Preparing for Influenza Season

Interim Report

Nadia Ordonez, October 2023

Project Overview

Motivation: The US 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 US, and the project will plan for the upcoming influenza season.

Hypothesis

If individuals are under 5 and over 65 years old, then they report similar death count ratios.

Data Overview

Mortality data

The Centers for Disease Control and Prevention (CDC) provides this administrative data. It contains monthly flu death counts in the US from 2009 to 2017. Counts are broken into two categories: state and age.

Census data

The US Census Bureau provides this survey data. It contains demographic data for states including gender and age groups in the US from 2009 to 2017.

Data Limitations

Mortality data

Imputing the "Suppressed" death count category: Since the "Suppressed" category represents 81.7% of the original data, any numerical replacement would result in different outcomes. In this project, the "Suppressed" category was imputed using CDC information reported for children's deaths per year and aimed not to overshadow the actual numerical death counts.

Census data

Time lag: This data set is not up-to-date and can only be used to create a medical staff plan for the 2018 flu season.

Descriptive Analysis

The vast majority of flu deaths occur among individuals over 65 years old (see Table 1). Annually, roughly 42163 individuals from this age group die due to flu complications. Children under 5 years old also died due to flu complications, however, their average death counts are lower compared to individuals between 5 and 64 years old and those over 65 years old.

Table 1. Annual flu-related deaths (estimates) in the US from 2009 to 2017

Age group	*Flu deaths*		*Population density*	
	Average	**Standard deviation**	**Average**	**Standard deviation**
Under 5 years	11.0	0.1	20.1m	308288.8
5-64 years	4120.0	935.0	24.9m	3628725.6
Over 65 years	42162.5	2339.0	42.3m	3379772.2

Additionally, the total number of flu deaths per state from 2009 to 2017 had a strong correlation (0.95) with their total population, meaning that largely populated states would need extra medical assistance during the coming flu season.

Results & Insights

In the provided Business Requirement Document, it was stated that vulnerable populations to the flu are individuals over 65 and under 5 years old. Based on my descriptive analyses, I observed that annual average death counts are particularly high among individuals over 65 years old in comparison to those under 5. Thus, the hypothesis testing is as follows:

Null hypothesis: The flu death ratios among children younger than 5 years old are not similar to those reported for over 65 years old.

$$H_0 = \mu_{\text{under 5-year-olds}} = \mu_{\text{over 65 years old}}$$

Alternative hypothesis: The flu death ratios among children younger than 5 years old are not similar to those reported for over 65 years old.

$$H_1 = \mu_{\text{under 5-year-olds}} \neq \mu_{\text{over 65 years old}}$$

At an alpha of 0.05, or a confidence level of 95 percent, there's a significant difference in death count rations between individuals under 5 and those over 65 years old. Since a larger number of individuals over 65 years old are dying from flu complications, more medical staff would be needed to attend to this age group. Moreover, additional solutions such as vaccination campaigns aimed at those over 65 years old might have a higher impact on reducing overall flu deaths.

Remaining Analysis and Next Steps

- Investigate if death count ratios for individuals over 65 years old are higher than the rest of the population. If this is correct, then update stakeholders about allocating extra medical resources to this age group. It would then be expected that most flu deaths in the coming flu season would occur among those over 65 years old.
- Investigate which states have larger older populations (over 65 years old) and if they correlate to flu death counts during the flu season.
- Create spatial and temporal visualizations per state.
    - Prioritize states based on reported flu death counts. A higher number of medical staff would be provided to states where more flu deaths are reported.
    - Investigate when the coming flu season might occur in each state. Schedule when would the medical staff be deployed to each of the states.

As a final step, a presentation would be delivered to summarize current and future insights.

Appendix

Appendix 1. Business Requirement Document

CAREERFOUNDRY

Preparing for Influenza Season

Stakeholder Identification

- Medical agency frontline staff (nurses, physician assistants, and doctors)
- Hospitals and clinics using the staffing agency's services
- Influenza patients
- Staffing agency administrators

Assumptions & Constraints

Assumptions:

- Vulnerable populations suffer the most-severe impacts from the flu and are the most likely to end up in the hospital.
- Flu shots decrease the chance of becoming infected with the flu.

