

Final Report

due November 16, 2021 by 11:59 PM

Harris Upchurch, Biniam Garomsa, Philemon Hailemariam

11/16/2021

Introduction

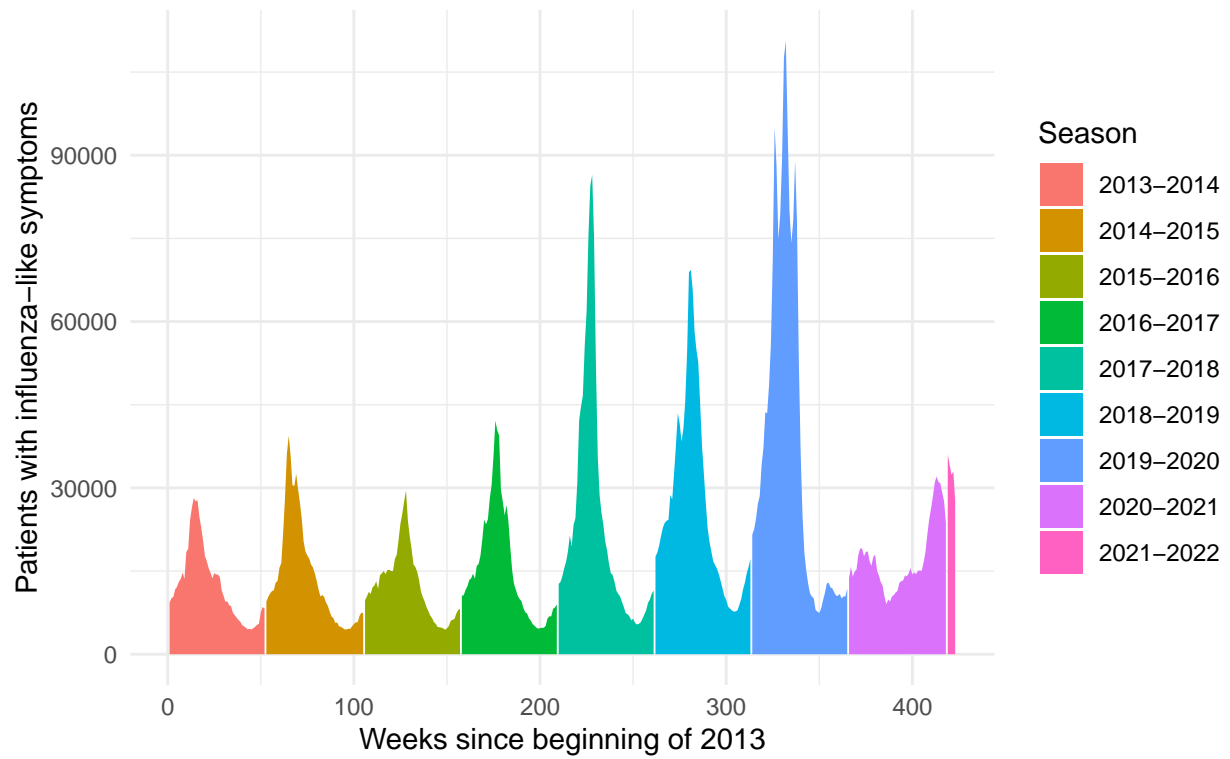
Starting in 2020, public health measures against the spread of respiratory disease were taken in an effort to slow or prevent the spread of COVID-19. These measures were much more significant than in previous years, as the primary strategies against the spread of respiratory disease in the United States are typically vaccination and encouraging frequent hand-washing, but measures against COVID-19 included shutting down large parts of society and moving online, as well as enforcement or encouragement of face coverings and vaccinations when available. This report studies how these public health measures may have affected the spread of other infectious disease by focusing on influenza hospitalizations and deaths in the United States, as the flu is one of the most common and deadly diseases each year in the U.S., and shares many characteristics with COVID-19.

