Preparing for Flu 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

Vulnerable populations are most likely to be hospitalized as a result of contracting the flu. If hospitals are appropriately staffed in areas that have a higher amount of vulnerable population, more deaths can be prevented.

Data Overview

Data Set 1: Influenza Deaths by geography, time, age, and gender

The data set contains data provided by the CDC's National Center for Health Statistics covering the years 2009 to 2017. The data is broken down by county, state, age groups. and deaths are categorized into age groups.

Data Set 2: US Census Data

This data set comes from the US Census Bureau based on census surveys collected covering the years 2009 to 2017. The data is broken down by county, state, male population, female population, and age groups.

Data Limitations

Data Set 1: Influenza Deaths by geography, time, age, and gender

There is a significant portion of the data listed as "suppressed" which is supposed to be less than 10 deaths occurred in that group. Entries marked as "suppressed" account for 81% of the data. An average of 5 was imputed to use this data set however the does impact accuracy. However, the focused age group (the vulnerable population) of 65+, "suppressed" only accounts for 39% of entries.

Data Set 2: US Census Data

Some of the numbers are estimates as the US Census is collected in 10 year intervals and therefore cannot be 100% accurate. Additionally, since the US Census is a survey relying on people to complete it accurately and return it, there is a lot of potential for human error.

Descriptive Analysis

	Influenza Deaths			US Census Data		
	65-74 Years	75-84 Years	85+ Years	65-74 Years	75-84 Years	85+ Years
Mean	151.31	268.95	476.54	448277.12	258539.43	109010.45
Standard Deviation	148.95	290.68	541.60	480494.33	281561.43	121607.14
Outlier Percentage	5%	4%	4%	7%	7%	6%
Correlation Variables	% of deaths vs total population			% of vulnerable population vs total population		
Correlation Coefficient	0.08					

There is a strong correlation between death rate and the size of the vulnerable population. As the ratio of vulnerable population increased, so did the rate of overall deaths. Because this has such a strong correlation, there should be a focus on areas with a high percentage of vulnerable populations when determining where to send additional staff.

Results & Insights

Null hypothesis: The vulnerable population has the same death rate as the rest of the population when contracting the flu.

Alternative hypothesis: The vulnerable population has a higher death rate when contracting the flu.

Type of test: One-tailed test—only interested in determining if vulnerable population death rates are higher than the rest of the population

Testing of these hypotheses resulted in rejection of the null hypothesis. This means the vulnerable population does have a higher rate of death when contracting the flu. With this information, the project can move forward basing staffing decisions on areas with higher amounts of vulnerable population in order to prevent more deaths overall.

Remaining Analysis & Next Steps

Analysis will continue by determining which states have the largest rates of vulnerable population in order to determine where to focus resources for allocation of medical staff during flu season. Current staffing levels need to be researched as well as what is considered “acceptable” staff to patient ratio. Other considerations would be, what areas of the country have historically been more affected by flu outbreaks. For example, if Florida has a high ratio of vulnerable population but does not generally get severely impacted by flu outbreaks, additional staff may not be needed there and that staff could be put to more use in areas more severely affected.

The data analysis will continue and composition, statistical, spatial, and temporal visualizations will be created as well as a final presentation.

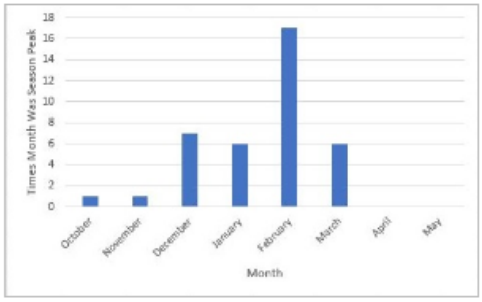
Appendix

Hypothesis Development:

- When is flu season?
    - https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)

      In temperate climates, seasonal epidemics occur mainly during winter, while in tropical regions, influenza may occur throughout the year, causing outbreaks more irregularly.
    - https://www.cdc.gov/flu/about/season/flu-season.htm

