Predicting drivers of savings in the Medicare Shared Savings Program: CS-109 Final Project - Milestone 2:

Obinna Ahanotu-Ejikeme, Aaron Coe, Ruizhe Kang, and Eshan Tewari

December 1, 2021

1 INTRODUCTION

In 2019, the United States collectively spent $3.8 trillion on healthcare—more than $11,000 per person. This is nearly 15% higher than it was at the start of decade, with no obvious signs of slowing; at current rates, our healthcare costs are expected to rise to almost $7 trillion by 2028. Much of these costs get passed on to consumers in the form of higher premiums, higher deductibles, and higher contributions to support dwindling Medicare funds. Needless to say, the state of our healthcare industry is very grim, and we need to take drastic cost-containment actions over the coming decade to slow or reverse the trend we are seeing[1].

The biggest drivers of our healthcare spending are hospital care (31% of overall expenditure) and physician services (20% of overall expenditure). Broadly, these two collectively encompass care that Americans receive from physician organizations. The vast majority of these providers are paid in a Fee for Service (FFS) fashion, wherein they are paid by insurers for completing services (for instance, conducting an annual physical, doing an MRI, completing a surgery, etc.). Put otherwise, physicians complete services, submit bills to insurers, and insurers reimburse them per negotiated rates. As one can imagine, this can create an incentive where physicians are reimbursed more for doing more, thereby creating a system that heavily rewards "over-treatment" and does not motivate accountability for good outcomes. There has been much dialogue about a new payment paradigm named "Value Based Care," wherein physicians assume different levels of risk for managing patient populations and face financial penalties for poor outcomes (for instance, if a patient is admitted to the emergency room for an avoidable reason, the physician group responsible for that patient's care is financially penalized). Our aim is to evaluate results from one such program: the Medicare Shared Savings Program (MSSP).

Medicare is a public insurance in the United States that primarily covers individuals 65 years and older, but also covers people with certain disabilities (chiefly, End Stage Renal Disease, i.e. kidney failure). The MSSP traces its roots to the George W. Bush administration, but was formally entrenched following the passage of the Affordable Care Act (ACA) in 2012. It is designed to promote a value based care arrangement called an Accountable Care Organization (ACO), in which groups of doctors, hospitals, and other health care providers, who come together voluntarily to give coordinated high-quality care to their Medicare patients. The details of the program are extremely complicated[2]. On a high level, if a physician group decides to partake in MSSP, CMS (the insitution that administers Medicare) computes the expected costs for the patients under the ACO's care, and then reimburses the physician group based on how much they save relative to that expected cost (also known as the "benchmark"). There are different risk tracks (Fig. 1), which enable providers to assume different levels of upside and downside risk. Generally, the more downside potential the ACO is willing to assume (i.e. in the case that they exceed expected costs, they will have to pay Medicare a portion of those losses), the more upside potential they have[3].

Characteristic	BASIC Track's Glide Path				ENHANCED Track (risk/reward)
	Level A & Level B (one-sided model)	Level C (risk/reward)	Level D (risk/reward)	Level E (risk/reward)	
Shared Savings (once Minimum Savings Rate (MSR) met or exceeded)[1]	1st dollar savings at a rate of 40% if quality performance standard is met; not to exceed 10% of updated benchmark	1st dollar savings at a rate of 50% if quality performance standard is met, not to exceed 10% of updated benchmark	1st dollar savings at a rate of 50% if quality performance standard is met, not to exceed 10% of updated benchmark	1st dollar savings at a rate of 50% if quality performance standard is met, not to exceed 10% of updated benchmark	1st dollar savings at a rate of 75% if quality performance standard is met, not to exceed 20% of updated benchmark
Shared Losses (once Minimum Loss Rate (MLR) met or exceeded)	N/A	1st dollar losses at a rate of 30%, not to exceed 2% of ACO participant revenue capped at 1% of updated benchmark	1st dollar losses at a rate of 30%, not to exceed 4% of ACO participant revenue capped at 2% of updated benchmark	1st dollar losses at a rate of 30%, not to exceed 8% of ACO participant revenue in 2019-2024, capped at 4% of updated benchmark. The loss recoupment limit is the percentage of revenue specified in the revenue-based nominal amount standard under the Quality Payment Program (QPP)[2] capped at 1 percentage point higher than the benchmark-based nominal risk amount[3]	1st dollar losses at a rate based on quality performance, with minimum shared loss rate of 40% and maximum of 75%, not to exceed 15% of updated benchmark

Fig. 1: The Medicare Shared Savings Program (Shared Savings Program) offers different participation options (tracks) that allow Accountable Care Organizations (ACOs) to assume various levels of risk. The table above summarizes the characteristics of the participation options under the BASIC track and ENHANCED track[3]

Our goal in this project is to use publicly available MSSP data to predict drivers of savings in the MSSP program. The MSSP public use file contains over 200 characteristics on each ACO that participates in the program, and our primary outcome variable will be the gross margin saving percentage (named Sav_rate below) that the ACO produced in the January 2020 to December 2020 period (Total Benchmark Expenditures Minus Assigned Beneficiary Expenditures as a percent of Total Benchmark Expenditures).

