

What predictive analytics can tell us about key drivers of MSSP results

Jill S. Herbold, FSA, MAAA
Anders Larson, FSA, MAAA
Cory Gusland, FSA, MAAA



Executive summary

In a quest to better understand drivers of gross savings and losses in the Medicare Shared Savings Program (MSSP), we applied a rigorous statistical approach to provide insight into the 2015 gross savings/(losses) achieved by accountable care organizations (ACOs) in their third performance year. We incorporated more than 190 objective features from publicly available sources into a machine learning algorithm and found that the following six features were most strongly associated with gross savings:

1. High baseline year 1 (BY1) per capita expenditures, after adjusting for geographic location and morbidity risk (i.e., ACOs that were historically less efficient)
2. National trends higher than local market Medicare fee-for-service (FFS) trends
3. Location in the Southeast and South Central regions designated by the Centers for Medicare and Medicaid Services (CMS)
4. Low annual expenditures for short-term inpatient admissions during the performance year
5. High baseline year 3 per capita expenditures for aged/non-dual beneficiaries, unadjusted
6. High baseline year 3 average CMS-Hierarchical Condition Category (HCC) risk score for aged/non-dual beneficiaries, i.e., populations with higher morbidity

Because more than half of gross savings/(losses) are not explainable by any the features we studied and only one of the top six features (#4) was influenced by an ACO during a performance year, we believe it reasonable to hypothesize that ACO population health efforts and operations can have a material impact on gross savings/(losses). Therefore, it is still possible for ACOs without the six features noted above to succeed in the program.

There are three primary changes to the MSSP financial benchmarking methodology being implemented between 2016 and 2018. We expect that the changes will result in baseline year expenditures and local Medicare FFS trends becoming less predictive of gross savings. However, we anticipate that higher morbidity in the baseline period and low annual expenditures

for short-term inpatient admissions in the performance year will continue to be associated with gross savings despite the policy changes. Because of the shifts toward local market trends and regional efficiency adjustments, it is difficult to assess whether ACOs in the Southeast and South Central CMS regions will continue to achieve higher gross savings than comparable ACOs in other areas.

Introduction

The MSSP has been in place for almost five years, and we are all eager to begin understanding the early results. A number of studies have explored whether the program is achieving savings.¹ Other studies have reviewed the associations between ACO results (gross and shared savings) and a limited number of characteristics and circumstances.² In this study, we applied a rigorous machine learning technique to identify and rank ACO characteristics most associated with results. Specifically, we explored more than 190 variables using a random forest regression model, applied to 2015 results for ACOs in their third performance year.

Our findings align with a few commonly held beliefs that baseline ACO efficiency, risk scores, and local market trends were strongly associated with MSSP financial performance.³ As we explain in this paper, our findings indicate that the top ACO performers in 2015 may not be the top performers in subsequent years because of changes in the MSSP's rebasing and benchmark update methodology. We also found minimal association between some key characteristics and results, including ACO size, whether the ACO was physician-led, demographic and entitlement category mix, most quality metrics, and the number of certain types of providers participating in the ACO. Despite some strong

1. McWilliams, J.M., Hatfield, L.A., Chernew, M.E., Landon, B.E., & Schwartz, A.L. (2016). Early performance of accountable care organizations in Medicare. *N Engl J Med.*; 374:2357–66.
2. Muhlestein, D., Saunders, R., & McClellan, M. (September 9, 2016). Medicare accountable care organization results for 2015: The journey to better quality and lower costs continues. Health Affairs Blog. Retrieved September 12, 2017, from <http://healthaffairs.org/blog/2016/09/09/medicare-accountable-care-organization-results-for-2015-the-journey-to-better-quality-and-lower-costs-continues/>.
3. Agreement Period 1, in 2015 for 2013 starters.

correlation between predicted and actual results (R-squared of approximately 34%⁴), a significant amount of performance is unexplained, which may indicate that ACO care management efforts are accounting for some of the remaining variation.

ACO features found to be strongly associated with gross savings

The table in Figure 1 summarizes the six ACO features determined to be the most strongly associated with gross savings, along with an indication of the relative predictive power of each feature. These six features capture most of the predictive power of the model, but other features were found to have modest association with gross savings. Many of the other predictive features were closely related to those in Figure 1 (for instance, geographic-risk-adjusted per capita costs from BY2).

