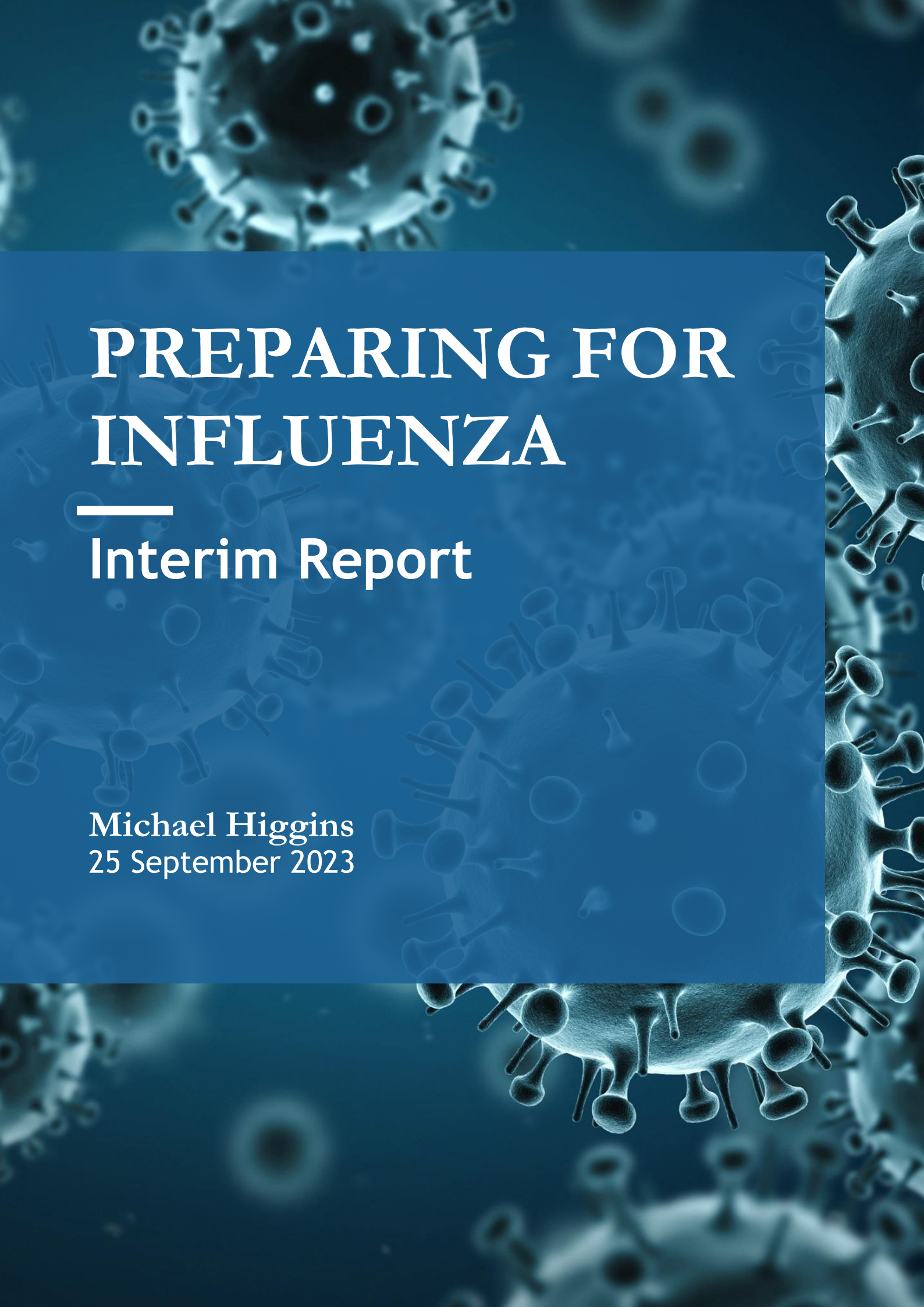

PREPARING FOR INFLUENZA

Interim Report

Michael Higgins
25 September 2023

Project Overview

Motivation

The United States has an influenza season where more people than usual suffer from influenza. 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 a person is in a higher age category, then their chance of mortality from influenza are increased.

Data Overview

CDC Influenza Deaths

These data show the number of influenza deaths each month from 2009 to 2017, organised by state and five-year age group.

US Census Population

These data list the estimated population in each US County for 2009 to 2017, organised by total population, sex, and ten-year age group.

Data Limitations

CDC Influenza Deaths

MISSING DATA

Data is suppressed for all persons under five-years-old.

If fewer than ten people died of influenza in any given category, the data are suppressed.

INACCURACIES AND BIAS

The data only records one source of death. Where influenza may be the pathogen that is recorded, an underlying health problem may have contributed to the patient's death. If there are multiple factors contributing to a death the coroner's recording of cause will be subjective and subject to potential pressures from their employer (for example to increase funding for specific diseases).

TIMELINESS

The data continues up to 2017 and does not have information after the Covid-19 pandemic. There are likely changes to demographics in vulnerable groups after a pandemic.

US Census Population Data

MISSING DATA

The data for some counties is incomplete and missing years.

INACCURACIES AND BIAS

The census is completed by survey of US residents, therefore could be completed inaccurately.

TIMELINESS

The census is completed every ten years, in years ending with zero. All data for years between these are estimates and will not be an exact value due to births, deaths and migration in individual counties.