2 EXPLORATORY DATA ANALYSIS

2.1 Data Characterization

Before commencing EDA, we grouped variables into the categories described below. We separated our EDA between these groups to obtain more meaningful insights. Specifically, there are 162 predictors in total, which we divided into 12 groups as follows:

1. Basic information for the ACOs, 12 in total:
    - ACO_ID
    - ACO_Name
    - 'ACO_State'
    - 'Agree_Type'
    - 'Agreement_Period_Num'
    - 'Initial_Start_Date'
    - 'Current_Start_Date'
    - 'N_AB'
    - What saving models have the ACOs participated:
        - 'Risk_Model'
        - 'Adv_Pay'
        - 'AIM'
        - 'SNF_Waiver'
2. Saving related (these are very likely to be our response variables), 5 in total:
    - 'Sav_rate'
    - 'MinSavPerc'

- 'BnchmkMinExp'
- 'GenSaveLoss'
- 'EarnSaveLoss'

3. Which track and BASIC Level the ACOs selected, 10 in total:
   - 'Current_Track_1'
   - 'Current_Track_2'
   - 'Current_Track_3'
   - 'Current_Track_1_Plus'
   - 'Current_BASIC_A'
   - 'Current_BASIC_B'
   - 'Current_BASIC_C'
   - 'Current_BASIC_D'
   - 'Current_BASIC_E'
   - 'Current_ENHANCED'
4. Per capita ESRD/DISABLED/AGED(DUAL)/AGED(NON-DUAL) expenditures in benchmark year 1,2,3 and performance year, 17 in total:
   - 'Per_Capita_Exp_ALL_ESRD_BY1'
   - 'Per_Capita_Exp_ALL_DIS_BY1'
   - 'Per_Capita_Exp_ALL_AGDU_BY1'
   - 'Per_Capita_Exp_ALL_AGND_BY1'
   - 'Per_Capita_Exp_ALL_ESRD_BY2'
   - 'Per_Capita_Exp_ALL_DIS_BY2'
   - 'Per_Capita_Exp_ALL_AGDU_BY2'
   - 'Per_Capita_Exp_ALL_AGND_BY2'
   - 'Per_Capita_Exp_ALL_ESRD_BY3'
   - 'Per_Capita_Exp_ALL_DIS_BY3'
   - 'Per_Capita_Exp_ALL_AGDU_BY3'
   - 'Per_Capita_Exp_ALL_AGND_BY3'
   - 'Per_Capita_Exp_ALL_ESRD_PY'
   - 'Per_Capita_Exp_ALL_DIS_PY'
   - 'Per_Capita_Exp_ALL_AGDU_PY'
   - 'Per_Capita_Exp_ALL_AGND_PY'
   - 'Per_Capita_Exp_TOTAL_PY'
5. Final, mean prospective CMS
   HCC risk score for ESRD HCC/DISABLED/AGED(DUAL HCC)/AGED(NON-DUAL HCC) enrollment type in benchmark year 1, 2 ,3 and performance year, 16 in total:
   - 'CMS_HCC_RiskScore_ESRD_BY1'
   - 'CMS_HCC_RiskScore_DIS_BY1'
   - 'CMS_HCC_RiskScore_AGDU_BY1'
   - 'CMS_HCC_RiskScore_AGND_BY1'
   - 'CMS_HCC_RiskScore_ESRD_BY2'
   - 'CMS_HCC_RiskScore_DIS_BY2'
   - 'CMS_HCC_RiskScore_AGDU_BY2'
   - 'CMS_HCC_RiskScore_AGND_BY2'
   - 'CMS_HCC_RiskScore_ESRD_BY3'
   - 'CMS_HCC_RiskScore_DIS_BY3'
   - 'CMS_HCC_RiskScore_AGDU_BY3'

- 'CMS_HCC_RiskScore_AGND_BY3'
- 'CMS_HCC_RiskScore_ESRD_PY'
- 'CMS_HCC_RiskScore_DIS_PY'
- 'CMS_HCC_RiskScore_AGDU_PY'
- 'CMS_HCC_RiskScore_AGND_PY'

6. Total number of assigned beneficiaries categorized by enrollment type, age, gender and race, 21 in total:
    - 'N_AB_Year_ESRD_BY3'
    - 'N_AB_Year_DIS_BY3'
    - 'N_AB_Year_AGED_Dual_BY3'
    - 'N_AB_Year_AGED_NonDual_BY3'
    - 'N_AB_Year_PY'
    - 'N_AB_Year_ESRD_PY'
    - 'N_AB_Year_DIS_PY'
    - 'N_AB_Year_AGED_Dual_PY'
    - 'N_AB_Year_AGED_NonDual_PY'
    - 'N_Ben_Age_0_64'
    - 'N_Ben_Age_65_74'
    - 'N_Ben_Age_75_84'
    - 'N_Ben_Age_85plus'
    - 'N_Ben_Female'
    - 'N_Ben_Male'
    - 'N_Ben_Race_White'
    - 'N_Ben_Race_Black'
    - 'N_Ben_Race_Asian'
    - 'N_Ben_Race_Hisp'
    - 'N_Ben_Race_Native'
    - 'N_Ben_Race_Other'
7. Expenditures per assigned beneficiary person years categorized by different purposes, 12 in total:
    - 'CapAnn_INP_All'
    - 'CapAnn_INP_S_trm'
    - 'CapAnn_INP_L_trm'
    - 'CapAnn_INP_Rehab'
    - 'CapAnn_INP_Psych'
    - 'CapAnn_HSP'
    - 'CapAnn_SNF'
    - 'CapAnn_OPD'
    - 'CapAnn_PB'
    - 'CapAnn_AmbPay'
    - 'CapAnn_HHA'
    - 'CapAnn_DME'
8. Total number of discharges per 1000 person years categorized by different services, 9 in total:
    - 'ADM'
    - 'ADM_S_Trm'
    - 'ADM_L_Trm'

