

MedCo Research

# Interim Report

Preparing for Influenza Season



This report gives an overview about the midterm findings of our project. We discovered a strong relationship between higher numbers of residents, older than 65 years, and a higher influenza deaths per state. Furthermore, it informs about the further analysis and the next steps.

## *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 population of residents above the age of 65 is high, then the influenza deaths are high as well.'*

### **Data Overview:**

*Influenza deaths, published by the CDC:*

It contains the death numbers of the population, separated by age group, and sorted by state and year.

*Population data, published by the US Census Bureau:*

It contains the population of specific age groups, sorted by state and year.

*Counts of influenza laboratory test results and hospital visits due to ILI (influenza-like illness):*

It contains the number of hospital visits due to ILI, the total number of hospital visits, and the data of flu tests, their total specimen, and the their results, sorted by state and year.

## Data limitations:

### *Influenza deaths, published by the CDC:*

Due to confidentiality constraints, death numbers below 10 are suppressed and therefore not available. In the data cleaning process those records were ignored.

### *Population data, published by the US Census Bureau:*

Since the census is done only once per decade, the rest of the numbers are estimates from the US Census Bureau.

### *Counts of influenza laboratory test results and hospital visits due to ILI (influenza-like illness):*

These data sets needed more manually interaction in the creating process and could therefore contain more human errors.

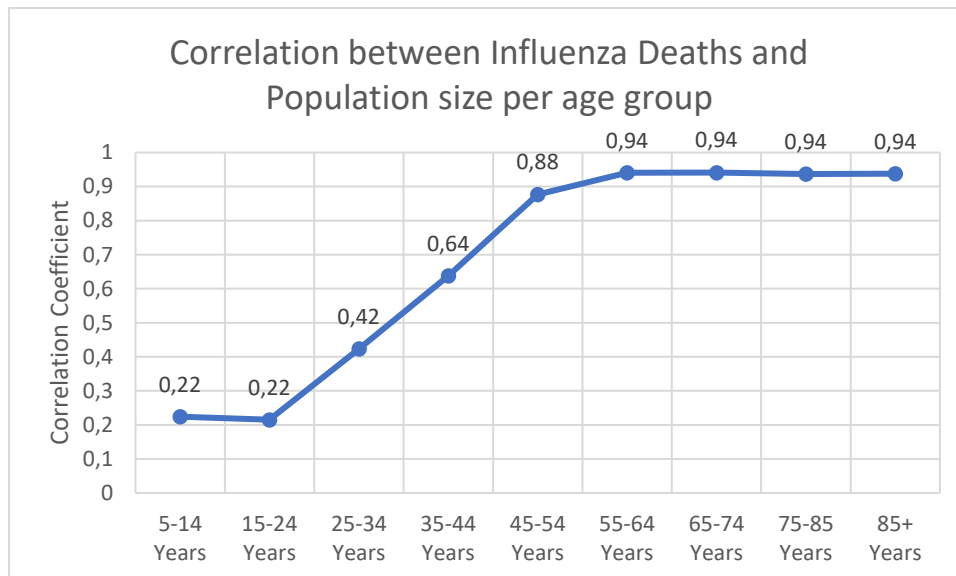
## Descriptive Analysis:

To check the hypothesis the two variables ('population of residents 65+' and 'number of influenza deaths') were tested for correlation. The numbers were gathered for every state and every available year and then compared.

	Population of Residents 65+	Number of influenza Deaths
Variance?	771479739446,10	1319833,10
Standard Deviation?	878339,1938	1148,84
Mean	801596	887,65
UL	2558274,529	3185,33
LL	0	0
Outlier Percentage	6,84%	4,27%

The correlation test resulted in a positive, strong correlation ( $r=0,95$ ). Therefore, we can assume an impact on the death numbers if the population of residents 65+ is high.

To put this in a better perspective, I tested every age group for correlation:



The graph excludes the 'below 5 Years' age group, due to the stated issues with the Influenza deaths data set, within the 'Data limitations' section. As we can see there's an increasing correlation with the population size of older age groups and the death numbers of said age group. Surprisingly this increase already starts at the age of 35+ to have a strong correlation. Further analysis may be needed. But for now, back to the main hypothesis.

## Results & Insights:

### *Null hypothesis:*

'The deaths of people above 65 is equal or lower than the deaths of people below 65.'

### *Alternative hypothesis:*

'The deaths of people above 65 is higher than the deaths of people below 65.'

	Deaths of Residents <65 due to Flu	Deaths of Residents >65 due to Flu
Mean	77,25	810,3974359
Variance	22579,77248	1021567,769
Observations	468	468
Hypothesized Mean Difference	0	
df	488	
t Stat	-15,5214865	

P(T<=t) one-tail	9,47E-45	
t Critical one-tail	1,647982077	
P(T<=t) two-tail	1,89316E-44	
t Critical two-tail	1,96483707	

The P-value is greatly below the significance level, therefore, I can reject the null hypothesis. The confidence level is at 95%. In conclusion, there is a significant difference in the number of deaths between residents of the age above 65 and below 65.

### Remaining analysis and next steps:

- The main age groups (65+ years) are known which require more medical attention, now we're going to use this data to determine in which states the need for the medical staff is greater.
- As already hinted, the vulnerability of the age groups already starts younger than in the hypothesis expected. This leaves room for further inspection and notice regarding the project task. How this information will be handled, is part of the further analysis and not determined yet.
- *Developing the visualisations in Tableau to show the findings.*
- *Creating a video presentation with the result for the stakeholders.*

### Appendix:

- The Excel files regarding the data set can be found in the cloud or under the following link: <insert link with excel files>
- Influenza deaths, published by the CDC
- Population data, published by the US Census Bureau
- Counts of influenza laboratory test results and hospital visits due to ILI (influenza-like illness)