Influenza Staffing: Interim Report

06.27.2021

Kyle Irish

Project Overview

Motivation: The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to adequately treat these extra patients. The medical staffing agency provides this temporary staff.

Objective: Determine when to send staff, and how many, to each state.

Scope: The agency covers all hospitals in each of the 50 states of the United States, and the project will plan for the upcoming influenza season.

Hypothesis

A state's influenza related deaths will be proportional to the state's vulnerable population

Data Overview

US Census Population integrated with US Influenza Deaths

- Contains the population count by age group, sex, and record of deaths with Influenza or Pneumonia as the primary cause, and year by US states.

CDC Influenza Visits integrated with CDC Lab Tests

- Contains a count of patients by year, week, and region admitted for influenza like symptoms along with the number of positive influenza specimens.

Data Limitations

Yearly Differences and Timeliness:

The population and influenza death data both range from 2009 - 2017. The influenza-like illness (ILI) Visit data ranges from 2010 -2019. The Lab Test data ranges from 2010-2015. While more recent data would be desirable this will still provide useful insight for relatively recent data. The yearly mismatch in visit and specimen data is not a concern due to week being the primary comparison.

Influenza Visit Data:

Relatively small sample size of providers enrolled in data collection program. 3000 providers are

enrolled in reporting this information in 2010-2011. Potential to be not representative, the current estimate is around 1 million active providers. However, consistent weekly reporting will still be a significant help in analyzing trends.

Descriptive Analysis

Variable	Mean (per state)	Standard Deviation (per state)
Influenza Deaths (from 2009 - 2017)	8145.5	10442.0
Population (from 2009 - 2017)	53762106.8	61771890.3
Positive Flu Specimens (from 2010-2015)	5675.8	5858.3
Patients Visit with Influenza like Illness (from 2010-2019)	127389	162767.8

Correlations:

The sum of ILI deaths and sum of total population had a very strong relationship (0.96). This helps confirm that with more people you will see a higher number pass away due to influenza.

The year and total population also had a very strong relationship (0.97) as population increased as time went on.

The year and the sum of ILI deaths have a moderate relationship (0.49). In addition, the year and the % of ILI deaths by state populations have a low strong relationship (0.53). ILI deaths are increasing as time goes but not nearly at the rate of population.

Results and Insights

Upcoming Influenza Season

Peak Season : 12/13/2021 - 3/3/2022 (week 50 to 9.5)
Lowest Flu Season : 6/21/2021 - 9/12/2021 (week 25 to 37)

We looked at the weekly sum of the 2 variables of patients reporting ILI symptoms and the sum of laboratory specimens testing positive for a strain of influenza for all states. The weekly values that had a patient visit or positive specimen sum higher than the 75% percentile for the variable were deemed to be peak flu season. Conversely weekly values with sums below the 25% were deemed to be the lower peak or the opposite of flu season. The week value ranges were averaged and assigned dates for both variables to find the dates above.

Influenza Season effect on ILI Visits and Positive ILI Specimens

After finding the national influenza season we wanted to test the difference between peak flu season and the rest of the year for the % of visits for Influenza symptoms and % of specimens that are positive for Influenza

For Visits with Influenza Symptoms:

> Null Hypothesis: The % of ILI visits to total visit during the weeks of peak flu season (weeks 50 to 9.5) will be = the % of ILI visits to total visits during the rest of the year (weeks 10 to 49)
>
> Alternative Hypothesis: The % of ILI visits to total visit during the weeks of peak flu season (weeks 50 to 9.5) will be > the % of ILI visits to total visits during the rest of the year (weeks 10 to 49)

We are able to reject the null hypothesis at a 95% confidence level. The 2% increase in the % of visits that are for Influenza symptoms during peak Influenza season is less than 1 % due to chance.

For Positive Influenza Specimens:

> Null Hypothesis: The % of positive ILI specimens to total specimens during the weeks of peak flu season (weeks 50 to 9.5) will be = the % of positive ILI specimens to total specimens during the rest of the year (weeks 10 to 49)
>
> Alternative Hypothesis: The % of positive ILI specimens to total specimens during the weeks of peak flu season (weeks 50 to 9.5) will be > the % of positive ILI specimens to total specimens during the rest of the year (weeks 10 to 49)

Again we are able to reject the null hypothesis at a 95% confidence level. The 18% increase in the % of specimens that test positive for Influenza during peak Influenza season is less than 1 % due to chance.

Vulnerable Populations and Influenza Deaths

Our main hypothesis regarded the comparison of the % of a state's population that is within the vulnerable age group and the % of a state's population that was an ILI related death.

> Null Hypothesis: States with a high vulnerable population (over 20% of pop) will have influenza deaths relative to state population <= states with a low vulnerable population (20% or less of pop).
>
> Alternative Hypothesis: States with a high vulnerable population (over 20% of pop) will have influenza deaths relative to state population > states with a low vulnerable population (20% or less of pop).

