charts show the estimates for relative mortality for each age group. The confidence intervals are shown in pink.


7
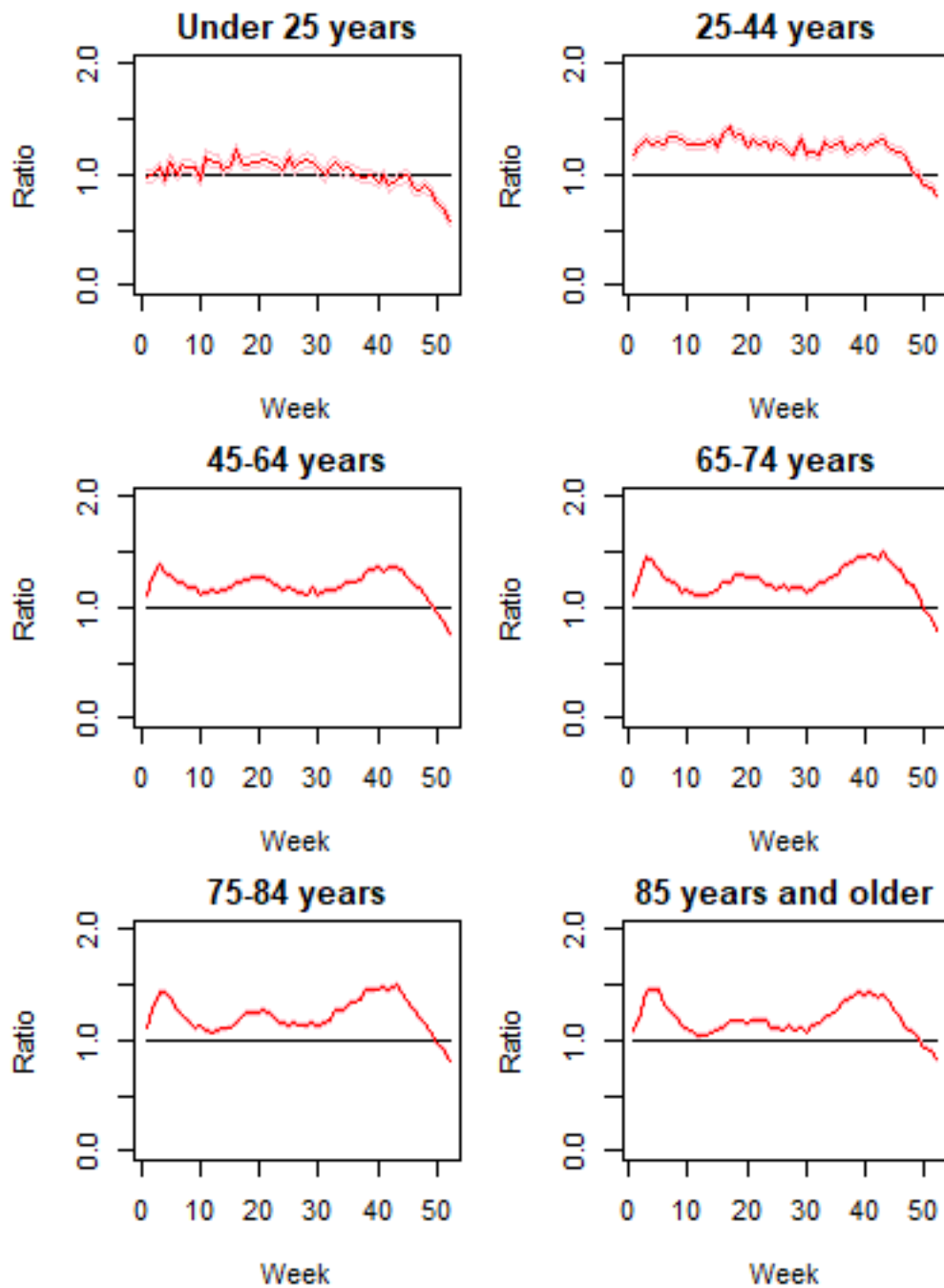
These next six charts show the estimates for cumulative relative mortality for each age group. This was calculated by taking the integral of the above figures.
