HUMANA-MAYS HEALTHCARE ANALYTICS 2020 CASE COMPETITION

TRANSPORTATION BARRIERS AS A SOCIAL DETERMINANT OF HEALTH

October 2020

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

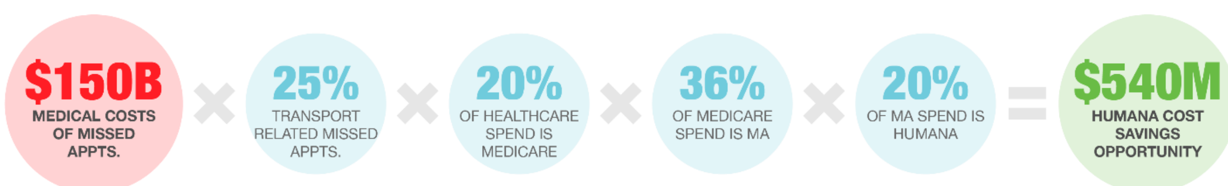
1.1. CONTEXT

Social determinants of health have received growing attention as managed care has grown in popularity throughout the United States' healthcare system. These determinants describe how an individual's daily life and surroundings impact their personal health and how they will interact with the healthcare system. An individual's economic stability, education, neighborhood, social status, and access to healthcare will greatly shape their behavior, their health, and the care they require.

Access to transportation is among these social determinants of health and can impact an individual's ability to access care. Missed appointments lead to increased rates of emergency care and hospital admissions, worse health outcomes, and higher costs.

The Bureau of Transportation Statistics reports that 11.2 million seniors have self-reported traveling disabilities.[1] Missed medical appointments are estimated to cost the healthcare system $150 billion per year, 25% of which is cause by transportation barriers.[2] Hospital admission rates are 57% higher for individuals who have recently missed medical appointments, emergency department admission rates are 47% higher, and complications are especially bad for individuals with chronic conditions like diabetes.[3] The broad scope and impact of this problem is clear and the opportunity to create value for patients as well as payors is significant. Medicaid has demonstrated this value through mandated non-emergency medical transportation benefits since the program's inception in 1965.

Using information from CMS, the Kaiser Family Foundation, and SCI Solutions we can estimate the total cost of transportation barriers to Humana.[4,5]



Humana and the Mays Business School at Texas A&M University are hosting a case competition in which they address this problem. The challenge is to build a predictive model which can ingest data on an individual's social determinants of health and output the likelihood that transportation challenges will negatively impact their ability to receive healthcare. The output and insights from the model should be applied to propose solutions which will help people overcome these transportation barriers and achieve their best health.

Humana is a health insurance company headquartered in Louisville, Kentucky. They are a leading administrator of Medicare Advantage health plans – federally funded health insurance programs which allow the administering payor to actively manage the care of the member to improve quality and cost outcomes. Humana manages 4.5 million Medicare Advantage lives, and they have provided anonymized member data to be used in model construction.

Texas A&M is the oldest and largest public university in the United States. The Mays Business School aims to “equip students with a holistic view of an enterprise, a business mindset, functional area expertise, and the full range of required technical skills that exist across all areas of business.”

1.2. MODELING

We approached this case competition methodically. First, we performed significant primary and secondary research to establish a foundational understanding of social determinants of health, especially those related to transportation challenges. Then, we selected Gradient Boosted Decision Trees as our classification model, implemented through XGBoost, because it was robust to the challenges posed by the dataset and has been proven to be very effective in producing accurate predictions.

Through a combination of feature selection, feature engineering, and a meticulous hyperparameter tuning process, we were able to train a classifier with an Area Under the Receiver Operator Characteristic Curve (AUC-ROC) of .787. More importantly, the specificity of the model is high, making it easy to operationalize and drive business value.

Through additional analysis of our model, we can show the impact of individual social determinants such as age, location, socioeconomic status, and ethnicity/race. We can also show the impact of medical factors such as CMS risk score, pharmaceutical utilization, and pre-existing conditions. In combination with our research, the output from our model can be effectively utilized to segment Medicare Advantage plans and optimize the Non-Emergency Medical Transportation (NEMT) benefit provided to each plan.

1.3. RECOMMENDATIONS

After constructing our model and interviewing five leaders in the NEMT industry, we have synthesized recommendations for Humana as they build and deploy NEMT benefits. Humana should apply our predictive model to each of their Medicare Advantage plans. **Due to the inability to discriminate benefit offerings to subpopulations within a plan, Humana should offer NEMT benefits broadly and scale the richness of the benefit to plans with high concentrations of identified enrollees with transportation challenges.** Those plans with the highest predicted likelihood of experiencing transportation difficulties should be given the most generous benefits in terms of round trips per year. Those with lower prediction scores should still be given NEMT benefits, but of lesser value. With this approach, all transportation-challenged enrollees have access to a transportation benefit, while Humana can control the costs of a program by preventing cost inflation from overutilization of those who may not necessarily need the benefit. We then recommend working with a tech-enabled legacy NEMT broker to scale the offering to plans nationally and use a per-ride payment model until more utilization data is collected in the coming years to shift payment to a capitated model. Finally, Humana can use existing marketing channels, as well as in-home care options already being utilized, to educate enrollees about the new NEMT offering. This strategy is meant to differentiate Humana MA plans from the competition by offering attractive benefits. In doing so, Humana can grow membership, increase access to healthcare, improve long-term health outcomes, while still controlling cost.

Even if the net savings of implementing our recommendations is only 10%, Humana could still achieve up to $54 million in incremental savings while attracting and retaining new members, improving member experience, and potentially strengthening their STAR ratings.

2. TECHNICAL ANALYSIS

2.1. MODEL OBJECTIVE

A major component of the case competition is to build a model that can predict "which members are likely struggling with transportation," and to make predictions for a set of holdout members. This model will be judged by the area under the curve for the receiver operator characteristic curve (AUC-ROC). This metric can be intuited as the probability that the predicted value for a positive record is greater than the predicted value for a negative record.

This metric can be intuited as the probability that an individual who did have transportation challenges will have a higher prediction score than an individual who did not.

2.2. THE DATASET

- Training data: 69,572 records by 825 variable columns, plus 'transportation_issues' response column
- Holdout data: 17,681 records by 825 variable columns

Humana provided a dataset of Medicare Advantage members, variable columns representing their social determinants of health, and a binary response variable 'transportation_issues.' This response variable represents each member's response to the question: "In the past 12 months, has a lack of reliable transportation kept you from medical appointments, meetings, work, or from getting things needed for daily living?" A value of "1" represents an answer of "Yes", and "0" represents an answer of "No." Given the nature of the response variable, this is a binary classification problem.

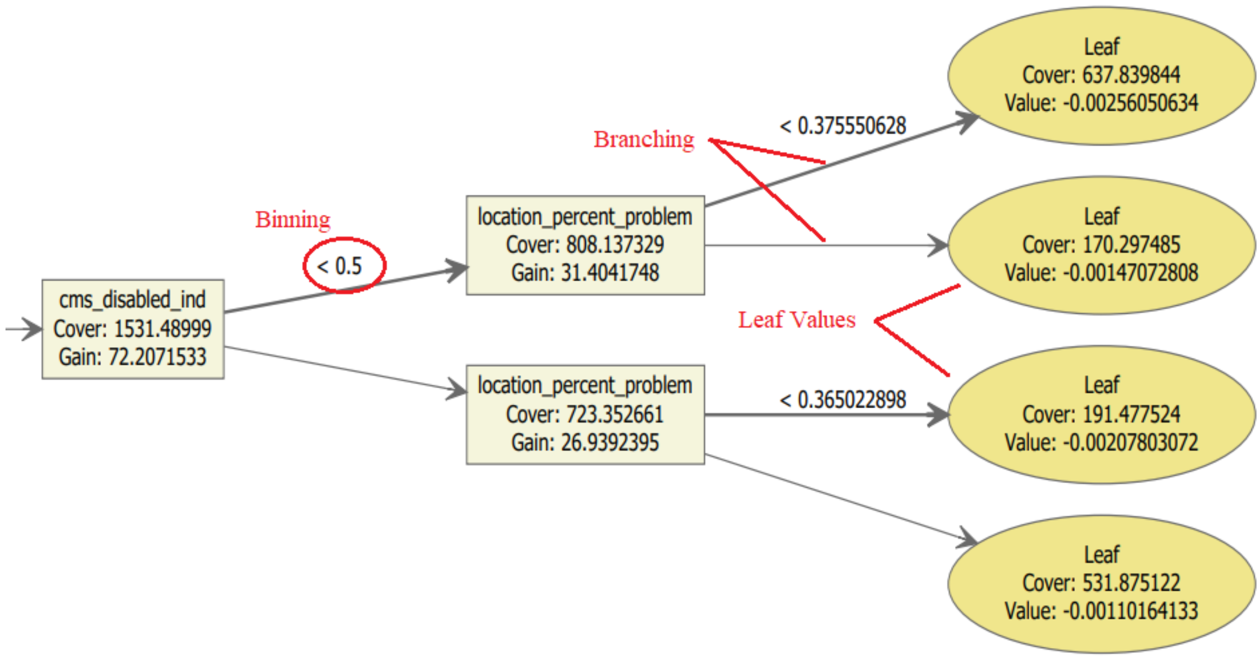
The social determinants of health variables included in the dataset range from medical and pharmacy claims information to credit and consumer data. They come in a variety of forms: scalar, integer, binary categorical, and multi-class categorical. The dataset is unbalanced relative to the 'transportation_issues' response, with only 14.66% of records in the training data being in the positive class, "1" ("The individual has missed medical appointments in the past 12 months due to unreliable transportation").

2.3. MODEL SELECTION

Because the data has so many dimensions of so many different types, the number of possible interactions between features is massive and cannot be explored thoroughly through traditional statistical methods like regression. If one were to check multiplicative interactions between just two variables, there would be more than 682,000 combinations. Additive, exponential, and inverse interactions create a feature space which is effectively infinite.