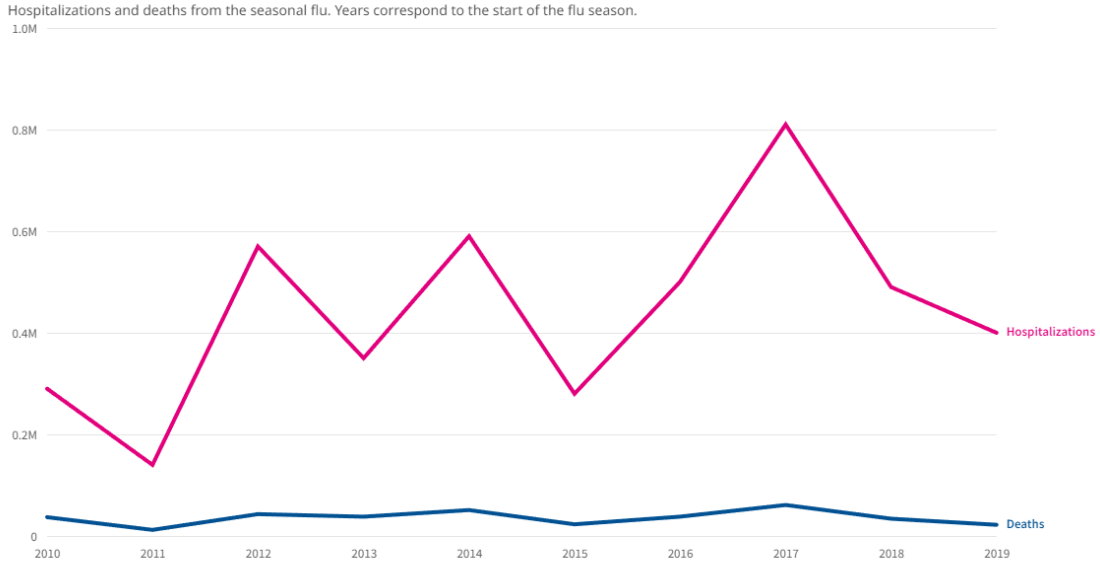
      Flu activity peak months in the U.S. from the 1982-1983 through 2019-2020 flu seasons

      In the United States, flu season occurs in the fall and winter. While influenza viruses spread year-round, most of the time flu activity peaks between December and February, but activity can last as late as May.
- Who is considered part of the vulnerable population?
    - https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)
      People at greater risk of severe disease or complications when infected are: pregnant women, children under 59 months, the elderly (65+ years), individuals with chronic medical conditions (such as chronic cardiac, pulmonary, renal, metabolic, neurodevelopmental, liver or hematologic diseases) and individuals with immunosuppressive conditions (such as HIV/AIDS, receiving chemotherapy or steroids, or malignancy).
- What are the rates of hospitalization and death from the flu?
    - https://usafacts.org/articles/how-many-people-die-flu/

About 61,000 people died in the 2017-2018 flu season, the worst in the past decade.

Hospitalizations and deaths from the seasonal flu. Years correspond to the start of the flu season.