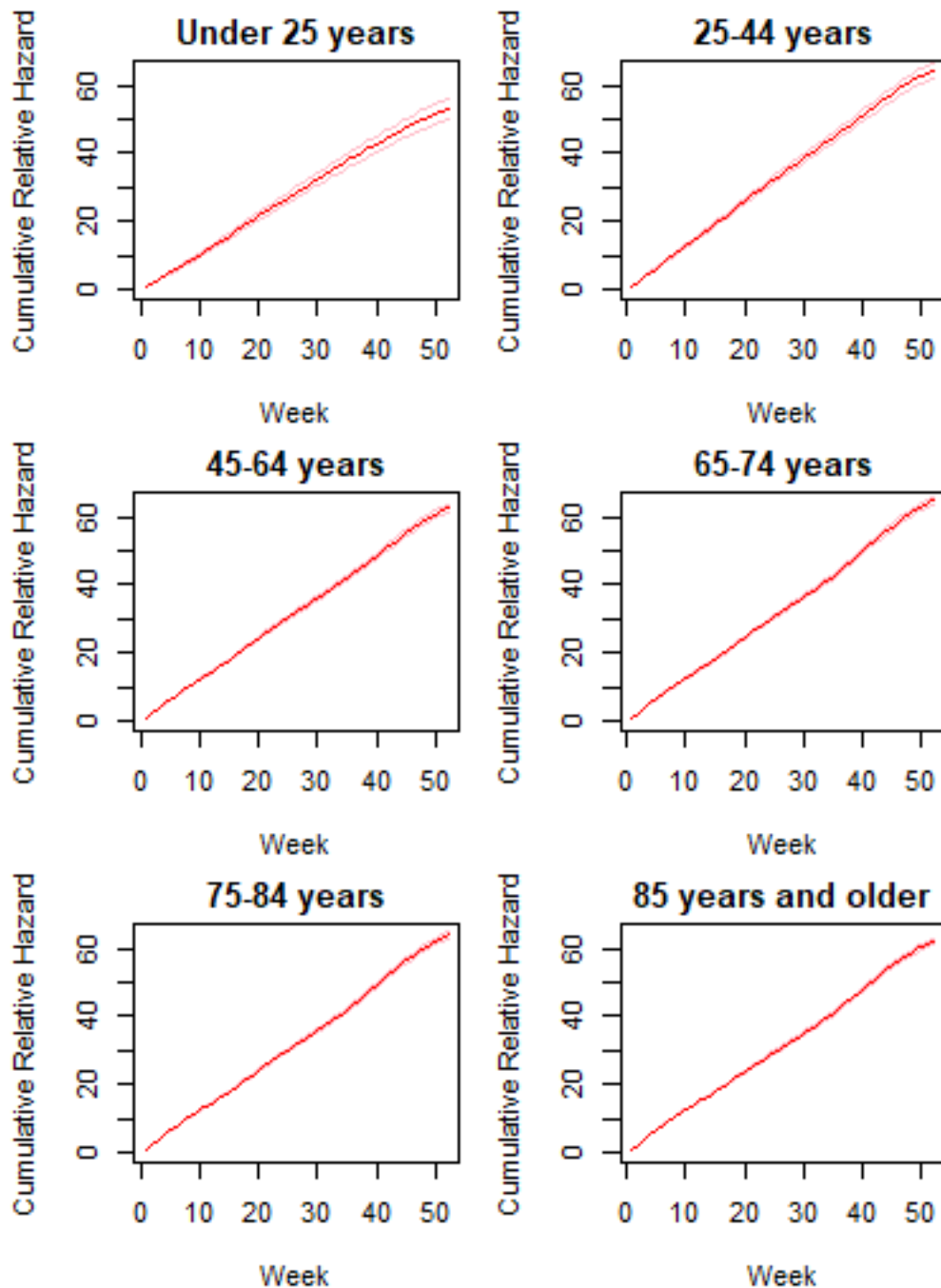
These next six charts show the estimates for excess mortality for each age group. The confidence intervals are shown in pink.
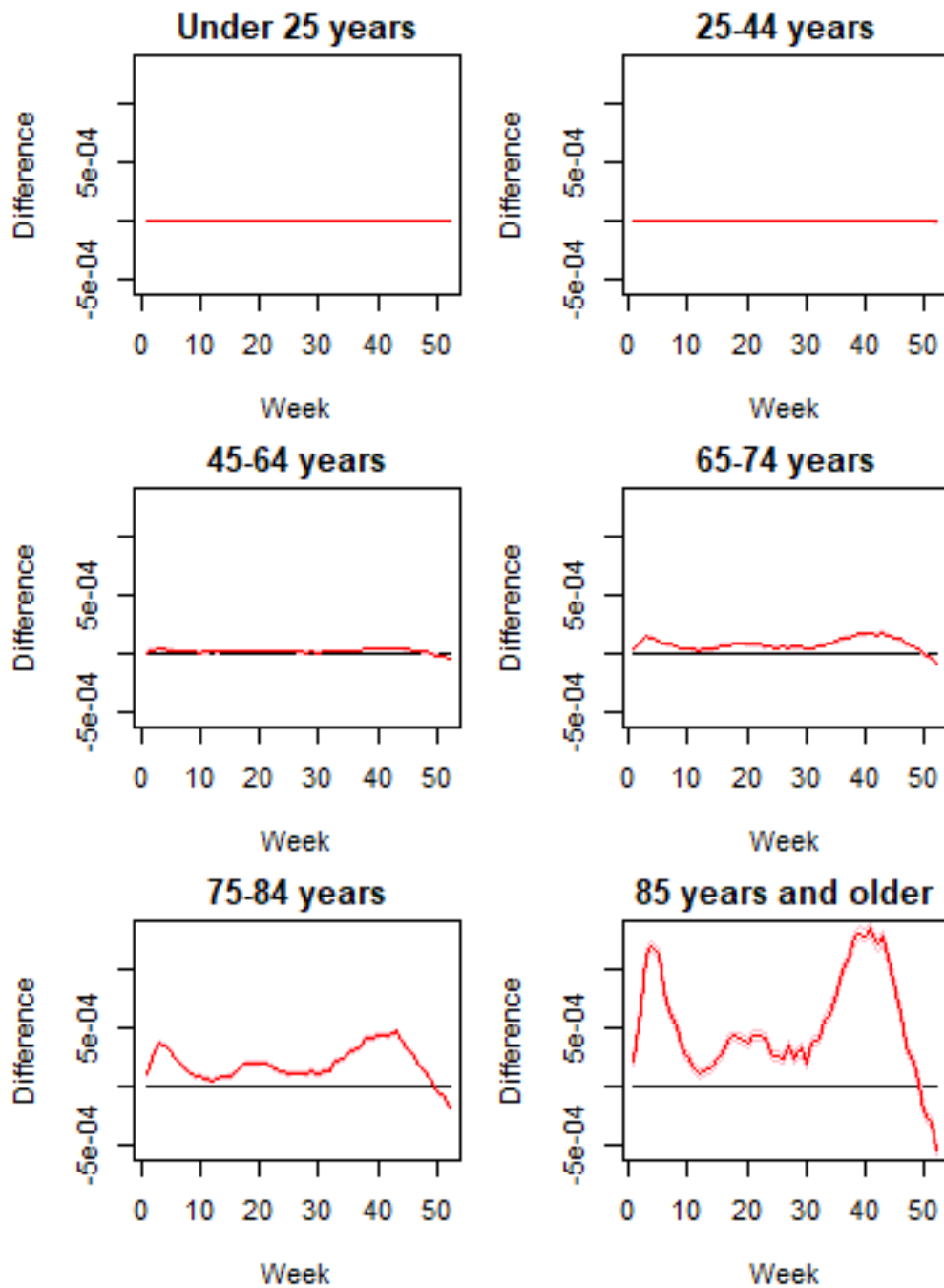
These next six charts show the estimates for cumulative excess mortality for each age group. This was calculated by taking the integral of the above figures.