The binary classification nature of the problem, the highly dimensional data, and the variety of data types led us to choose gradient boosted decision trees as our model. Decision trees employ many "if-then" statements to group records together, creating decision branches which culminate in "leaves." Each leaf represents a prediction that the records in that leaf have a certain value, making them ideal for classification problems.

The decision tree algorithm is "greedy," which means it will naturally seek out the variables which carry the most signal to the response. They can map many complex, non-linear interactions between variables through their binning (creating thresholds in variable distributions) and branching (continuing to explore new interactions) processes. Below is an example sourced from one of the decision trees in our model.



Boosted decision trees build many decision trees (a "forest") in sequence, each tree learning from the residual errors of the last. The results from each tree are weighted so that the most powerful trees have the most influence. The weights are then optimized to minimize the error of the model.

We chose to build our model in R 4.0.2 using the XGBoost ("eXtreme Gradient Boosting") package and an RStudio IDE. XGBoost has become a staple in machine learning since it was first presented by Tianqi Chen and Carlos Guestrin in 2016. It is lightweight and fast relative to other boosted models, with implementations in R, Python, Java, Spark, etc., making it easy to deploy for many businesses in many different situations. The R interface to XGBoost provides significant control over the algorithm through hyperparameters, which allows us to build a tightly tuned model to fit the problem at hand. All XGBoost models built during this study use the logistic loss function for binary classification.

Before proceeding to the *Data Cleaning and Imputation* section, it should be noted that we originally intended to build a stacked ensemble model. This is a technique where multiple "base learner" models are built, each one extracting different information from the training data. The predictions of these base learners are then used to create a new training data set. An "ensemble learner" model is then used to fit the new predictions data set to the response.

Stacked ensemble models generally perform very well in data science competitions and in situations where the most accurate prediction possible is required. In theory, they combine the best insights from many models to create the best possible model. When applied to this problem, however, no variation of strict stacking improved model performance over a single, tuned XGBoost model. Instead, we found that elements of stacking did provide value in feature engineering for our XGBoost model, to be covered later.

2.4. DATA CLEANING AND IMPUTATION

The dataset provided to us was not ready to be ingested by a machine learning model. There are many missing or null values, and many values that required cleaning. For this purpose, we split the data into three categories: scalar variables, binary numeric variables, and categorical variables.

SCALAR VARIABLES

Null values were replaced by the median of values for that variable.

BINARY NUMERIC VARIABLES

Null values represent unknown or non-applicable information about the member. Rather than impute that information, we treated it as a third classification so that no information was lost. For any binary variable which also has null values, the variable was one-hot encoded into three binary variables representing the information available for each record (I.e. 'variable_y', 'variable_n', and 'variable_unk'). The new values are "1" and "0".

CATEGORICAL VARIABLES

Categorical variables were also one-hot encoded – they were given one binary variable for each possible character value, including null/unknown.

NOTABLE EXCEPTIONS

In the raw data provided by Humana, the "pdc" (or "percentage of days covered") variables already had null values imputed to a value of 1.1. This is not logical because no individual can have pharmaceutical coverage for more than 100% of days. Some of these variables have significant correlations with the response variable, so we attempted to re-impute these values by using the rest of the data and an XGBoost GBLinear model. Unfortunately, this re-imputation did not measurably improve the model's predictive capability in any controlled studies. We chose to leave the variables as they were.

'zip_cd', 'state_cd', and 'cnty_cd' were also removed from the dataset. There are 800+ categorical zip codes represented, and approximately 66% of all geographical values are filled with "other." This is not an ideal format for machine learning models. However, zip code and geographical information are key to social determinants of health, so we will revisit them in the *Research and Feature Engineering* section of the case.

Binary variables where an equivalent scalar variable exists (i.e. 'betos_m5b_ind' and 'betos_m5b_pmpm_ct') are redundant and were removed. They are always equal to zero where the scalar variable is equal to zero; the binning process of the XGBoost algorithm automatically splits the scalar variable by a threshold which carries more signal than the binary indicator could.

2.5. FEATURE SELECTION

Very few machine learning models built on structured data require 825 variables. Beyond variables which are the most predictive of the response, additional variables will cost processing time and add noise that can damage the power of the model. Because of this, we needed to reduce the number of variables and select those which will act as the best features for our model. Filter methods are a great way to do this. We performed Pearson correlation tests between each individual variable and the response to create metrics to filter by: the magnitude of the correlation and the significance of the correlation.

We then built many XGBoost models, each using a different set of features determined by combinations of correlation magnitudes and significance values. For example:

- Feature_set_1 consisted of variables from the training dataset where the absolute correlation magnitude is greater than or equal to .013, and the absolute correlation significance is less than or equal to .0005
- Feature_set_2 consisted of variables from the training dataset where the absolute correlation magnitude is greater than or equal to .001, and the absolute correlation significance is less than or equal to .0000002

We used the validation-set AUC metric from 5-fold cross-validation to evaluate the predictive power of each model during training.

Bayesian optimization was used to tune the correlation and significance thresholds with the goal of maximizing the AUC. It was also used to find the hyperparameters which were used to build each of these exploratory models. Bayesian optimization as a tool for hyperparameter tuning will be covered in the *Final Model Construction* section of the case.

The most predictive model from this search process was built using 207 features having absolute correlation magnitude to the response of greater than or equal to ~0.0391 and an absolute significance of correlation of less than or equal to ~.0094. This model was able to achieve measurably (albeit, slightly) higher AUC values than a model built using all provided variables.

2.6. RESEARCH AND FEATURE ENGINEERING

To produce the most powerful machine learning model possible, it is also necessary to engineer new features from existing and new data sources. We performed significant primary and secondary research to develop the subject matter expertise required to develop these new features.

First, we conducted five interviews with business and policy managers in the Non-Emergency Medical Transportation ("NEMT") industry. We also made use of *Traveling Towards Disease: Transportation Barriers to Healthcare Access*, a meta-analysis research paper which aggregates 61 research papers on the topic of how transportation barriers impact the delivery and reception of healthcare services.[6] This research influenced our initial hypotheses and guided early data exploration (histograms, correlation tests, transformations, etc.).

Below is a non-exhaustive list of our findings.

LEARNING	PUBLISHED STUDIES	TEAM EVALUATION
There is no significant difference in transportation difficulties between rural and urban populations	• Blazer et al.	Confirmed through analysis of 'rucc_category'
Not having access to a car, a friend with a car, or another form of private transportation increases transportation difficulties	• Giambruno et al. • Cunningham et al. • Flores et al. • Etc.	Confirmed through analysis of 'cons_n2pmv' (KMB_Census % Motor Vehicle Ownership)
Needing to walk or use public transportation to reach the care location increases transportation difficulties	• Rask et al.	Potentially confirmed through study of available public transit databases, although results could be disputed
Increased distance to care providers has an inconclusive effect on transportation difficulties	• Canupp et al. • Kruzich et al. • Nemet et al. • Etc.	No provided variable directly identifies distance, although it is implicitly encoded into 'zip_cd' which has a significant relationship
Ethnicity and race are significant factors in predicting transportation difficulties	• Borders et al. • Call et al. • Guidry et al. • Etc.	Confirmed through analysis of 'cons_n2pbl' (KBM-Census % Black), which is an incomplete measurement of ethnicity and race
Transportation assistance would improve medication adherence	• Kripalani et a. • Levine et al.	Established a relationship between Rx utilization and transportation difficulties through the 'rx_' schema fields
Lack of available finances increases transportation difficulties	• Malmgren et al. • Wallace et al. • Branch et al.	Confirmed through the 'cms_low_income_ind' and the 'credit_' schema fields
Lack of transportation is associated with patients receiving emergency room care rather than at their usual place of care	• Hoffman et al	Confirmed through 'betos_o1a_pmpm_ct' (betos count for ambulance utilization) and 'med_ambulance_visit_ct_pmpm'

Our research led us to believe there are three interrelated areas of influence which could cause an individual to experience transportation difficulties:

- Location – Density of healthcare provider relative to density of healthcare consumer; status and usage statistics of local public transportation; socioeconomic demographics of the location
- Financial – Personal and local vehicle ownership; unemployment/median income; ability to afford insurance/medical care; etc.
- Health condition – Behavioral health condition of the member; pharmaceutical plan adherence of the member; physical disabilities; etc.

These three "buckets" are effectively a generalization of the topics most often considered in the Social Determinants of Health as described by the Kaiser Family Foundation: economic stability, neighborhood and physical environment, healthcare system, social and community context, food, and education.[7]

Our filtered dataset already contained many features which describe these three general topics. To improve the predictive power of our model, we designed new variables and a basic experiment to determine whether they should be added to the final model:

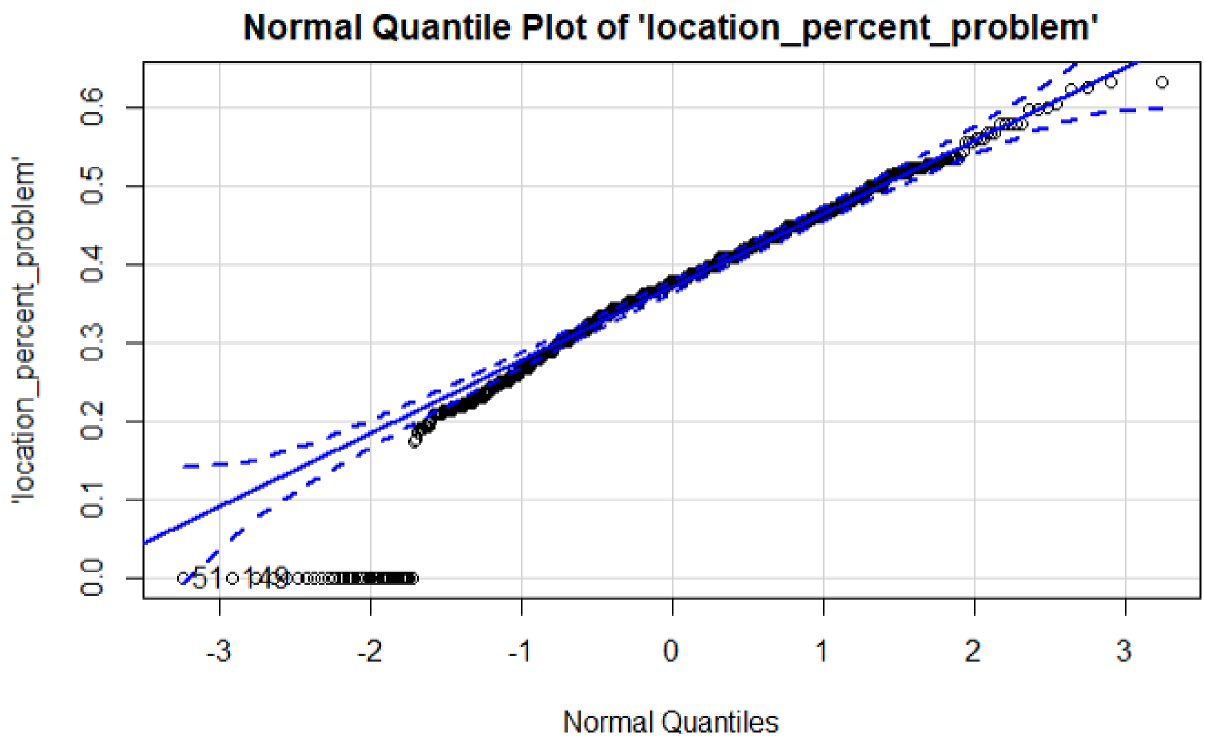
Given the same learner (an XGBoost model with hyperparameters tuned to our filtered ~207 feature dataset, initialized with a controlled random number generator seed), does the predictive power of the model, as measured by AUC, improve with the addition of the new variable as a feature? If so, the resulting dataset and predictive power become the new benchmark. If not, the variable is not added.

LOCATION

'location_percent_problem' – Because 'zip_cd' is a categorical variable with 800+ values, it was not reasonable to one-hot encode the values as individual binary columns. Instead, we chose to aggregate 'zip_cd' into a scalar variable representing the percentage of members from a given zip-state combination (to preserve unique values) with reported transportation difficulties. This aggregation was only performed for members whose 'zip_cd' and 'state_cd' were not of value "other" because those values would dilute the information gained from this aggregation.

Exploration of 'location_percent_problem' showed that the variable distribution is approximately normally distributed, but skewed left. A square root transformation of the variable corrected for this, notwithstanding some outliers.

Below is the normal quantile plot after the transformation.



This new 'location_percent_problem' field was then joined onto the dataset by 'zip_cd', and any null values (which were originally of value "other") were imputed by an XGBoost GBLinear model. We were careful not to include the response 'transporation_issues' in a non-aggregate form anywhere in this process, as that would have undermined the validity of the variable. 'location_percent_problem' ended up being one of the most predictive features for the model and was added as a permanent feature. It should be noted that we tested this variable with and without the outliers of value "0", and the variable was more predictive with the outliers included. Also, it should be noted that transformations to normal distributions are not necessary when building an XGBoost model, however, achieving more normal distributions of variables cannot hurt.

Public transit data – No data on public transit systems was included in the original data set, so we incorporated the National Transit Database, a set of local public transit statistics maintained by the Federal Transit Administration. From this dataset (and a publicly available crosswalk of FIPS code to zip code for convenience), we were able to derive three variables:

- 'miles_per_person' – The average number of miles ridden by an individual on their local public transit system.
- 'trips_per_person' – The annual number of trips utilized by a member of the serviced population.
- 'fare_per_person' – The average cost of a single member's public transit trip/ride.

These variables were each aggregated to the 'zip_cd' level and lose detail, such as the mode of transportation (i.e. rail, bus, etc). Each variable was then joined onto the dataset and tested individually for marginal predictive power. Null values after the join were imputed with an XGBoost GBLinear model. While, independently, each variable held some predictive power, none provided a marginal improvement over 'location_percent_problem'. In fact, marginal AUC was negative for all three of the variables, implying that they are each at least partially collinear with 'location_percent_problem'. None of these three variables were added as permanent features.

LOCATION / FINANCIAL

'unemployment_rate' – No traditional measures of local economic performance were provided in the original dataset, so we added the Local Area Unemployment Statistics maintained by the U.S. Bureau of Labor Statistics. This data was also aggregated to the 'zip_cd' level and joined to the dataset. Like the public transit data, 'unemployment_rate' independently added a small amount of predictive power, but marginal AUC was negative when added with 'location_percent_problem', implying some collinearity. Thus, 'unemployment_rate' was not added as a permanent feature.

'med_income' – Median income is another economic measurement that is included in the data from the U.S. Bureau of Labor statistics. The data was treated and joined in the same fashion as 'unemployment_rate'. Unfortunately, this variable behaved the same as the past four experimental variables, as it improved performance independently but was collinear with 'location_percent_problem'. 'med_income' was not added as a permanent feature of the model.

SYNTHETIC VARIABLES

The provided dataset is quite thorough in its coverage of the potential variable space and what pertinent variable interactions might exist. For example, the credit "dpd" (days past due) and "severederog" (severe derogatory acts) have "total" variables, which sum related groups of variables. Rx variables also have aggregated versions in the form of "pdc" (percent of days covered) variables for multiple disease spaces. Variables without these associated aggregate variables have little immediately reasonable relationships to one another.

As previously mentioned, the distributions of individual variables and interacted variables have no impact on the predictive power of the XGBoost algorithm, so identifying potential transformations will not add value. Because of this, and the thorough variable interactions already provided, there were few avenues for traditional feature engineering. Instead, we turned to building "sub-models" to generate synthetic scores which have pertinent relationships encoded into them.

For example, we took all the 'credit' variables from the original 825 variables and used them to build a tuned XGBoost model with a response of 'transportation_issues'. The predictions then act as a score for that member's financial health and encodes the relationships between variables that influence the score. This score was then added to our filtered dataset of 207 features and tested for predictive power. This process allows us to encode most of the pertinent information we have about an individual's social determinants of health without adding too much noise.

We performed this process for each different variable group in the original dataset: 'betos', 'bh', 'ccsp', 'cms', 'cons', 'credit', 'hedis', 'lab', 'med', 'prov', 'rev', 'rx', 'submcc', and 'total'.

Because each of these sub-models had been exposed to the response variable during training, predictive power was tested on a holdout test partition, for which the individual score values were predicted and appended. None of the models added predictive power, and a few reduced predictive power. This implies that our variable filtering process was effective in distilling the dataset down to the relevant information. The remaining variables carried most of the signal from their respective categories. Therefore, none of these synthetic scores were added as permanent features of the model.

STACKING

We had initially planned for our final model to be a stacked ensemble model. After preparing our final feature set, we trained multiple models, including:

- Boosted Decision Tree (XGBoost package)
- Deep Neural Network (Neuralnet package)
- Logistic Regression (glm function, base R)
- Naïve Bayes (e1071 package)
- Random Forest (randomForest package)
- Support Vector Machine (e1071 package)

The support vector machine simply took too long to complete training to produce any sort of tuned result. The random forest algorithm is relatively similar to boosted decision trees but produced inferior results. Oddly, the neural network also produced inferior results to logistic regression, despite being an evolution of the same principles.

Given the predictions from each of these models, no ensemble learner (randomForest, XGBoost, or glm) produced results which outperformed our base XGBoost model in AUC. Overall accuracy improved very slightly, but AUC was always worse. This was a surprising result, but nothing is guaranteed in machine learning.

Rather than give up on stacking, we chose to experiment with adding the predictions of other models as variables in our XGBoost model. The predictions for the logistic regression and the naïve Bayes model added significant predictive power to the model. This makes sense because these two models' predictions were the least correlated with the predictions from the initial XGBoost model, meaning that they carry different signal than the initial XGBoost model. Given that these predictions add power to the model only in the presence of the original features, it implies that certain patterns in the original variables give the model indication that the prediction from the logistic regression or naïve Bayes model would be correct.

In summary of our feature engineering, all of the signal that we were able to extract is present in the original dataset, filtered to 207 features achieving absolute correlation magnitude and absolute correlation significance thresholds, as well as our calculated ‘location_percent_problem’ field. Adding the predictions from the logistic regression and the naïve Bayes model allowed us to extract further signal.

2.7. FINAL MODEL CONSTRUCTION

With the model features selected and engineered, the XGBoost model needed to have its hyperparameters finely tuned to achieve the maximal predictive power. Those hyperparameters are as follows:

- ETA – the learning rate for the gradient decent algorithm which optimizes the base learner model weights. Different learning rates will be able capture different amounts of signal depending on the local and global topology of the loss function relative to each variable.
- Nrounds – the number of iterations that the algorithm should perform. Combined with ETA, this represents how "far" the model should train.
- Max_depth – the maximum depth of each decision tree to be used by the model. A measure of the complexity of interactions to be allowed. Values which are too high can cause the model to overfit to the training data.
- Subsample – The percentage of records to be used in each iteration of gradient decent to train the model. Reducing this from 100% can help the model generalize to unseen data, but reducing it too low can cause the model to miss out on signal.
- Colsample_bytree – The percentage of features to be used in the growth of each tree. Reducing this from 100% also helps the model to generalize.
- Min_child_weight – A threshold by which to prune branches. If a node does not have a high enough weight, it will not be split into further branches. This is a measure by which to control the complexity and generalizing capability of the model.
- Scale_pos_weight – a multiplicative weight to be applied to errors made when the true value of the record response is the positive class. This allows us to balance the impact of positive and negative response records in the classification problem.
- Gamma – like min_child_weight, gamma is the minimum loss reduction required for a leaf node to continue branching. This is also a form of regularization to aid the generalization of the model.