Flu seasons during the COVID-19 Pandemic vs. past flu seasons

Have flu deaths and hospitalizations changed during the COVID-19 pandemic compared to past years? If so, is this difference statistically significant?

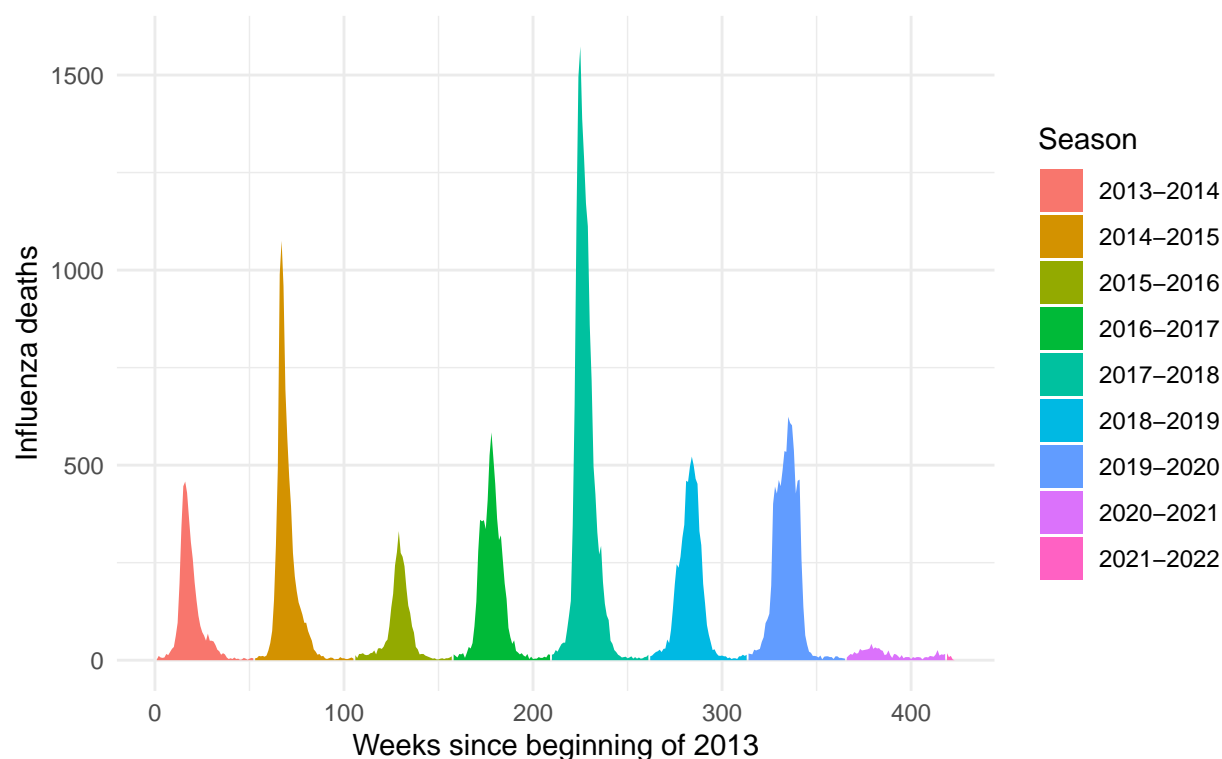
U.S. Influenza Patients

CDC FLuView, 2013–2021



U.S. Influenza Deaths

CDC FLuView, 2013–2021

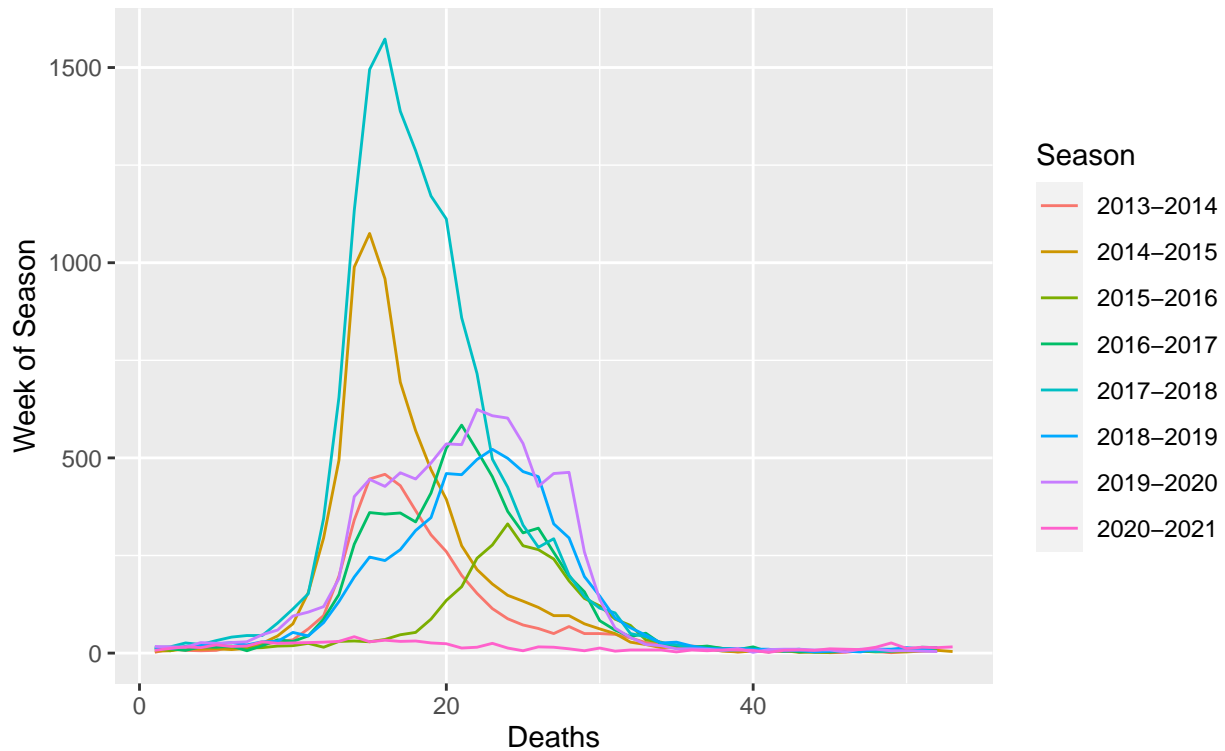


```
## # A tibble: 9 x 3
##   season    total_flu_deaths total_flu_patients
##   <chr>          <dbl>          <dbl>
## 1 2013–2014         4255          649058
## 2 2014–2015         7921          751851
## 3 2015–2016         3237          644170
## 4 2016–2017         6672          802314
## 5 2017–2018        14968         1215966
## 6 2018–2019         6835         1384249
## 7 2019–2020         8914         1976525
## 8 2020–2021          839          917038
## 9 2021–2022          39          163186
```

Based on the data, the hospitalizations for the 2020-2021 flu season were similar to several seasons in past years, but noticeably lower than the most recent 2017-2019 seasons, with a total of 917,038 hospitalizations, which was somewhat higher than the 2013-2017 seasons, but about half of the most recent season. The 2021-2022 season is not complete, but appears to be on track to be similar to years before the 2017-2019 period, like the 2020-2021 season. The deaths of both the 2020-2021 season and the 2021-2022 season were and have been noticeably less than any of the previous years used for this comparison. The 2020-2021 season had 25.9% of the deaths of the next least deadly year and 11.1% of the average deaths (7,543.14) of previous seasons, and the 2021-2022 season is on track to follow a similar pattern. These findings indicate some noticeable differences between influenza effects before and during COVID-19, but statistical tests should be done to prove and quantify their difference. Due to the variance between weeks within season, a test should be used which tests for significance either across seasons in total or mean or within analogous weeks across seasons. However, because the peak and distribution within each season is not consistent, comparing along weekly lines causes some issues. This significantly reduces the sample size that may be used within a test from approximately 400 weeks to about 9 seasons. One potential way to adjust for the seasonal patterns is