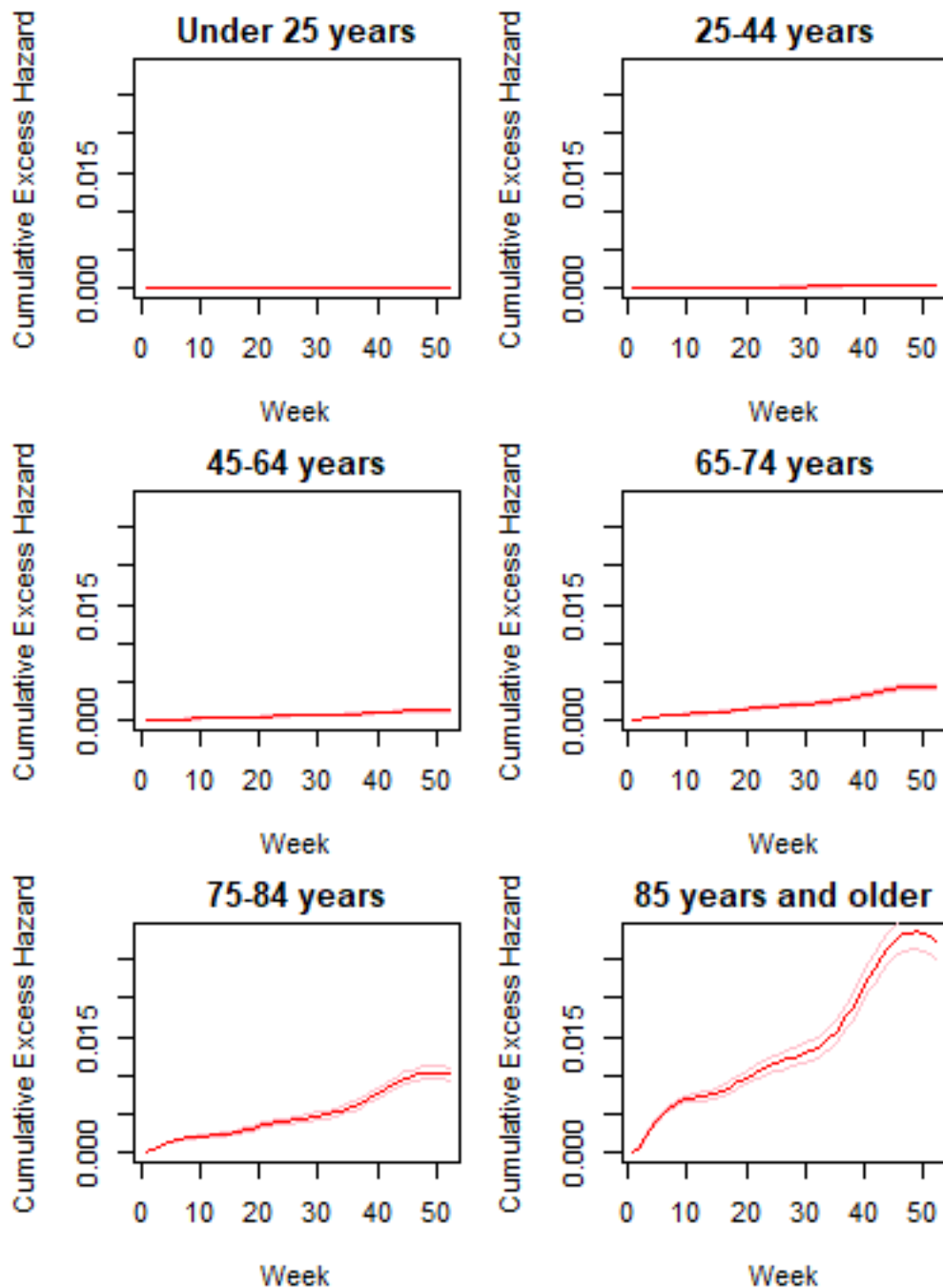
The same process was applied to the race and ethnicity data set. This first chart shows the number of deaths per week for each racial group before and during the pandemic.
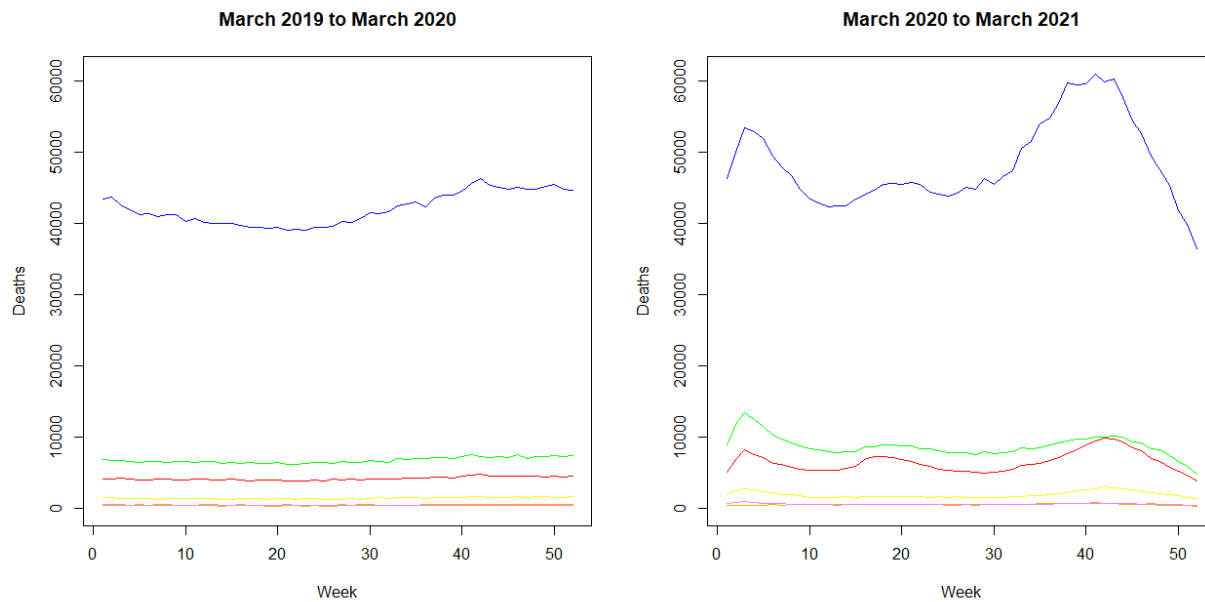
March 2019 to March 2020

