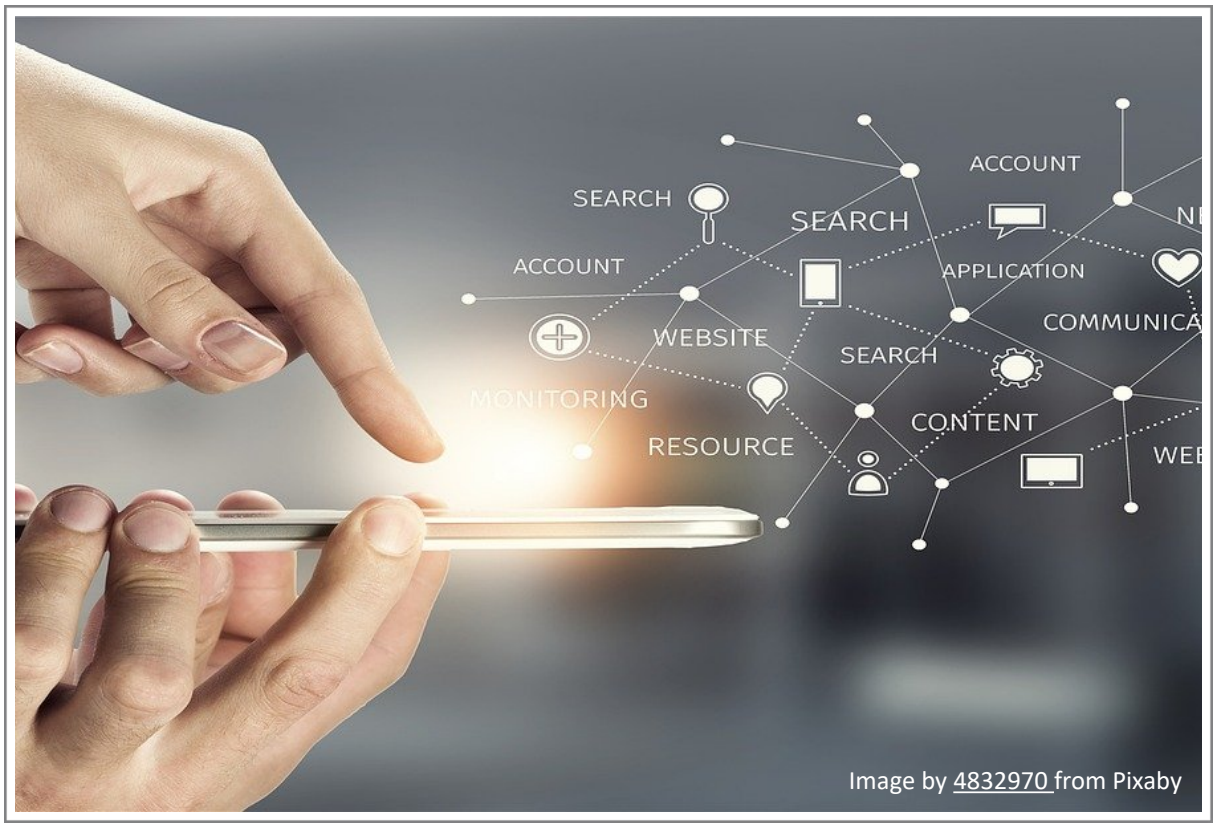
Med Staff Influenza Season Planning

- **Data Overview**



Elsa Ekevall

15 December 2021

Summary Of Data Sets

Project Overview:

Hospitals and clinics need additional staff to adequately treat particularly vulnerable populations. Determine when to send medical agency staff, and how many, to each state.

My hypothesis:

If states have a high proportion of vaccinated health care workers (nurses, physician assistants and doctors), then fewer vulnerable people will develop serious complications and end up in the hospital.

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.

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.

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.

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 prioritise 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.

Population Data by Geography:

The file doesn't mention which survey the population data by geography is from, therefore I am assuming it is a 2009 population estimate based on the Decennial Census of Population and American Community Survey.

Data Sourcing

This is an external data source provided by the U.S. Census Bureau, the US federal government's largest statistical agency. The U.S. Census Bureau's mission is to serve as the nation's leading provider

of quality data about America's people, places and economy. This is government data and can be verified as a trustworthy data source.

Data Collection

The data is administrative data collected as part of the American Community Survey (ACS) and the decennial census which takes place every 10 years (e.g. 2000, 2010 and 2020). The ACS a sample survey sent to about 3.5 million addresses in the 50 states, District of Columbia and Puerto Rico asks about age, sex, race, owner/renter status and topics such as education, employment, internet access, and transportation. The ACS selects a random sample of addresses focusing on quality geographic coverage. Households are asked to complete the survey and if they do not respond an additional survey questionnaire is sent, and if also not completed then personal follow-up visits ensue. The census provides an official count of every person living in the 50 states, District of Columbia, and the five U.S. territories. All U.S. residents must respond within a set period to the questionnaire and if not completed then personal follow-up visits ensue.