Descriptive Analysis

Key Statistical Indicators by Age Group

Age Group	Mean	Standard Deviation
<65-years-old (population)	5,167,038	5,938,164
≥65-years-old (population)	806,989	886,050
<65-years-old (mortality)	438.4	889.9
≥65-years-old (mortality)	125.6	974.8

Vulnerable Deaths and Population

The correlation coefficient for 2017 normalised deaths per population over 65-years and normalised population over 65-years is -0.36. This represents a moderate negative correlation, where 0 represents no correlation and ±1 represents perfectly correlated data.

From this we can infer that influenza is impacted by the population of the State, where the smaller the population the larger the mortality rate (Deaths per 100,000 population).

Possible reasons for this could include lower access to healthcare in less populous regions, or fewer resources focus on vulnerable people's needs in areas with low levels of vulnerability. However, the weak correlation means that other factors may have greater influence on the rate of mortality for vulnerable people, for example such as the state's economic strength.

Results and Insights

Null Hypothesis

A person aged 65 or over has an equal or lower probability of death from influenza than that of someone under 65.

Alternative Hypothesis

A person aged 65 or over has a higher probability of death from influenza than that of someone under 65.

Hypothesis Testing

With greater than 95% confidence we can say that there is a significantly higher mortality rate (deaths per 100,000) for persons aged 65 and over, than those under 65. Therefore, we can reject the null hypothesis.

Analysis and Next Steps

As we are now confident that being aged over 65-years is a defining factor for mortality, and that age population and mortality negatively correlate, the following must be determined:

i. The states that have the highest population over 65-years.
ii. If there are different categories of vulnerability within the over and under 65 groups.
iii. When influenza season is, and if influenza season is variable by state.
iv. Investigate which counties have smaller populations and where they are, and how this will impact the distribution of healthcare staffing.

To complete these, and present to stakeholders, the following tasks are to be completed:

i. Create data visualisations:
    a. Graphs of state mortality by month comparing and highlighting trends, commonality, and seasonality.
    b. Mortality rate changes over time for age groups, to determine if vulnerability is increasing or decreasing.
    c. Heat maps of vulnerable population by state and normalised vulnerable population by state population.
    d. Categorising states by high and low levels of vulnerability need to build up a national picture.
ii. Determine conclusions and recommendations.
iii. Create a video report presentation.

Appendix

Project Brief

Hypothesis Development

The hypothesis was developed through asking the following clarifying questions: What fraction of people are considered vulnerable? Which states have the most vulnerable people? And what type of vulnerability is prevalent in each area (young/old/medical)? How does the vulnerability correlate with population?

Questions regarding vaccination rates and underlying medical conditions were also asked, but these data were either unavailable or suppressed due to ethical reasons surrounding privacy.

Data Overview

US CENSUS DATA

This is an external data source; the data is owned and provided by the US government's Bureau of the Census. As US Government data it can be verified as trustworthy data.

The collection of census data is legally mandated in the US constitution to be conducted every ten years. The data is collected through surveying each household.

Variable	Min	Max	Mean	St. Dev
Year	2009	2017	2012.9	2.6
Total population	41	10105722	107854.4	331600.3
Male Total population	23	4979641	53027.7	162821.5
Female Total population	15	5126081	54826.7	168822.5
Under 5 years	0	733897.125	6966.0	22515.5
5 to 9 years	0	665400.06	7034.9	21933.4
10 to 14 years	0	724111.83	7162.8	22311.5
15 to 19 years	0	753656.519	7504.6	23162.3
20 to 24 years	0	777987.834	7631.4	24325.1
25 to 29 years	0	831276	7427.8	26044.5
30 to 34 years	0	762619	7113.2	24401.7
35 to 39 years	0	753467.715	7010.4	23297.4
40 to 44 years	1.001	733897.125	7304.0	23452.5
45 to 49 years	0	704717.784	7615.5	23377.6
50 to 54 years	0	683886.54	7625.2	22590.0
55 to 59 years	2.976	628513	6920.8	19906.7
60 to 64 years	0	535357	5947.8	16633.4
65 to 69 years	0	415243	4591.6	12551.0
70 to 74 years	0	295420	3427.1	9315.6
75 to 79 years	0	215181	2624.9	7207.0
80 to 84 years	0	161647	1999.8	5670.4
85 years and over	0	177493	1949.8	5706.3

INFLUENZA MORTALITY DATA

This is an external data source. The data is collected and collated by the Centers for Disease Control and Prevention. As government data, this as a trustworthy data source.

The data is administrative data collected as part of the National Vital Statistics Cooperative Program.

Each of the U.S. states is required to record all deaths with cause. However, death certificates only list a single cause of death. This could create some inconsistencies within vulnerable populations with underlying health conditions.

	Minimum	Maximum	Mean	St. Dev
State Code	1	56		
Year	2009	2017		
Deaths	0	512	10.0	27.7

Results and Insights

Results of the two-sample t-test with a one-tailed test, assuming unequal variables of mortality per population of non-vulnerable and vulnerable people:

t-Test	Mortality Rate <65	Mortality Rate ≥65
Mean	0.21	1.26
Variance	0.05	0.19
Observations	51	51
Hypothesized Mean Difference	0	

df	73
t Stat	-15.5
P(T≤t) one-tail	4.16E-25
t Critical one-tail	1.67

t-Stat of -15.5 shows a high difference between mortality rates for the two groups, with a P(T≤t) value showing that the age variable is highly significant to these differences in mortality rate.

The 51 samples are the average mortality rate in each state for the period 2009-2017.