- 'ADM_Rehab'
- 'ADM_Psych'
- 'chf_adm'
- 'copd_adm'
- 'P_SNF_ADM'
- 'prov_Rate_1000'

9. Total number of different types of services per 1000 person years in the performance year, 9 in total:
   - 'P_EDV_Vis'
   - 'P_EDV_Vis_HOSP'
   - 'P_CT_VIS'
   - 'P_MRI_VIS'
   - 'P_EM_Total'
   - 'P_EM_PCP_Vis'
   - 'P_EM_SP_Vis'
   - 'P_Nurse_Vis'
   - 'P_FQHC_RHC_Vis'
10. Total number of different types of medical groups or individuals participating in the ACO in the performance year, 11 in total:
    - 'N_CAH'
    - 'N_FQHC'
    - 'N_RHC'
    - 'N_ETA'
    - 'N_Hosp'
    - 'N_Fac_Other'
    - 'N_PCP'
    - 'N_Spec'
    - 'N_NP'
    - 'N_PA'
    - 'N_CNS'
11. Consumer Assessment of Healthcare Providers and Systems (CAHPS), 23 in total:
    - 'ACO1'-'ACO28'
12. Other predictors, 17 in total:
    - 'DisAdj'
    - 'DisAffQual'
    - 'Met_QPS'
    - 'QualScore'
    - 'RecvdMean'
    - 'RegTrndUpdt'
    - 'PosRegAdj'
    - Benchmark expenditures:
      - 'UpdatedBnchmk'
      - 'HistBnchmk'
      - 'ABtotBnchmk'
      - 'ABtotExp'
    - Advanced payment related:
      - 'Adv_Pay_Amt'

  - 'Adv_Pay_Recoup'
- Sharing related:
  - 'QualPerfShare'
  - 'FinalShareRate'
  - 'RevLossLimit'
  - 'Rev_Exp_Cat'

We also identified columns with missing data:

Variable	**Number of Missing Entries**
DisAdj	512
Adv_Pay_Amt	488
Adv_Pay_Recoup	488
PosRegAdj	101
ACO40	5
ACO42	2
ACO17	1

Table 1: Columns with missing values.

as well as those with the greatest entropy:

Variable	**Entropy**
ABtotBnchmk	2.247277e+08
ABtotExp	2.161871e+08
BnchmkMinExp	1.299521e+07
GenSaveLoss	1.205110e+07
EarnSaveLoss	6.936331e+06
Adv_Pay_Amt	8.186884e+05
Adv_Pay_Recoup	4.089678e+05
N_AB	2.146341e+04
N_AB_Year_PY	2.094654e+04
N_Ben_Race_White	1.913646e+04
N_AB_Year_AGED_NonDual_PY	1.793429e+04
N_AB_Year_AGED_NonDual_BY3	1.710155e+04
Per_Capita_Exp_ALL_ESRD_BY1	1.281000e+04
Per_Capita_Exp_ALL_ESRD_BY2	1.259939e+04
Per_Capita_Exp_ALL_ESRD_BY3	1.246288e+04
N_Ben_Female	1.221305e+04
Per_Capita_Exp_ALL_ESRD_PY	1.167248e+04
N_Ben_Age_65_74	1.014679e+04
N_Ben_Male	9.264029e+03
N_Ben_Age_75_84	6.481909e+03
Per_Capita_Exp_ALL_AGDU_PY	3.599659e+03

Table 2: Columns with greatest variation.

2.2 Visualization

Following our data grouping and overview, we looked at high level program results, focusing on Group 2 - the outcome variables related to shared savings. At the program level, 513 ACOs participated in the shared savings programs (these are our observations), cumulatively serving

1,0614,589 beneficiaries. On average, each ACO served 20,691 patients. The ACOs cumulatively generated \$4,144,939,915 in savings, \$1,861,215,125 of which was captured by CMS and the remainder by the participating ACOs. On avearge, each ACO generated \$8,079,805, leading to \$390 in savings per beneficiary. These are substantial savings that highlight the high level success of the program.