March 2020 to March 2021

Color code:

- Hispanic
- Non-Hispanic American Indian or Alaska Native
- Non-Hispanic Asian

- Non-Hispanic Black
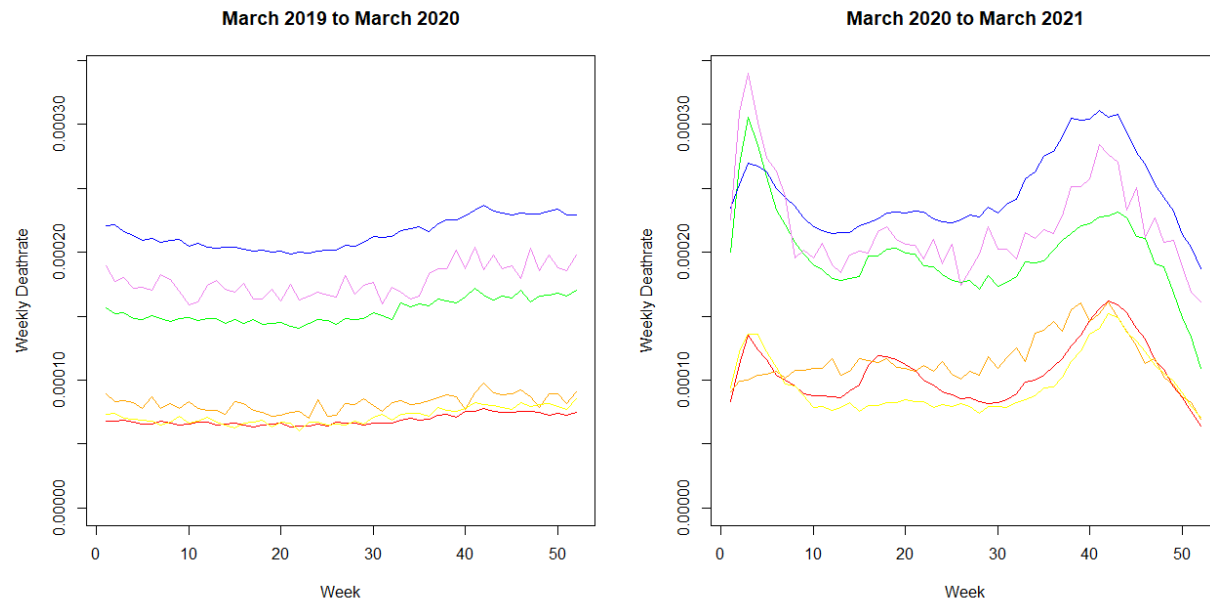- Non-Hispanic White
- Other