FIGURE 1: TOP ACO CHARACTERISTICS MOST INFLUENTIAL ON GROSS SAVINGS, 2015 MSSP ACOS IN PERFORMANCE YEAR 3

RANK	ACO FEATURE	VALUES ASSOCIATED WITH HIGHER GROSS SAVINGS	RELATIVE IMPORTANCE ⁵
1	GEOGRAPHIC-RISK-ADJUSTED PER CAPITA COST, BY1	HIGHER ADJUSTED BASELINE COSTS	2.94
2	NATIONAL-TO-LOCAL TREND DIFFERENCE	HIGHER NATIONAL-TO-LOCAL TREND DIFFERENCE	1.75
3	CMS REGION	SOUTHEAST AND SOUTH CENTRAL	1.59
4	CAPPED ANNUAL COSTS FOR SHORT-TERM INPATIENT ADMISSIONS, PY	LOWER PERFORMANCE YEAR INPATIENT COSTS	1.26
5	UNADJUSTED PER CAPITA COSTS, AGED/NON-DUAL, BY3	HIGHER BASELINE COSTS	1.25
6	CMS-HCC RISK SCORE, AGED/NON-DUAL, BY3	HIGHER RISK SCORES	1.03

BASELINE COSTS

Three types of baseline per capita cost measures are included in this analysis: unadjusted, risk-adjusted, and geographic-risk-adjusted. Risk-adjusted per capita costs have been adjusted to account for each ACO’s average risk score and mix of entitlement categories. Geographic-risk-adjusted per

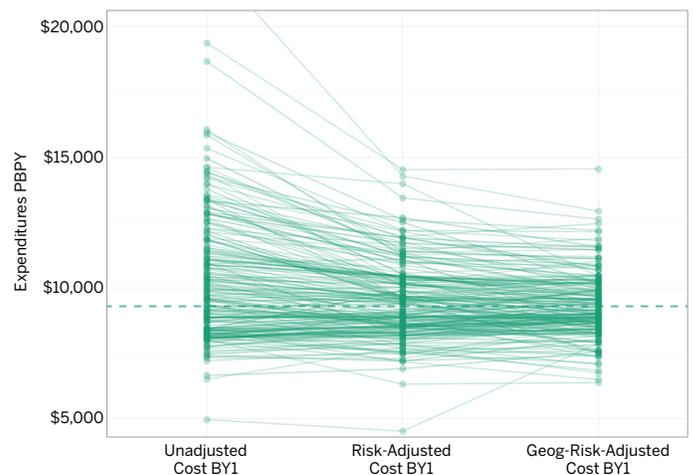
4 The R-squared is estimated from the “out-of-bag” predictions, which are used to approximate the predictive power of the algorithm if applied to new data. We could optimize certain parameters of the algorithm to improve the R-squared, but doing so would reduce the ability to make inferences about such a wide range of features. Achieving maximum predictive power was not the goal of this analysis.

5 The relative importance roughly measures how each feature contributed to the predicted accuracy. The number represents the increase in error if the feature was randomized (therefore, rendering it useless). The relative magnitude of each number is more important than the actual number itself.

capita costs have also been adjusted to account for Medicare reimbursement levels in each ACO’s area. Although both unadjusted and adjusted costs in the baseline period were associated with gross savings, we found that geographic-risk-adjusted per capita costs were more predictive than the other cost measures. One hypothesis is that geographic-risk-adjusted costs are a better measure of “efficiency” in the baseline period, and ACOs that were less efficient in the baseline period have more ability to improve and create low trends. Many ACOs with high unadjusted baseline costs may have had above-average morbidity or above-average Medicare reimbursement levels and were not necessarily less efficient.

Figure 2 illustrates the impact of the geographic adjustment and risk adjustment on BY1 costs for each ACO. Note that there is much less variance among ACOs after the adjustments than there had been prior to the adjustments. However, there is still a relatively wide gap between the most efficient and least efficient ACOs.

FIGURE 2: ACO PER CAPITA COSTS IN BASELINE YEAR 1, BEFORE AND AFTER RISK AND GEOGRAPHIC ADJUSTMENTS, 2015 MSSP ACOS IN PERFORMANCE YEAR 3



To illustrate the finding that geographic-risk-adjusted per capita costs in BY1 were more influential than other measures of baseline cost, we grouped ACOs based on their geographic-risk-adjusted per capita costs in BY1 and their historical benchmarks. The results are shown in Figure 3. Not surprisingly, ACOs that had high costs under both metrics had the most favorable results and ACOs that had low costs under both metrics had the least favorable results. However, we found that ACOs with low historical benchmarks and high geographic-risk-adjusted per capita costs in BY1 had considerably better results (1.3% gross savings) than ACOs with high historical benchmarks and low geographic-risk-adjusted per capita costs (0.2% gross losses) in BY1.