to select the most impactful weeks from each season, so that there are more data points for each category than only using seasons, but the range of values does not include many weeks where the flu is not present. Additionally, the 2021-2022 season must be adjusted to be able to be fairly compared and used for statistical significance. The mean will not allow comparison on its own, due to the peak pattern of each season, so the best way to compare 2021-2022 to other seasons is most likely to cut off each season to the first few weeks, which would further reduce sample size. The statistical test that is used will test for how the binary variable for whether or not the COVID-19 pandemic is occurring explains the observations for the numerical variables of influenza hospitalizations and deaths.

Seasonal Influenza Deaths CDC, 2013–2021



In the line graph above, some of the issues posed by the data in finding statistical significance can be observed. Peaks and patterns do not line up consistently across years and the current definition of seasons includes many weeks that are not actually affected by the flu. Restricting the definition of a season by only taking the top weeks for deaths or hospitalization can allow for more effective significance testing. This figure also shows that the distribution of the data across weeks in a season is approximately normal, which means that it may be used for modelling.

```
## # A tibble: 2 x 7
##   term      estimate std.error statistic  p.value conf.low conf.high
##   <chr>      <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)    528.      36.2     14.6 3.34e-24    456.     600.
## 2 covidstatus  -498.     94.7     -5.26 1.19e- 6   -686.    -309.

## # A tibble: 2 x 7
##   term      estimate std.error statistic  p.value conf.low conf.high
##   <chr>      <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)  42691.    2619.     16.3 3.77e-27  37480.   47903.
## 2 covidstatus -25191.    6846.     -3.68 4.22e- 4 -38814. -11568.
```