11

This next chart shows how the populations for the six racial groups changed over time.
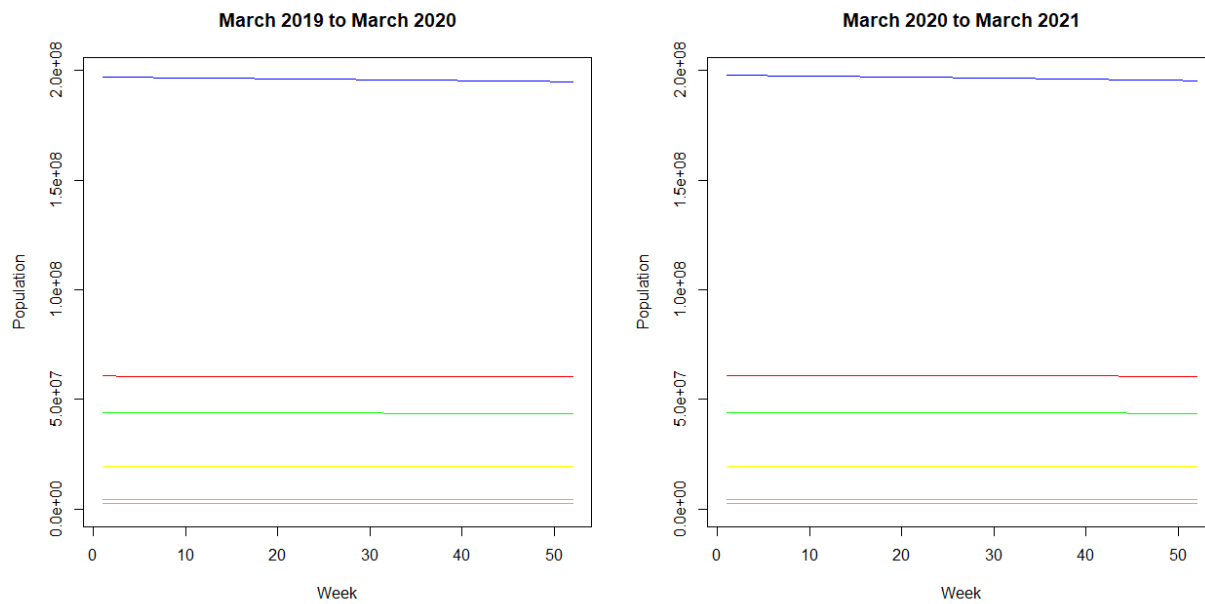
Color code:

- Hispanic
- Non-Hispanic Black
- Non-Hispanic American Indian or Alaska Native
- Non-Hispanic White
- Other
- Non-Hispanic Asian

This next chart shows the weekly death rate for the same time periods.

March 2019 to March 2020

March 2020 to March 2021

Color code:

