Image by tungnguyen0905 from Pixaby

Med Staff Influenza Season Planning - Data Analyst Interim Report

Prepared for: Project Sponsor, Project Team and all relevant stakeholders

Prepared by: Elsa Ekevall

3 February 2022

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.

Summary

This analysis sought to help plan for influenza season, by determining where to send temporary workers on an as-needed basis. Following the hypothesis that in states with a high proportion of vaccinated healthcare workers fewer people will develop serious complications and end up in hospital. Although in this analysis no link between vaccinated healthcare workers and hospitalisation was found — nor was a link with visits for influenza like illnesses or death ratio — there is a connection with the proportion of people vaccinated. When there is a higher percentage of vaccinated healthcare workers, the percentage of the vaccinated populations is higher.

The Business Requirements Document including the assumptions and constraints can be found here and the Project Management Plan is provided here.

HYPOTHESIS

If states have a high proportion of vaccinated health care workers (nurses, physician assistants and doctors), then fewer vulnerable people will develop serious complications and end up in the hospital.

DATA OVERVIEW

Influenza Deaths:

- The data contains monthly death counts for influenza-related deaths in the United States from 2009 to 2017. Counts are broken into two categories: state and age.

New Population Data:

- The data contains population counts for each state in the United States from 2010 to 2017. Counts are broken into categories: total population and age 5 year age groups (e.g., under 5, etc).

Influenza Laboratory Patient Visits

- The data tracks patient visits to a medical provider for influenza. It counts the number of visits, number of providers, and total patients seen by week and state from late 2010 to early 2019.

Vaccination Coverage Among Health Care Personnel (HCP) and the General Population

- The data provides seasonal state influenza vaccination coverage estimates for the general population (2010-2017) and health care personnel (2013-2017) split by month and age groups.

Hospitalisation Rates

- The data contains Influenza Hospitalisation Rates (per 100,000 population) — cumulative rate and weekly rate — from over 70 counties covering 2010 to 2018 by state, year, week and age group.

Further information can be found in the Data Overview document here.

DATA LIMITATIONS

The data does not take into account all the vulnerable populations, only those who are under 5 years and 65+ years (for more info see hypothesis development section below).

Influenza Deaths: this may be an incomplete count due to deaths on a death certificate only list one cause of death which may create some discrepancies. 80% of this data set was suppressed and the values were imputed using the mean.

Population Data by Geography: assumed to be a complete and accurate count/estimate of the population. However there were too many missing values in the initial data set and it was replaced by a more comprehensive data set from the same source.

Influenza Laboratory Tests: this survey data is prone to error and there may be inconsistencies, plus there are no values for the state of Florida.

Vaccination Coverage Among Health Care Personnel (HCP) and the General Population: these data sets are based on a number of national opt-in sources; and had been checked and cleaned then weighted and adjusted to take into account any bias.

Hospitalisation Rates: this survey data is prone to error and there may be inconsistencies. It is also limited in that it only comes from approx 9% of the US population and does not cover each of the 50 states. Records for five of the 17 states do not cover the full eight year period. The data is used in the absence of other suitable data with the caveat that it may not be generalisable and rates are likely to be underestimated.

DESCRIPTIVE ANALYSIS

Data required to answer the hypothesis

Vulnerable populations: patients likely to develop flu complications requiring additional care, adults over 65 years and children under 5 years. Therefore need the percentage of Health Care Professionals vaccinated, the percentage of the population who developed Influenza Like Illnesses (ILI Total), the percentage of these populations who were hospitalised with influenza and in the absence of reliable hospitalisation data — how many in these vulnerable populations died (Death Ratio). The data set was filtered to only include the years 2013 to 2017 when the Health Care Personnel vaccination estimates are available.

Dependent Data Variable and Independent Data Variables

Variable	Average	Standard Deviation
Percentage of Vaccinated Health Care Workers	86.58%	6.91%
Percentage hospitalised in the under 5 years age group	0.05%	0.02%
Percentage hospitalised in the 65+ years age group	0.21%	0.12%
Average Percentage of visits for Influenza Like Illnesses per Month	2.39%	3.48%
Death ratio in the under 5 years age group	0.09%	0.09%
Death ratio in the 65+ years age group	0.12%	0.05%
Death ratio in all age groups	0.04%	0.03%
Percentage hospitalised in all age groups	0.55%	0.34%

The population under 5 years had a strong correlation (0.53) with visits for Influenza Like Illnesses, meaning that states with higher populations in this age group appear to report a higher number of visits. This population also had a strong negative correlation (-0.52) with death ratio under 5 years, meaning that states with higher populations in this age group appear to have fewer deaths. [NB. 80% of the death data was imputed.]