Based on the model, there is a statistically significant difference between the number of deaths at the peak 10 weeks of each season when comparing weeks in non-COVID times with weeks in COVID-affected times. The p value for covidstatus is 1.19×10^{-6} , which is less than .05 and the confidence interval of -686.46 to -309.41 does not include the null hypothesis of 0, which would indicate no difference in the data. The model indicates that moving from non-COVID to COVID time correlates with a decrease in deaths of 497.94 from the estimate of 527.69. Based on this data, something that has happened during the pandemic has most likely caused a decrease in deaths due to the flu. The model also indicates that there was a statistically significant decrease in hospitalizations when compared to the norm, by 25,190.74, from an estimate of 42,691.49, with a p value of 4.22×10^{-4} , which is less than 0.05, and a confidence interval of -38,813.74 to -11,567.73, which does not include the null hypothesis of 0. This is less than the difference in deaths, in terms of significance, but it is interesting because the hospitalizations were similar to past hospitalization levels, but the model still determine that its difference was significance due to the upward trend in hospitalization in recent years. The data for deaths was not similar to anything recorded in the past for the time period tested.

```
##
## Welch Two Sample t-test
##
## data: covid_flu_table$total_flu_deaths by covid_flu_table$covidstatus
## t = 12.754, df = 69.148, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## 420.0507 575.8207
## sample estimates:
## mean in group 0 mean in group 1
## 527.6857 29.7500

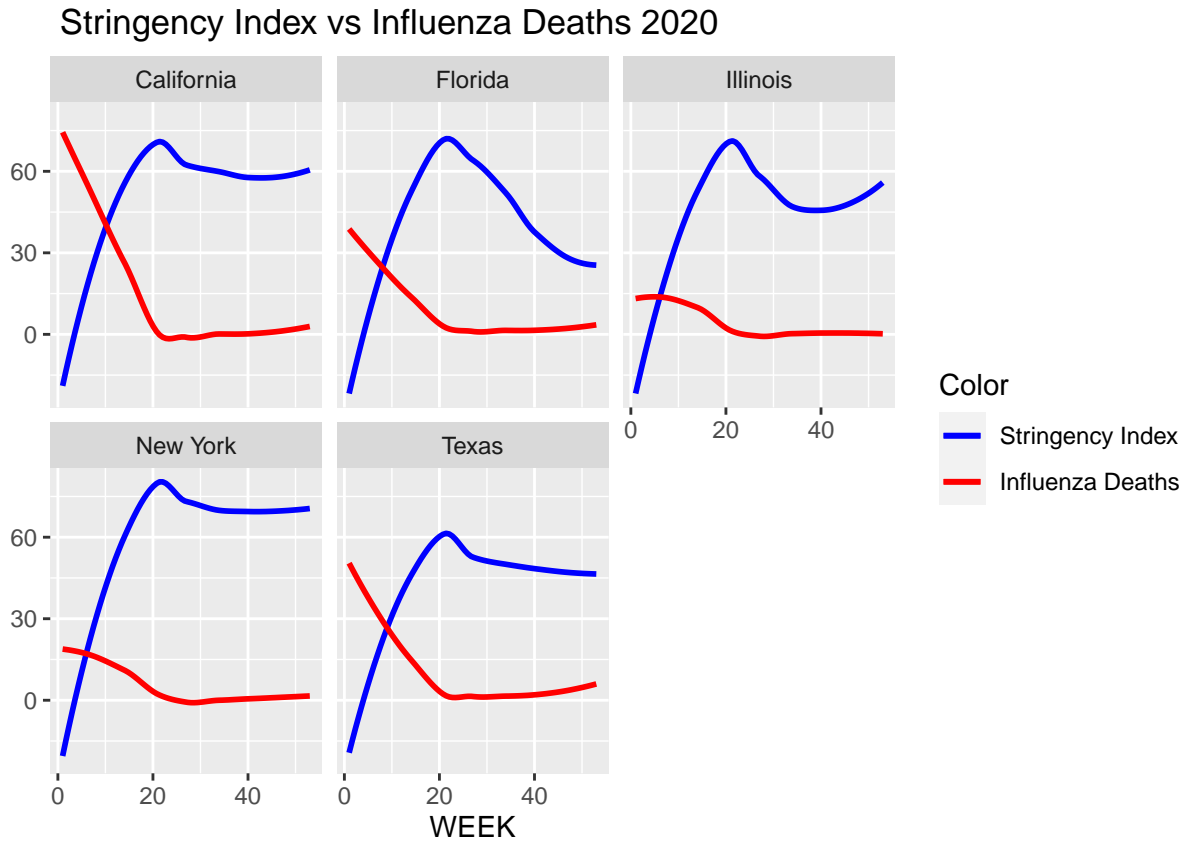
##
## Welch Two Sample t-test
##
## data: covid_flu_table$total_flu_patients by covid_flu_table$covidstatus
## t = 8.0363, df = 77.744, p-value = 8.165e-12
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## 18949.89 31431.58
## sample estimates:
## mean in group 0 mean in group 1
## 42691.49 17500.75
```

These t-tests confirm the results of the models that the impact of the flu has been different during the pandemic as compared to before. The data was already confirmed to be approximately normal, so the t-test compares the distributions of data before and during the pandemic. The means are clearly much less for both deaths and hospitalizations during the pandemic, although the death difference is greater. The p-values of 2.2×10^{-16} and 8.17×10^{-12} are both less than .05, and the confidence intervals of 420-576 and 18950-31432 do not include the null hypothesis of 0, confirming the results of the model that there is a statistically significant difference during the pandemic in the impact of the flu.

Our hypothesis is that the change in number of deaths and hospitalizations due to influenza is highly related to the restrictions enforced as a result of the COVID-19 pandemic. In order to examine this hypothesis, a variable that quantitatively describes the restriction policies imposed during the pandemic was used from the Oxford COVID-19 policy tracker data. This variable is called **Stringency Index** and is calculated as a mean of sub-indices, which describe restriction policies such as school closures, work space closing, cancellation of public events, public and international travel controls and mask and vaccination policy. These sub-indices paired with additional flags corresponding to whether the policy has been applied locally, in specific areas/circumstances, or generally, nationwide are used in calculating the **Stringency Index** ("COVID-19 Government Response Tracker" 2021). Thus, the **Stringency Index** is a good measure for the number and strictness of government policies in response to COVID-19 pandemic.