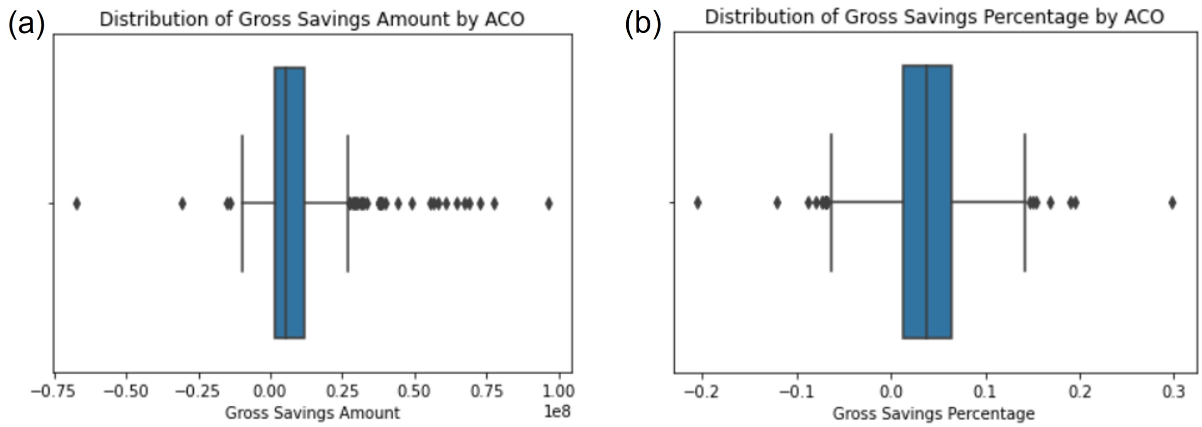

Fig. 2: (a) Gross Savings per ACO (b) Gross Savings Rate per ACO

Next, we look at how ACOs are distributed across risk tracks. These risk tracks determine whether or not ACOs are exposed to downside risk (vs upside-only reward). Opting into downside risk increases the potential upside that ACOs are able to receive, in theory suggesting that ACOs opting into downside risk are more motivated to generate higher savings. Our EDA demonstrates this hypothesis: ACOs in two-sided models tend to generate significantly more savings than ACOs in upside-only models. In the figures below, Track 1 and Basic Levels A/B are upside only, while the rest of the tracks are two-sided.

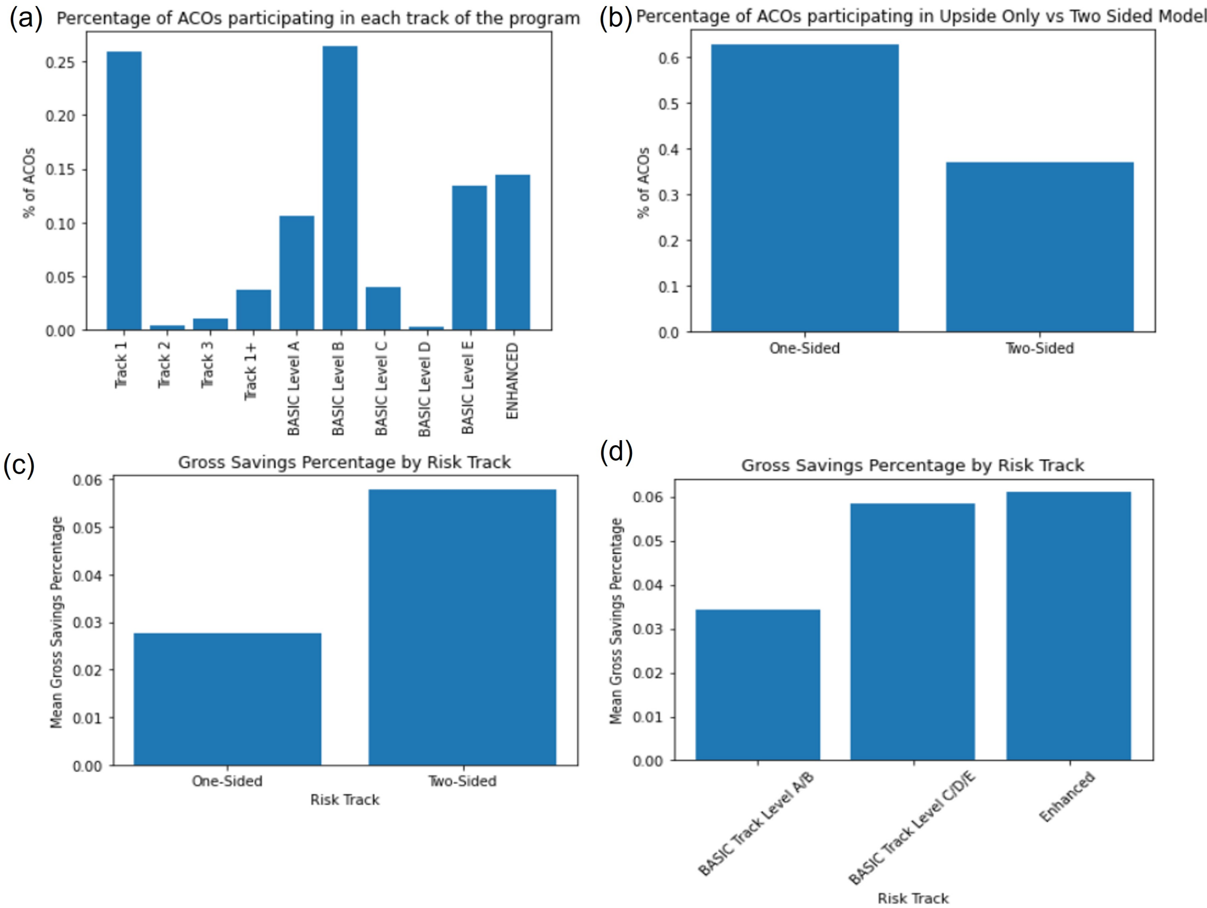

Fig. 3: (a) Percentage of ACOs participating in each track of the program (b) Percentage of ACOs participating in Upside Only vs Two Sided Model (c) Gross savings percentage by risk models (d)Gross savings percentage by risk track

We then moved to Group 6, which primarily contains characteristics of patients served by ACOs and found the following: (1) From Fig. 4, patients that are aged/non-dual comprise an overwhelming majority of beneficiaries. (2) There are slightly more female patients than male patients and (3) patients between 65 and 74 year of age represent the majority of the patient data. (4) The number of white beneficiaries is more than the sum of every other race, which could be a source of the bias for this data set.