At a confidence level of 95% we are unable to reject the null hypothesis. There is only a 50% chance that the difference in means (a 0.0012% increase of state population flu death) is due to

the proportion of the state vulnerable population being over 20%. To get an idea of this difference, we multiplied the average state population (53762107) by the % difference in means (0.0012%). The increase in the number of deaths in a state would be 665.

We do want to highlight that vulnerable age populations still account for the significant majority of influenza deaths as noted in the project assumptions. Vulnerable age groups account for 91% of Influenza deaths. Furthermore, deaths of this vulnerable age group for Influenza only are recorded for the above 65 age sub group. The under 5 age group reported no Influenza deaths.

To further look at vulnerable populations in relation to influenza deaths, we tested the relationship between vulnerable age groups and influenza deaths without normalizing deaths as a % of state population.

> Null Hypothesis: States with a high vulnerable population (over 20% of pop) will have Influenza deaths <= deaths of states with a low vulnerable population (20% or less of pop).
>
> Alternative Hypothesis: States with a high vulnerable population (over 20% of pop) will have influenza deaths > deaths of states with a low vulnerable population (20% or less of pop).

We are unable to reject the null hypothesis at a 95% confidence level. We found an 88% chance that a decrease of 5609 flu deaths is due to the proportion of the state vulnerable population being over 20%.

State Populations and Influenza Deaths

We wanted to investigate the influenza deaths with state population to see if it would be a better indicator than % vulnerable population.

> Null Hypothesis: States with a larger population (over average across states) will have total influenza deaths <= states with a lower population (under or equal to average across states).
>
> Alternative Hypothesis: States with a larger population (higher than average across states) will have total influenza deaths > states with a lower population (under or equal to average across states).

We are able to reject the null hypothesis at a 95% confidence level with a P(T<=t) two-tail value (0.00015) and difference in means (15151 deaths). Having an above average sized population of vulnerable or an above average size state population is a significant indicator in an increase in influenza deaths.

State Risk Levels

We found that population was a significant indicator of an increase in Influenza deaths. However since the majority of Influenza deaths occur in the vulnerable age group we wanted to use both of these variables and categorized a state's risk level as a combination of population size and % vulnerable age group.

The highest risk level will have both an above average population size and above average % of their population made up of the vulnerable age groups. The lowest risk level will be below average in both categories. The middle risk level will contain the remaining states but indicate which was their above average category.

High Risk	Medium Risk (High % Vulnerable)	Medium Risk (High Total Pop)	Low Risk
Arizona	Alabama	California	Alaska
Florida	Arkansas	Georgia	Colorado
Michigan	Connecticut	Illinois	District of Columbia
New Jersey	Delaware	Indiana	Louisiana
Ohio	Hawaii	Massachusetts	Maryland
Pennsylvania	Idaho	New York	Minnesota
Tennessee	Iowa	North Carolina	Nevada
	Kansas	Texas	New Hampshire
	Kentucky	Virginia	Utah
	Maine	Washington	Wyoming
	Mississippi		
	Missouri		
	Montana		
	Nebraska		
	New Mexico		
	North Dakota		
	Oklahoma		

Oregon
Rhode Island
South Carolina
South Dakota
Vermont
West Virginia
Wisconsin

Remaining Analysis and Next Steps

We are continuing to investigate the data for more information on the remaining question below in order to develop a staffing plan that can address needs for flu season by state.

Do any states vary from the national flu season dates found?
The results of state flu dates based on the process used in the results and insights section can be found in the appendix at the end of this report. We will continue to investigate differences with a spatial analysis of state flu season variance as part of data visualizations noted below.

We will also be moving forward with the additional deliverables of the project.

- Build data visualizations via Tableau storyboard (by week 5)
- Deliver presentation of report in meeting to stakeholders (by week 6)

Appendix:

Business Requirements Document

Upcoming Influenza Season Per State

Note - Florida not included due to lack of results for Lab Test or ILI Visit data

State	Week Flu Low Start	Week Flu Low End	Week Flu High Start	Week Flu High End
Alabama	23.5	37	48.5	10
Alaska	25	39.5	47.5	17
Arizona	26	40.5	50.5	10.5
Arkansas	22	35.5	51	10.5
California	26	37.5	51.5	11
Colorado	25.5	37	50	9.5
Connecticut	23.5	36.5	1	14
Delaware	23	38.5	51.5	11.5
Georgia	23.5	36.5	48	9
Hawaii	26.5	42.5	1	13.5
Idaho	20	40	51	10.5
Illinois	24	37.5	49	11
Indiana	21.5	37.5	49.5	9.5
Iowa	22	35	49.5	11
Kansas	21.5	37	50.5	10.5
Kentucky	23.5	36.5	49.5	10
Louisiana	23.5	38	48.5	9