Further information on the variables can be found in the Integrated Data Set Dictionary Excel file here.

See the Appendix section below for more information on the testable hypothesis and results. The full descriptive statistical analysis can be found here and the hypothesis testing here.

RESULTS AND INSIGHTS

When the healthcare worker vaccination estimate is above the mean at a confidence level of 5 percent there was no significant difference between the following sample groups: Hospitalisation Rate and Death Ratio in the vulnerable age groups (under 5 years and 65+ years) and in all age groups; nor in the percentage of visits for Influenza Like Illnesses. The average vaccination rate for health care workers is high 86.6% and this may contribute to the lack of statistical difference. Another contributing factor may be the less than accurate Hospitalisation Rate data.

REMAINING ANALYSIS WORK AND NEXT STEPS

Although healthcare worker vaccination estimate is not useful as a direct measure for hospitalisations, there was a significant difference between the percentage of the general population vaccinated sample groups at a confidence level of 5 percent. When the proportion of vaccinated healthcare workers is higher, the proportion of the population vaccinated in the vulnerable and in all age groups is higher.

- It is known that vaccination can lessen the severity of influenza illness and hence reduce the hospitalisation rate[1]. **Therefore it is recommended that an analysis of the proportion of the population vaccinated with hospitalisation rate and death ratio be undertaken.**

As expected States with higher populations in the vulnerable age group also report higher percentage of patient visits with Influenza like Illnesses. However as the percentage of visits for Influenza Like Illnesses increases the correlation shows a decrease in the death ratio.

- Perhaps the more people who report Influenza Like Illnesses means that they are treated early on resulting in fewer deaths? **Thus it would also be worthwhile carrying out an analysis of patient visits for Influenza like illnesses with hospitalisation rate and death ratio.**

Recommendations for the Next Steps:

- Explore with the stakeholders if they are able to source more accurate hospitalisation rate data or a suitable alternative.
- Perform further analyses as described in bold above in the Remaining Analysis Work and Next Steps.
- Examine the seasonal trends in influenza (e.g., start and end date across the US, peaks and troughs of the percentage of visits for Influenza Like Illnesses and hospitalisation rates) and how they can be used to proactively plan for staffing needs across the country.
- Provide composition, statistical, spatial, and temporal visualisations in the presentation of the final results to the stakeholders.

1 Understanding the unique characteristics of seasonal influenza illness to improve vaccine uptake in the US, Rafik Bekkat-Berkani, Luis Romano-Mazzotti, Vaccine, 2018 Nov 19 - https://pubmed.ncbi.nlm.nih.gov/30366802/; accessed 8 December 2021.

APPENDIX

[The folder and files in the Google drive folder are restricted to the stakeholders named in the Project Management Plan. If you experience any issues please contact the Project Manager - f.lastname@HealthcareInc.com tel: (xxx) xxx-xxxx (xxx)]

Business Requirements Document and Project Management Plan

- The Business Requirements Document can be viewed and downloaded here.
- The Project Management Plan can be viewed and downloaded here.

Hypothesis Development

Key questions asked that formed the hypothesis development:

Who are the vulnerable populations, and which ones have data available that can be analysed?

- pregnant women, **children under 59 months, the elderly**, individuals with chronic medical conditions and individuals with immunosuppressive conditions.[1 & 2]

Are there factors other than vulnerable populations that contribute to the high rate e.g. people mixing indoors due to colder temperatures, higher population density?

- Seasonal influenza spreads easily, with rapid transmission in **crowded areas** including schools and **nursing homes**.[1]
- **Health care workers are at high risk of acquiring influenza due to increased exposure to infected patients and risk further spread particularly to vulnerable individuals.[1]**

Data is readily available on two of these vulnerable populations — under 5 years and 65+ years. Heath care workers are also a group at high risk of catching influenza and then spreading the disease to their patients, most notably to the populations who are most vulnerable. Another reason for looking at vaccination rates in health care workers is the assumption that if a large number of workers have influenza this will result in higher staffing shortages.

Further information on the key questions identified from the Business Requirements Document and the references consulted can be found on Page 6 of the Project Management Plan.

1. WHO website, Influenza (Seasonal), 6 November 2018: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal); accessed 8 December 2021.
2. CDC website, Flu Season: https://www.cdc.gov/flu/about/season/flu-season.htm; accessed 8 December 2021.

Data Overview

- The Data Overview document be viewed and downloaded here.