There are multiple ways to find the best set of hyperparameters for a machine learning problem. Among them are:

- Grid search – trying each possible combination of hyperparameters
- Random grid search – randomly sampling the grid space, rather than the entire space
- Bayesian optimization – creating a surrogate function for your model relative to a chosen performance metric and optimizing the solution given the result of prior trials.

Because of the huge space of this particular hyperparameter search, grid search would consume far more computational resources than we had available. Random grid search would not guarantee that we find an effective area of the hyperparameter topology. Thus, we decided to use Bayesian optimization.

Our search followed the following process:

1. Create a function to build XGBoost models, for which the optimizer will input values for each hyperparameter, and the output is an AUC measurement of cross-validated predictions:
    a. We used k-fold (5-fold) cross-validation on the training dataset, with unlimited iterations, and early_stopping_rounds = 200, which controlled the total required

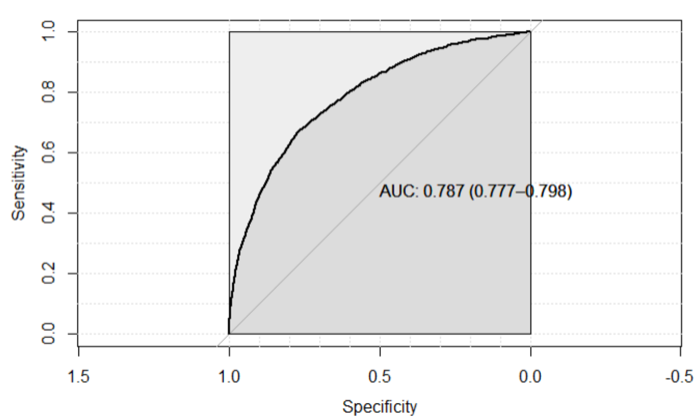
compute time, prevented over-training, and reported the iteration where the validation AUC was the highest (as an average of the 5 models built during cross-validation).

    b. Because the training data had predictions from the glm and naïve Bayes model which were directly influenced by the response, the cross-validated data AUC could not be used as an accurate measurement of the model's performance. We built a singular XGBoost model on the entire training dataset using the nrounds identified in the cross-validation, then applied it to the holdout test data to measure its performance on an unbiased data.

2. Apply a Bayesian optimization search to this function using broad ranges for each hyperparameter:
    a. We used the package rBayesianOptimization to do this, using the "probability of improvement" methodology. This was mostly due to technical limitations with the "expected improvement" and "GP Upper Confidence Bound" methods. The optimization is run for 40 iterations, with the first 6 being based on parameters known to be relatively effective (from prior Bayesian optimization runs, established during feature selection and engineering). This creates a strong set of priors to maximizes the probability that the optimizer will converge on the true optimal solution.
3. Repeat the Bayesian optimization process on a narrower range of values for each hyperparameter, centered around the best hyperparameter values found in the last search. Extract the best hyperparameter values from this search. These are the final values.
4. XGBoost models incorporate some random properties when they are built. Until this point, we have used random seeds to control the initialization of each model so that they are directly comparable. Now, we build 20 XGBoost models using the identified optimal hyperparameters, each with different random seeds, and test the performance of each model on the holdout test data. This will allow us to sample the prediction space and select the best possible model.

This process led us to a model having an AUC of .787 on the test dataset. The confusion matrix and a more detailed breakdown of model performance is below.

```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 11330  1503
         1   527   554

               Accuracy : 0.8541
                 95% CI : (0.8481, 0.8599)
    No Information Rate : 0.8522
    P-Value [Acc > NIR] : 0.2639

                  Kappa : 0.2797

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.26932
            Specificity : 0.95555
         Pos Pred Value : 0.51249
         Neg Pred Value : 0.88288
             Prevalence : 0.14784
         Detection Rate : 0.03982
   Detection Prevalence : 0.07769
      Balanced Accuracy : 0.61244

       'Positive' Class : 1
```



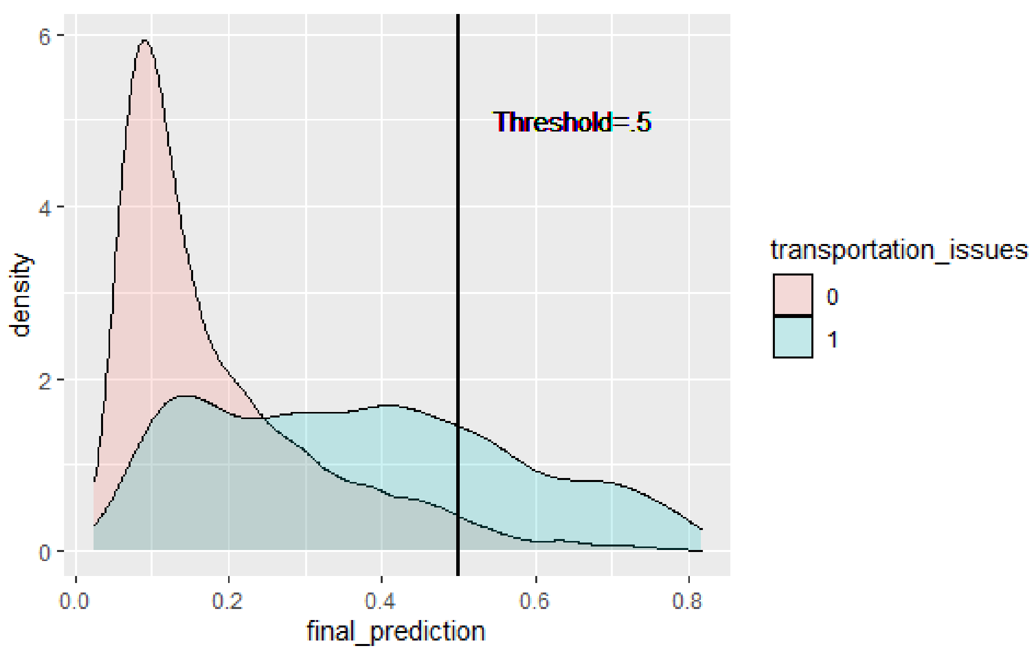
2.8. POST MODELING ANALYSIS

AUC is one measure of predictive power, but we need to examine the model in more detail to understand how it works and how to use the model appropriately.

The confusion matrix points out several important metrics:

- Sensitivity (also known as recall): A value of .269 is not very high and can be interpreted as correctly identifying 26.9% of cases where the true response is positive.
- Specificity (also known as precision): A value of .956 is very strong and can be interpreted as the model being correct 95.6% of the time when it predicts that a record will be positive.
- Accuracy: A value of .854 is not particularly strong when considering that 14.66% of records in the training set are positive. This score can be interpreted as correctly predicting the response value 85.4% of the time, or almost exactly the same result as if you were to guess every result is false.

Looking at these three metrics paints a specific picture of this predictive problem. Whether or not an individual will miss a medical appointment due to transportation issues has some inherent randomness. Considering the problem intuitively, this is logical: life is simply too complicated to map every possible input into such an abstract outcome. We can investigate this further using a predicted probability distribution plot.



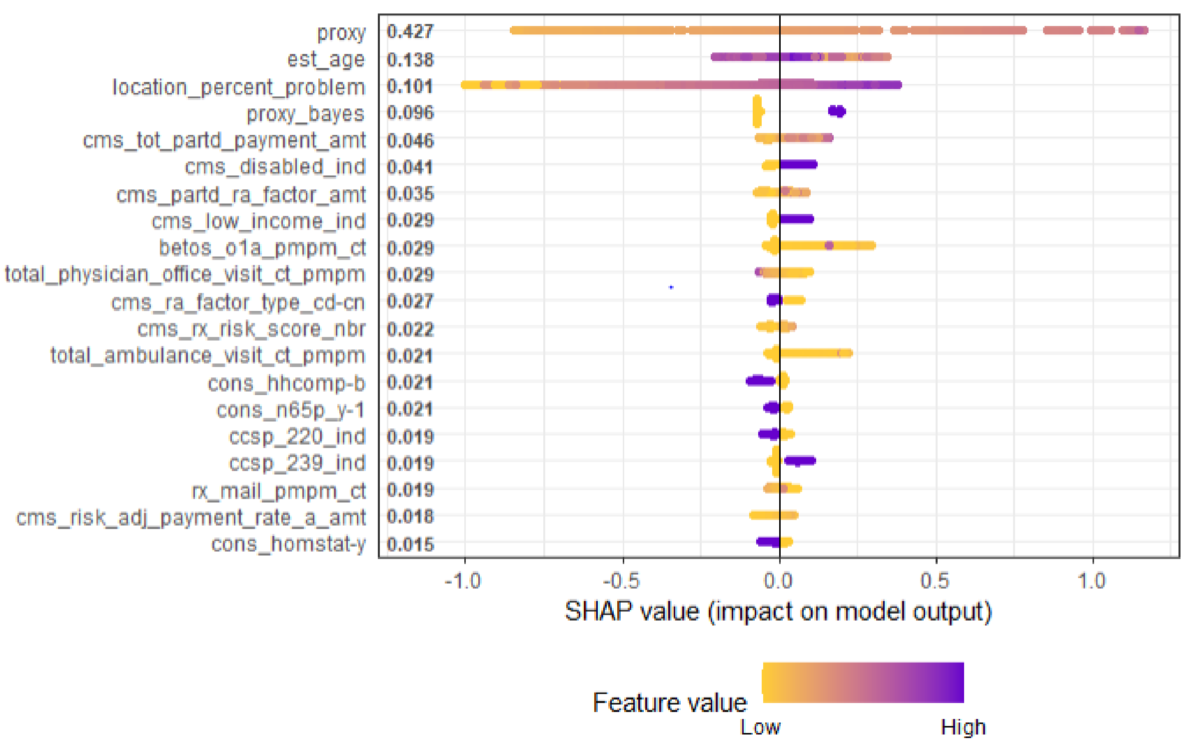
Our model is generally able to separate out individuals who do not have transportation issues, however, the predictions for true cases of the response are much more evenly distributed – adding validity to the conjecture that the positive class is difficult to predict. However, our model excels at selecting records which are very likely to have transportation difficulties.

