Polynomial Regression

for CDC data with Application for Predicting RSV in USA

Arun & Zhongyi

11/18/2022

Table of contents

Preface

This is the book to highlight our proposal of new statistical regression tool, Polynomial Regression Model, to investigate the dynamic of RSV in USA.

We have seen a dramatic growth of RSV cases this year, but there aren't any articles founded yet to give prediction of the RSV cases. Based on the previous researches and the curve of RSV data current analysis has been carried out.

1 Introduction

Polynomial Regression for CDC data with Application for Predicting RSV in USA

1.1 Background

Respiratory Syncytial Virus (RSV) was discovered in 1956 and has been recognized as one of the most common causes of childhood illness. RSV symptoms usually look like a common cold, but it can be serious, leading to bronchiolitis (inflammation of the small airways in the lung) and pneumonia, especially for infants and older adults.

According to the Centers for Disease Control and Prevention, RSV results in around 58,000 annual hospitalizations and 100 to 300 deaths among children under 5.[1] In most regions of the United States, RSV circulation starts in the fall and peaks in the winter. With mask-wearing and physical distancing for COVID-19, there were fewer cases of RSV in 2020. However, RSV cases began to increase in spring 2021 when safety measures relaxed with the arrival of COVID-19 vaccines. This year, RSV in multiple U.S. regions are nearing seasonal peak levels. [2] It is unclear to anyone where RSV will lead us, a reliable prediction model for RSV data is very crucial for policymakers to implement fast actions to curb the spread of the infection, and for public health professionals to understand trends in virus circulation, estimate disease burden, and respond to outbreaks.