Responses to the ACS and Census are required by law and can be assumed to be a complete and accurate count/estimate of the population. Data must meet quality standards set by the agency and where bias may occur this is reported.

Data Contents

The data contains population counts for each county in the United States from 2009. County counts are broken into categories: total population, male total population, female total population and age (split into 5 year age groups e.g. under 5, 5 - 9 years, 10 - 14 years, etc).

Data Relevance

The population data by geography shows the population count by age and sex across the United States in the year 2009. It can be assumed to be the most trustworthy and complete version of the data available.

The data set can be used to normalise other data such as influenza deaths and is critical to addressing the project objective.

N.B. It was discovered during the analysis that the above data set was incomplete and it was replaced by a more comprehensive data set from the same source. The relevant columns from the American Community Survey - DPO5 ACS Demographic and Housing Estimates tables were selected and the data sets for the individual years (5 year estimates) were dowloaded in excel format from: https://data.census.gov/cedsci/table?q=Age%20and%20Sex&g=0400000US01,02,04,05,06,08,09,10,11,12,13,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,44,45,46,47,48,49,50,51,53,54,55,56&tid=ACSDP1Y2019.DP05

Influenza Laboratory Tests and Patient Visits:

Data Sourcing

This is an external data source provided by the Centers for Disease Control and Prevention (CDC) through CDC FLuView a web based application on the CDC website maintained by the National Center for Immunization and Respiratory Diseases (NCIRD). NCIRD whose mission "is the prevention of disease, disability, and death through immunisation and by control of respiratory and related diseases" is part of CDC, a major operating component of the Department of Health and Human Services in the U.S., This is government data and can be verified as a trustworthy data source.

Data Collection

The data is administrative data collected as part of the U.S. influenza surveillance system - "a collaborative effort between CDC and its many partners in state, local, and territorial health departments, public health and clinical laboratories, vital statistics offices, healthcare providers, hospitals, clinics, emergency departments, and long-term care facilities."

The data is part of the government's influenza surveillance system program and there may be inconsistencies due to the following reasons:

- influenza activity reporting by public health partners and health-care providers is voluntary;
- the national influenza surveillance system consists of nine complementary surveillance components in five categories, which includes reports from more than 350 laboratories, approximately 3,000 outpatient health care providers, vital statistics offices and the National Center for Health Statistics, research and healthcare personnel at FluSurv-NET sites, hospitals, long-term care facilities and influenza surveillance coordinators and state epidemiologists from all state, local, and territorial health departments;
- and weekly reporting of preliminary data may change as more data is received.

This survey data will be prone to error. There may be typos if entered manually or information may be incorrectly interpreted by the software when scanned. With such a large number of organisations involved on a voluntary basis there may also be inconsistencies in the way the data is classified. There may be some discrepancies due to Influenza-like-illnesses being reported that have not been confirmed in the laboratory and are in fact not influenza.

- *The data for the download Influenza Visits comes from 3,500 outpatient healthcare providers.*
- *The data for Lab Tests comes from 100 public health providers and over 300 clinical laboratories located throughout the United States and its territories.*

Both the Influenza Visits and the Lab Tests downloads are part of the survey data described above and not complete counts of all influenza visits or laboratory tests in the United States.

Data Contents

The data contains information on where, when, and what influenza viruses are circulating and can be used to determine if influenza activity is increasing or decreasing; as well as the impact it has on illness, hospitalisation and deaths. It does not directly provide the number of influenza illnesses.

- *The download Influenza Visits tracks patient visits to a medical provider for influenza. It counts the number of visits, number of providers, and total patients seen by week and state from late 2010 to early 2019.*
- *The download Lab Tests counts the number of positive influenza laboratory tests by week and state from late 2010 to early 2015.*

Data Relevance

The data provides a comprehensive system for influenza surveillance across the United States over multiple years. As it was collected via the government influenza surveillance system program, it can be assumed that it's the most trustworthy and the most complete version of the data available.

Historical trends often mirror upcoming trends and the historical influenza data can be used to predict future influenza seasons for planning purposes. The project overview asks when and where to send staff, and this data can address where and when influenza is circulating and the impact it has on illness, hospitalisation and deaths. This data set is critical to addressing the project objective.

Children Flu Shots:

Data Sourcing

This is an external data source provided by the Centers for Disease Control and Prevention (CDC) through their National Center for Immunization and Respiratory Diseases (NCIRD). NCIRD whose mission "is the prevention of disease, disability, and death through immunization and by control of respiratory and related diseases" is part of CDC, a major operating component of the Department of Health and Human Services in the U.S., This is government data and can be verified as a trustworthy data source.