Because the specificity of the model is so high, it can be very effective as a tool to identify individuals who will need non-emergency medical transportation benefits or some other kind of intervention. Since a positive prediction is almost always correct, it can be used to make sure that any benefits or intervention offered are optimally utilized and directed to those who may need transportation assistance the most.

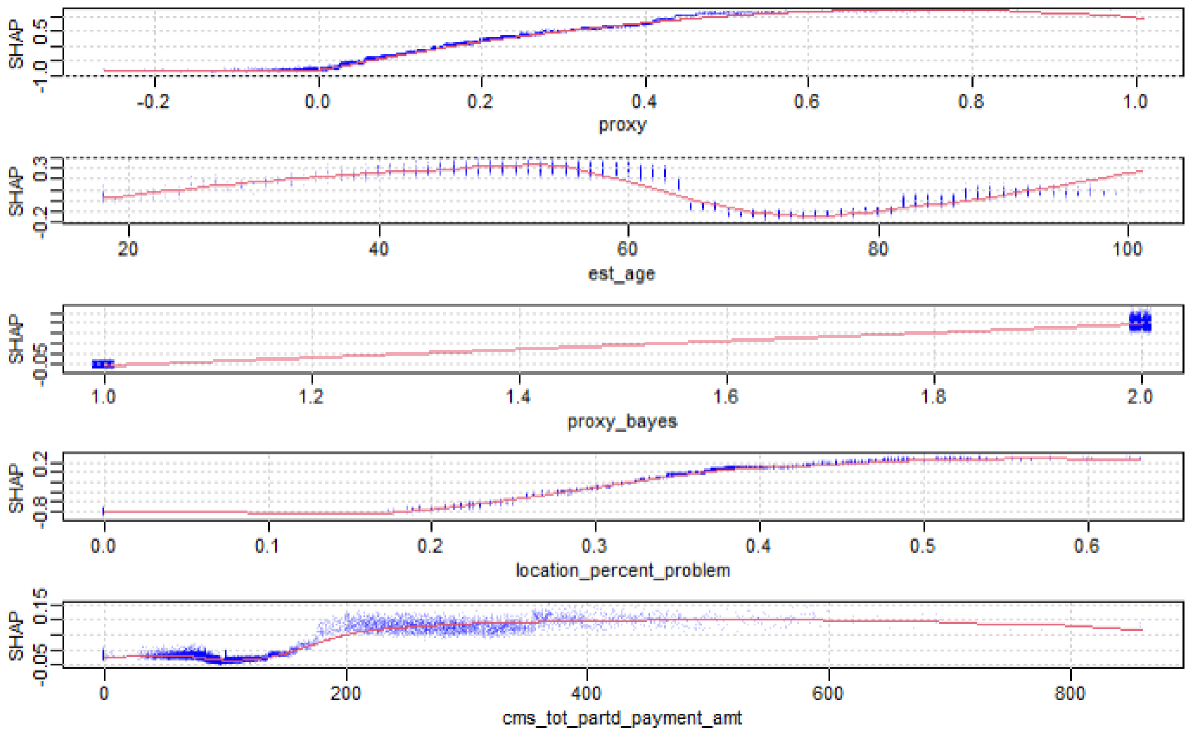
AUC as a measure of model performance is agnostic to the decision threshold probability assigned to each class – it does not matter if a positive prediction is classified by the prediction value being >.5 or >.7, AUC remains the same. However, changing the threshold will change sensitivity or specificity, therefore changing the potential applications that the model can be applied to.

There is also more that we can learn from the model. While machine learning models can be "black boxes," methodologies have been developed to gain an intuitive understanding of how each variable will impact a prediction. One of the stronger tools in this interpretive machine learning toolbox is the Shapley value.

In short, a Shapley value shows how a prediction is impacted by a particular feature having a relatively high or low value. Plotting many records of these Shapley values can give us a broad understanding of how the model is working. Below is a Shapley value plot for the twenty most influential variables in our final XGBoost model.



Below are the individual Shapley plots for the five variables with the most influence.



Using these plots, we can learn some important things:

- ‘proxy’, which is the name assigned to the predictions of the logistic regression, has a large impact on our model. The positive correlation between feature value and Shapley Value is strong, however, there is some non-linearity towards the highest values. This adds some validity to the earlier conjecture that the logistic regression model has a non-linear/conditional interaction with the rest of the features.
- ‘est_age’ is an important factor. Counterintuitively, older individuals tend to have less problems with transportation. This could be because younger individuals can gain access to Medicare through being disabled (it appears most problems happen under the Medicare applicability age of 65), or that older members have more severe conditions which require more urgent care, therefore missing less care.
- ‘location_percent_problem’ is also one of the most powerful predictors. Because ~66% of values in this field are imputed, this shows that we were successfully able to encode a lot of geographical data into this field. This also adds validity to the prior conjecture that the public transportation and economic data is collinear with it.
- ‘proxy_bayes’ represents the predictions from our naïve Bayes model. While they add to the predictive power of the model, it is difficult to interpret them because the predictions are binary, showing no variation. However, it can be observed that positive predictions from the Bayes model hold more influence than negative predictions.

- Many of the CMS reported values add predictive power as well. These include the total pharmaceutical spend of the member under Part D, the member's disabled indicator, their CMS risk adjustment factor, their low-income indicator, their Rx risk score, and their risk adjustment payment rate amount. While the correlation trends for each of these features are mixed, each of them is influential. These features also represent concepts from our initial hypothesis that overall/specific health conditions and financial health influence transportation difficulties.
- 'betos_o1a_pmpm_ct' is a measure of how many ambulance rides that a member takes in a month on average. It shows that any utilization of ambulances is an indicator that the member will potentially miss future appointments. 'total_ambulance_visit_ct_pmpm' is a similar feature and also contributes.
- 'cons_hhcomp_b' is a binary variable stating whether the household is "married with no children present." It shows that households fitting this description are less likely to experience transportation difficulties.
- 'cons_n65p_y-1' is a binary variable stating whether the household has a family member who is 65+ years of age. Households fitting this description are less likely to experience transportation difficulties. This is difficult to intuit – our guess is that this is a flag for multiple 65+ household members, in which case a high need for medical transportation might increase the likelihood of having access to transportation, or that younger members of the household would be able to provide transportation.
- 'ccsp_220_ind' is a binary variable stating whether the individual experienced intrauterine hypoxia or birth asphyxiation – meaning that they were deprived of oxygen during birth. These individuals have a higher likelihood of experiencing transportation difficulties, possibly because of associated birth defects and disabilities. While this condition is extremely specific, it shows that similarly specific conditions can carry a large amount of signal for the small affected populations.
- 'ccsp_239_ind' is a binary variable stating whether the individual has any superficial injuries or contusions. Those who did are more likely to experience transportation difficulties. This is easy to intuit.
- 'rx_mail_pmpm_ct' is a scalar variable representing how many prescriptions the member receives my mail, thereby reducing their need for healthcare transportation. The correlation trend is not conclusive. While higher claim counts indicate utilization of mail order pharmacy, it may also indicate more severe health conditions.
- 'cons_homestat-y' is a binary variable classifying whether the member is a homeowner. Homeowners are less likely to experience transportation issues.

Further analysis could be performed to understand the features which influence an individual's likelihood to experience transportation difficulties. The coefficients of the logistic regression would shed some light on the average influence of features, as could the prior probability distributions used in the naïve Bayes model. However, we feel that the Shapley analysis gives a strong intuitive understanding for what features have influence, and their inherent nuances.

3. BUSINESS ANALYSIS AND RECOMMENDATIONS

3.1. TRANSPORTATION BENEFITS OVERVIEW

Supplemental benefits have always been an attractive feature to draw Medicare beneficiaries to Medicare Advantage (MA) plans, offering additional products that would not fall in the realm of traditional Medicare. According to the Kaiser Family Foundation[8], in 2020 93% of all MA plans offer a fitness benefit, 88% offer a dental benefit, 87% offer eye exams and glasses benefit, and 83% offer a hearing aid benefit. Today, with more focus being placed on social determinants of health, payors are finding value in using supplementary benefits to grant plan enrollees access to additional services that may overcome these factors, lowering costs to these populations and increasing health outcomes. Support for these types of services is still less than common today, with 46% of MA plans offering a meal benefit and only 33% offering a transportation benefit. With a large portion of the senior population unable to drive, lacking means of taking public transportation, and without the excess income to afford private transportation, transportation access frequently becomes a barrier to an enrollee receiving medical care. Missed physician appointments can lead to poor clinical adherence, as well as increased medical complications and higher hospitalization rates, therefore providing transportation is important for maintaining member health and containing plan costs. Patient no-shows due to the transportation challenges cost the U.S. healthcare system an estimated $37.5 billion annually.[2] An often-cited 2005 study by the Altarum Institute estimated that 3.6 million people miss or delay medical care because of challenges accessing transportation each year.[9] This number has surely increased as the population aged and income inequality increased in the last 15 years. Indeed, the Bureau of Transportation Statistics reports that 11.2 million seniors have self-reported traveling disabilities.[1] If the dataset provided by Humana is representative of the payor's 4.5 million enrolled MA lives[10], then 655,000 Humana enrollees likewise face a similar transportation challenge.