- Hispanic

- Non-Hispanic American Indian or Alaska Native

- Non-Hispanic Asian

- Non-Hispanic Black

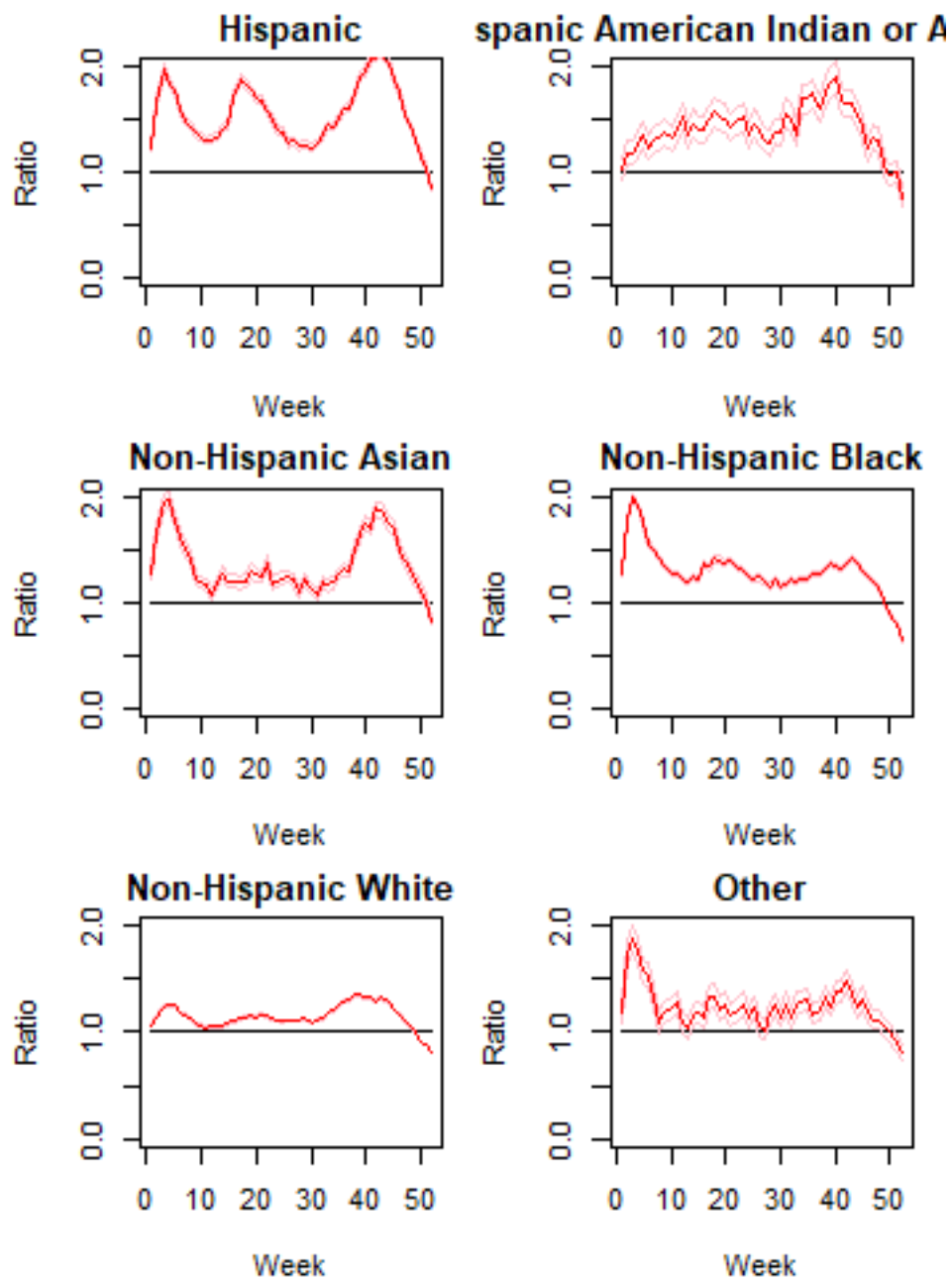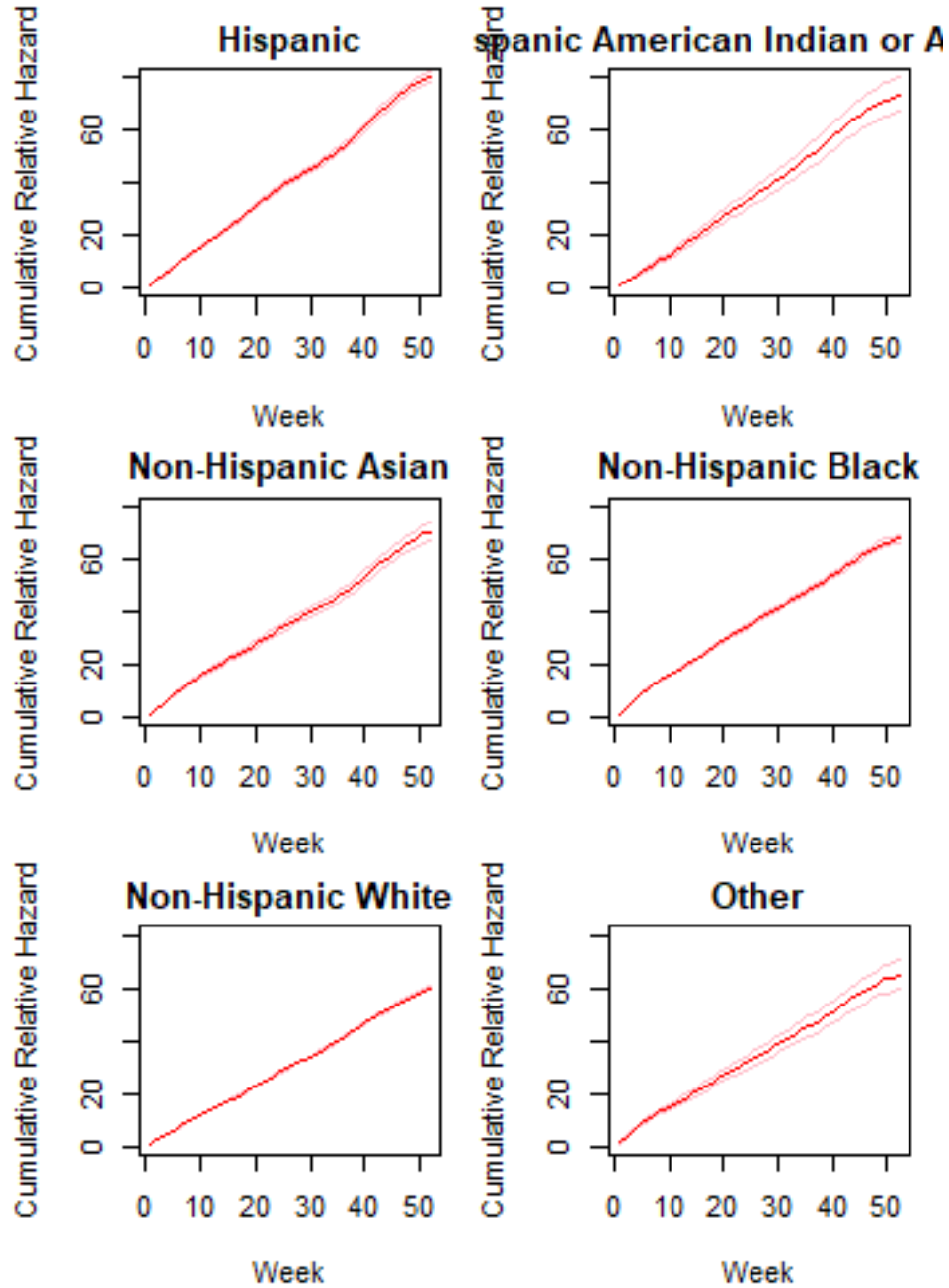