Data Collection

The data is administrative data collected as part of the National Immunization Surveys (NIS). NIS is a group of telephone surveys that use two phases of data collection to obtain vaccination information for a large national probability sample of young children. First a randomly sampled telephone survey identifies households with children aged 19-35 months for household interviews with parents or guardians, and if the parent or guardian provides consent and contact information this is followed by a provider record check survey to obtain their provider-reported vaccination history. *The University of Chicago runs the surveys on behalf of the Centers for Disease Control (CDC).*

The data will have some bias as it is self-reported by only the parents who wish to take part. However, the data collected from the health care providers can be considered accurate. There will also be sampling and non sampling errors. There may be typos if entered manually or information may be incorrectly interpreted by the software when scanned. The data will have been checked and cleaned then weighted and adjusted to take into account any bias.

Data Contents

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

Data Relevance

The data provides national, state, and selected local area estimates of vaccination coverage rates for U.S. children 19–35 months (NIS-Child) and for U.S. adolescents 13–17 years (NIS-Teen) from 2015 to present. As it was collected via the government NIS program, it can be assumed that it's trustworthy.

Historical trends often mirror upcoming trends and the historical vaccination data can be used to predict future vaccinations for planning purposes. The project overview mentions prioritising vulnerable populations and children under 5 are one of these vulnerable groups. However on its own this vaccination data is less reliable.

The CDC use this data and other sources to provide influenza vaccination coverage estimates for the general population, health care personnel, nursing home residents, and pregnant women on the FluVaxView website. The other sources include the National Immunization Survey-Flu (NIS-Flu), National Health Information Survey (NHIS), the Behavioral Risk Factor Surveillance System (BRFSS), the Pregnancy Risk Assessment Monitoring System (PRAMS), the Minimum Data Set (MDS), and Internet panel surveys. These influenza vaccination coverage estimates based on a combination of data sources provide a more comprehensive view and will be critical to addressing the project objective. In particular the Vaccination Coverage Among Health Care Personnel data set based on national opt-in survey sources will be critical in addressing my hypothesis.

Hospitalisation Rates:

Data Sourcing

This is an external data source provided by the Centers for Disease Control and Prevention (CDC) through CDC FluSurv-NET Interactive a web based application on the CDC website maintained by the Influenza Hospitalization Surveillance Network (FluSurv-NET). This is government data and can be verified as a trustworthy data source.

Data Collection

The data is administrative data collected as part of the Influenza- Associated Hospitalisation Surveillance Network (FluSurv-NET). Laboratory-confirmed influenza-associated hospitalisation rates are collected through population-based surveillance for laboratory-confirmed influenza-associated hospitalisations among children (persons younger than 18 years) and adults through a network of acute care hospitals in 14 states. A case is defined as a person who is a resident in a defined FluSurv-NET catchment area and tests positive for influenza by a laboratory test ordered by a health care professional within 14 days prior to or during hospitalisation. Rates from 2010-11 onwards reflect cases hospitalised during influenza season (October 1 - April 30).

The data is part of the government's influenza-associated hospitalisations surveillance network program and there may be inconsistencies due to the following reasons:

- the information might not be generalisable to the entire country;
- The rates provided are likely to be underestimated as influenza-associated hospitalisations might be missed due to test availability and provider or facility testing practices;
- although staff are trained the surveillance network consists of over 70 counties in the 10 Emerging Infections Program (EIP) states and four additional states through the Influenza Hospitalization Surveillance Project (IHSP);
- and weekly reporting of preliminary data may change as more data is received.

This survey data will be prone to error. Data are collected for a random sample of cases stratified by age and surveillance site. There may be typos if entered manually or information may be incorrectly interpreted by the software when scanned.

The data for Hospitalisation Rates comes from over 70 counties (more than 29 million people and approx. 9% of US population) in the 10 Emerging Infections Program (EIP) states and four additional states through the Influenza Hospitalization Surveillance Project (IHSP).

- Participating states include: California, Colorado, Connecticut, Georgia, Iowa, Maryland, Michigan, Minnesota, New Mexico, New York, Ohio, Oregon, Tennessee and Utah.

Data Contents

The data contains weekly hospitalisation rates calculated as the number of residents of a defined area who are hospitalised with a positive influenza laboratory test divided by the total population within the defined area.

- The data contains the Influenza Hospitalization Rates (per 100,000 population) - cumulative rate and weekly rate from late 2010 to early 2018 from over 70 counties by state, year, week and age category.

Data Relevance

As it was collected via the government FluSurv-NET program, it can be assumed that it’s trustworthy.

Historical trends often mirror upcoming trends and the historical hospitalisation rates data can be used to predict future hospitalisations for planning purposes. Although this data is less reliable in the absence of other data it should be used with the caveat that it may not be generalisable and rates are likely to be underestimated.