Factors Associated with Medicaid Eligibility Processing Time

Word Count: 3195

Caroline Adams, PPOL 564

Introduction

Medicaid provides critical health coverage for individuals with low incomes and disabilities in the United States.[1] However, in some states it is harder to gain coverage than in others. Processes to determine eligibility can be burdensome and time-consuming, which can prevent individuals from receiving coverage in a timely manner. Wide variation exists in the time it takes for states to process Medicaid applications and determine if an individual will receive coverage.

This project aimed to assess factors at the state level that are predictive of how long it takes for a state to make Medicaid eligibility determinations. Visual explorations of the data were conducted to visualize associations between different independent variables and the outcome variable. Machine learning methods were applied to try and predict if a state was able to make most eligibility determinations in a timely manner. While this analysis is preliminary, the findings have important theoretical policy implications, as states look for ways to reduce burdens to accessing Medicaid coverage for residents.

This report provides an overview of the problem, outlines the data used to complete this analysis, discusses methods employed to explore and analyze the data, and summarizes the results. The report concludes with a discussion on the project's success, limitations and potential next steps.

Problem Statement and Background

The Medicaid program is an important safety-net program that is jointly financed by the federal government and individual states.[1] It was designed to support individuals with low incomes and/or disabilities access health care services through the provision of government-funded health insurance.[1] The federal government sets broad requirements for state programs to meet, but states have the flexibility to design their program elements to meet the needs of their enrollees.[2] Given this flexibility, state eligibility criteria and determination processes vary widely across the nation, leading to inequities in the ability and time it takes to enroll in

Medicaid across the United States.[3]

Multiple barriers exist to individuals receiving Medicaid coverage, even when eligible. Many of these barriers relate to burdensome and lengthy application and eligibility determination processes that are set by state programs.[3] These burdens can make it difficult for individuals to complete applications for Medicaid coverage and result in delays in receipt of Medicaid coverage. Delays in coverage provision can force eligible individuals to forgo or postpone needed care or lead to unnecessary financial burden.

Each state has a unique Medicaid program and incentives to process eligibility determinations quickly or slowly. For example, states with higher incomes may have more money to invest in technology that supports fast eligibility determinations. Conversely, other states may want to conserve funds and it may be in their interest to delay the provision of Medicaid coverage for as long as possible. Political motivations may factor into this process as well. For example, conservatives are not often in favor of government entitlement programs and may intentionally increase burdens associated with the application process or refrain from improving application processing times to make it more difficult for individuals to receive coverage.

Previous research has qualitatively examined factors associated with the time it takes for states to make eligibility determinations, as well as what changes states have made to improve their eligibility determination processes.[4,5] However, no analyses have been completed for data beyond 2019 or to assess what factors may be predictive of eligibility determination time. Understanding which factors are associated with longer processing times at the state level may help identify policy solutions that can reduce barriers to eligible individuals accessing Medicaid coverage and ultimately aid in reducing inequities in health care access and outcomes in the United States.

Data

To complete this assessment, I used a variety of data sources to understand what factors are associated with eligibility determination times. The unit of analysis for all data was set as the state, and if applicable, the observations were also organized by month and year.

Eligibility Processing Time First, I used data from the Centers for Medicare and Medicaid Services (CMS) for my outcome of interest, the time it takes for states to make Medicaid eligibility determinations.[5] This data is collected by state for three months of each year (February, March, and April) for the years 2018, 2019, and 2020. States report information regarding total number of applications, application processing time, and total enrollment by month and year. The data is organized by the percentage of eligibility determinations that are completed within 24 hours, between 1 and 7 days, between 8 and 30 days, between 30 and 45 days, and beyond 45 days. Eligibility determinations that are completed within 24 hours are considered to be "real-time" eligibility determinations.

The data was downloaded as a CSV file and manipulated using pandas.[6] Two versions were kept for each year of the data: a full version with the percentage of determinations that were processed in each of the time categories noted above for the months February, March, and April, and a version that was grouped by state to find the average percentage of determinations that were processed in each of the time categories for each year. Ultimately, the separate dataframes for each year were appended together for the purposes of modeling.