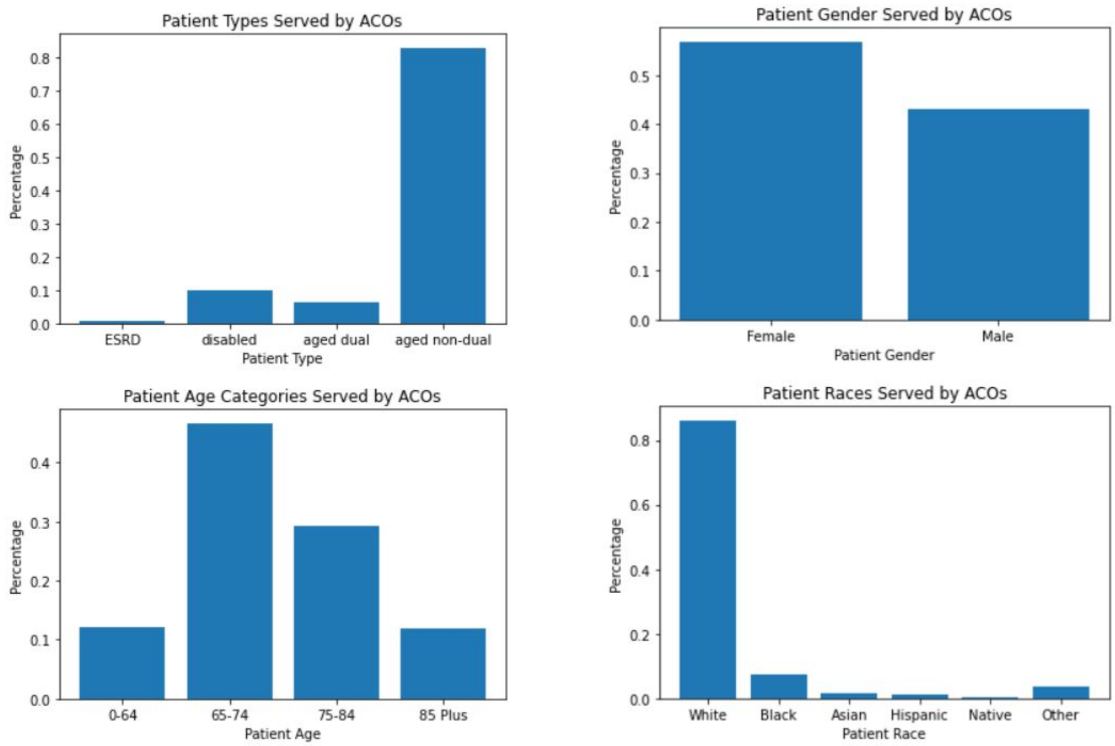

Fig. 4: Patient types categorized by disease, gender, age, and races.

We also analyzed how total expenditure changes with time for each type of patient, shown in Fig. 5. For the four different types of patients, the distribution of the total expenditure doesn't change significantly with time and ESRD patients account for the highest expenditures per capita.