State	Week Flu Low Start	Week Flu Low End	Week Flu High Start	Week Flu High End
Maine	24.5	40	52.5	11.5
Maryland	25	38.5	50.5	10.5
Massachusetts	24.5	40.5	52	14.5
Michigan	24.5	39	51.5	12.5
Minnesota	24	37.5	50	12
Mississippi	23	36	48.5	8.5
Missouri	23	36	49	10.5
Montana	23	42	51	11
Nebraska	23	38	51	11
Nevada	26	39	51	11
New Hampshire	24	37	1	14
New Jersey	26	37	51	11
New Mexico	24	36	51	11
New York	28	39	52	12
North Carolina	26	39	52	12
North Dakota	26	39	51	12
Ohio	22.5	39.5	50	11
Oklahoma	23.5	34	50.5	10.5
Oregon	26	42	51.5	12
Pennsylvania	24	37	51	10.5
Rhode Island	22	34	2	15
South Carolina	24.5	37.5	49.5	9.5

State	Week Flu Low Start	Week Flu Low End	Week Flu High Start	Week Flu High End
South Dakota	21.5	37	52.5	12
Tennessee	22.5	36.5	49.5	9.5
Texas	24.5	36	49	9
Utah	27.5	40.5	50	10.5
Vermont	22.5	42.5	52.5	12.5
Virginia	25	40.5	50.5	10
Washington	26	39.5	50.5	12.5
West Virginia	23	37	51	10.5
Wisconsin	23.5	37.5	50.5	11.5
Wyoming	21	37.5	50	10

Data Profiles:

Influenza Visits and Lab Test Integrated Data

Data Grain: The data grain will be a 3 variable combo of State, Year, and Week

Data Changes:

- Updated the Region Type column for the Regions (District of Columbia, New York City, Virgin Islands, Puerto Rico, Commonwealth of the Northern Mariana Islands)
- Removed the Age group variables.
- Removed all rows with region type that was not State.
- Recalculated %UNWEIGHTED ILI as "ILITOTAL" divided by the "TOTALPATIENTS"
- Removed 465 rows of data for Florida region where % WEIGHTED ILI, %UNWEIGHTED ILI, ILITOTAL, NUM. OF PROVIDERS, and TOTAL PATIENTS that were all X values.
- Replaced all X values with 0 in the columns TOTAL SPECIMEN, PERCENT POSITIVE, A (2009 H1N1), A (H1), A (H3), A (Subtyping not Performed), A (Unable to Subtype), B, H3N2v
- Added 'Sum of ILI Positive Types' column to sum all the different specimen columns where ILI positive tests were found
- Recalculated PERCENT POSITIVE as % Positive column to deal with formatting issues

Influenza Deaths and Census Population Integrated Data

Data Grain: The data grain will be a 2 variable combination of state and year.

Data Changes:

- Changed 'A?asco Municipio, Puerto Rico' to 'Anasco Municipio, Puerto Rico'.
- Changed 'R?o Grande Municipio, Puerto Rico' to 'Rio Grande Municipio, Puerto Rico'
- Changed 'Pe?uelas Municipio, Puerto Rico' to 'Penuelas Municipio, Puerto Rico'
- Changed 'Lo?za Municipio, Puerto Rico' to 'Loiza Municipio, Puerto Rico'
- Changed 'Do?a Ana County, New Mexico' to 'Dona Ana County, New Mexico'
- Changed 'Can?vanas Municipio, Puerto Rico' to 'Canovanas Municipio, Puerto Rico'
- Changed 'Rinc?n Municipio, Puerto Rico' to 'Rincon Municipio, Puerto Rico'
- Changed 'Cata?o Municipio, Puerto Rico' to 'Catano Municipio, Puerto Rico'
- Changed 'Comer?o Municipio, Puerto Rico' to 'Comerio Municipio, Puerto Rico'
- Changed 'Manat? Municipio, Puerto Rico' to 'Manati Municipio, Puerto Rico'
- Changed 'Mayag?ez Municipio, Puerto Rico' to 'Mayaguez Municipio, Puerto Rico'
- Changed 'Bayam?n Municipio, Puerto Rico' to 'Bayamon Municipio, Puerto Rico'
- Changed 'Las Mar?as Municipio, Puerto Rico' to 'Las Marias Municipio, Puerto Rico'
- Changed 'Juana D?az Municipio, Puerto Rico' to 'Juana Diaz Municipio, Puerto Rico'
- Changed format for age category columns to integers. Data brief noted they are estimates and it doesn’t make sense to have part of people.
- Separated County into 2 columns for State and County
- Removed 3278 Duplicates on Data Grain of State, County, and Year
- Replace 144 values in State column that had #N/A values with District of Columbia as they all had 11 as their state code. In addition the frequency of District of Columbia was 144 below the frequency of the other states.
- Replaced Year values of 20133 with 2013
- Removed year from Month column and removed entire Month Code column