Data Profiles and Processes

- An Excel file containing the profiles of all the data sets can be viewed and downloaded here.
- The full checking and cleaning process of all the data files can be downloaded *[NB the Excel file 'Data_Profiles_Checks_Cleaned' is 154.5 MB and too large to be previewed]* here.
- An Excel file containing the Data Transformations that created the Integrated data Set can be viewed and downloaded here.
- An Excel file containing the Integrated Data Set Dictionary can be viewed and downloaded here.

Results and Insights

The analysis is based on the average weekly influenza like illness visits total over the month, the max vaccination estimates for Health Care Personnel (HCP) and the General Population. The analysis samples come from a subset of the integrated data set covering the years 2013 to 2017 when the Health Care Personnel vaccination data is available. The first 113 (108 for percentage of visits for Influenza Like Illnesses) samples were randomly selected from the HCP Vaccinated above mean data set to make both sample sets even.

The Hospitalisation Rate data set is a further subset covering only the states that have hospitalisation rate data and the years 2013 to 2017. The first 28 samples from the HCP Vaccinated below mean data set were also selected randomly. The Hospitalisation Rate data set only has 60 records covering 12 states.

Testable Hypotheses			
Null Hypothesis	There is no difference on the Death Ratio if the percentage of Health Care Professionals vaccinated is higher or lower.	H0 : µDeath Ratio HCP vaccination high = µDeath Ratio HCP vaccination low	Two-tailed test
Alternative Hypothesis	If the percentage of Health Care Professionals vaccinated is higher the Death Ratio is higher or lower than the mean.	H0 : µDeath Ratio HCP vaccination high <> µDeath Ratio HCP vaccination low	alpha = 0.05
Null Hypothesis	There is no difference on the percentage of visits for Influenza Like Illnesses if the percentage of Health Care Professionals vaccinated is higher or lower.	H0 : µILI Total HCP vaccination high = µILI Total HCP vaccination low	Two-tailed test
Alternative Hypothesis	If the percentage of Health Care Professionals vaccinated is higher the percentage of visits for Influenza Like Illnesses is higher or lower than the mean.	H0 : µILI Total HCP vaccination high <> µILI Total HCP vaccination low	alpha = 0.05
Null Hypothesis	There is no difference on the percentage of populations vaccinated if the percentage of Health Care Professionals vaccinated is higher or lower.	H0 : µPopulation Vaccinated HCP vaccination high = µPopulation Vaccinated HCP vaccination low	Two-tailed test
Alternative Hypothesis	If the percentage of Health Care Professionals vaccinated is lower the percentage of the populations who are vaccinated is higher or lower than the mean.	H0 : Population Vaccinated HCP vaccination high <> µPopulation Vaccinated HCP vaccination low	alpha = 0.05

Two-tailed test — would like to know if the Death Ratio or percentage of visits for Influenza Like Illnesses is higher or lower when vaccination rates are high.

Results			
	Variable	p-value	t-Test: Two-Sample Assuming Unequal Variances
Death Ratio	A Under 5 years	68.60%	When the Health Care Personnel vaccination rate is above the mean (86.58%) the Death Ratio is neither higher or lower in the vulnerable age groups (under 5 years and 65+ years) and in all age groups. At a confidence level of 5 percent there was no significant difference between the sample groups. [NB A large amount (80%) of this data was imputed due to suppressed values and the distributions are right skewed.]
	L 65+years	21.71%	
	K All Age Groups	84.23%	
Influenza Like Illnesses	Average Weekly ILI Total	27.70%	When the Health Care Personnel vaccination rate is above the mean (86.58%) the percentage of visits for Influenza Like Illnesses is neither higher or lower. At a confidence level of 5 percent there was no significant difference between the sample groups.
Hospitalisation Rate	A Under 5 years	33.39%	When the Health Care Personnel vaccination rate is above the mean (86.58%) the hospitalisation rate is neither higher or lower in the vulnerable age groups (under 5 years and 65+ years) and in all age groups At a confidence level of 5 percent there was no significant difference between the sample groups. [NB Sample size (28) is below 30 and distributions are right skewed.]
	L 65+years	93.15%	
	K All Age Groups	98.655%	
All Vaccinated	A Under 5 years	**0.004%**	When the Health Care Personnel vaccination rate is above the mean (86.58%) the mean vaccination rate in the vulnerable age groups (under 5 years and 65+ years) and all age groups is also higher. At a confidence level of 5 percent there was a significant difference between the sample groups.
	L 65+years	**0.06%**	
	K All Age Groups	**3.85%**	

- An Excel file containing the Statistical Analysis can be viewed and downloaded here.
- An Excel file containing the Hypothesis Testing can be viewed and downloaded here.

Google Drive Folder (Includes all Project Data)

https://drive.google.com/drive/folders/10qhzcQjTbdWEBbHxxzBsJjmh2pJvrXtq?usp=sharing