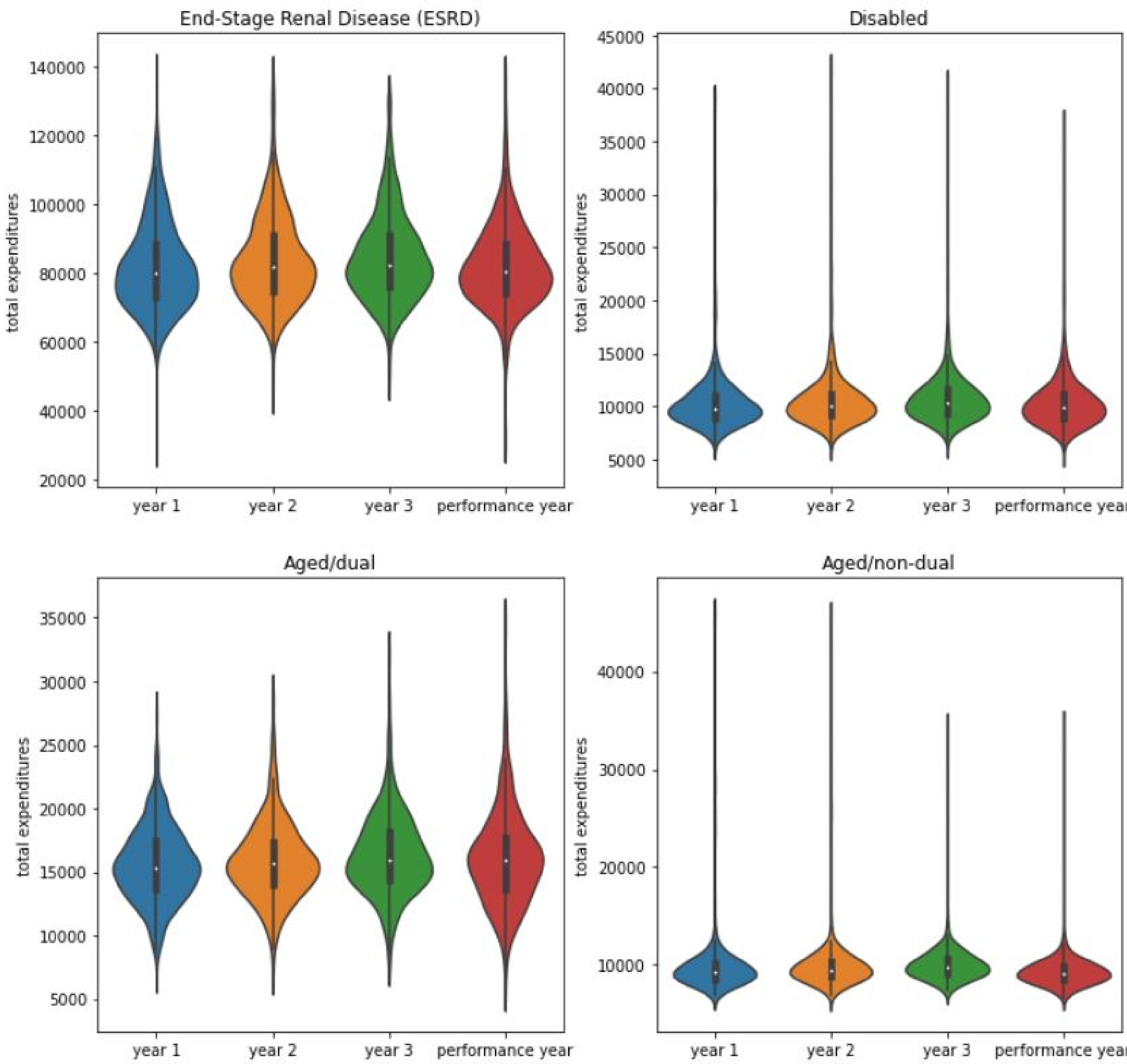

Fig. 5: Violin plots of total expenditures per capita across program participation years for different types of patients.

Provider Services (group 9) provide the most compelling glimpse into drivers of expenditures. On a top-level, short-term acute visits constitute the highest share of hospitalizations and hospital-related expenditures. Skilled Nursing admissions are the second leading driver of hospital-related expenditures, with the remaining categories representing fairly low amounts of expenditure. On a top level, it does seem like more overall hospitalizations lead to fewer shared savings, and more PCP visits correspond to greater shared savings, suggesting that substituting PCP visits and performing preventative care rather than treating patients in the hospital can generate overall program savings.

