

# Analysis of United States Medical Benchmark for Disease Outcome

Vijay Rajagopal<sup>1</sup>, Jarod Jelinek<sup>2</sup>, Abhishek Ravi<sup>3</sup>, and Tyler Nguyen<sup>4</sup>

**Abstract—This proposal paper seeks to establish a clear objective, procedure, and timeline in order to come to a fruitful analysis of a given disease outcome based on medical benchmarks published by each state in the United States.**

## I. INTRODUCTION

With the current COVID-19 pandemic situation, it has become evident that the distribution of infections in the United States has been heavily uneven in terms of locality, race, and sex. This distribution not only puts an uneven strain on different portions of the United States, but it also adversely affects the ability for governments to keep vital infrastructures in place due to the reliance on "essential" workers. Due to these factors, there is a need to develop methodologies that can determine the potential complications from COVID-19 that a person could face based on various conditions. This project will attempt to develop a method of evaluating an

individual's relative mortality by utilizing state medical benchmarks as well as factoring in patient data (race, sex, etc.). Furthermore, we intend to develop a front-end application or website that will allow individuals to enter their general patient data and receive a report that showcases potential factors that could inhibit proper treatment of COVID-19.

## II. PROJECT PROCEDURES

### A. Data Collection Procedures

As mentioned before, a important part of this project is collecting medical datasets that are both up-to-date and contain appropriate granularity in its content. With this criteria in mind, the team members of this project have decided that a primary repository of our information will be coming from the **Agency for Healthcare Research and Quality** (AHRQ). The AHRQ contains overall quality of health reports of each state, and each of these reports contain sub-reports (example: a report on the at-home treatment of dialysis patients in Tennessee). These sub-reports are classified as "Far away from Benchmark", "Close to Benchmark", and "Achieved Benchmark or Better". The categorization of each of these sub-reports can indicate how well a state's medical infrastructure is helping and treating their patients.

We intend to gather sub-reports that are appropriate for our COVID-19 evaluation and parse them into usable CSV files. Furthermore, we would like to learn more

\*This work was not supported by any organization

<sup>1</sup>V. Rajagopal is with Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37916, USA `vrajago2 at vols.utk.edu`

<sup>2</sup>J. Jelinek is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37916, USA `jjelinek at vols.utk.edu`

<sup>3</sup>A. Ravi is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37916, USA `aravi at vols.utk.edu`

<sup>4</sup>T. Nguyen is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37916, USA `tnguye69 at vols.utk.edu`

about the raw material capacity of each state's medical infrastructure. By raw material capacity, we mean metrics such as number of total hospitals per state, the overall funding, and the amount of beds, medicine, and doctors available in each hospital (**note: the prior list is not exhaustive**). For this type of information, we are looking into using the "Hospital General Information"<sup>1</sup> dataset from the United States Department of Health and Human Services. This dataset is accessible via Google BigQuery and contains detailed info about Medicare-supported hospitals on a state-by-state basis.

### *B. Team Member Responsibilities*

The project will require the following software development categories:

- Data Collection (raw)
- Data Parsing and Interpreting
- Generate Programmable Conclusions from Data
- Website Development Using Conclusions

Delegation of responsibilities will be distributed based on each member's aptitude. **Abhishek Ravi** and **Tyler Nguyen**, based on their previous work in website software development, will be taking up the task of designing and implementing the website portion of this project. **Vijay Rajagopal** and **Jarod Jelinek** will be tasked with raw data collection and generating the majority of the code for parsing this data. All members will help in the task of interpreting the medical benchmark data since all members have had previous experiences in that line of work.

### *C. Project Timeline*

The following list the anticipated timeline for major project milestones (weeks start by the second following week of the submission of this proposal):

- **Week 1** - Establish list of reliable, parsable datasets related to state medical evaluation, state hospital data, and state demographics.
- **Week 2** - Start development of code that can parse and store chosen datasets from **Week 1**. If dataset code is easily human-readable, then start sampling data points for some kind of correlation.
- **Week 3** - Continue development of code from **Week 2** and debug any issues that occur.
- **Week 4** - Start development of website design and wrap up the majority of the dataset parsing
- **Week 5** - Continue development of website and have a rough MVP done by the end of this week. **This week is also when we hopefully see some solid correlations between the data we have scrapped and processed.**
- **Week 6** - Start compiling conclusions into usable code for website.
- **Week 7** - Finalize website design and functionality for showcase.

## III. EXPECTED OUTCOMES

Based on previous studies and articles produced on the demographic disparities of COVID-19 infections and outcomes, we anticipate average median incomes to heavily influence all other aspects of our medical scoring system. This includes things such as general AHRQ Composite Benchmark scores as well as distribution of hospital and medical institutions within a given state. Additionally, states with a large minority population are expected to score lower and show more potential COVID-19 complications versus states without that type of demographic situation. This is due to a number of reasons including institutional racism, historical mistrust of medical institutions, and lack of medical infrastructure.

<sup>1</sup>Source: <https://bit.ly/3i0YSrQ>