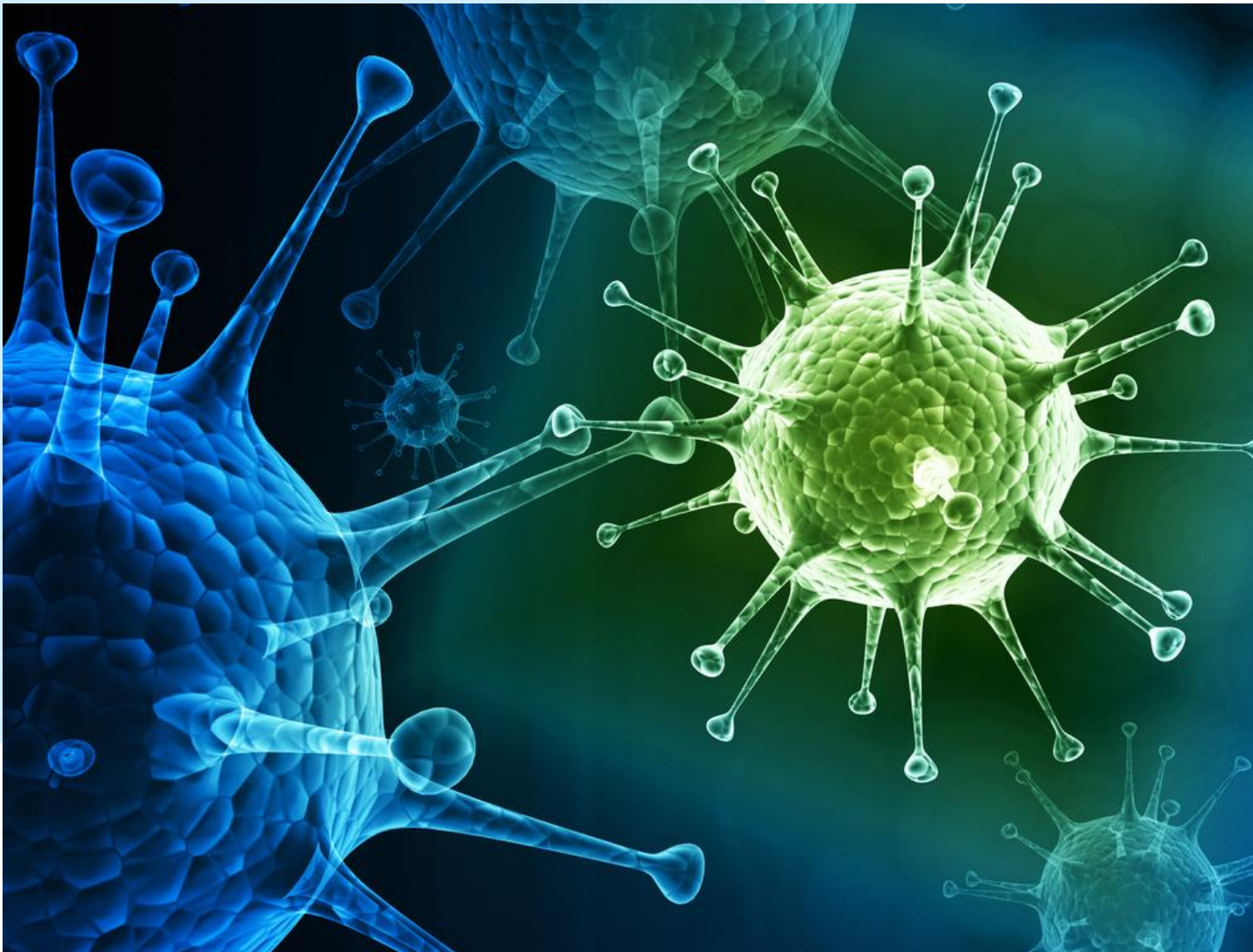# PREPARING FOR INFLUENZA SEASON

INTERIM REPORT

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

If the state has higher population in vulnerable age (under 5 and over 65 years old), then death rate of influenza will be higher.

Data overview

Influenza Deaths Data Set

This data shows the geographic and monthly spread of influenza across the United States and 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*.

US Census data set

This data shows yearly population counts in the United States from 2009 to 2017, grouped by categories: *states, total population, population by gender and by age*.

Data limitations

Influenza Deaths Data Set

Time lag:	No

Inaccuracies:	No
Bias:	Death records come from death certificates which only list one cause of death. This could create some discrepancies within vulnerable populations whose decline in health is due to influenza, but their cause of death is listed as another reason such as AIDS.

US Census data set

Time lag:	Carried out once per 10 years. Presented data is yearly estimation as it does not include immediately automatic updates as death, birth, moving population to another state etc.
Inaccuracies:	No
Bias:	No

Descriptive analysis

The influenza death rate is calculated from normalized data that considers the population size of each US states.

	Influenza Death Rate Under 5 years	Influenza Death Rate 5-64 years	Influenza Death Rate 65+ years
Average	0.10%	0.03%	0.16%
Standard deviation	0.001111162	0.000241	0.000614

Table 1: Statistics of Core Variables

There is a strong positive correlation between flu deaths and population size for people aged 65 and over.