Medicaid has offered non-emergency medical transportation (NEMT) benefits since program inception in 1965 to provide low income beneficiaries means of getting to sites of care, especially when a reliable mode of transportation is not readily available. An industry has been built around facilitating this service, and the Transit Cooperative Research Program estimates NEMT spending at $3 billion annually, or less than 1% of total Medicaid expenditures.[11] As Medicaid is administered at the state level, rather than federal, state Medicaid agencies have outsourced benefit coordination to third party brokers, who act as a middleman between the health plan, care providers, transportation providers, drivers, and beneficiaries. Numerous studies have been conducted as to the return on investment (ROI) states receive for investing in comprehensive NEMT programs. One 2008 study of the state of Florida's transportation program found that if 1% of medical trips funded result in the avoidance of a hospital stay, the payback to the state Medicaid agency would be 1,108%, or $11.08 for every dollar spent on the benefit.[12] More recently, in 2018, the Medical Transportation Access Coalition concluded that using NEMT as a care management strategy for people with chronic diseases created a positive ROI of over $40 million per month per 30,000 Medicaid beneficiaries.[13] While the Medicaid cost savings are evident for investing in such a program, MA plans have been relatively constrained until the last few years in designing transportation benefits

Thankfully, changes to federal regulation and the Centers for Medicare & Medicaid Services (CMS) policy have made it simpler for MA plans to construct sophisticated non-health-related benefits, and find opportunities to target specific demographics identified by chronic medical conditions or social determinants of health. After reviewing the recent evolution of the regulations around supplemental benefit construction, and specifically the Value-Based Insurance Design (VBID) Model, our case will identify several alternatives for building NEMT benefits and trade-offs between models. After weighing trade-offs, we recommend using our model predictions to construct a variable NEMT program built around relative population of transportation-challenged enrollees. We also recommend ideas for selecting a broker and payment model, as well as how to market this program to the targeted populations and how to measure the success of the program. Finally, we weigh the risk factors of implementing such a program, specifically looking at the potential of an NEMT program to inflate costs to Humana and the potential for Medicare Trust Fund payment cuts.

3.2. RECENT EVOLUTION OF SUPPLEMENTARY BENEFITS FOR MA PLANS

Prior to 2019, payors were fairly limited in the nature of the supplemental benefits they could offer to beneficiaries through Medicare Advantage plans.[14] Three factors have helped lessen these constraints, namely CMS redefining what supplemental benefits can look like, the expansion of supplemental benefits to patients with chronic conditions through the CHRONIC Care Act of 2017, and the expansion of the Value-Based Insurance Design Model.

First, CMS has historically looked at three requirements for structuring supplemental MA benefits: they must not be covered by original Medicare, they must be primarily health related, and they must incur a direct medical cost. It was challenging to structure in transportation-related benefits before this, as payors needed to demonstrate that providing non-emergency medical transportation was primarily health related. Beginning in 2019, CMS expanded the requirements of supplemental benefits to include better coverage for products that address social determinants of health. The new requirements include that a benefit can diagnose, prevent, or treat an illness or injury, can compensate for physical impairments, can ameliorate the impact of injuries or conditions, or can reduce emergency and healthcare utilization.[15] This greatly improves the opportunity for products or services that are non-medical in nature but can assist in maintaining or promoting good health to be structured into benefits packages. This regulatory change alone has had the greatest influence in expanding the creative benefits MA plans offered over the last two years in addressing the social determinants of health, and specifically driving one third of plans today to offer some sort of a transportation benefit.

Additionally, the CHRONIC Care Act of 2017 sought to improve care for Medicare beneficiaries with chronic conditions by expanding the supportive benefits Medicare and MA plans can offer to help with daily non-medical needs. Specifically, beginning in 2020, plans can now add benefits that "have a reasonable expectation of improving or maintaining the health or overall function of the chronically-ill enrollee and would not be limited to primarily health-related services."[16] This bill passed with bipartisan support, and further gives plans the mandate to incorporate supplemental benefits to address social determinants of health and chronic illness. While specific benefits are not listed in the bill, new transportation benefits, such as NEMT, fall within this scope.

Finally, through the Medicare Advantage Value-Based Insurance Design (VBID) Model, CMS' Innovation Center aims to test "health plan innovations designed to reduce Medicare program expenditures, enhance the quality of care for Medicare beneficiaries, including those with low incomes, such as dual-eligibles, and improve the coordination and efficiency of healthcare service delivery."[17] While the program has historically looked at cost-sharing initiatives and promoting high-value clinical services in a relatively small number of states, VBID has since expanded to all 50 states with new innovations, including the ability to test specific non-primary health related benefit designs based on the demographics or chronic conditions of enrollees. Benefits, such as transportation assistance, can now be structured in to specifically alleviate social determinants of health pressures for certain demographics in participating plans. While only 1.6 million beneficiaries are expected to be targeted with new supplemental benefits through the VBID Model in 2021, Humana by far has the largest presence of the 19 participants, bringing new innovations to plans in 41 states and Puerto Rico. This provides Humana greater flexibility than any other payor when experimenting with new transportation benefit design, as we will later propose. While structuring specific benefits to demographics is confined to the VBID Model today, we expect that innovations emerging from the VBID Model will be adapted to broader plans to the extent that those innovations fulfill the goals of enhancing quality of care, improving care service delivery, and reducing program expenditures.

3.3. OPTIONS FOR STRUCTURING NEMT BENEFITS

When it comes to creating a transportation benefit, payors historically have had two options: operate the benefit in-house by directly managing transportation vendors or outsource the operation of the benefit to a transportation broker. First, by operating the benefit in-house, payors are taking on the responsibility of contracting with transportation fleets to operate transit rides, providing vouchers to enrollees for public transportation, or managing the enrollee reimbursement for finding transportation on their own means. It is a difficult model to scale, especially to cover Humana's population, and is an option that works best when beneficiary needs are similar and predictable. Lyft and Uber Health have begun to offer NEMT solutions, but drivers will not have the training or door-to-door service level that a specialized transportation service could offer. While this could potentially be the cheaper of the two options, it works best for smaller payors with a homogenous enrollee pool.[18]

Payors have the option to outsource the operation of a transportation benefit to a broker, who connects the plan and beneficiary with transportation providers and drivers. The largest brokers have built legacy businesses supplying NEMT services to Medicaid plans because these plans need cost containment, quality control, safety, and budget predictability. Due to the regulatory changes discussed above, many brokers have recently begun to adopt and sell to Medicare Advantage plans. With a 55-year history of NEMT benefits to Medicaid, these brokers have already built national networks to schedule rides and support beneficiaries. These national networks include sophisticated fleets of vehicles and drivers capable of delivering room-to-room transit and accommodate wheelchairs or stretchers if needed. On the broker side, comprehensive services include identifying the eligibility of an individual, call center operation, scheduling rides, ride-tracking, provider payment, and fraud, waste, and abuse detection – all capabilities a payor would have to build out should they choose to self-operate. Brokers are typically paid a capitated rate by the health plan for offering these services. While the legacy providers have been established for decades, the service itself has become slightly commoditized, with few points of differentiation between them. With the rides only experienced by beneficiaries at the local fleet level, the transportation networks themselves look similar and most support service capabilities are indistinguishable. Legacy brokers include LogistiCare (subsidiary of Providence Service Corporation), MTM (private), Access2Care (subsidiary of Global Medical Response / AMR), and SoutheasTrans (private).

Unfortunately, the legacy broker market has been slow to adopt new technologies such as real-time ride tracking, passenger scheduling optionality, event-based interventions, and sophisticated fraud, waste, and abuse monitoring that go a long way towards improving the benefit for all parties involved. A new generation of brokers, such as Veyo, Kaizen Health, Ride Health, Roundtrip, and Hitch Health have all seen recent success bringing technology-enabled tools to managing an NEMT benefit. While these companies are usually smaller in scale, they offer more flexible solutions to a payor looking to implement a transportation benefit. For example, if a payor only wanted to outsource a fraction of the management of the benefit, these newer brokers can individually offer dispatching, scheduling, routing, or billing services. From an enrollee side, members benefit from having greater flexibility on how they can book transportation. Where legacy brokers typically require scheduling through a call center, these tech-enabled brokers can offer scheduling through call centers, web portals, or mobile applications. Additionally, these brokers can better match enrollees with the appropriate vehicle type given a real-time view of where drivers with certain capabilities are at a given time. Finally, from a fraud, waste, and abuse perspective, the ability to track in real-time how often enrollees are using the benefit and where their destinations are can help a plan detect ride abuse and rides for non-medical needs outside of the scope of the benefit. While these businesses may not provide the network scale of a legacy broker, the transparency and convenience provided to the beneficiary and plan can lead to greater outcomes and overall satisfaction.

Of course, the other way of assisting enrollees who have transportation issues in accessing medical care is to bring medical care to those enrollees. Through health risk assessments, CMS' Chronic Care Management program, medication therapy management offerings, telemedicine, and other care management initiatives, plans have been able to use third party vendors to bring care directly to a patient's home, both in person and telephonically. Humana already works with Signify Health, Matrix Medical Network, and Your Home Advantage to bring care to enrollees where they live, including wellness screens, diagnostic testing, and comprehensive medication reviews. While these initiatives are difficult to scale to cover every enrolled member, they do provide a way to keep members engaged with their health. Home health services will never be a perfect replacement for care done at a hospital or physician office, but they are a complementary offering that still provides an alternative for those with transportation issues. The rest of our case will focus on our recommendation for constructing and implementing an NEMT benefit.

3.4. RECOMMENDATIONS FOR BUILDING AN NEMT BENEFIT

We believe that in order to overcome transportation challenges as a social determinant of health for a significant portion of Humana's Medicare Advantage population, Humana should institute a comprehensive NEMT benefit across its plans. We will discuss how Humana should target specific sub-populations, design the richness of the benefit, select an NEMT broker, choose payment models to NEMT brokers, and market the benefit.

As a payor, Humana needs to balance providing transportation to those who need it most with potential cost inflation from offering a benefit to those who do not need it. Medicare Advantage plans are not yet able to discriminate benefit construction to sub-populations within plans, and therefore must offer universal access to the same benefit within each plan. Humana should utilize our predictive model to segment their MA plans by average prediction score of transportation challenges per member. We propose that plans with a low concentration of transportation-challenged receive a simple benefit of 12 roundtrips per year or one trip per month. This can serve as both a marketing tool to draw healthier enrollees into a plan, while offering the transportation-challenged access to rides that can be used for medical or non-medical purposes. For plans with higher concentrations of individuals that our model identifies as transportation-challenged, we propose offering a fuller benefit of 24-36 roundtrips per year or two-three trips per month. This allows Humana to offer more frequent transportation options to those who need it most, while avoiding cost inflation of program utilization by healthy enrollees.