Unfortunately, many state observations were missing from this data. In the 2018 data, all observations for Arkansas, Louisiana, New York, and Tennessee were missing. Additionally, data for Kansas in February was missing. In 2019, all observations for Tennessee were missing, in addition to Vermont in February. In 2020, all observations for Tennessee were again missing. Given that no outcome variable data was available for Tennessee in any year, Tennessee had to be dropped from the analysis.

Medicaid Expansion Status Multiple additional data sources were used to obtain predictor variables of interest, the first of which was Medicaid expansion status by state. The Patient Protection and Affordable Care Act (ACA) allowed states the opportunity to expand their Medicaid programs to populations beyond those that were previously required (e.g., children, pregnant individuals, parents and caregivers); this included adults ages 19 to 64 that did not have disabilities and were not parents or caregivers to children under the age of 6, but had incomes that fell below 138 percent of the federal poverty level. States that have not expanded their programs typically have Republican governors or legislatures that are ideologically opposed to extending Medicaid coverage. For this analysis, I theorized that states that chose to expand their Medicaid programs would on average have faster eligibility determination times.

Medicaid Enrollment Numbers The second predictor variable was total Medicaid enrollment numbers. I theorized that for states with higher Medicaid enrollment, on a monthly basis, they may have higher numbers of applicants and therefore it may take longer to process eligibility determinations. Both Medicaid expansion status and total Medicaid enrollment numbers by state and year were obtained from CMS as a CSV file for the years 2018, 2019, and 2020.[7] The enrollment data file had comprehensive information about each state's Medicaid and CHIP enrollment by month. It also contained information from preliminary and final reports. The CSV files for each year were manipulated using pandas. Only observations corresponding to a final report were kept. All variables were dropped except state, date of report, Medicaid expansion status, and total Medicaid enrollment. Medicaid expansion status values were recoded from "Y" and "N" to 1 and 0 respectively. Variable names were recoded for clarity.

Managed Care The third predictor variable included in this analysis was if a state contracted with managed care organizations (MCOs) to manage the care of Medicaid enrollees. Many states rely on MCOs to improve health outcomes and the cost-effectiveness of their programs. I theorized that states that have managed care would have faster eligibility determination times. Total numbers of Medicaid managed care organizations in each state were obtained from the Kaiser Family Foundation (KFF) as a CSV file.[8] Again, manipulations of the data were completed using pandas. If a state had greater than zero MCOs, their value was converted to a one to indicate presence of managed care and zero otherwise.

Real-Time Eligibility Determination Technology The fourth predictor variable was the ability of states to make real-time eligibility determinations (within 24 hours). States are increasingly adopting technological solutions and upgrades to improve their eligibility determination process. As of January 2019, 46 states had the ability due to technology to complete an eligibility determination within 24 hours. Despite this, most of these states do not complete most of their eligibility determinations in real-time. Some states rely more heavily on the technology to complete determinations and others still involve case workers in addition to the online system to complete the process. I theorized that the states with technology that enables automatic, real-time determinations to be made would have faster average eligibility determination times.

Data on state ability to conduct real-time eligibility determinations was obtained from KFF and downloaded as a CSV file with state and real-time determination ability as variables.[9] Data manipulation was conducted using pandas. An extra index variable was dropped and variable names were renamed for clarity. If the value for real-time eligibility determination was "yes," this was replaced with a one, and if the value was "no," this value was replaced a zero. The variable was then converted to all floats.

State Political Leaning The sixth predictor variable included was political leaning by state. As mentioned earlier, political leaning greatly influences the structure of a state's Medicaid program. Medicaid programs in Democrat-leaning states typically have higher eligibility cut offs and cover more benefits, while the opposite is true in Republican-leaning states. Data from FiveThirtyEight on the partisan lean of each state was used for this variable.[10] Partisan lean values are the average margin difference between how each state votes and how the country votes overall in congressional and gubernatorial elections. The values diverge in two directions for Democrat and Republican lean and converge at a value of 0 for no partisan lean.

To obtain this data, I scraped a table from the FiveThirtyEight webpage that contained partisan lean values and used pandas to convert the scraped table into a concatenated pandas dataframe. The scraped table was slightly disjointed, with the top half of state values falling into two columns for Republican and Democrat lean, and the bottom half of the state values falling into two additional columns. I renamed the matching columns and appended them together to fix this issue. I then created one single variable called Partisan Lean to contain all values, Democrat or Republican. The old variables were then dropped from the table. Values in this one variable were edited to remove instances of "R+" and "D+" before the numbers and turned into a scale of negative to positive numbers, with the most Republican-leaning state having the most negative value. For data visualization purposes, I created one additional version of this dataframe that turned partisan lean into a categorical variable. Partisan lean values above 0 were converted to "Democrat" and values below 0 were converted to "Republican."