- Non-Hispanic White
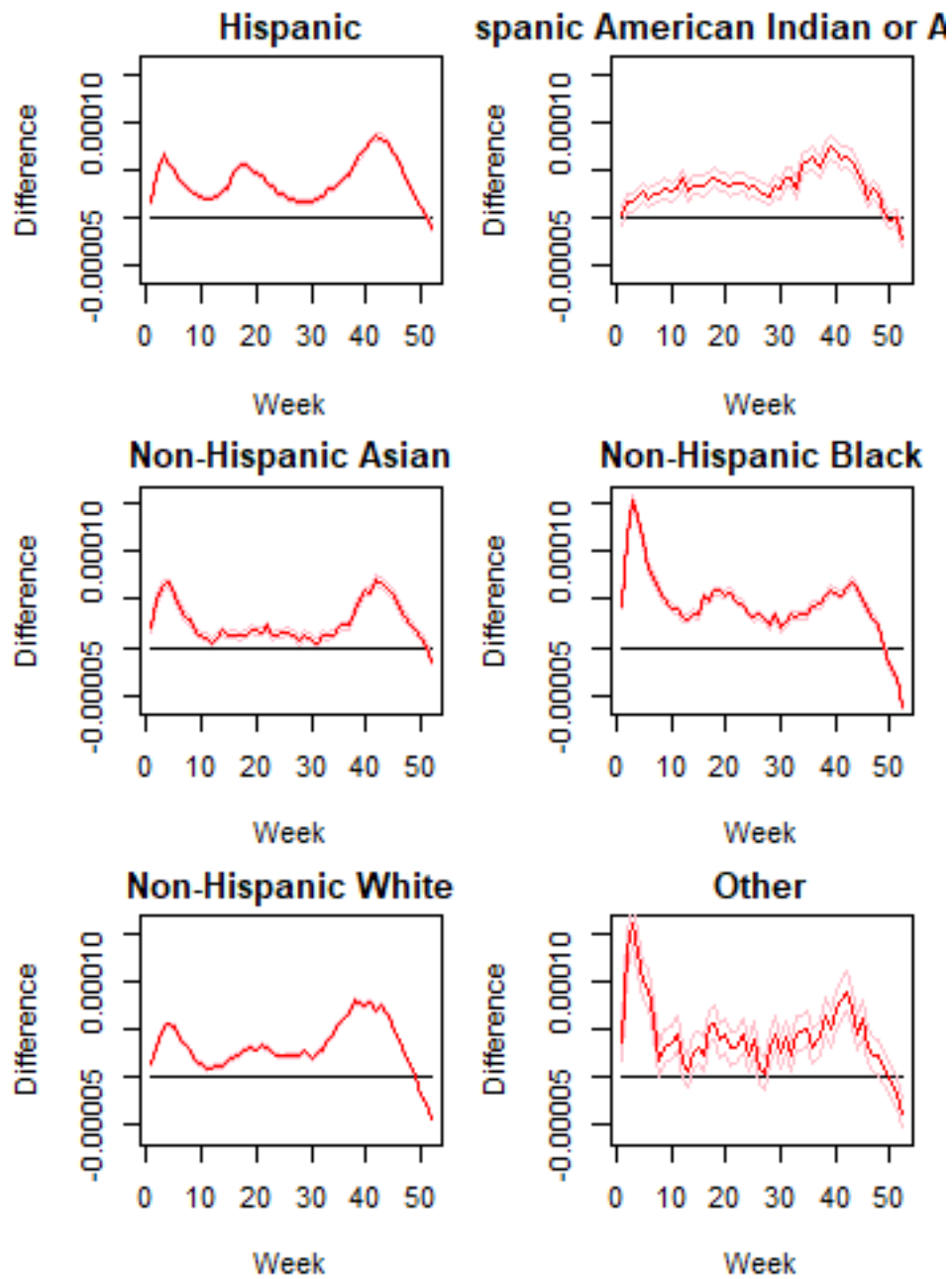
- Other
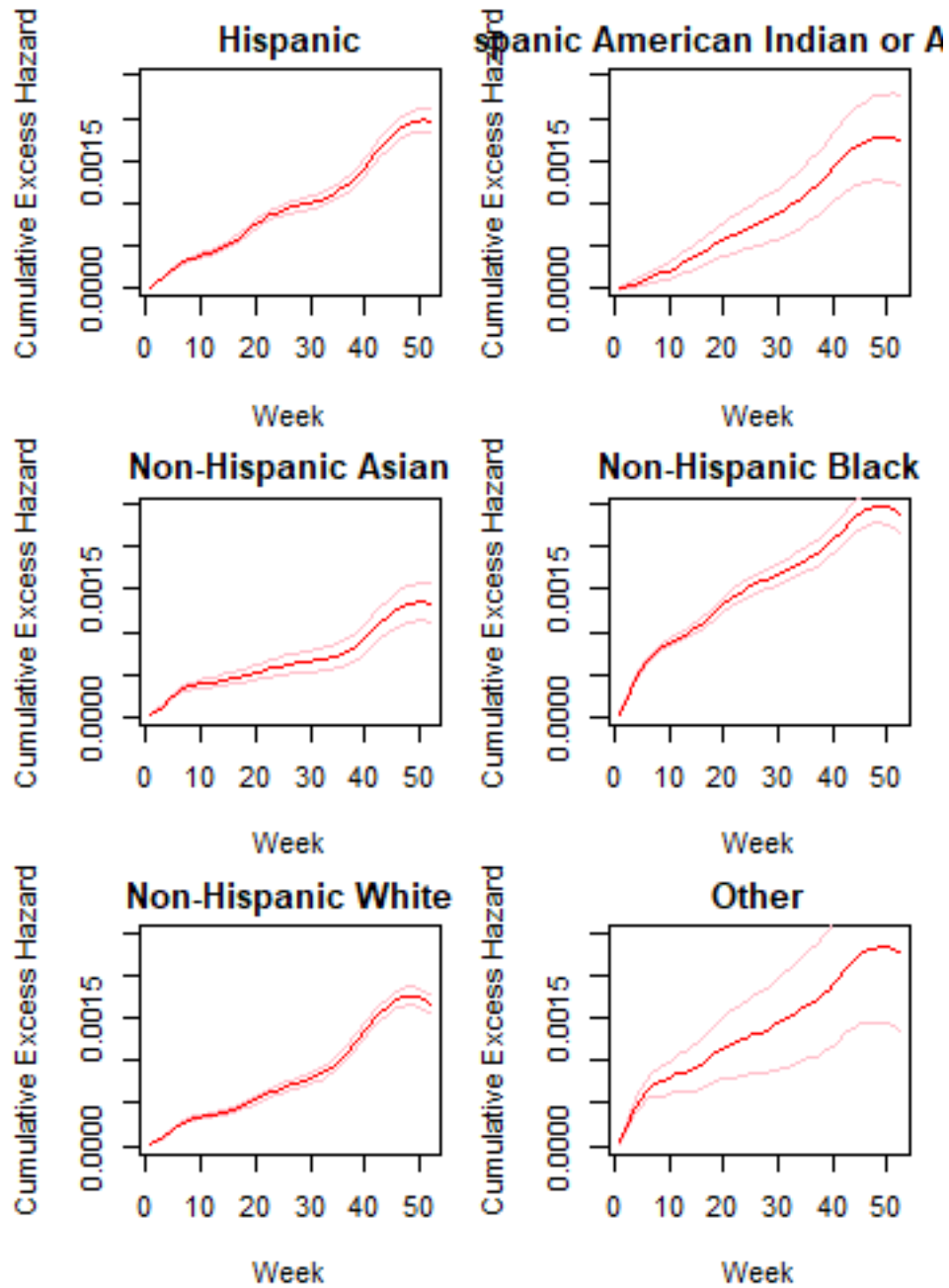
The relative mortality for the six racial groups