The decision threshold for a plan to receive 12 versus 24 versus 36 rides could be determined by the mean and standard deviation of average prediction score among Humana's MA plans. While we do not currently have access to such data, we can make educated assumptions from the dataset provided for the case. The average prediction score for the holdout data is .214. Assuming a standard deviation of ~.04, plans could be segmented into three groups:

- Plans with average predictions below the mean receive 12 rides per year
- Plans with average predictions between .214-.254 receive 24 rides per year
- Plans with average predictions above .254 receive 36 rides per year

Additionally, through the VBID program, Humana has the opportunity to target specific demographics or individuals suffering from chronic illnesses. We recommend that Humana use the VBID Model to experiment with a more generous benefit of 48 rides per year. Through the inclusion of certain features, our predictive model suggests that diabetes, cardiovascular disease (i.e. CHF), bipolar disorder, and major depressive disorder are among chronic conditions that increase likelihood of transportation challenges. This program truly allows Humana to discriminate benefit construction to those who need transportation the most, and it should be used to study the longer-term effects and ROI of investing in the NEMT benefit. As only a small fraction of Humana's enrolled members is participating in the VBID Model, it serves as a useful population to test innovative benefit construction like this without risking too much cost inflation. The number of rides per year that we are recommending are influenced by our interviews with industry leaders, from which a consensus emerged that 36 rides per year represents an average NEMT benefit.[19] We believe that our recommended plan will optimize cost, utilization, medical outcomes, member attraction and retention, as well provide opportunities to study the benefit ROI.

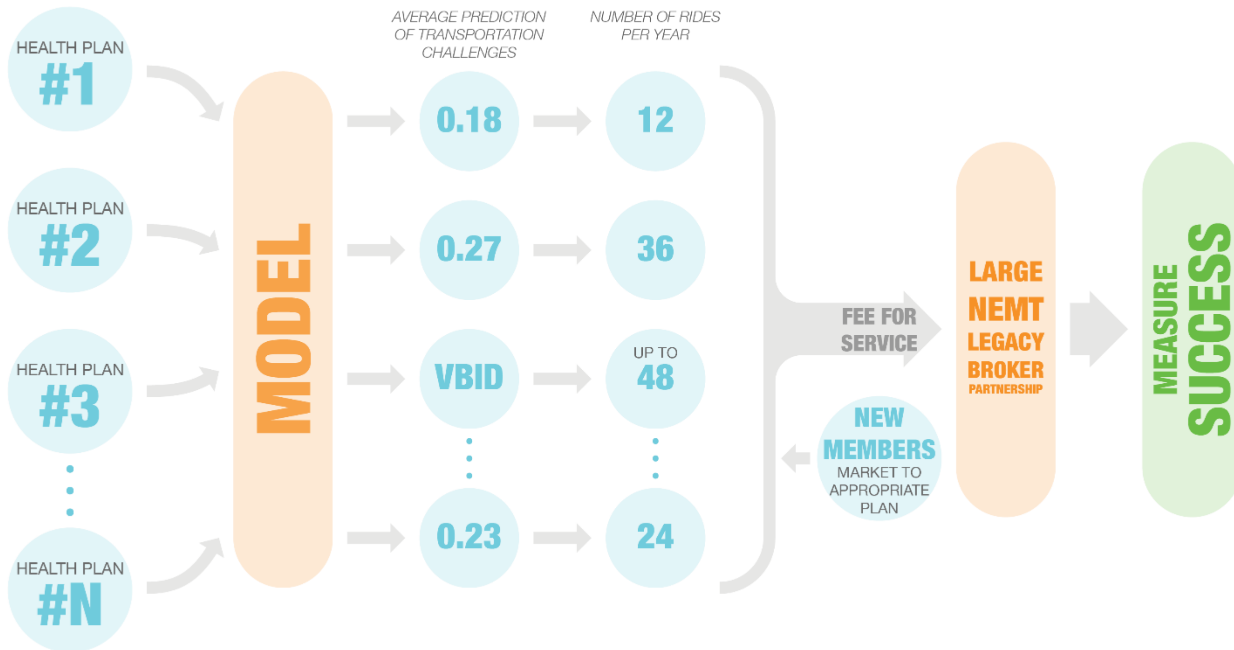
Humana will need an NEMT broker in order to outsource some or all of the operation of the benefit. We suggest that Humana contract with a major, legacy broker, as these firms are the ones who will have the scale and driver network to help roll out the benefit to 4.5 million members. While the new generation of brokers have differentiated themselves on the data tools they have built to collect real-time ride information, the legacy brokers have built or acquired similar tools themselves in order to provide all relevant information back to a health plan. This data is instrumental in measuring the success and utilization of the benefit, but Humana needs the network to be able to roll this out broadly. Another decision to be made is whether to outsource all the benefit operations or keep some services in-house. We interviewed two employees at UnitedHealthcare that led efforts to coordinate United's insurance offerings with National MedTrans' NEMT brokerage following the 2017 acquisition.[19,20] While United wanted an in-house business running the NEMT management it would have otherwise outsourced, it became clear to us that the effort failed as the business lost money under pressure to offer a more favorable but expensive experience to riders. United ultimately sold National MedTrans to LogistiCare in May 2020, just three years after the acquisition.

Transportation benefits are "a unique animal," and while some payors may want their care management team working the call centers of ride scheduling, there is an impressive amount of coordination needed between the plan, broker, dispatcher, driver, and patient. There is an inherent tension for brokers to want to lower costs while payors are supporting favorable patient experience at a lower cost, and the best way to keep member satisfaction with the benefit high is to outsource this to someone with the established capabilities to run it properly.

Humana will then need to determine how it wants to pay brokers for running the benefit. Medicaid plans have historically paid a capitated PMPM rate to brokers for the benefit, as this is a required offering for the plan and is sound from a cost containment perspective for the broker to go at risk for utilization. However, some brokers have the reputation of being fierce managers of utilization in order to profitability manage the benefit and may inconvenience the patient to save money while hurting the reputation of the plan. Medicare Advantage plans, on the other hand, are in the early stages of adding NEMT as a supplemental benefit and do not yet have significant data to understand how utilization looks. While some Medicare Advantage plans have opted to pay NEMT brokers a PMPM rate with risk corridors, we recommend using a fee-for-service model to pay brokers on a per ride basis, at least initially. While Humana bears the risk for the elevated cost associated with higher utilization of the benefit, structuring the number of NEMT rides by plan based on relative need helps to ensure that those who need the rides the most have the greatest access. The broker can work to ensure a comprehensive and favorable benefit for the patient, hopefully strengthening enrollee satisfaction ratings without Humana paying a higher expense than needed. After several years, once the utilization data is known, it may be in Humana's interest to restructure the payment model to a PMPM rate.

Finally, Humana should actively market the benefit through plan websites and promotional materials. With only 33% of Medicare Advantage plans offering a transportation benefit in 2020, this supplemental benefit could be a differentiator to both health and transportation-challenged enrollees. Acting early in the transportation benefits lifecycle and executing well on the implementation could draw both new enrollees to Medicare Advantage as well as transferred enrollees to Humana and create a stronger reason to retain those enrollees. Humana should supplement a broader information campaign with specific information to those identified or predicted to be transportation-challenged. Sending mail information packets, email campaigns, and directly calling these individuals creates awareness of the benefit and drives utilization for those who will see the greatest improvement in health outcomes. Additionally, Humana could use its network of home health patient providers (Signify Health, Matrix Medical Network and Your Home Advantage, among others), to promote the benefit and serve as a live resource for those wishing to learn more.

Our recommendations can be summarized in the following visualization.



3.5. MEASURING SUCCESS OF PROPOSED PROGRAM

When evaluating the success of the recommended NEMT program, Humana should measure how well transportation-challenged enrollees can access the benefit, how much cost containment the benefit generates, the plan membership growth and retention rates, and overall member experience and satisfaction with the plan. First, Humana can look at how well the transportation-challenged members can access the benefit and are therefore able to go to sites of care. Working with an NEMT broker that collects real-time data for ride utilization, the payor can track how much the benefit is being utilized on an enrollee, sub-population, or plan basis. If the transportation-challenged enrollees are able to utilize this benefit, it can be an indicator of success. Other performance metrics could include the number of unique individuals using the transportation benefit as an indicator for how well patients know about and understand the benefit. While a higher number of unique individuals using transportation may be cost inflationary, a higher number of transportation-challenged individuals using the benefit could be seen as successful implementation to those who need transportation the most. Humana can also track the changes in new member adoption rates, as well as member retention, for each plan after implementing an NEMT benefit.

Secondly, though it is difficult to tie transportation benefits to clinical adherence due to lags in the data for a broader population, it is still possible to look more closely at how clinical adherence, and therefore longer-term cost mitigation, is improved for sub-populations. By tying claims data to a specific ride, Humana can analyze how provider visits and clinical adherence improves for those who are transportation-challenged over a long period of time. Across the broader plan, a reduction in the number of emergency room visits and number of hospitalizations could also be indicative of success. Assuming transportation is used for medical purposes, and these new provider visits improve health status over the long term, the plan should see a decline in the number of emergency room visits and hospitalizations, specifically among transportation-challenged individuals. It is through these performance metrics that Humana will see the greatest ROI for implementing an NEMT program.

Finally, Humana can track overall member experience and satisfaction with the plan. With much of an MA payor's incentive payments tied to plan rating, this would be important to optimize. According to the CEO of Ride Health, transportation benefits for Medicaid populations, while being such a small portion of total plan spend at 1%, can contribute as much as 60-70% of complaints and grievances.[21] Enrollees have a certain expectation for how transportation should work, and if that expectation is not met then complaints are filed. Again, by working with a tech-enabled NEMT broker, real-time data helps to mitigate this when event-based status updates can allow the broker to intervene with the passenger, coordinator, or driver if there are issues, reducing total number of complaints. Additionally, by paying on a per ride basis, this removes the incentive for the broker to actively manage utilization to the detriment of member experience, and the broker can focus on creating a good benefit for the beneficiary. By frequently surveying beneficiaries of their experience with the benefit, as well as seeing how member retention is improved or hurt by the addition of the benefit between years, Humana will have an indicator of the benefit's success.