	Death/population Under 5 years	Death/population 35-44 years	Death/population 65+ years
Correlation coefficient	0.01	0.44	0.90
Strength of Correlation	No relationship (<0.1)	Moderate (0.3-0.5)	strong (0.5-1.0)

Table 2: Correlation results

Results & Insight

Null hypothesis: Influenza death rate of population in vulnerable age (<5, 65+ yr.) is less or same as influenza death rate of population of non-vulnerable age (5-64 yr.).

Alternative hypothesis: Influenza death rate of population in vulnerable age (<5,65+ yr.) is higher as influenza death rate of population of non-vulnerable age (5-65 yr.).

	Vulnerable population (<5, 65+ yr.)	Non-vulnerable population (5-65 yr.)
Mean	0.00131878	0.000251625
Variance	3.20471E-07	5.80392E-08
Observations	459	459
df	619	
t Stat	37.16164283	
P(T<=t) one-tail	**4.42E-160**	
t Critical one-tail	1.647319	
P(T<=t) two-tail	8.8374E-160	
t Critical two-tail	1.963803787	

Table 3: T-test results

The p-value for this one-tail test is significantly less than the significance level of 0.05. This rejects our null hypothesis, and we can say with 95% confidence that the population in vulnerable age (<5, 65+ yr.) has higher influenza death rate than non-vulnerable age (5-64 yr.).

Remaining analysis and next steps

• Identify states with the largest population sizes for ages 65 and older

• Determine whether influenza occurs seasonally or throughout the entire year

• Create composition, statistical, spatial, and temporal visualizations

• Make a Final report/Presentation to give to stakeholders

Appendix I: Business Requirements Document

Goal

To help a medical staffing agency that provides temporary workers to clinics and hospitals on an as-needed basis. The analysis will help plan for influenza season, a time when additional staff are in high demand. The final results will examine trends in influenza and how they can be used to proactively plan for staffing needs across the country.

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

 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%)

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.

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.

Appendix 2: Hypothesis development

The following questions were considered to aid in the process of developing the hypothesis.

 Clarifying questions:

1. Which states have the most flu vaccinated populations?
2. Why some states more effected by influenza?
3. What staff-to-patient ratio is adequate for each state?

Funneling questions:

1. Which states has the most admissions from influenza?
2. Which states has the highest death rates?
3. Which states have the most residents' adults over 65 years?
4. Which states have the most children under 5 years?
5. Which states have the most pregnant women?
6. Which states have the most individuals with chronic medical conditions/immunosuppressive conditions?
7. Are flu records, admissions, or death rates from influenza higher in states with more vulnerable population?
8. Are flu records, admissions or death rates lower in states with flu vaccinated populations?
9. Are death rates or admissions from influenza lower among vaccinated vulnerable populations?
10. What staff-to-patient ratio is adequate?

Privacy and ethics questions:

1. Is there privacy law we need to adhere to related to collecting, storing, and analyzing medical patient data?

2. Is there privacy law we need to adhere to related to collecting, storing, and analyzing data from

3. Since some patients are minors, do we need to ask for their parents' consent to use their data?

Appendix 3: Profiles of data

Influenza Deaths Data Set

Variables	time -variant/-invariant	structured/ unstructured	qualitative/ quantitative	nominal/ordinal discrete/continuous
State	time -invariant	structured	qualitative	nominal
State Code	time -invariant	structured	qualitative	nominal
Year	time -invariant	structured	quantitative	continuous
Month	time -invariant	structured	qualitative	nominal
Month Code	time -invariant	structured	qualitative	ordinal
Ten-Year Age Groups	time -invariant	structured	qualitative	nominal
Ten-Year Age Groups Code	time -invariant	structured	qualitative	ordinal
Deaths	time -variant	structured	quantitative	discrete

US Census data set

Variables	time -variant/-invariant	structured/ unstructured	qualitative/ quantitative	nominal/ordinal discrete/continuous
Country	time -invariant	structured	qualitative	nominal
Year	time -invariant	structured	qualitative	continuous
Total population	time -variant	structured	quantitative	discrete
Male population	time -variant	structured	quantitative	discrete
Female population	time -variant	structured	quantitative	discrete
Under 5 years	time -variant	structured	quantitative	discrete
5-9 years	time -variant	structured	quantitative	discrete
10-14 years	time -variant	structured	quantitative	discrete
15-19 years	time -variant	structured	quantitative	discrete
20-24 years	time -variant	structured	quantitative	discrete
25-29 years	time -variant	structured	quantitative	discrete
30-34 years	time -variant	structured	quantitative	discrete
35-39 years	time -variant	structured	quantitative	discrete
40-44 years	time -variant	structured	quantitative	discrete
45-49 years	time -variant	structured	quantitative	discrete
50-54 years	time -variant	structured	quantitative	discrete
55-59 years	time -variant	structured	quantitative	discrete
60-64 years	time -variant	structured	quantitative	discrete

65-69 years	time -variant	structured	quantitative	discrete
70-74 years	time -variant	structured	quantitative	discrete
75-79 years	time -variant	structured	quantitative	discrete
80-84 years	time -variant	structured	quantitative	discrete
85 and over	time -variant	structured	quantitative	discrete

Influenza Laboratory Tests and Patient Visits data sets

This data source was not relevant for analysis because it is limited to the number of patients and clinics/laboratories who participating in survey. Contains inaccuracies (contains manual errors) and bias as it is survey.

Children Flu Shots data set

This data source is not relevant for analysis, because is limited by year 2017 only and one part of vulnerable population (children 19-35 month). Beside that as it is survey can contain bias as it is survey and time lag (no information about frequency of interview).