Constraints:

- The staffing agency has a limited number of nurses, physician assistants, and doctors on staff.
- There's no money to hire additional medical personnel.

Requirements

- Provide information to support a staffing plan, detailing what data can help inform the timing and spatial distribution of medical personnel throughout the United States.
- Determine whether influenza occurs seasonally or throughout the entire year. If seasonal, does it start and end at the same time (month) in every state?
- Prioritize states with large vulnerable populations. Consider categorizing each state as low-, medium-, or high-need based on its vulnerable population count.
- Assess data limitations that may prevent you from conducting your desired analyses.

Success Factors

The project's success will be based on:

- A staffing plan that utilizes all available agency staff per state requirements, without necessitating additional resources
- Minimal instances of understaffing and overstaffing across states (a state can be considered understaffed if the staff-to-patient ratio is lower than 90% of the required ratio and overstaffed if greater than 110%)

Glossary

Influenza: a contagious viral infection, often causing fever and aches.

Vulnerable populations: patients likely to develop flu complications requiring additional care, as identified by the Centers for Disease Control and Prevention (CDC). These include adults over 65 years, children under 5 years, and pregnant women, as well as individuals with HIV/AIDs, cancer, heart disease, stroke, diabetes, asthma, and children with neurological disorders.

Additional Context

A count of the historical influenza deaths gives an indication of the severity of flu in an area. Deaths can be prevented with flu shots and adequate medical staff. In the United States, each state has a different population composition, meaning that some states will have more vulnerable populations. In this project, you should pay particular attention to influenza deaths, vulnerable populations, and (optionally) flu-shot rates—particularly in vulnerable populations—to determine medical staffing needs.

Appendix 2. Hypothesis development

The remaining analyses would target research questions 1 and 2 and subsequent funneling questions, mainly through visualizations. Research question 3 can't be answered since no data set regarding medical staff numbers was originally provided. Stakeholders can be consulted whether such data exists, so I can provide a more specific number of how many medical personnel would be needed per state.

1) Does flu season occur at the same time in each state?
    a. How frequently does flu season occur per year in each state?
    b. When does flu season start per year in each state?
    c. Does flu season frequency and starting season drastically vary over the years in each state?
2) Which states have a large flu-vulnerable population?
    a. What are the parameters to classify states as large, medium, or low vulnerability?
    b. How do flu-related deaths/admissions correlate to states classified as large/medium/low vulnerable populations?
    c. Do states with large vaccination rates among children have fewer deaths/admissions related to flu?

3) How many hospitals/clinics per state do we have?
    a. What is the staff-to-patient ratio that we want to achieve?
    b. How many temporary personnel (nurses, physician assistants, doctors) do we have?
    c. How many in-house medical staff dealing with flu complications currently exist per state?
    d. How many patients do the hospitals/clinics admit due to flu complications during flu season annually?
    e. How do flu-related deaths correlate to the staff-to-patient ratio?