3.6. RISK FACTORS

One risk factor for adding an NEMT benefit is that the benefit potentially becomes cost inflationary for Humana. As a payor cannot discriminate in its plan structure to target specific demographics or patients with chronic conditions, an NEMT benefit must be universally available to all within a plan, including the healthiest patients (outside of the VBID Model). If relative utilization skews towards healthy patients who want to take advantage of having free trips once or twice a month, Humana will see higher costs of the benefit without longer-term outcomes improvement. The true success of adding the benefit relies on those who are transportation-challenged fully utilizing the benefit in order to more actively access care or non-medical destinations that can improve their health. Without that utilization, an NEMT benefit could be seen as an expensive marketing tool to draw enrollees into a plan. Our proposed strategy of segmenting plans by concentration of transportation-challenged individuals helps mitigate this risk. Adjusting the size of the benefit to target plans with the highest share of transportation-challenged members creates a greater opportunity for those who need it most to actually use the greatest number of rides.

A second risk factor is that there may be no definitive benefit from providing NEMT services to the transportation-challenged. A clear ROI has been established for Medicaid NEMT benefits due to the federal requirements of the plan and the share of disabled patients enrolled. After speaking with United's National MedTrans team, as well as operators of new brokers, the Medicare Advantage transportation benefit is so new that there is limited proof tying the benefit to improved clinical outcomes and lower longer-term cost of care. Due to the incredible amount of coordination required between the plan, broker, provider, and patient to align care, seeing success over a long period of time is difficult. In fact, the benefit may be more aligned with a healthcare provider than with the plan, as in a fee-for-service environment, healthcare providers are only paid when patients arrive for appointments.[22] While the benefit will still be appreciated by members as an extra incentive to enroll in a Humana plan, the correct partner for implementation is crucial in order to correctly monitor in real-time the utilization of patients that need this the most. Our plan structuring allows Humana to attract the healthiest people to the platform who may use the benefit the least by scaling the benefit based on relative share of transportation-challenged enrollees. However, only by working with a sophisticated broker that can supply the data to tie claims and provider visits to specific rides will an ROI be established in the long run.

A final risk factor of expanding supplemental benefits, including a transportation benefit, is related to the projected insolvency of the Medicare Trust Fund.[23] Medicare is financed through a Hospital Insurance (HI) trust fund and a Supplementary Medical Insurance (SMI) trust fund, each of which is financed through payroll taxes for the former and monthly premiums and general revenues for the latter. Medicare Advantage is paid for by the government to payors through a combination of these funds. Unfortunately, due to rapid growth in program spending, the 2020 Medicare Trustees Report projects that the HI trust fund will be insolvent by 2026. This does not account for the COVID-19 pandemic, which has reduced the payroll taxes paid into the Medicare Trust Fund by reducing employment and wages. Unless Congress can find new sources of funding to help fund the Medicare program, CMS may have to cut rates or cut benefits in order to curb expenses. This is a major issue for benefits consultants as identified by a Senior Director of Policy Affairs at Willis Towers Watson we interviewed, as supplemental benefits for Medicare are value-additive in nature.[24] There is a risk that the recent expansion of supplemental benefits for Medicare Advantage plans, including transportation benefits, could be rolled back to fund core hospital, physician, and prescription drug payments. While the cost containment of an NEMT benefit is generally observed in the long run, a potential loss from investing in such a program could hamper these efforts.

4. CONCLUSION

Humana is interested in helping their Medicare Advantage members overcome transportation barriers to receiving medical care, with the goal of improving health and economic outcomes. In our case study, we developed a machine learning model capable of identifying transportation-challenged members with high confidence. We also proposed a benefit design structure to attract and engage members, optimize health outcomes, and control the cost of the benefit.

We believe Humana should deploy NEMT benefits across its Medicare Advantage plans. Humana should apply our model to each of these plans to determine the relative concentrations of transportation-challenged members. Plans with high concentrations of transportation challenges should receive more robust benefits of up to 48 rides per year, while plans with low concentrations should receive up to 12 rides per year. The benefits should be operated by a large, legacy NEMT broker and marketed to individuals identified as transportation-challenged.

This plan is designed to maximize measurable ROIs, such as membership growth and retention, member satisfaction, benefit utilization ratios, while minimizing cost. It will also provide the time and data required to study the more abstract ROIs, such as reduced hospital admissions rates and improved health outcomes. As a whole, our plan will enable Humana to extract significant value from the $540 million savings opportunity that transportation barriers pose to their business, while improving the care and experience provided to the member.

5. REFERENCES

[1] "Travel Patterns of American Adults with Disabilities." *Bureau of Transportation Statistics*, 11 Dec. 2018, www.bts.gov/topics/passenger-travel/travel-patterns-american-adults-disabilities.

[2] "Circulation Health: On-Demand NEMT for Hospitals, Health Plans." www.circulation.com/.

[3] Nuti, Lynn A, et al. "No-Shows to Primary Care Appointments: Subsequent Acute Care Utilization among Diabetic Patients." *BMC Health Services Research*, 2012, www.ncbi.nlm.nih.gov/pmc/articles/PMC3470968/pdf/1472-6963-12-304.pdf.

[4] "National Health Expenditures 2017 Highlights." *CMS*, 2017, www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/highlights.pdf.

[5] Jacobson, Gretchen, et al. "A Dozen Facts About Medicare Advantage in 2019." *Kaiser Family Foundation*, 6 June 2019, www.kff.org/medicare/issue-brief/a-dozen-facts-about-medicare-advantage-in-2019/.

[6] Syed, Samina T., et al. "Traveling Towards Disease: Transportation Barriers to Health Care Access." *J Community Health*, Oct. 2013, www.ncbi.nlm.nih.gov/pmc/articles/PMC4265215/pdf/nihms646723.pdf.

[7] Artiga, Samantha, and Elizabeth Hinton. "Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity." *Kaiser Family Foundation*, 10 May 2018, www.kff.org/racial-equity-and-health-policy/issue-brief/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/.

[8] Jacobson, Gretchen, et al. "Medicare Advantage 2020 Spotlight: First Look - Data Note." *Kaiser Family Foundation*, 24 Oct. 2019, www.kff.org/report-section/medicare-advantage-2020-spotlight-first-look-data-note/.

[9] Hughes-Cromwick, P., and R. Wallace. *Cost Benefit Analysis of Providing Non-Emergency Medical Transportation*. Transit Cooperative Research Program, Transportation Research Board of the National Academies, Oct. 2005, altarum.org/sites/default/files/uploaded-publication-files/05_project_report_hsd_cost_benefit_analysis.pdf.

[10] "Humana Reports Second Quarter 2020 Financial Results; Continues Proactive Relief Efforts Amid Pandemic." *Humana*, 5 Aug. 2020, humana.gcs-web.com/static-files/d1483126-54dd-4bd2-ad3a-ede66a1c5f50.

[11] Musumeci, MaryBeth, and Robin Rudowitz. "Medicaid Non-Emergency Medical Transportation: Overview and Key Issues in Medicaid Expansion Waivers." *Kaiser Family Foundation*, 24 Feb. 2016, www.kff.org/medicaid/issue-brief/medicaid-non-emergency-medical-transportation-overview-and-key-issues-in-medicaid-expansion-waivers/.

[12] Cronin, J. Joseph. "Florida Transportation Disadvantaged Services: Return on Investment Study." *The Marketing Institute - Florida State University College of Business*, Mar. 2008.

[13] Adelberg, Michael, et al. "Non-Emergency Medical Transportation: Findings from a Return on Investment Study." *Medical Transportation Access Coalition*, 2018, mtaccoalition.org/wp-content/uploads/2018/07/NEMT-ROI-Methodology-Paper.pdf.

[14] "CMS Changes Pave the Way for Medicare Advantage Transportation Benefits in 2020." *SafeRide Health*, 13 Feb. 2019, www.saferidehealth.com/blog/medicare-advantage-transportation-in-2020.

[15] Isgur, Benjamin, et al. "CMS Expands Medicare Advantage Coverage for Social Determinants of Health." *Health Research Institute - Regulatory Insight: January 2019*, PricewaterhouseCoopers, Jan. 2019, www.pwc.com/us/en/health-industries/health-research-institute/pdf/CMS-expands-MA-social-determinants_PwC_Jan2019.pdf.

[16] *S.870 - Creating High-Quality Results and Outcomes Necessary to Improve Chronic (CHRONIC) Care Act of 2017*. 115th Congress (2017-2018), 9 Feb. 2018, www.congress.gov/bill/115th-congress/senate-bill/870/text.

[17] "Medicare Advantage Value-Based Insurance Design Model." *Centers for Medicare & Medicaid Services Innovation Center*, innovation.cms.gov/innovation-models/vbid.

[18] "The Current State of Medical Transportation: a Landscape Analysis." *SafeRide Health*, 13 Mar. 2019, www.saferidehealth.com/blog/medical-transportation-landscape.

[19] "Interview with Former Vice President of Product at UnitedHealth Group." 8 Oct. 2020.

[20] "Interview with Director of Emerging Benefits at UnitedHealth Group / Former Director of Business Performance at National MedTrans." 7 Oct. 2020.

[21] "Interview with CEO of Ride Health." 6 Oct. 2020.

[22] "Interview with Former Chief Revenue Officer at Circulation." 7 Oct. 2020.

[23] Davis, Patricia A. "Medicare: Insolvency Projections." *Congressional Research Service*, 29 May 2020, fas.org/sgp/crs/misc/RS20946.pdf.

[24] "Interview with Senior Director of Policy Affairs at Willis Towers Watson." 2 Oct. 2020.