The cumulative relative mortality for the six racial groups

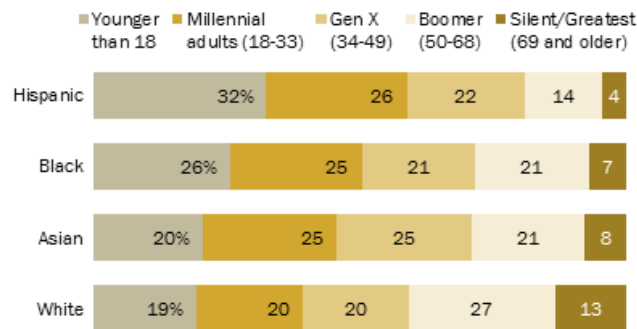The estimated excess hazard for the six racial groups

16

The cumulative excess hazard for the six racial groups

The age distribution among different ethnicities is not the same. The Hispanic population is overall much younger than other demographics, which may explain why the relative mortality for this population is so high. The following image from Pew Research shows the age distributions for different races in 2014.

Nearly six-in-ten Hispanics are Millennials or younger

	Younger than 18	Millennial adults (18-33)	Gen X (34-49)	Boomer (50-68)	Silent/Greatest (69 and older)
Hispanic	32%	26	22	14	4
Black	26%	25	21	21	7
Asian	20%	25	25	21	8
White	19%	20	20	27	13

Note: Whites, blacks and Asians include only those who are single race and not Hispanic. Hispanics are of any race. Figures may not add to 100% due to rounding.
Source: Pew Research Center analysis of 2014 American Community Survey (IPUMS).
"The Nation's Latino Population Is Defined by Its Youth"

PEW RESEARCH CENTER

https://www.pewresearch.org/hispanic/2016/04/20/the-nations-latino-population-is-defined-by-its-youth/

Conclusion

The estimates for relative and excess mortality allow us to see the extent to which different age groups were affected by the coronavirus pandemic from March 2020 to March 2021. People aged 25 years and under experienced almost no change, while the death rates for people over 65 were significantly higher. The death rates for all racial and ethnic groups saw increases, although non-Hispanic blacks were noticeably more affected.

We can very clearly see the three phases of the pandemic in the spring, summer, and winter. The death rate shows a clear decline towards the end of March 2021. The death rate is actually lower than baseline now. The vaccine is certainly playing a role, but I speculate that this decline is also attributable to the fact that we lost so many people over 85 in 2020.

References

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1992). Statistical Models Based on

Counting processes. Springer-Verlag, New York.