Appendix 3. Data Profiles

Mortality data set: Total observations: 66097

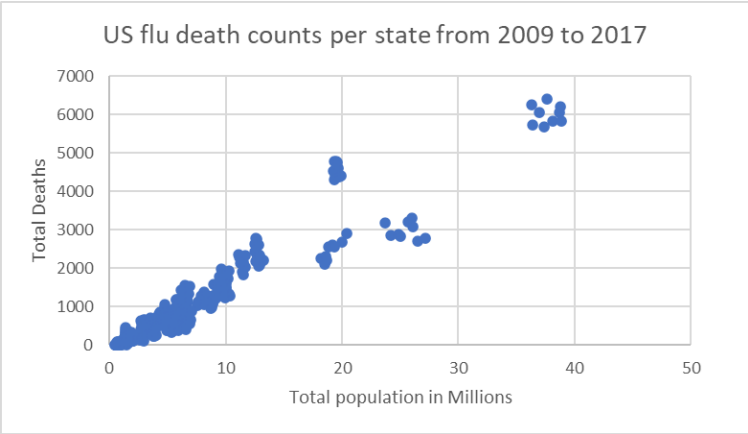
Variables	***Change***	***Qualitative: Nominal / Ordinal Quantitative: Discrete / Continuous***
State	Rename rows by cross-reference with the state code variable	Nominal
State Code	No changes were made	Ordinal
Year	17 observations stated as 20133 were changed in 2013	Discrete
Month	No changes were made	Ordinal
Month Code	No changes were made	Ordinal
Ten-Year Age Groups	No changes were made	Ordinal
Ten-Year Age Groups Code	No changes were made	Ordinal
Deaths	The "Suppressed" category imputed from annual children deaths reported by the CDC	Discrete

Source Link: https://www.verywellhealth.com/deaths-from-flu-2633829
Based on the information provided by the above link, specifically at the Number of Pediatric Flu Deaths in the U.S. From recent flu seasons 2003–2019, I calculate that **103 children** (under 18 years old) died from flu-related issues per year in the U.S.

Census data set: Total Observations: 28986

Variables	***Changes***	***Qualitative: Nominal/Ordinal Quantitative: Discrete/Continuous***
County	The state was extracted from this variable	Nominal
Year	No changes were made	Discrete
Total population	No changes were made	Discrete
Male Total population	No changes were made	Discrete
Female Total population	No changes were made	Discrete
Under 5 years	No changes were made	Continuous
5 to 9 years	Age categories from 5 to 64 were grouped	Continuous
10 to 14 years	Age categories from 5 to 64 were grouped	Continuous
15 to 19 years	Age categories from 5 to 64 were grouped	Continuous
20 to 24 years	Age categories from 5 to 64 were grouped	Continuous
25 to 29 years	Age categories from 5 to 64 were grouped	Continuous
30 to 34 years	Age categories from 5 to 64 were grouped	Continuous
35 to 39 years	Age categories from 5 to 64 were grouped	Continuous
40 to 44 years	Age categories from 5 to 64 were grouped	Continuous
45 to 49 years	Age categories from 5 to 64 were grouped	Continuous
50 to 54 years	Age categories from 5 to 64 were grouped	Continuous
55 to 59 years	Age categories from 5 to 64 were grouped	Continuous
60 to 64 years	Age categories from 5 to 64 were grouped	Continuous
65 to 69 years	Age categories over 65 were grouped	Continuous
70 to 74 years	Age categories over 65 were grouped	Continuous
75 to 79 years	Age categories over 65 were grouped	Continuous
80 to 84 years	Age categories over 65 were grouped	Continuous
85 years and over	Age categories over 65 were grouped	Continuous

Appendix 4. Correlation between total flu death counts per state and total population from 2009 to 2017 in the US



States with the largest populations have the largest flu death counts. This finding was to be expected, since the higher the number of people in a given state, then the higher the chance that more flu-related deaths are reported. For our project, it means that states with the largest populations would need more medical staff to reduce flu deaths.

Appendix 5. Hypothesis testing

Two-tailed test: I care whether death counts are higher or lower for those under 5 years old.

Two-samples: I am using the flu death ratios among under 5 and over 65 years old

Significance level: 0.05

Variables	**Under 5 years**	**65+ years**
t-Test: Two-Sample Assuming Unequal Variances		
	Under 5 years	*65+ years*
Mean	1.54665E-06	0.000816946
Variance	2.85621E-12	1.73854E-07
Observations	459	459
Hypothesized Mean Difference	0	
df	458	
t Stat	-41.89675367	
P(T<=t) one-tail	4.3914E-159	
t Critical one-tail	1.648187415	
P(T<=t) two-tail	**8.78E-159**	
t Critical two-tail	1.965157098	