FIGURE 3: AVERAGE GROSS SAVINGS FOR ACOS GROUPED BASED ON TWO MEASURES OF BASELINE COSTS, 2015 MSSP ACOS IN PERFORMANCE YEAR 3

GEOGRAPHIC-RISK-ADJUSTED PER CAPITA COSTS IN BY1	HISTORICAL BENCHMARK PER CAPITA	
	BOTTOM HALF	TOP HALF
BOTTOM HALF	-1.3%	-0.2%
TOP HALF	1.3%	4.7%

Interestingly, we found that geographic-risk-adjusted costs in BY1 (the earliest baseline year) were found to be more predictive of gross savings than costs in other baseline years, despite this year having only 10% weight in an ACO’s historical benchmark (costs are weighted 10%/30%/60% among BY1/ BY2/ BY3, respectively). This could be due to the fact that the three-year benchmark effectively rewards ACOs that have a low trend across the three baseline years. Therefore, based on our analysis, it is always preferable to have a high BY1 cost, whereas a lower BY3 cost can be an indication that the ACO’s costs are already trending in the right direction heading into the performance period. In fact, one of the features in our analysis was the ratio of risk-adjusted costs in BY3 to risk-adjusted costs in BY1. Although this feature was not as predictive as some others discussed in this paper, we found that ACOs with lower ratios (i.e., lower risk-adjusted trends from BY1 to BY3) tended to have higher gross savings.

REGIONAL EFFECTS

Our analysis indicates that the gap between national FFS trends and the FFS trends in an ACO’s local region⁶ were a key driver of ACO success. This result is intuitive in the context of the MSSP benchmarking methodology. Because ACO benchmarks are updated using national FFS trends, if there are factors in an ACO’s region that lead to lower trends than the national FFS population (for instance, changes in wage indices), this will benefit the ACO.

Other studies had noted that when ACOs were grouped by CMS region, gross savings varied widely among regions.⁷ We also found this to be true, but our analysis indicates that much of the variation is explained by geographic-risk-adjusted per capita costs in BY1 and national-to-local trend differences. Figure 4 shows a map with the average gross savings in each CMS region, and Figure 5 shows the geographic-risk-

adjusted per capita costs in BY1 and the national-to-local trend differences in each CMS region. Note that the average gross savings in each CMS region is highly correlated with the other two variables. Our modeling indicated that CMS region was still associated with gross savings after accounting for other factors, but CMS region alone was not as predictive as geographic-risk-adjusted per capita costs in BY1 or national-to-local trend differences.

FIGURE 4: MAP OF AVERAGE GROSS SAVINGS BY CMS REGION, 2015 MSSP ACOS IN PERFORMANCE YEAR 3

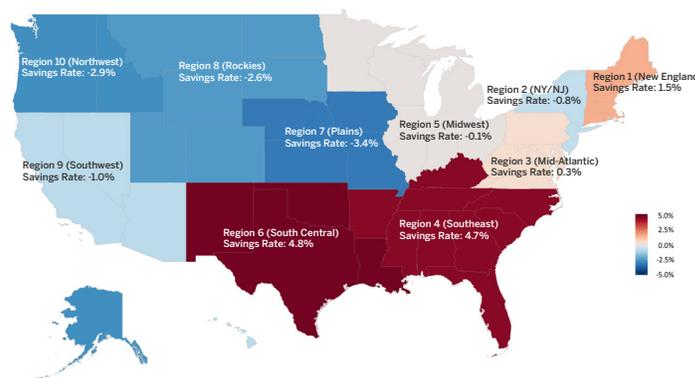


FIGURE 5: GEOGRAPHIC-RISK-ADJUSTED BY1 COSTS, NATIONAL-TO-LOCAL TREND DIFFERENCES, AND AVERAGE GROSS SAVINGS BY CMS REGION, 2015 MSSP ACOS IN PERFORMANCE YEAR 3

REGION	2015 RESULTS FOR 2012/2013 STARTERS		
	NATIONAL VS. LOCAL TREND GAP	GEOGRAPHIC-RISK-ADJUSTED COST, BY1	AVERAGE GROSS SAVINGS PERCENTAGE
REGION 6 (SOUTH CENTRAL)	3.2%	\$10,187	4.8%
REGION 4 (SOUTHEAST)	4.2%	\$9,346	4.7%
REGION 1 (NEW ENGLAND)	1.8%	\$9,490	1.5%
REGION 3 (MID-ATLANTIC)	2.0%	\$8,888	0.3%
REGION 5 (MIDWEST)	2.7%	\$9,347	-0.1%
REGION 2 (NY/NJ)	-0.3%	\$8,775	-0.8%
REGION 9 (SOUTHWEST)	0.0%	\$9,422	-1.0%
REGION 8 (ROCKIES)	0.0%	\$8,845	-2.6%
REGION 10 (NORTHWEST)	-1.3%	\$7,544	-2.9%
REGION 7 (PLAINS)	-0.8%	\$9,458	-3.4%