Analysis

The methods used in this project fell into three overarching categories: data cleaning and wrangling, data visualization, and modeling. In the first phase, I obtained all sources outlined above and used pandas to manipulate the data in ways that were discussed previously. I created one joint dataset for the purposes of modeling, including all predictor variables and the outcome variable, using state as the primary unit of the analysis. Observations were organized by month and year. This dataframe initially included 459 observations. Additional steps were taken to wrangle the data in the visualization phase to alter variables when necessary.

I assessed for missing observations using the package missingno and noticed that the state of Tennessee would have to be dropped from the analysis entirely given that the state had not reported any eligibility determination time to CMS.[11] Additionally, other states were missing observations in this variable and others and these observations were dropped as well. Post-data wrangling, my sample size was reduced to 450.

The second category of methods included data visualization. I completed multiple exploratory data visualizations to understand how predictor variables were associated with the outcome variable. Visualizations were completed to communicate preliminary findings from this phase of the project as well as results from the modeling phase.

Lastly, the third category of methods included modeling and the application of machine learning techniques. Given the low sample size of my data, all available data was used as training data. Although not in the timeline of this project, the test data will be the future data that is released by CMS for 2021.[1] The Scikit-learn package was utilized to estimate multiple predictive models on my data including Naive Bayes, K-Nearest Neighbors (KNN), Decision Tree, and Random Forest.[12] A machine learning pipeline was developed to incorporate pre-processing steps, define which models should be estimated, and outline which hyper-parameters should be tuned for each type of model.

Pre-processing steps included dropping observations with missing values and creating dummy variables for states and month/year. Given that my data was time series data, the dummy variables for state and month/year were created to account for within-state temporal variation over time. Additionally, I logged the variable for total Medicaid enrollment, as it was heavily right-skewed. I also made the decision to restructure the outcome variable from the initial data format. I recoded the data to indicate whether a state completed most (greater than 50 percent) of their eligibility determinations within a week with a 0 or 1.

GridSearchCV was used to identify the model with the highest predictive accuracy.[13] The measure that was used for this was an AUC-ROC score. The AUC-ROC score serves as a measure of performance for classification problems. The score indicates how well a model is able to tell the difference between classes 0 and 1. The modeling pipeline was generated and a grid search was conducted. The training data was permuted 25 times to identify the variables with most influence in the best-fitting model. The top three most influential variables were identified and plotted to show relative reduction in AUC ROC (see Figure 7 below).

Results

Preliminary results were obtained from the data exploration and visualization phase of this project. Figure 1 below depicts total Medicaid enrollment numbers by state in February, 2020 to demonstrate variation by state in the number of covered individuals.

The remaining data visualizations resulted in four preliminary conclusions. Data for the year 2020 is displayed, as it has the least number of missing observations. First, the percentage of eligibility determinations completed in real-time (within 24 hours) has increased since 2018. Additionally, in 2019 and 2020, most eligibility determinations (greater than 50 percent) were completed within a week across all states, which demonstrates an improvement from 2018. Figure 2 depicts the percentage of eligibility determinations in 2018, 2019, and 2020 across all states that were completed within 5 different time categories.

Figure 1. Total Medicaid Enrollment by State (February 2020)

