# RESEARCH PROPOSAL

## TEAM 5- DATA 602 (SPRING - 2024)

**(Raja Pavan Kalyan, Bala Sai Santhoshi, Venkata Sai Swaroop, Kesari Nandan and Ruchita Das)**

## TITLE: PREDICTIVE MODELING OF RESPIRATORY VIRUS HOSPITALIZATION TRENDS IN THE USA : A STUDY

### Introduction

The COVID-19 pandemic has posed unprecedented challenges to healthcare systems worldwide, with pediatric populations emerging as a significant concern due to evolving disease dynamics and potential long-term implications. Given the circumstances, precise prediction of pediatric COVID-19 hospitalizations has become crucial to allocate resources proactively and devise efficient healthcare management strategies. This literature review aims to examine existing research efforts in pediatric COVID-19 hospitalization forecasting, identify key findings, and highlight gaps in knowledge to justify further exploration in this critical area.

### Literature Review

Different kinds of studies have been conducted regarding COVID-19 hospitalization. One such study analyzed COVID-19-related hospitalizations among children and adolescents in the US revealing a notable increase in hospitalization rates, particularly after the emergence of the Delta variant. Unvaccinated adolescents aged 12-17 years were significantly more likely to be hospitalized compared to fully vaccinated counterparts [1].

Another study aimed to evaluate racial disparities in COVID-19 hospitalization and ICU admission among children in the United States. Using pediatric COVID-19 data from the CDC in 2020, the analysis revealed significant racial differences, with Black/African American children having a higher likelihood of hospitalization and ICU admission compared to White children. Asian and American Indian/Alaska Native children also showed increased risk. These findings underscore the need for health equity initiatives to address disparities in pediatric COVID-19 outcomes [2].

A similar study has investigated the epidemiological characteristics and transmission patterns of pediatric COVID-19 cases in China, comprising 2,135 cases reported from January 16, 2020, to February 8, 2020. Findings revealed that children of all ages were susceptible to COVID-19, with most cases being asymptomatic or mild. Transmission spread rapidly from Hubei province to surrounding regions. Future studies could longitudinally analyze pediatric COVID-19 cases, identify risk factors for severe outcomes, conduct geospatial analysis to identify high-risk areas, explore transmission dynamics, and conduct comparative studies with adult cases to understand age-specific differences [3].

The findings of the predictive modeling are very important which can emphasize the urgent need for comprehensive pandemic response planning to support worldwide. Keeping this in mind a study was aimed to model the increase in children affected by COVID-19-associated orphanhood and caregiver death during the 6 months following April 30, 2021. By utilizing updated mortality and fertility data, the study estimated a 90.0% increase in affected children, totaling approximately 5.2 million globally [4].

While there is a limited number of strategies or predictive models explored in research on pediatric and adolescent COVID-19 hospitalizations, our goal is to create a predictive model that can contribute to the development of effective strategies within the healthcare system of the USA. By utilizing the data that is continuously updated and controlled by the Centers for Disease Control and Prevention (CDC) under the program code of HHS: Immunization and Respiratory Diseases and falls under the category of Public Health Surveillance.

## Objective

This research seeks to devise predictive frameworks for pediatric COVID-19 hospitalizations utilizing CDC data governed by the HHS Immunization and Respiratory Diseases program, categorized under Public Health Surveillance. Employing advanced predictive analytics, the aim is to forecast pediatric COVID-19 admissions across diverse age groups, facilitating proactive healthcare resource allocation and management. Furthermore, these models will augment our understanding of COVID-19 hospitalization dynamics, aiding in the enhancement of preparedness strategies for addressing the pandemic.

## Dataset

Taken from https://data.cdc.gov/Public-Health-Surveillance/Respiratory-Virus-Response-RVR-United-States-Hospi/9t9r-e5a3/about_data

This dataset {consists of Rows **84.1K** and **115** Columns} and encompasses hospitalization data and metrics at various levels of aggregation, including national, HHS regional, and state/territory levels. The data are sourced from reports to the CDC's National Healthcare Safety Network, which tracks trends in healthcare system strain, capabilities, and local disease prevalence across roughly 6,000 hospitals in the United States. The information provided by hospitals to NHSN and featured in this dataset consists of aggregated counts, detailing metrics related to hospital admissions, as well as occupancy levels for inpatient and ICU beds. Specifically, the dataset includes fields related to new admissions of pediatric patients confirmed to have COVID-19, categorized by age groups: 0-4 years, 5-11 years, and 12-17 years. It's important to note that hospitals were not required to report data for these age-specific admissions until February 2022. As a result, the dataset contains data for these specific fields starting from March 1, 2022, to accommodate the initial reporting delay.

