# Influenza Death Data Sourcing

## DATA SET DESCRIPTIONS

### POPULATION DATA BY GEOGRAPHY: DATA SOURCING

This is an external data source. The medical staffing agency does not have this information, so it's relying on government data provided by the US Census Bureau. As this is government data, it can be verified as a trustworthy data source.

### POPULATION DATA BY GEOGRAPHY: DATA COLLECTION

The data is an interview/survey data collection method via internet, phone, or by paper questionnaire. Due to the source of data coming from the US Census Bureau, it can be assumed that it is a complete and accurate count of the population in each state. However, the data is only collected annually which can lead to a lag in timeliness. Depending on the survey method, there could be manual entry which could lead to errors.

### POPULATION DATA BY GEOGRAPHY: DATA CONTENTS

The data contains information about county & state population from 2009 to 2017 in terms of total population in relation to gender, and age broken down into 4 year increments.

### POPULATION DATA BY GEOGRAPHY: DATA RELEVANCE

This data shows the geographic and age spread of the population across the United States over multiple years. As this data set is directly related to the hypotheses for the 65+ population and 6 months to 5 years of age, this data set will be used in the analysis. There are limitations as it's only calculated annually, thus suffering timeliness but can be assumed to be a trustworthy and complete version of the data.

---

### INFLUENZA VISITS AND LAB TEST: DATA SOURCING

This is an external data source. The reporting comes from outpatient healthcare providers but the medical staffing agency does not have this information, so it's relying on government data collected by the Centers for Disease Control and Prevention (CDC). As this is government data, it can be verified as a trustworthy data source.

### INFLUENZA VISITS AND LAB TEST: DATA COLLECTION

Both data sets are survey data collected via manual entry on a weekly basis at the outpatient clinic level. Due to the initial input coming from manual entry at specific locations, there are possibilities for bias for funding, as well as typos and errors. However, since it's being sourced through the CDC, it adds an additional level of reporting resulting in higher validity and more likely a trustworthy source.

**INFLUENZA VISITS AND LAB TEST: DATA CONTENTS**
The first data set tracks patient visits, number of providers, and total patients seen by week and state from late 2010 to early 2019. Age was not included so another source of data will be needed. This reporting comes from 3,500 outpatient healthcare providers.

The second data set tracks weekly positive influenza results by subgroup by state and year between late 2010 and early 2015. This reporting comes from 100 public health providers and over 300 clinical laboratories located throughout the US.

**INFLUENZA VISITS AND LAB TEST: DATA RELEVANCE**
The first data set tracks patient visits, number of providers, and total patients seen by week and state from late 2010 to early 2019. This data is helpful in supporting the hypotheses in regards to vulnerable populations when looking at how many patients were seen in a certain geographical location. However, age was not included but it's a good starting place to understand the relationship between number of visits and how many staff are needed to adequately staff an outpatient clinic.

The second data set tracks weekly positive influenza results by subgroup by state and year. This data set is not relevant to the hypotheses as the subtype of influenza is not important to the overall project goal.

_____

**CHILDREN FLU SHOT DATA: DATA SOURCING**
This is an external data source. The National Immunization Surveys (NIS) provided the survey via phone interviews through the University of Chicago on behalf of the Centers for Disease Control (CDC). All flu shot information was verified with health providers. As there are many levels of reporting and third-party organizations, this results in higher validity and more likely a trustworthy source.

**CHILDREN FLU SHOT DATA: DATA COLLECTION**
The data is an interview/survey data collection method via phone interviews. Due to the survey being provided by NIS and being conducted via a third-party through the CDC, it can be assumed that it is a complete and accurate count of flu shot data for children between 6 months and 17 years. However, the data has limitations as all entries are manually entered which can lead to typos and errors. Depending on the random sampling picked and the return questionnaire from the provider, there could be a delay in data collection. This survey is also only collected on an annual basis which can lead to some time delays.

**CHILDREN FLU SHOT DATA: DATA CONTENT**

The data contains flu shot data for children 6 months to 17 years. It is categorized by geographic state, and contains family demographics including poverty level, race, and marital status.

**CHILDREN FLU SHOT DATA: DATA RELEVANCE**

The data contains flu shot data for children 6 months to 17 years. It is categorized by geographic state, and contains family demographics including poverty level, race, and marital status. As this data set is directly related to the hypotheses for the 6 months to 5 years of age vulnerable population, this data set will be used in the analysis.