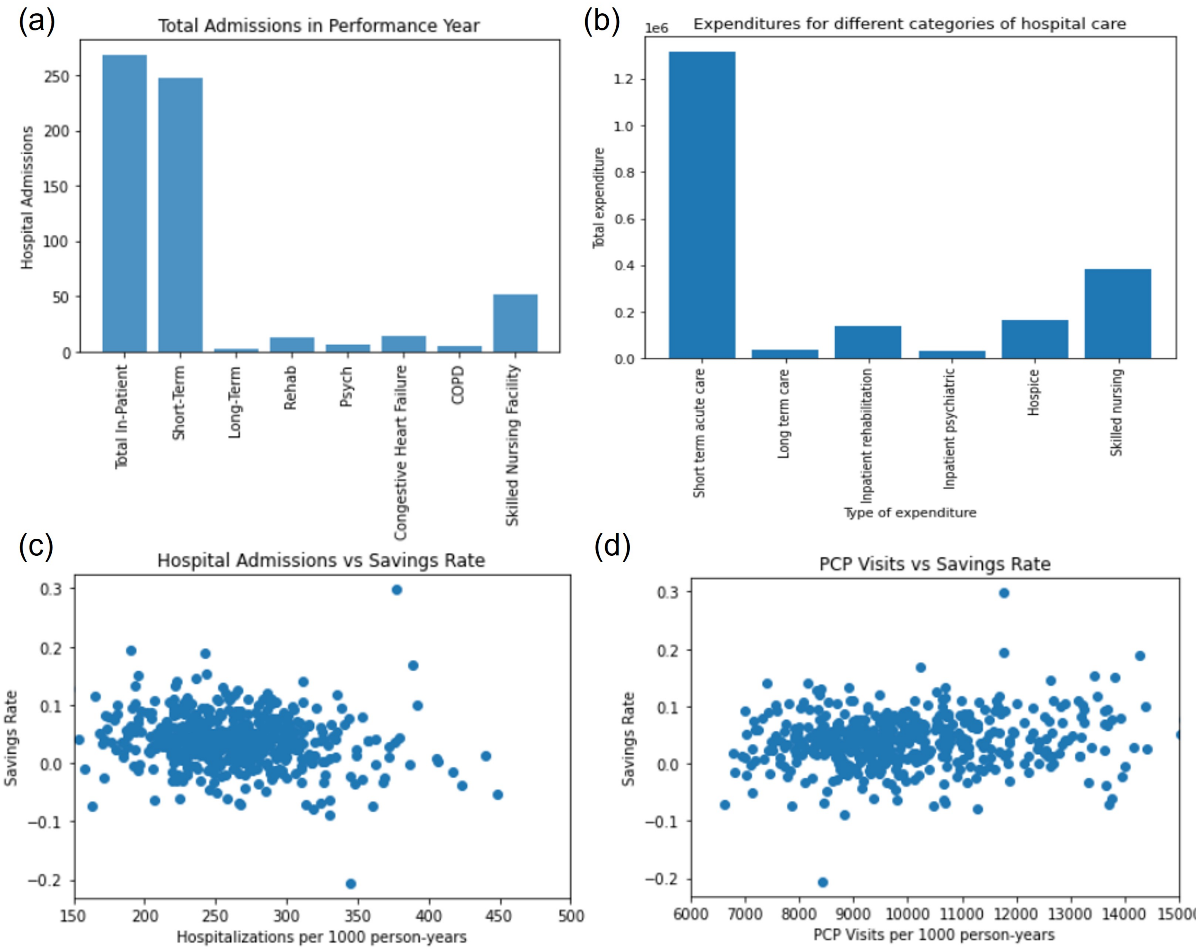

Fig. 6: (a) Admissions to different types of hospital services (b) Expenditures from different types of hospital services (c) Savings Rate vs Hospitalizations (d) Savings Rate vs PCP Visits.

Many variables in Group 11 concern themselves with measures of quality and patient care. As discussed earlier, some of these measures have a considerable amount of missing data, but nevertheless we wanted to explore if any of them could be related to savings rates. While no metric is obviously correlated to savings rates, it does appear that a number of them are positively associated (e.g. ACO43, ACO14, ACO18, and ACO19).

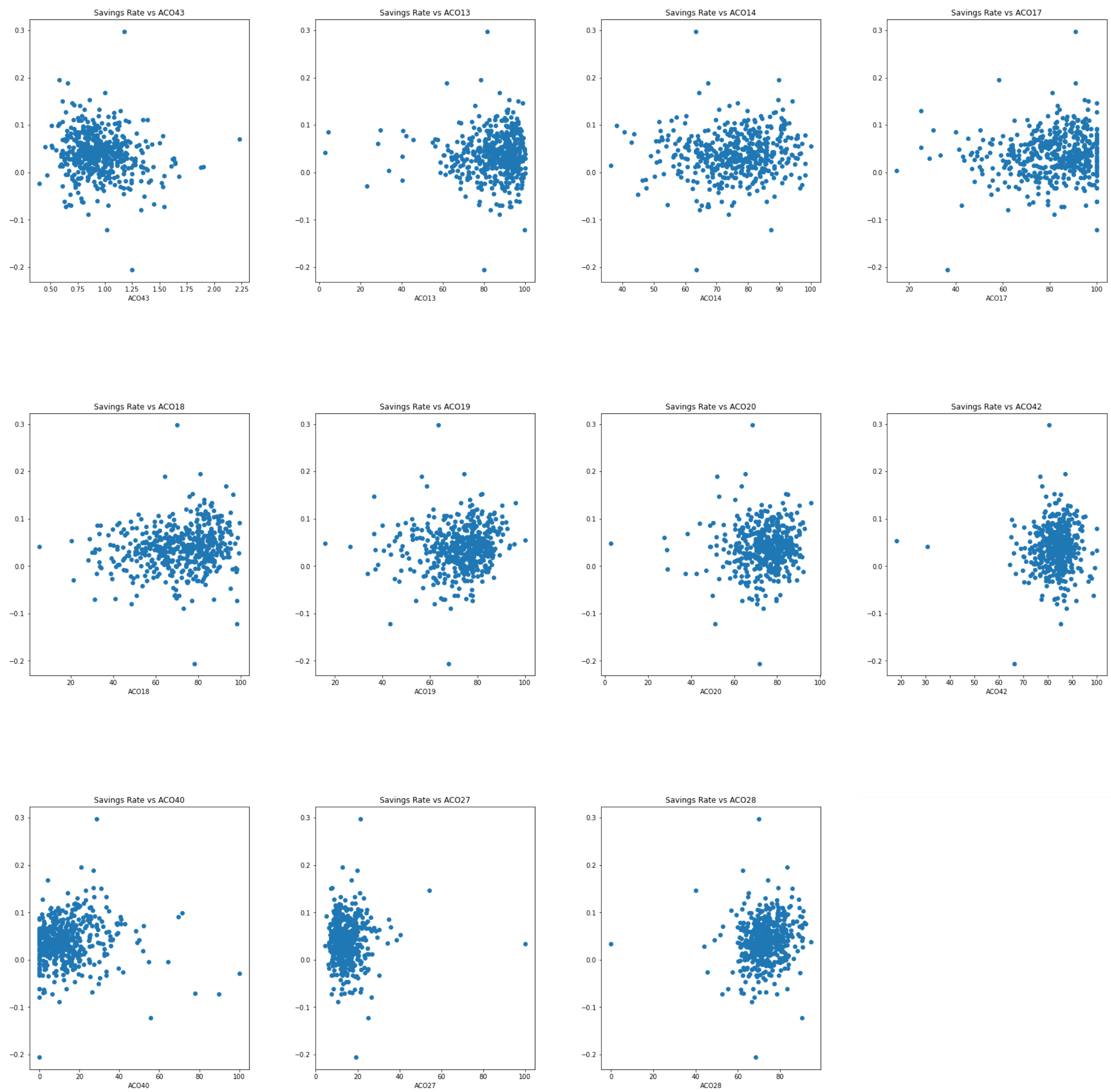

Fig. 7: Scatter plots of different ACO quality measures to savings.