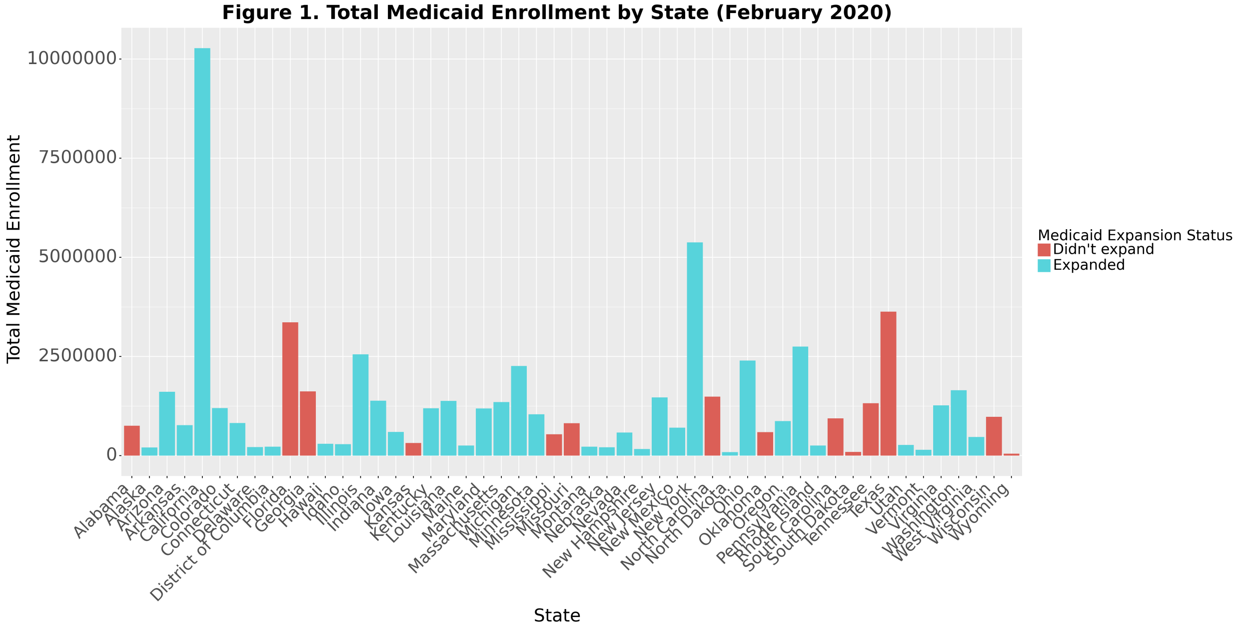
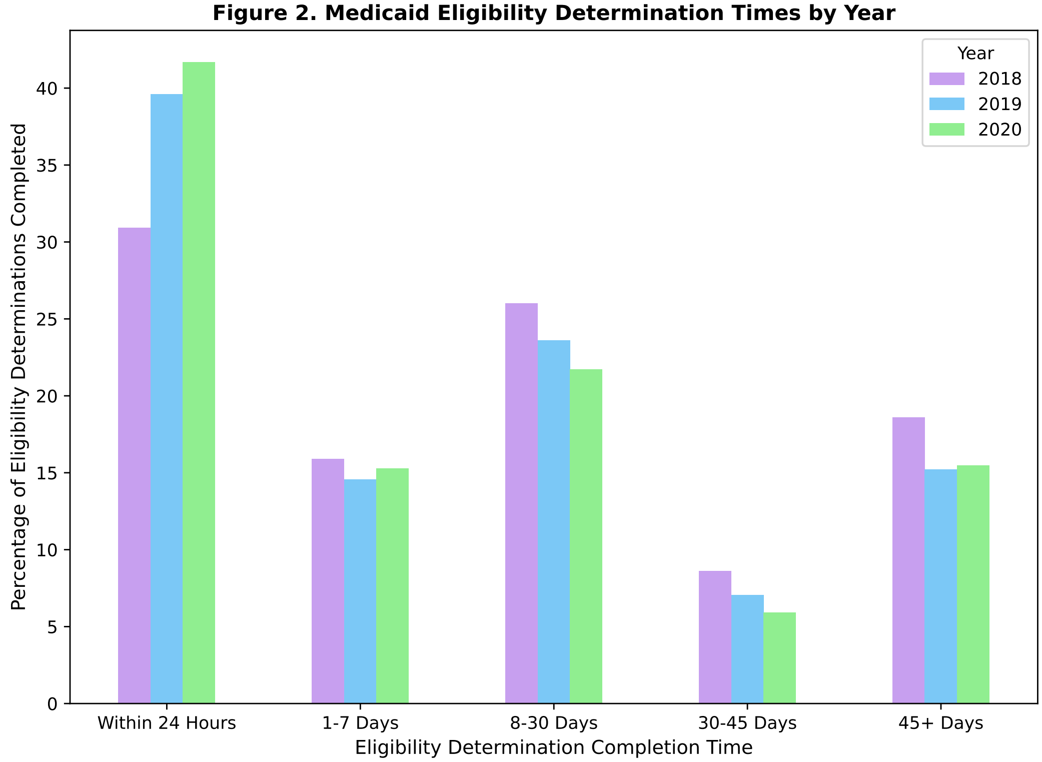


Figure 2. Medicaid Eligibility Determination Times by Year



Second, it is clear that states with technology that allows for automatic eligibility determinations completed determinations faster on average. Figure 3 compares the percentage of eligibility determinations completed within the time categories for states with and without this technology. However, the presence of this technology does not guarantee that all or most determinations are completed in real-time. In fact, less than 40 percent of all determinations were made in real-time by states with this technology in 2020.