6 Under the MSSP Final Rule published in June 2016, each ACO’s “region” was defined based on the counties in which ACO beneficiaries reside. These regions are unique to each ACO, and we used this method for estimating regional trends, i.e., local market trends. This definition of “region” is different from the CMS region, which is also used in this report and is defined strictly based on the ACO’s primary state. CMS region is not used directly in the MSSP benchmarking methodology.

7 National Association of ACOs (September 2016). A Look at MSSP ACO Performances Years 1-3. Retrieved September 12, 2017, from <https://naacos.memberclicks.net/comprehensive-analysis-on-performance-year-3-results>.

We also found that CMS-HCC risk scores were predictive of ACO gross savings, although less so than the characteristics discussed earlier. In particular, ACOs that had higher CMS-HCC risk scores in the baseline period tended to have higher gross savings. We found that this was true even after accounting for other features, such as geographic-risk-adjusted per capita costs in BY1. One hypothesis is that all else being equal, ACOs with sicker populations generally have more opportunity to reduce costs than ACOs with healthier populations.

Figure 6 shows average ACO results after grouping ACOs into buckets based on geographic-risk-adjusted per capita costs in BY1 and CMS-HCC risk scores for aged/non-duals in BY3. Note that average gross savings generally increase as the CMS-HCC risk score increase (left-to-right) and as the geographic-risk-adjusted per capita costs in BY1 increase (top-to-bottom).

FIGURE 6: AVERAGE GROSS SAVINGS BY GEOGRAPHIC-RISK-ADJUSTED PER CAPITA COSTS IN BY1 AND CMS-HCC RISK SCORES FOR AGED/NON-DUAL BENEFICIARIES IN BY3, 2015 MSSP ACOS IN PERFORMANCE YEAR 3

GEOGRAPHICAL-RISK-ADJUSTED PER CAPITA COSTS IN BY1	PERCENTILE OF CMS-HCC RISK SCORE, AGED/NON-DUAL, BY3				
	0-20%	20-40%	40-60%	60-80%	80-100%
BOTTOM HALF	-2.3%	-1.0%	-3.0%	2.4%	1.4%
TOP HALF	-0.9%	1.5%	3.3%	4.6%	6.6%

The final ACO characteristic that was determined to be strongly predictive of gross savings was per capita (unadjusted) inpatient costs during the performance year. Inpatient admissions generally account for around 30% of costs for Medicare ACOs,⁸ so it is not surprising that, all else being equal, ACOs with lower inpatient costs in the performance year had higher gross savings. When ACOs were ranked based on inpatient costs in the performance year, ACOs in the lowest quintile had average gross savings of 4.3% and ACOs in the highest quintile had average gross savings of 0.3%.⁹ Unlike most other characteristics in our analysis, inpatient costs are something that ACOs can control during the performance year. ACO management efforts are therefore reflected to some extent in this characteristic. Service-category metrics from the baseline period were not available in the MSSP Public Use File (PUF) and were not included in our analysis.

8 Based on analysis of Annual Aggregate Expenditure/Utilization Report, 2015.

9 We also found that quality metric 36, risk-standardized acute admission rate for patients with diabetes, was mildly predictive of gross savings (lower readmission rates were associated with higher savings).

Other notable ACO features

By including more than 190 ACO features in our analysis, we were also able to find notable features that were not strongly associated with gross savings (after accounting for other features). Some of these features had modest association with gross savings, but their predictive power was much lower than the features described earlier in this paper.

PHYSICIAN-LED ACOS

One of the most interesting features was whether the ACO was physician-led. The random forest algorithm indicated that this feature had relatively low predictive power,¹⁰ but we did find that in general physician-led ACOs tended to have higher gross savings than other ACOs. Overall, physician-led ACOs had average gross savings of 3.4%, compared with 0.2% for ACOs that were not physician-led. However, physician-led ACOs tended to have higher baseline costs, higher national-to-local trend differences, and higher risk scores than other ACOs. Based on these metrics, we would expect physician-led ACOs to have gross savings of approximately 2.1%. This would indicate that physician-led ACOs do tend to have better results, even after accounting for other key features.

