

A5 – Extension Plan

Motivation

I am planning to investigate how availability of hospital beds & vents in Marion County, Indiana impacted COVID-19 outcomes. This will include a look at confirmed COVID-19 cases, hospital occupancy & capacity rates within the county, and COVID-19 deaths. This analysis will attempt to answer whether limited hospital capacity had an impact on COVID-19 outcomes for Marion County residents.

This is a human-centered problem because it can better inform staffing & resourcing decisions related to future pandemic or epidemic events with the potential to overwhelm hospital capacity. This remains a present concern – as recently as September 2021, Indianapolis Emergency Medical Services (an ambulance transport service operating primarily in the largest metropolitan area in Marion County) [reported](#) seeing more than 33% higher daily call volumes than pre-pandemic. This has resulted in some Marion County hospitals [suspending elective procedures](#) through this fall and could potentially have an impact on excess deaths in the region.

As an aside, I hope to also investigate whether rate of vaccination increased surrounding periods of increased COVID-19 cases or deaths. This will explore whether a perceived feeling of “safety” or lack thereof impacted preventative measures (in this case vaccination) against COVID-19. The goal is to determine whether a more real awareness of the impact of COVID-19 can promote vaccination – this is related to the “scared straight” phenomenon relating to vaccination discussed [here](#).

Research Question

I am hoping to discover whether there was an impact on COVID-19 death rate during period of high ICU bed or ventilator usage at Marion County, Indiana hospitals during the COVID-19 pandemic. I hypothesize that periods of peak ICU capacity will be following by an increase in COVID-19 death rate among Marion County residents. I expect some lag time between increased hospitalization rates and increased death rates from COVID-19, but I hope to still identify some statistically significant impact.

I am also interested in determining whether Marion County, Indiana saw an uptick in vaccination rates during periods of increased hospitalizations or deaths. I expect that individuals are more likely to get vaccinated when the perceived threat of negative outcomes from a potential COVID-19 infection are higher.

Data Used

From the original A4 analysis – I plan to continue using the below datasets:

- [US County Assignments](#)
 - This dataset gives the county population from the 2020 US Census. It does not come with any specific license or terms of use and will allow me to determine prevalence among the population.
- [JHU COVID-19 Cases](#)
 - This dataset comes from a Kaggle repository and covers confirmed COVID-19 cases at the county level. It comes with an *Attribution 4.0 International* license which allows for users to share and adapt the dataset provided appropriate credit is given. The main considerations here include remaining cognizant that due to testing & reporting concerns, the data present here may not be complete or inclusive of all COVID-19 cases.

In addition to the previous sources, I also plan to include data from:

- [JHU COVID-19 Deaths](#)
 - This data is subject to the same license agreement and ethical considerations as described above in *JHU COVID-19 Cases* as it comes from the same repository.
- Indiana Department of Health (IDOH) – [COVID-19 Bed & Vent Usage by Day](#)
 - The IDOH provides daily state-level ICU bed & ventilation statistics – this dataset will allow me to make some generalizations from state to county-level hospitalization rates and discover periods of low or high hospital capacity. This information is made available under a *Creative Commons Attribution* which requires appropriate credit to the creator in the case of re-distribution or re-use.
- IDOH – [COVID-19 Vaccination Demographics by County & District](#)
 - This is also made available by the IDOH under the same *Creative Commons Attribution* as described above. It will allow me to explore vaccination rates over time in Marion County, Indiana.
- Commercial Appeal – [COVID-19 Hospital Capacity](#)
 - This dashboard contains data for Marion County, Indiana COVID-19 hospitalizations by hospital by day. However, it is not in a format that is easy to download or aggregate at the county level. This data will mainly be used to help validate (or not) the assumption that IDOH state hospitalization data can be scaled to the county-level. This is a potential dependency which will be explored further in A6-7. Commercial Appeal
- Definitive Healthcare: [Hospital ICU Beds](#) via c3.ai datalake
 - This can be used to determine staffed, licensed, and adult ICU beds by county. This will be combined with the IDOH COVID-19 Bed & Vent Usage data along with some simplifying assumptions to determine county level hospital utilization

rate. This data is available for educational and academic research and should provide attribution to c3.ai and the original source (i.e. Definitive Healthcare).