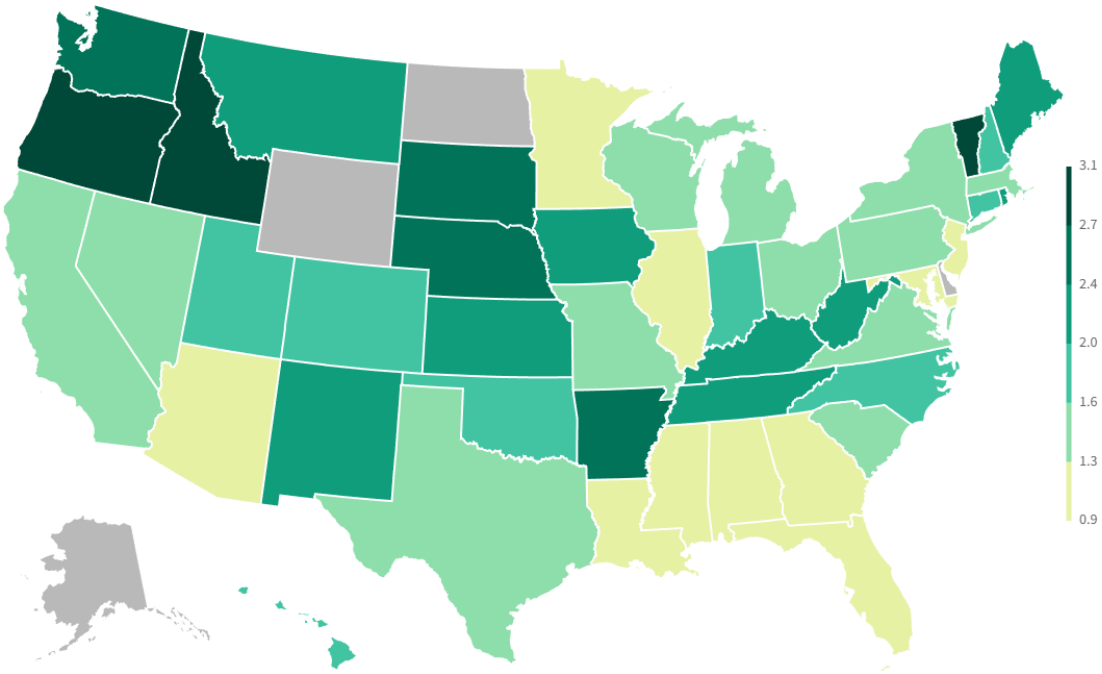


Sources: **Centers for Disease Control and Prevention.** *see more*

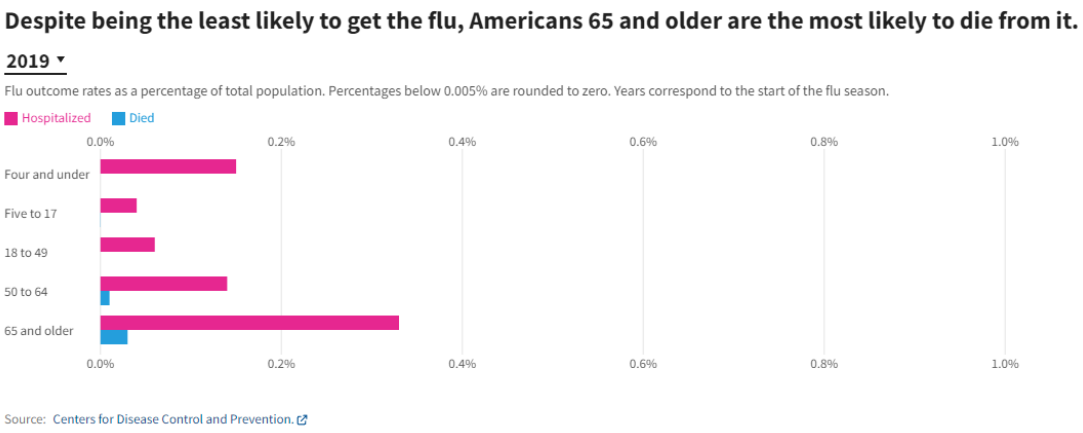
Vermont, Oregon, and Idaho had the highest flu death rates in 2019.

2019

Age-adjusted flu deaths per 100,000 people, according to death certificate data. The CDC cautions that flu deaths are severely underreported on death certificates.



Sources: **Centers for Disease Control and Prevention.** *see more*



Data Profiles:

Influenza Deaths Data Set

Variable	Time Variant/ Invariant	Structrured/ Unstructured	Qualitative/ Quantitative	Nominal/Ordnial or Discrete/Continuous
State	Time Invariant	Structured	Qualitative	Nominal
State Code	Time Invariant	Structured	Qualitative	Ordinal
Year	Time Invariant	Structured	Qualitative	Ordinal
Month	Time Invariant	Structured	Qualitative	Ordinal
Month Code	Time Invariant	Structured	Qualitative	Ordinal
Ten-Year Age Groups	Time Invariant	Structured	Qualitative	Ordinal
Ten-Year Age Groups Code	Time Invariant	Structured	Qualitative	Ordinal
Deaths	Time Variant	Structured	Quantitative	Discrete

US Census Data Set

Variable	Time Variant/ Invariant	Structrured/ Unstructured	Qualitative/ Quantitative	Nominal/Ordnial or Discrete/Continuous
County	Time Invariant	Structured	Qualitative	Nominal
Year	Time Invariant	Structured	Qualitative	Ordinal
Total population	Time Variant	Structured	Quantitative	Discrete
Male Total population	Time Variant	Structured	Quantitative	Discrete
Female Total population	Time Variant	Structured	Quantitative	Discrete
Under 5 years	Time Variant	Structured	Quantitative	Discrete
5 to 9 years	Time Variant	Structured	Quantitative	Discrete
10 to 14 years	Time Variant	Structured	Quantitative	Discrete
15 to 19 years	Time Variant	Structured	Quantitative	Discrete
20 to 24 years	Time Variant	Structured	Quantitative	Discrete
25 to 29 years	Time Variant	Structured	Quantitative	Discrete
30 to 34 years	Time Variant	Structured	Quantitative	Discrete
35 to 39 years	Time Variant	Structured	Quantitative	Discrete
40 to 44 years	Time Variant	Structured	Quantitative	Discrete
45 to 49 years	Time Variant	Structured	Quantitative	Discrete
50 to 54 years	Time Variant	Structured	Quantitative	Discrete
55 to 59 years	Time Variant	Structured	Quantitative	Discrete
60 to 64 years	Time Variant	Structured	Quantitative	Discrete
65 to 69 years	Time Variant	Structured	Quantitative	Discrete
70 to 74 years	Time Variant	Structured	Quantitative	Discrete
75 to 79 years	Time Variant	Structured	Quantitative	Discrete
80 to 84 years	Time Variant	Structured	Quantitative	Discrete
85 years and over	Time Variant	Structured	Quantitative	Discrete

Statistical Testing:

Non-Vulnerable Population Death Rate	Vulnerable Population Death Rate
0.10%	0.38%
0.10%	0.37%
0.10%	0.35%
0.10%	0.35%
0.10%	0.32%
0.11%	0.33%
0.10%	0.28%
0.09%	0.26%
0.09%	0.25%
0.03%	0.12%
0.03%	0.12%
0.03%	0.12%
0.03%	0.11%
0.03%	0.12%
0.03%	0.12%

t-Test: Two-Sample Assuming Unequal Variances

	<65 Years	*65+ Years*
Mean	0.000446506	0.001641391
Variance	2.15157E-07	3.69504E-06
Observations	459	459
Hypothesized Mean Difference	0	
df	511	
t Stat	-12.94593833	
P(T<=t) one-tail	1.18101E-33	
t Critical one-tail	1.647841009	
P(T<=t) two-tail	2.36202E-33	
t Critical two-tail	1.964617222	