Third, states that expanded their Medicaid programs under the ACA completed a much higher percentage of determinations in real-time (almost 40 percent) than states that did not expand (just over 25 percent). Figure 4 compares the percentage of eligibility determinations completed within the time categories for states that did and did not expand their programs.

Fourth, states that lean Democrat politically completed approximately half of all determinations in real-time, while states that lean Republican only completed approximately 25 percent in real-time and less than half within a week. Figure 5 compares the percentage of eligibility determinations completed within the different time categories for Democrat- and Republican-leaning states.

Figure 3. Determination Times for States with Real-Time Determination Technology (2020)

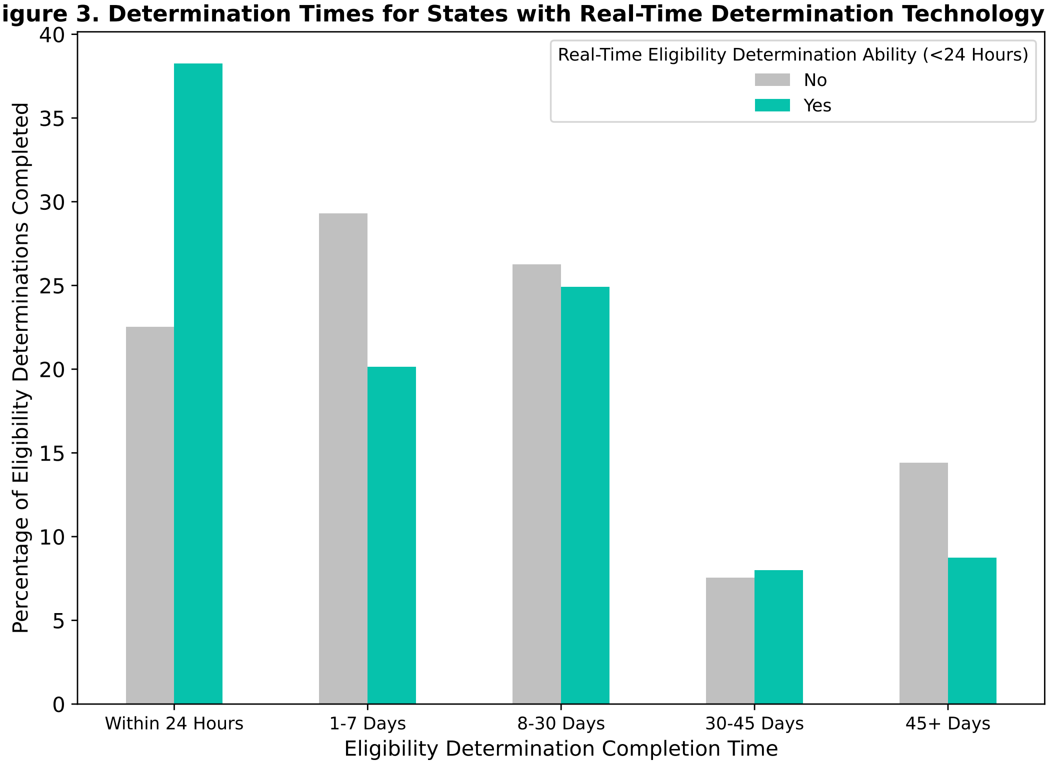
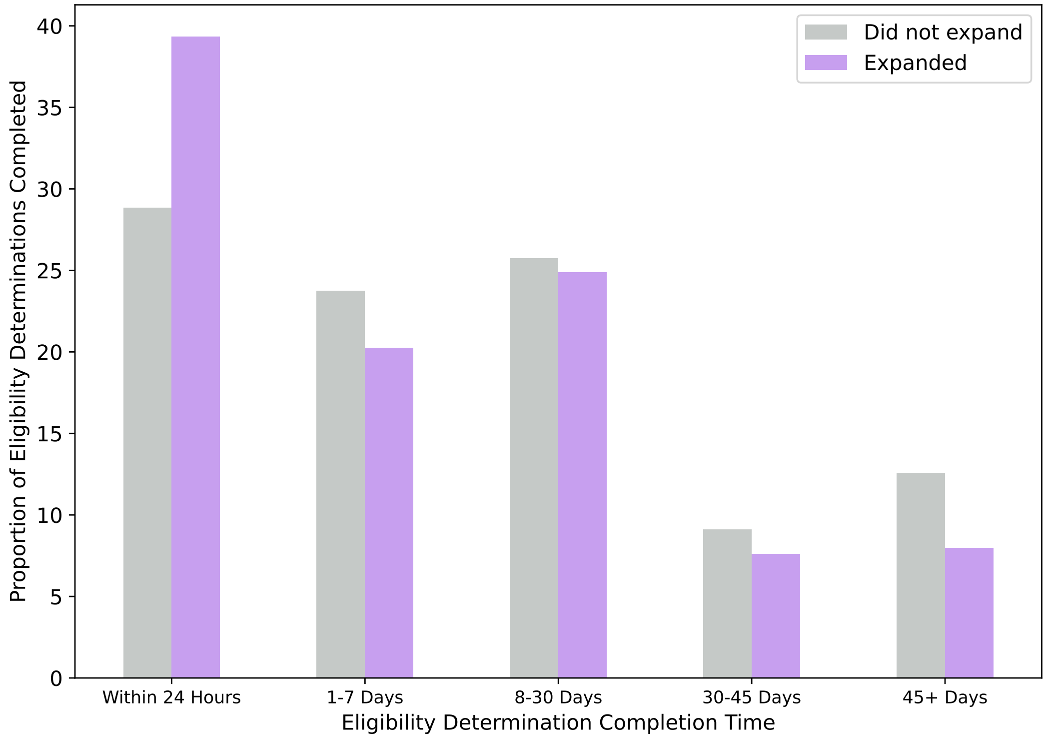
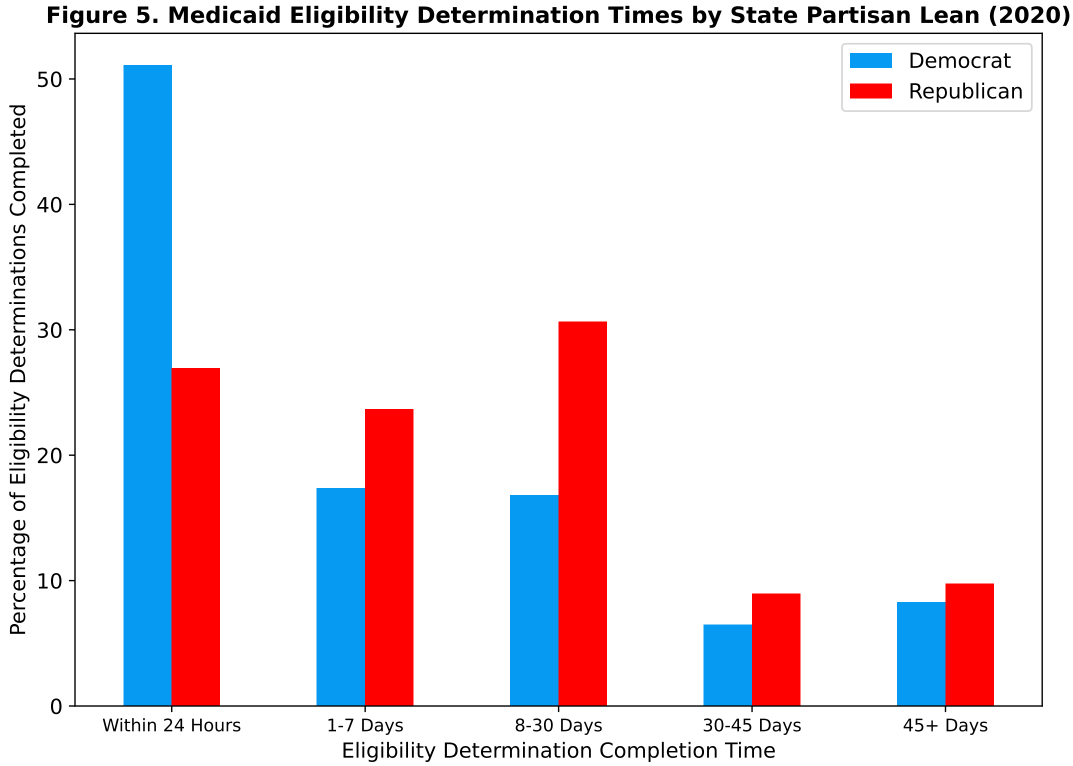


