Data Immersion

1.4 Sourcing the Right Data

Objective & Hypotheses

Objective: Determine when to send staff, and how many, to each state.

Hypotheses:

- If an individual is older than 65 years of age or less than 5 years of age, then that individual has a greater chance of hospitalization or death as a result of infection with the flu.
- If more staff is sent to the areas with more vulnerable people in these age brackets, then hospital and staff shortages can be lessened or avoided.

Population by Geography: US Census Bureau

US Census Bureau: Data Sourcing

This is an external data source because it comes from outside the staffing agency or hospital and clinics. These are not official annual population estimates, but they are demographic analysis estimates that are created by calculating the change in population relative to the most recent census population measures. This data comes from a government source, the US Census Bureau, so it is assumed this is a trustworthy data source.

US Census Bureau: Collection

The Census Bureau uses a demographic accounting process to produce the DA estimate of the total US population. First, the Census Bureau begins with birth records to get their first estimate, then the deaths during that time period are subtracted from births. Then migration or those living overseas as well as US Armed Forces population living overseas is accounted for in the estimate. Finally, Medicare records are used to estimate the population born before 1945 to calculate the most recent DA estimate.

US Census Bureau: Content

The data contains DA population estimates for each year by state and county by gender for each age cohort.

US Census Bureau: Data Relevance

This data is relevant to the objective and hypothesis because it would allow the calculation of death as a percentage of the population or age cohorts in each region, which could help evaluate the need for additional staff. Although this is not as current as it could be, it does span nine years from 2009 to 2017. However, historical data is always helpful when estimating future trends.

Influenza Visits Lab Tests by State: CDC

Influenza Visits and Lab Tests: Data Sourcing

This is an external data source because it comes from outside the staffing agency or hospital and clinics. The flu lab test data comes from collaborating WHO and National Respiratory and Enteric Virus Surveillance System (NREVSS) laboratories. Outpatient visits data comes from the US Outpatient Influenza-like Illness Surveillance Network (ILINet). The outpatient visits data is collected through a collaboration effort between the CDC, state and local health departments, and healthcare providers. Because it is a government source, it is assumed to be trustworthy; however, there are limitations to be discussed in other sections.

Influenza Visits and Lab Tests: Data Collection

There are two data sources here, the flu-like outpatient visits and the flu lab tests. For the outpatient visits data, the effort began in 1997-1998 flu season in which there were 250 providers. By 2010-2011 flu season, there were more than 3,000 providers. There are some limitations to this data that not all people with the flu will visit the doctor, so estimates are not very accurate. Also, the data includes flu-like symptoms, not necessarily flu-positive individuals. Finally, the other limitation is that children tend to be more symptomatic than adults, so there would be a large underestimation of flu-like illnesses of elderly and adult populations in general.

For the lab tests, at the start of the effort in 1997-1998 flu season, there were 43 state public health laboratories. Now all state public health laboratories participate. The limitations in this data include that each region and season accounts for different testing and triaging practices, so the percent positive rates between regions cannot be compared.

Also, for both data sets, not all people who have flu-like symptoms during a physician's visit or have a positive test for the flu will be reported to public health laboratory or government agency. Overall, the purpose of both data sets seem to be more of a marker of flu strains present in a season and a general marker for the beginning and ending of flu seasons and not for accurate counts of illness or positive flu tests.

Influenza Visits and Lab Tests: Data Content

The content of the outpatient visits by state, year, and week of the year. Along with patient visits, it shows the number reported to have flu-like symptoms and number of providers. It has age categories, but these are typically blank or null.

The content of the flu lab tests includes the state, year, and week of year along with total specimens tested for the flu, number of positive outcomes, and subtype information.

Influenza Visits and Lab Tests: Data Relevance

While in theory, a complete and accurate assessment of lab visits and positive flu tests by age could have value. However, the various limitations in this data, even stated by the CDC, is not meant to be able to compare regions for the number of infections or severity, so this data is not relevant for the purpose of obtaining accurate metrics for the extra staff needed in each region.

Flu Vaccination Rates in Children: CDC

Flu Vaccination Rates in Children: Data Sourcing

This data source is one of the National Immunization Surveys (NIS). The flu vaccinations rates of children ages 6 months to 17 years come from phone surveys. The surveys are sponsored and conducted by the National Center for Immunization and Respiratory Diseases (NCIRD) of the CDC and authorized by the Public Health Service Act. Also, administering clinic and vaccination data are given to the survey taker to verify. Because the data comes from a government source and data is verified or has a chance of being verified, it is assumed to be trustworthy. However, there are limitations in this data regarding this project to be discussed further.

Flu Vaccination Rates in Children: Data Collection

The data surveys are conducted by NORC at the University of Chicago under the direction of the CDC. The surveys are current, population-based, state, and local area estimates of vaccination coverage. The phone surveys interview parents and guardians of the children in all 50 states, the District of Columbia, and some US territories. Cell phone numbers are randomly selected and include one or more eligible child or teen from each household.

Flu Vaccination Rates in Children: Data Content

This data includes the region, age, and family demographics, including some socioeconomic data of the family along with if the child or children received a flu vaccine.

Flu Vaccination Rates in Children: Data Relevance

This is not likely relevant to the project, but it may be helpful supplemental information. It only includes data on children and not the elderly. Plus, this data does not directly focus on those likely to be hospitalized as a percentage of the population (in the hypothesis), but focuses on another issue, vaccination rates. This may be helpful, although indirectly, to determine staff needed to give vaccinations in areas with large population of children and low vaccination rates.

Influenza Deaths: CDC

This data was provided by CareerFoundry. I did not write.

Influenza Deaths: Data Sourcing

This is an external data source. The medical staffing agency doesn't have this information, so it's relying on government data. The data is provided by the Centers for Disease Control and Prevention (CDC) through their National Center for Health Statistics. As government data, you can verify this as a trustworthy data source.

Influenza Deaths: Data Collection

The data is administrative data collected as part of the National Vital Statistics Cooperative Program. Each of the U.S. states and territories is required to record all births, deaths, marriages, and divorces within their jurisdiction. Death records come from death certificates, in which a doctor codes the primary cause of death as "Influenza" or "Pneumonia" (ICD-10 codes J09-J18).

Because this data is part of the government's vital statistics program, it's similar to a census, meaning that you can assume a complete and accurate count of deaths. The one caveat, however, is that deaths on a death certificate only list one cause of death. This could create some discrepancies within vulnerable populations, such as those with AIDs—while the cause of death may be related to AIDs, their decline in health may have been initiated by influenza.

Influenza Deaths: Data Contents

The data 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*.

Influenza Deaths: Data Relevance

The data shows the geographic and monthly spread of influenza across the United States over multiple years. As it was collected via the government vital statistics program, you can assume that it's the most trustworthy and complete version of the data available.

Historical trends often mirror upcoming trends. For this reason, the historical influenza data can be used to predict future influenza seasons for planning purposes. The project objective asks when and where to send staff, which is something this data can address. Additionally, one of the project requirements is to prioritize vulnerable populations. The included demographic data can help illuminate vulnerable populations that require additional care when planning for influenza. For these reasons, this data set is critical to addressing your project objective.

You can also add relevancy based on your specific project hypothesis. Suppose you had a hypothesis that focuses on a specific vulnerable population (i.e., children under five years). Your data relevancy section would need to address whether the data source includes age as a variable.