This **Stringency Index** was calculated on a daily basis. The weekly average indices were used for merging data with the influenza records. In order to visualize and better understand the relationship between the stringency index and influenza infections, we selected the five US states with the most COVID-19 cases recorded.

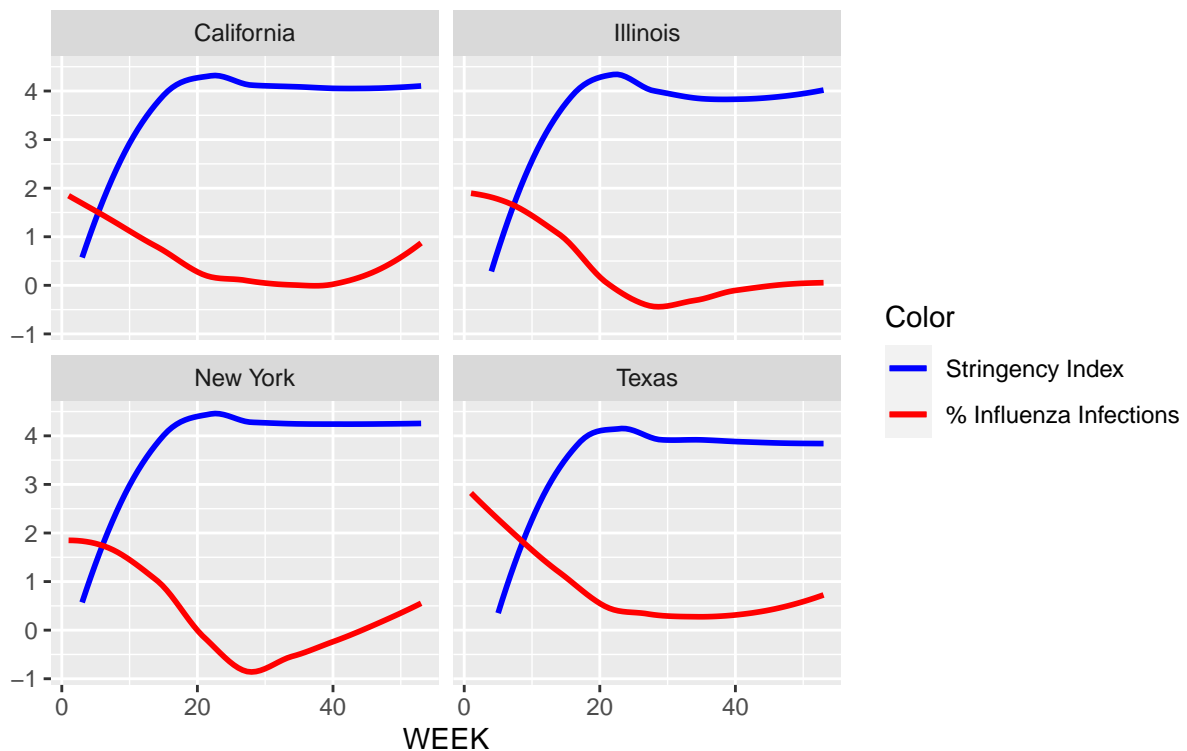
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Percentage of Influenza Infections in Total Patients 2020