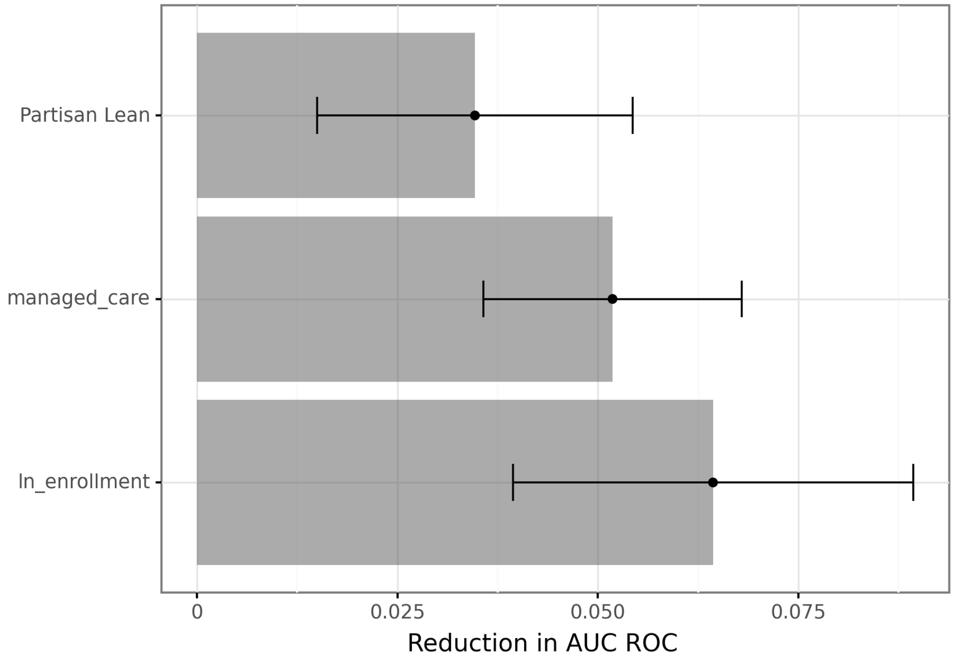
Figure 4. Medicaid Eligibility Determination Times by Expansion Status (2020)





From the machine learning component of the project, the best-fitting model produced an AUC ROC score of 0.953. This means that the model was able to distinguish between states that completed over 50 percent of eligibility determinations within in a week and those that did not 95.3 percent of the time. The best fitting model was the K Nearest Neighbors classifier, with a number of neighbors of 5. When the predict method was utilized on the training data, an R2 score of 0.745 was calculated. The model's overall predictive accuracy score was 0.937. While this seems highly predictive, it is possible that the model was over fit. Figure 6 below depicts the top three variables that had the highest influence (reduced the AUC ROC by the greatest amount) in the best fitting model. The variables logged-total Medicaid enrollment, the presence of managed care, and partisan lean were most influential in predicting whether a state completed greater than 50 percent of eligibility determinations within a week.

Figure 6. Top 3 Variables with Most Importance in Best-Fitting Model



Discussion

In my project proposal, I defined success as completing all required project elements. I also added that success would involve gaining a deeper understanding of the methods learned in class and how to apply them to new, real-world data. I hoped to have a comprehensive understanding of the limitations of the data used in my project and the applicability of methods from class to complete the intended analyses.

I believe that I have accomplished the goals that I set for myself at the beginning of this project. I performed extensive data cleaning and manipulation on the numerous data sources I obtained and merged together. I leveraged the data in a variety of ways to examine variation in the data over time to inform potential modeling mechanisms and to pull out visual insights related to my question of interest. I also carefully thought through different ways to apply machine learning methods to my project and set up a modeling pipeline that had a high level of predictive accuracy.