**The most relevant fields in the dataset are:**

- **adm_00_04_covid_confirmed:** Admissions of patients aged 0-4 years with confirmed COVID-19.

- **avg_adm_00_04_covid_confirmed:** Average admissions of patients aged 0-4 years with confirmed COVID-19.

- **avg_adm_00_04_covid_confirmed_per_100k:** Average admissions per 100,000 population of patients aged 0-4 years with confirmed COVID-19.

- **adm_05_11_covid_confirmed**: Admissions of patients aged 5-11 years with confirmed COVID-19.

- **avg_adm_05_11_covid_confirmed:** Average admissions of patients aged 5-11 years with confirmed COVID-19.

- **avg_adm_05_11_covid_confirmed_per_100k**: Average admissions per 100,000 population of patients aged 5-11 years with confirmed COVID-19.

- **adm_12_17_covid_confirmed**: Admissions of patients aged 12-17 years with confirmed COVID-19.

- **avg_adm_12_17_covid_confirmed**: Average admissions of patients aged 12-17 years with confirmed COVID-19.

- **avg_adm_12_17_covid_confirmed_per_100k**: Average admissions per 100,000 population of patients aged 12-17 years with confirmed COVID-19.

## Methodology (Results of Basic EDA)

**Target Variable**- admissions_all_covid_confirmed, this is so because it measures an outcome of interest that can be predicted based on other variables directly—the number of confirmed COVID-19 admissions. It is Numeric and Continuous.

**Potential Features**:

**Jurisdiction** - Because of differences in healthcare access, population density, and public health policies, the geographic area may have an impact on COVID-19 admission rates.

**Collection_date -** Seasonal factors and time-related patterns may have an impact on COVID-19 admissions.

**Number_hospitals_reporting_today -** The reported admissions on any day may vary depending on the number of hospitals reporting.

**Total_hospitals -** In each jurisdiction, the number of hospitals may have an impact on the ability to admit and report COVID-19 cases.

To find important missing data, we have computed missing value percentages for every column in a dataset. To simplify the analysis and concentrate on more comprehensive data, columns with more than 5% of missing values are flagged for removal. Specific imputation techniques are used for the remaining columns, where the column median is used to fill in any missing values. The integrity of the analysis is ensured by this thorough method of data preservation that also preserves the quality and integrity of the dataset.

## Conclusion

The foremost ones are analyses like the trends of COVID-19 admissions and categorizing admissions as Low, Medium, High, and Very High. These offer insight into how the pandemic has taken place and its different levels of stress on healthcare services. Such findings are essential in health planning as they allow for a prioritization of beds, personnel, and medical requirements during peak periods. Furthermore, correlation analysis reveals linkages between various healthcare indicators which may also indicate co-existing factors that would help guide policies or healthcare strategic steps. Moreover, the distribution analysis helps in understanding how common admission rates are needed during real-time planning resources.

The “COVID-19 Admissions Over Time” analysis is probably the most important followed by “Categorization of Daily Admissions”. Their purpose is to show how COVID-19 has changed over time and give a basis for determining its severity. This becomes an invaluable framework for healthcare provisioning in real-time, emergency planning as well as public health policy by providing a data-driven basis for informed choices regarding resource allocation, policy enactment, and public health intervention. It helps maximize medical care delivery through future surges with greater accuracy while mitigating the impact of the pandemic through specific interventions.

## References -


1. Hospitalizations Associated with COVID-19 Among Children and Adolescents - COVID-NET, 14 States, March 1, 2020-August 14, 2021, Miranda J Delahoy, et al. MMWR Morb Mortal Wkly Rep, 2021 Sep 10;70(36):1255-1260.
2. Black-White Risk Differentials in Pediatric COVID-19 Hospitalization and Intensive Care Unit Admissions in the USA, Laurens Holmes Jr, J Racial Ethn Health Disparities, 2023 Jun;10(3):1187-1193.
3. Epidemiology of COVID-19 Among Children in China, Yuanyuan Dong, Xi Mo, Yabin Hu, Xin Qi, Fan Jiang, MD, Zhongyi Jiang, Shilu Tong, Journal of American Academy of Pediatrics, vol145, issue 6.
4. Global, regional, and national minimum estimates of children affected by COVID-19-associated orphanhood and caregiver death, by age and family circumstance up to Oct 31, 2021: an updated modeling study, H Juliette T Unwin, Lancet Child Adolesc Health. 2022 Apr; 6(4): 249–259.