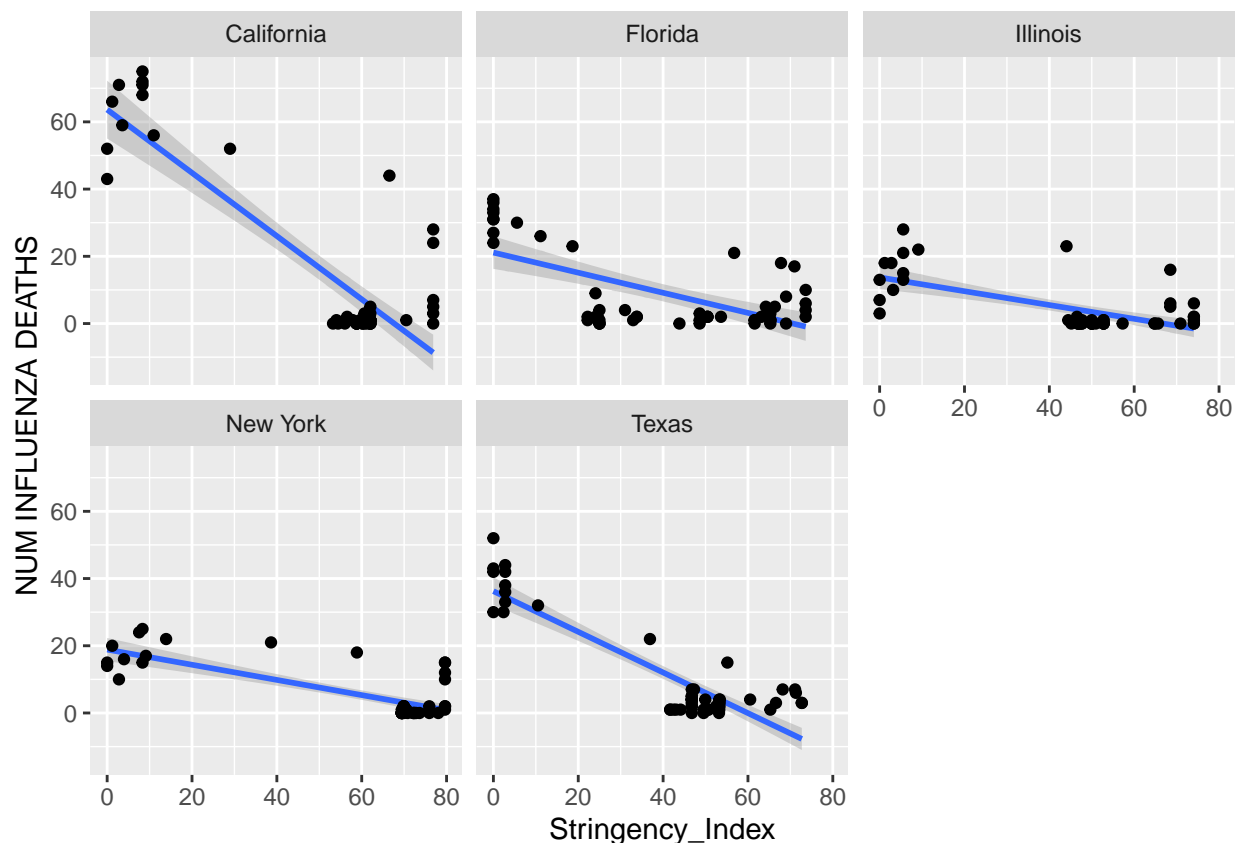
Log normalized



The stringency index was zero until COVID-19 restrictions began in the few early months in 2020. According to the plots above, the number of deaths due to influenza started to decrease as the stringency index began to increase in early 2020. Once the stringency index has hit a peak, which means a very strict lock-down, the number of influenza infections and deaths were relatively low and stable. This implies the existence of a relationship between COVID-19 restrictions and influenza infections.

```
merged_5_states%>%
  filter(YEAR == 2020)%>%
  ggplot(aes(`Stringency_Index`, `NUM INFLUENZA DEATHS`))+
  geom_smooth(method = "lm")+
  geom_point()+
  facet_wrap(~STATE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



The scatter plot of stringency index and number of deaths due to influenza, as shown above, indicates that a negative correlation exists. A linear regression model was chosen to further investigate this relationship. This linear regression model was fit with stringency index as an explanatory variable and the weekly number of deaths due to influenza as a response variable.