Initially, I had planned to include state-level demographic data (e.g., average income, race/ethnicity breakdowns) from the American Community Survey in my analysis. If given more time, I would see how this information is associated with my outcome variable of interest. Additionally, I would reach out to CMS to see if there is additional data available that I could use to increase my sample size. For example, if they collect eligibility determination data for each state every month, but only publish data for three months of the year, I would ask if it would be possible to use the data from the unpublished months or years. I would also include the 2021 data in my analysis, as it should be released in the next few months, and this would also increase my sample size. Lastly, I would see if I could use existing Medicaid application data to create a variable for number of new Medicaid applications by month. I believe there is currently data for total new applications by month for Medicaid and CHIP combined, but I would like to investigate if there is a way to take out the applications geared towards CHIP. I believe that number of new applications may be more strongly correlated with application processing time than total Medicaid enrollment and may strengthen the predictive accuracy of my best-fitting model.

Work Cited

1. Rudowitz, R., Garfield, R., & Hinton, E. (2019). 10 Things to Know about Medicaid: Setting the Facts Straight. Retrieved from https://www.kff.org/medicaid/issue-brief/10-things-to-know-about-medicaid-setting-the-facts-straight/

2. Medicaid.gov. (n.d.). Eligibility. Retrieved from https://www.medicaid.gov/medicaid/eligibility/index.html

3. Brooks, T., Roygardner, L., & Artiga, S. (2019). Medicaid and CHIP Eligibility, Enrollment, and Cost Sharing Policies as of January 2019: Findings from a 50-State Survey. Retrieved from https://www.kff.org/report-section/medicaid-and-chip-eligibility-enrollment-and-cost-sharing-policies-as-of-january-2019-findings-from-a-50-state-survey-medicaid-and-chip-enrollment-and-renewal-processes/

4. Artiga, S. & Diaz, M. (2019). How Quickly are States Connecting Applicants to Medicaid and CHIP Coverage? Retrieved from https://www.kff.org/medicaid/issue-brief/how-quickly-are-states-connecting-applicants-to-medicaid-and-chip-coverage/

5. Wishner, J., Hill, I., Marks, J., & Thornburgh, S. (2018). Medicaid Real-Time Eligibility Determinations and Automated Renewals. Retrieved from https://www.urban.org/sites/default/files/publication/98904/medicaid_real-time_eligibility_determinations_and_automated_renewals_1.pdf

6. Centers for Medicare and Medicaid Services. (n.d.). Medicaid Modified Adjusted Gross Income & Children's Health Insurance Program Application Processing Time Report. Retrieved from https://www.medicaid.gov/state-overviews/medicaid-modified-adjusted-gross-income-childrens-health-insurance-program-application-processing-time-report/index.html

7. Data structures for statistical computing in python, McKinney, Proceedings of the 9th Python in Science Conference, Volume 445, 2010.

8. Medicaid.gov. (n.d.). Monthly Medicaid & CHIP Application, Eligibility Determination, and Enrollment Reports & Data. Retrieved from https://www.medicaid.gov/medicaid/national-medicaid-chip-program-information/medicaid-chip-enrollment-data/monthly-medicaid-chip-application-eligibility-determination-and-enrollment-reports-data/index.html

9. Kaiser Family Foundation. (2018). Medicaid Enrollment in Managed Care by Plan Type. Retrieved from https://www.kff.org/medicaid/state-indicator/enrollment-by-medicaid-mc-plan-type/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

10. Kaiser Family Foundation. (2020). Medicaid Eligibility Determinations, Applications, and Online Accounts. Retrieved from https://www.kff.org/health-reform/state-indicator/features-of-online-and-telephone-medicaid-applications/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

11. FiveThirtyEight. (2021). How red or blue is your state? Retrieved from https://data.census.gov/cedsci/profile?q=United%20States&g=0100000US

12. Bilogur, (2018). Missingno: a missing data visualization suite. Journal of Open Source Software, 3(22), 547, https://doi.org/10.21105/joss.00547

13. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

14. LaValle, S. M., Branicky, M. S., & Lindemann, S. R. (2004). On the relationship between classical grid search and probabilistic roadmaps. The International Journal of Robotics Research, 23(7–8), 673–692.