3 BASELINE MODEL

We decided to use a simple Random Forest model as our baseline model based on the work by Larson *et. al.*[4]. The default parameters for the scikit-learn[5] RandomForestRegressor were employed except for the number of features used in each estimator which was set to be the square root of the number of predictors. EarnSaveLoss, GenSaveLoss, and BnchmkMinExp, were removed from the dataset as these variables are used to calculate the savings rate. Additionally, current track variables (such at Current_Track_1) were dropped to remove correlations with the risk model variable. Rev_Expt_Cat, Risk_Model, and Agree_Type are categorical variables and one-hot encoding was used to prepare the data for the model. The training data used 80% of the 2020 data and the rest was used for test data.

The R^2 score for the train data and test data was 0.909 and 0.409, respectively. The variable importance, calculated as the total amount that the mean squared error is decreased due to splits over a given predictor is shown in Fig. 8. The model appears to value expenditures over program structure predictors. The predictors with Per_Capita in the name, such as Pre_Capita_Exp_ALL_AGND_PY, correspond to per capita expenditures on different populations such as non-dual, disabled, etc. CapAnn_OPD and CapAnn_SNF refer to mean expenditures for outpatient expenditures and skilled nursing facilities, respectively. P_SNF_ADM is the total number of skilled nursing facility discharges and P_CT_VIS in the total number of tomography scans. HistBnchmk is the historical benchmark per capita. N_SPEC represents the number of participating specialists in an ACO and is the only program structure variable rated within the top ten features.

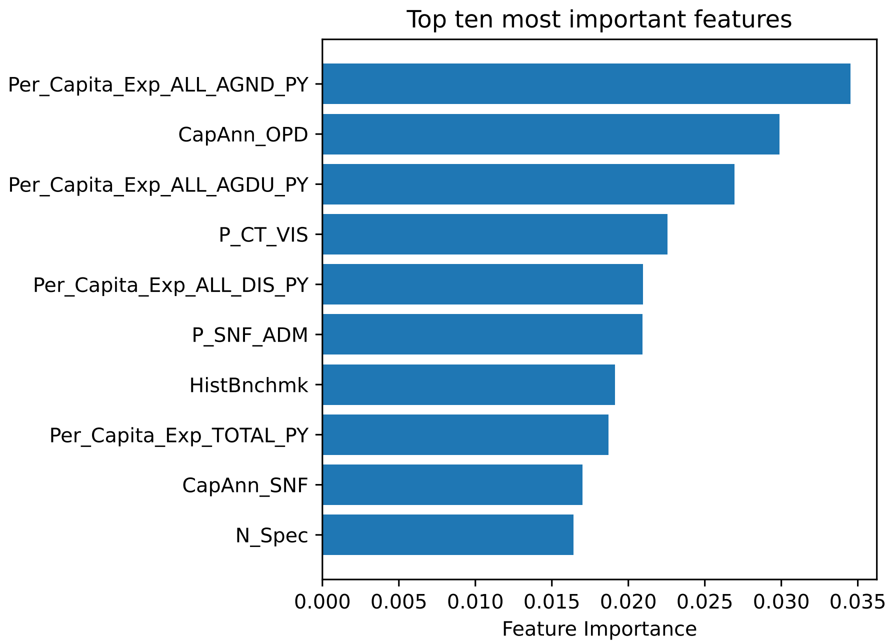

Fig. 8: Baseline Random Forest Model feature importance. Variables with Per_Capita in the name refer to per capita expenditures on certain populations for specific program years. CapAnn variables give the mean for a certain expenditures. P_SNF_ADM and P_CT_VIS are related to the number of skilled nursing facility admissions and total CT scans, respectively. N_Spec is the number of specialists in the ACO program.

Interestingly, the Random Forest model did not rank the risk model choice of the ACO as one of the most important features. This is in contradiction to our EDA analysis that found a noticeably higher average savings rate for ACOs participating in tracks with downside. Our model also ranked the number of specialists participating in the ACO highly, in agreement with the Random Forest model by Larson *et. al.*[4], even though we see minimum correlation in our EDA.

4 REFERENCES

1 J. Probasco, Why do healthcare costs keep rising? (2021).

2 M. S. S. Program, Shared savings and losses and assignment methodology (2018).

3 M. shared savings program, Shared savings program participation options for performance year 2022 (2021).

4 E. M. R. R. G. C. Larson, A., What predictive analytics can tell us about key drivers of mssp results: 2021 update (2021).

5 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Journal of Machine Learning Research **12**, 2825 (2011).