Unknowns and Dependencies

It is still unknown whether the state-level hospital capacity data is generalizable to the county-level at all points in time. It is unlikely that this is the case simply because Marion County is the most populous county in Indiana and contains the largest metropolitan area in the state. There is some dashboard level aggregate data that can potentially provide more nuance, but it is not in an easily digestible format – this will have to be further investigated and could be a major hurdle to this analysis.

I anticipate the “time series” nature of this analysis to be challenging – there will likely be some subjective decisions made regarding what constitutes an acceptable “lead time” from a spike in hospital utilizations to increased mortality. It is anticipated there will be some lag, but the exact timeline can be highly variable. The [Regenstrief](#) COVID-19 dashboard notes that length of ICU hospital stay for all age groups in Marion County, Indiana is 11.8 days (15.8 days for non-ICU admissions).

Additionally, while there is some indication that increased deaths can result from overutilization ([here](#)) as well as evidence that Marion County hospitals saw capacity surges ([here](#)), it is possible that these so-called surges aren’t dramatic enough to show up in the data or have a meaningful impact on COVID-19 mortality.

Methodology

I will gather and preprocess the data in python using similar techniques to those I employed in A4 – Common Analysis. Most of this data will be filtered to the county level with some simplifying assumptions being made on the relationship between state and county level hospital occupancy. There will also be some assumptions made relating to case duration, possibility of reinfection, and length of hospital stay – much of this data is unavailable due to privacy concerns and will instead generalize either the mean or mode value to the population where appropriate. After these assumptions are implemented and the data is filtered to the county-level, it will be joined on date for this time series analysis and [smoothed](#) to mitigate any reporting inconsistencies.

I plan to start by identifying relative minima and maxima of COVID-19 hospital occupancy in Marion County, Indiana throughout the course of the pandemic. This will hopefully create a baseline estimate of COVID-19 mortality in the region and identify periods where an increase in mortality may have occurred. This will be visualized through a standard [time series](#). I will also look at how Case Fatality Ratio (CFR) has changed over the same period, hopefully at this point

there are some discernable trends and time periods that call for further analysis. I hope to use [peak picking](#) to identify [synchrony](#) between the time series for hospital occupancy and COVID-19 mortality rates to identify whether any discernible relationship exists. This [article](#) outlines methods for assessing synchrony between time series data.

I am admittedly less familiar with these techniques than I would prefer but hope to gain a better understanding of time series analysis through this process – some potential [missteps](#) are outlined here. I hope to use the [Pearson correlation](#) to determine if global synchrony exists between these time series. This approach could yield a signal as to whether any constant relationship exists between occupancy and mortality. I will then use [dynamic time warping](#) (DTW) or [cross-correlation](#) techniques which are also able to account for lag and distance to hopefully identify the amount of lag that is present between these time series.

This process and results will then be documented and shared via a modified [PechaKucha](#) (12 slides for 20 seconds each) with the class. I hope to analyze whether a relationship exists between these time series and present findings that back this conclusion. This presentation and analysis will be fully reproducible and documented in a GitHub repository that builds from the Common Analysis relating to the impact of mask mandates on COVID-19 case rates to this extension plan investigating the relationship between hospital occupancy and COVID-19 death rate.

Timeline

- 11/13 - Data Collection & Preprocessing
- 11/16 - Assumption Validation
- 11/23 - Model & Statistical Test Design
- 11/30 - Analyze Model & Visualize Results
- 12/3 - Process Documentation
- 12/7 - [PechaKucha](#) Presentation
- 12/14 - Final Report