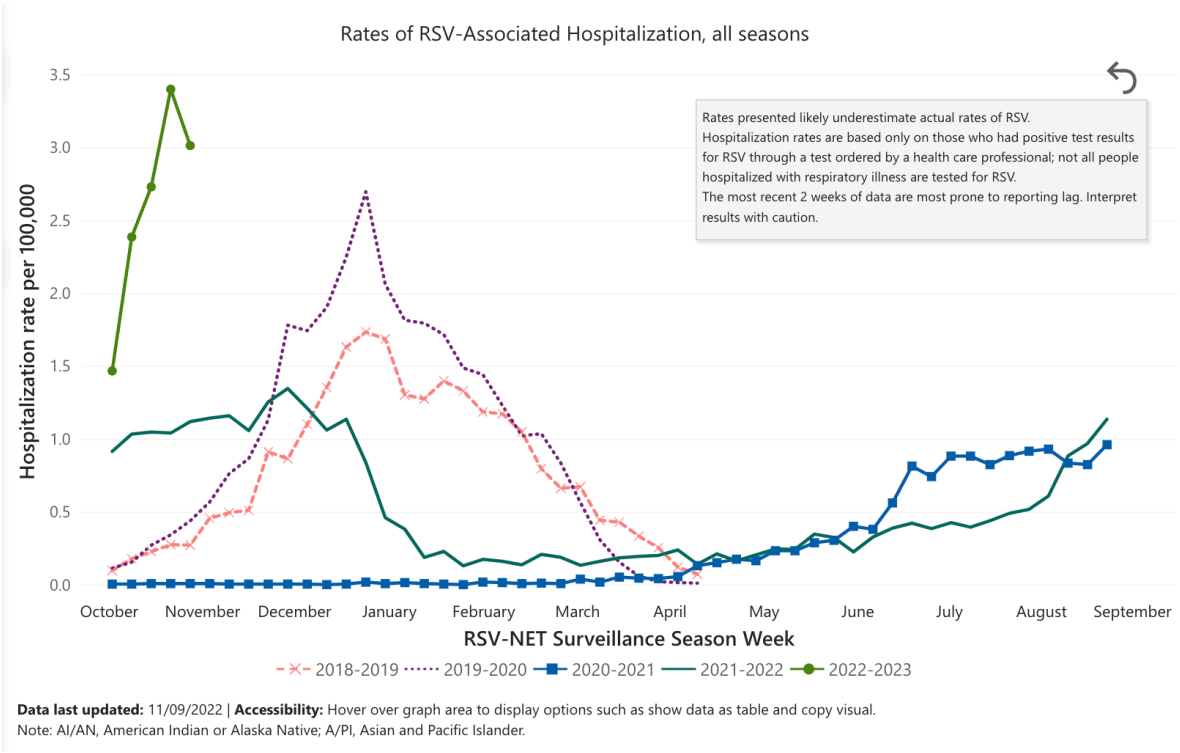

Figure 1.1 RSV-NET Interactive Dashboard | CDC

1.2 Literature Review

Respiratory syncytial virus (RSV) infection trend has gained many researchers' concerns globally. Thongpan, Ilada etc. applied multivariate time-series analysis to show the possible prediction of RSV activity based on the climate in Thailand. [3] The other researchers tracked RSV through internet search engine data. Oren, Eyal etc. highlighted the use of search filters and domain adaptation techniques to assist in identifying spread of both local and more widespread RSV transmission where comprehensive epidemiological data is not easy to collect.[4] Manuel, Britta etc. applied logistic regression to develop a prediction model and developed a web-based application to predict the individual probability of RSV infection.[5] Furthermore, researchers are using different modeling approaches to predict the RSV trend. A research done by Reis, Julia etc. tried to built a real-time RSV prediction system using a susceptible-infectious-recovered (SIR) model in conjunction with an ensemble adjustment Kalman filter (EAKF) and 10 years CDC data[6]. Bayesian stochastic susceptible-infected-recovered-susceptible (SIRS) model was presented by Corberán-Vallet etc. to understand RSV dynamics in the region of Valencia, Spain. However, this continuous-time deterministic model is not suitable when the initial number of infected individuals is small[7].

Leecaster, Molly etc. use simple linear regression to explore the relationship between three epidemic characteristics (final epidemic size, days to peak, and epidemic length) and exponential growth calculated from four weeks of daily case data. They find out exponential growth was correlated to epidemic characteristics[8].

1.3 Problem Statement

We have seen a dramatic growth of RSV cases this year, but there are not any articles founded yet to give prediction of the RSV cases now. Based on the previous researches and the curve of RSV data, we propose a new statistical regression tool—**Polynomial Regression Model** — to investigate the dynamics of RSV in USA.

2 Methodology

The data set for this research is from **RSV Hospitalization Surveillance Network (RSV-NET)** (one of CDC research and surveillance platforms) ,which conducts population-based surveillance system for laboratory-confirmed COVID-19, RSV, and influenza-associated hospitalizations in the US among children younger than 18 years of age and adults.[8]

RSV-NET has been collecting RSV-associated hospitalizations in adults and children since 2018-2019 season from 58 counties in 12 states, including California, Colorado, Connecticut, Georgia, Maryland, Michigan, Minnesota, New Mexico, New York, Oregon, Tennessee, and Utah. Almost 9% of the U.S. population is covered and reported by the RSV-NET.

2.1 About this Dataset

1. **Time frame:** In season 2018-2019, 2019-2020, data collected is from October 1 to April 30. In season 2020-2021, 2021-2022, 2022-2023, data collected is from October 1 to October 1 next year.
2. **How an entry is made in data set:**

   A case is defined by laboratory-confirmed RSV in a person who Lives in a defined RSV-NET surveillance area AND Tests positive for RSV within 14 days before or during hospitalization. Evidence of RSV infection can be obtained through several laboratory tests.
3. **Variables:**

   In the original data set, it has 8 columns which are described in below table

Column Name	Description
State	Represents the state name from which data was collected. Entire state means all 9 states considered in this dataset.
MMWR Year	Represents Year
MMWR Week	Represents week of that year. MMWR - *Morbidity and Mortality Weekly Report,* is prepared by the Centers for Disease Control and Prevention (CDC).