**Project Description:**

Over the course of each flu season the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC) in conjunction with public health organizations worldwide monitor flu severity and ascertain the prevalent strains. According to data provided by the World Health organization, seasonal respiratory illnesses are linked to approximately 650,000 deaths each year worldwide. In the United States the flu is responsible for approximately 42,000 deaths per year and ranks 13th most frequent cause of death. As such, influenza is the subject of extensive public health surveillance and mediation efforts.[1, 2]

During a typical season flu outbreak, it may be possible and useful to track the geographic and temporal origin of the responsible virus. Knowledge of the general pattern of viral infection and spread on both the worldwide and local levels may make it possible to better plan public health interventions such as vaccine planning. Strains that are responsible for the largest disease burden are typically H1N1 (such as A/Victoria/4897/2022 (H1N1)pdm09-like virus), H3N2 (such as A/Darwin/9/2021 (H3N2)-like virus), and a B strain (B/Austria/1359417/2021 (such as B/Victoria lineage)-like virus).[3, 4] My project will be to attempt to track one or more of these strains from the beginning of the flu season, typically week 38-40 of the year through the end of the flu season, typically week12 to 14 of the year.

**Available Data:**

Detailed data on flu strains, outbreaks, and general epidemiological information is available via several sources. The CDC maintains a database of information about influenza outbreaks in the United States (<https://www.cdc.gov/flu/index.htm>). Data on these infectious diseases is available from sites such as GISAID (<https://gisaid.org/>) and the Global Health Data Exchange (<https://ghdx.healthdata.org/>). These databases are highly detailed and can provide data that includes information about lineage and geographic data. The WHO also maintains a database of influenza vital statistics. (<https://www.who.int/teams/global-influenza-programme>) Genetic data is available from sites such as the NCBI Virus database (<https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/>), and the NCBI Influenza Database (<https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database>).

**Data Retrieved and Preliminary Workup:**

Retrieved data and preliminary workup of that data is available in a public Github repository:

<https://github.com/ClarkeMiller/Miller-EPID-8200-Project.git>

Data includes:

* Data from the WHO concerning flu incidence for the defined WHO regions.
  + **FluNet Dataset:** A a global web-based tool for influenza virological surveillance first launched in 1997. The virological data entered into FluNet, e.g. number of influenza viruses detected by subtype, are critical for tracking the movement of viruses globally and interpreting the epidemiological data. The data at country level are publicly available and updated weekly.
  + **FluID Dataset:** A global platform for data sharing that links regional influenza epidemiological data into a single global database. The platform provides connections between existing databases and can also be used to directly enter data by the national focal points through a web-based interface. It complements the FluNet database.
  + **Data Dictionary:** A codebook for the data from FluNet and FluID.
* Sequence Data
  + The initial sequence set was downloaded from the NCBI Influenza Database. The query was restricted as follows:
    - Nucleotide
    - Human
    - USA
    - PB1
    - Week 42, 2019 to Week 52, 2019
    - Restricted to complete sequences.
  + This resulted in a dataset of 906 sequences, 383 of these were unique sequences.

Preliminary workup includes:

* Initial import and cleaning of XXX data.
* Trends per WHO region examined.
* Initial (read down and dirty) multiple sequence alignment and phylogenetic tree for the 383 unique sequences using MUSCLE online tool.

Preliminary Results (Proof of Concept):

Preliminary phylogenetic tree: A preliminary phylogenetic tree was prepared using 383 unique DNA sequences with the online MUSCLE alignment tool available at the NCBI Influenza resource site and at <https://www.ebi.ac.uk/jdispatcher/msa/muscle>. This work shows that the earliest reported flu strain in the US for this set of data from the 2019 season starts in Florida. The progression appears to be throughout the US from the southeastern united states. Likewise, the dates/locations of the reported sequences also correspond to this pattern.

As further proof of concept data was downloaded from the WHO influenza monitoring effort. The data was cleaned and distilled to WHO region, county, date, total influenza cases, total influenza A cases, and total influenza B cases. Total influenza cases for each WHO region were graphed vs. time. The resulting graphs showed a clear progression of outbreak starting in the South East Asia Region (SEAR) in July followed by subsequent outbreaks every 4 weeks or so in the Eastern Mediterranean Region (EMR), the Western pacific Region (WPR) the Americas Region (AMR) and finally the European Region (EUR). (**Figure 1.**) An interesting outcome from examining these graph is an apparent well defined dependence of peak flu cases in AMR and EUR stimulated by the outbreak in WPR. The results of the EUR graph seem to show a bimodal trend in the number of peak cases influenced by both WPR and AMR cases. The number of cases in AMR also shows bimodal characteristics, possibly influenced by the first EUR peak.

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated

**Figure 1:** Regional pattern of flu transmission. The figure above shows a graph of total flu cases per WHO region throughout the 2023-2024 flu season. The results show a clear domino effect in the timing of peak flu cases.

A graph of different sizes and colors

Description automatically generated with medium confidence

**Figure 2.** Data for WPR, AMR, and EUR created using online tools at the WHO Influenza website. These graphs match the results generated in R.

1. Nair, H., et al., *Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis.* The Lancet, 2011. **378**(9807): p. 1917-1930.

2. Thompson, W.W., et al., *Estimates of US influenza-associated deaths made using four different methods.* Influenza and Other Respiratory Viruses, 2009. **3**(1): p. 37-49.

3. Fischer, W.A., 2nd, et al., *Global burden of influenza as a cause of cardiopulmonary morbidity and mortality.* Glob Heart, 2014. **9**(3): p. 325-36.

4. Gordon, A. and A. Reingold, *The Burden of Influenza: a Complex Problem.* Current Epidemiology Reports, 2018. **5**(1): p. 1-9.