```
lmodel<-lm(`NUM INFLUENZA DEATHS`~`Stringency_Index`, data = merged_5_states)
lmodel%>%
  tidy(conf.int=TRUE)
```

```
## # A tibble: 2 x 7
##   term                estimate std.error statistic  p.value conf.low conf.high
##   <chr>                <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)         15.4      1.10     14.0 2.50e-37    13.2    17.5
## 2 Stringency_Index   -0.232    0.0231   -10.1 1.24e-21   -0.278   -0.187
```

The result of our model shows that the stringency index has negative correlation to the number of deaths due to influenza. This result is statistically significant because the p-value is 1.2^{10-21} , which is less than 0.5 and the confidence interval does not include the value zero. Thus, an increase of 1 in the value of stringency index corresponds to an average decrease of 0.23 in the weekly number of deaths due to influenza. When the stringency index is zero, which means no restrictions, the weekly number of deaths due to influenza is 15.38.

```
lmodel<-lm(`%UNWEIGHTED ILI`~`Stringency_Index`, data = merged_5_states)
lmodel%>%
  tidy(conf.int=TRUE)
```

```
## # A tibble: 2 x 7
##   term                estimate std.error statistic  p.value conf.low conf.high
##   <chr>                <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
```


## 1 (Intercept)	4.21	0.201	20.9	7.30e-65	3.81	4.61
## 2 Stringency_Index	-0.0490	0.00407	-12.0	2.02e-28	-0.0570	-0.0410

A similar linear regression model was fitted using the percentage of influenza infections out of total patients as a response variable. The result also shows that the stringency index has negative correlation to the response variable. This result is statistically significant because the p-value is 2.02×10^{-28} , which is less than 0.05 and the confidence interval does not include the value zero, which is the null hypothesis.

The results of the two linear models indeed support our initial hypothesis that the number of influenza infection is highly related to the COVID-19 restrictions.

Conclusion

The data demonstrate that the spread of influenza has been affected during the COVID-19 pandemic, and that this difference is explainable by the public health measures put in place to stop COVID-19. Both influenza deaths and influenza hospitalizations were predictable to a statistically significant degree when using linear models by the binary presence of the COVID-19 pandemic, indicating that the flu seasons during COVID were different from previous flu seasons. Additionally, within COVID, influenza deaths and hospitalizations were able to be explained by the explanatory variable of stringency index from Oxford University, a variable which draws on various anti-coronavirus measures to create a general measure of lockdowns, masking policies, and related measures. Each of these linear model terms had p values of less than .05 and confidence intervals which did not include the null hypothesis of 0 difference, indicating statistical significance. Each of the terms also had negative coefficients, the impact of the flu was reduced by being during the coronavirus pandemic and by increased stringency, when compared to normal years or no stringency measures. These results were expected, as influenza is a respiratory virus which spreads similarly to COVID-19, and many of the practices measured by the stringency index would logically also affect COVID-19. However, one unexpected result was that the number of deaths appeared to be more significantly affected by these variables than the hospitalization rate. The mortality of flu viruses varies from year to year, as observed in the data from previous years, but one potential reason for this difference is that hospitalization from COVID and the flu are non-exclusive, and dying from COVID and the flu are exclusive. This means that while measures related to the coronavirus may have prevented the spread of the flu, coronavirus itself likely did not, but populations which were already susceptible to dying to the flu may have been either reduced by coronavirus before they could be affected by the flu, or they were uniquely protective in a way that populations that were likely to be hospitalized but not die were not.

The implications of this data could be explored further by comparing effects of measures across countries and globally, as well as comparing the effects of these measures on other diseases. Social distancing may have been less effective in many low-income countries or more effective in countries with collectivist cultures, and it may have prevented the spread of STIs or TB, while promoting mental health issues. These data could be combined with rates of different diseases to generate similar models and find the effects of lockdown on our health as a whole.

References

“COVID-19 Government Response Tracker”. 2021. Bsg.Ox.Ac.Uk. <https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker>.