This effect is somewhat analogous to the impact of gender in predicting healthcare costs. In aggregate, there are clear differences in expected costs for men and women, but simply knowing a person's gender does not significantly improve predictive accuracy. Other factors, including age and clinical conditions, are much more predictive than gender. In the same way, while physician-led ACOs tended to generate higher savings, this characteristic alone does not significantly improve predictive accuracy.

OTHER FEATURES

Many of the other features in the model had little to no association with gross savings. These remaining features can be summarized into the following six groups:

1. Number of assigned beneficiaries
2. Quality score and most quality metrics
3. Mix of beneficiaries by entitlement category
4. Mix of other demographic characteristics (gender, age, race)
5. Mix of ACO provider types, including hospitals, primary care physicians (PCPs), Federally Qualified Health Centers (FQHCs), etc.
6. Utilization metrics other than inpatient admissions during performance year

10 Some studies (including the one at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1796903/>) have found that measures of feature importance in random forests tend to be biased toward categorical variables with many choices (such as CMS region). Because there were only three choices for the physician-led feature (Yes, No, or Unknown), the random forest may be understating its importance.

Recent policy changes and implications on ACO decision-making

In June 2016, CMS made several key updates to the MSSP rules for ACOs renewing their CMS agreements in January 2017 and beyond.¹¹ Among them was an adjustment to the ACO's historical rebased benchmark based on its performance versus a regional average benchmark. Additionally, CMS moved from a national absolute dollar trend to a regional percentage trend when updating the benchmark. Again, this is for renewing ACOs (Agreement Period 2 and beyond starting in January 2017). Based on the results of our analysis, we believe that these changes will have a material impact on an ACO's results as compared with the original Agreement Period 1 rules. The two rule changes described above impact the two most influential ACO characteristics from our analysis.¹² High historical costs and low local market trends will likely not benefit an ACO in the second agreement period.

Despite what may be two positive changes for efficient ACOs, CMS still has the challenge of setting a benchmark methodology that will work for all ACOs. One challenge for policy makers is how to provide adequate, ongoing incentives for efficient ACOs located in lower-cost regions. If an efficient ACO is part of a region with lower average per capita costs, the regional benchmark adjustment may not provide much relief from the challenges of trying to outperform its own (relatively favorable) historical experience. In addition to lower average per capita costs, an efficient region may have a lower trend rate. This too could cause ACOs in its region to struggle to hit their benchmarks, even if they are performing efficiently.

The results of our analysis will help ACOs when trying to evaluate and understand past performance. While ACO population health efforts are likely contributing to successes, our findings demonstrate that a number of factors outside of ACO clinical performance were predictive of gross savings and losses. At the same time, our findings do not mean that a loss position is entirely due to unfavorable uncontrollable circumstances. While we demonstrated that some factors outside of an ACO's control contributed to poor financial performance, the factors we analyzed only explained some of the results.

The findings of our analysis, in conjunction with policy changes related to second and subsequent agreement periods, also suggest that an ACO's recent results may not be predictive of future results. As noted above, some of the factors that were associated with positive gross savings in the first agreement period will likely not benefit an ACO in the second agreement period, which is due to changes in the benchmarking methodology. Conversely, some ACOs that struggled initially may have more favorable results in the second agreement period.

11 42 CFR Part 425, CMS-1644-F, viewed at <https://www.gpo.gov/fdsys/pkg/FR-2016-06-10/pdf/2016-13651.pdf>.

12 Geographic-risk-adjusted per capita cost, BY1 and national-to-local trend difference.

Data sources and methodology

This analysis was based primarily on data from the 2015 MSSP Public Use File (PUF). We incorporated quality metric information from the 2015 MSSP ACO Performance Results, also made publicly available by CMS. We excluded any variables that were directly related to the performance year gross savings calculation, such as shared savings amounts and performance year costs. Based on these two sources alone, we engineered more than 50 additional features. They included, but were not limited to, risk-adjusted costs, percentage of the ACO population by entitlement category, change in entitlement category mix from baseline to performance year, number of PCPs per capita, number of specialists per capita, and CMS region (based on primary state).

We also added additional features using analysis of outside data sources:

- For the geographic-risk-adjusted costs, we also developed ACO-specific geographic reimbursement factors. They were developed using area factors from the Milliman Health Cost Guidelines™, weighted based on each ACO's mix of assigned beneficiaries by metropolitan statistical area (MSA). Each ACO's mix of beneficiaries by MSA was estimated using county-level assignment information from the 2015 Number of ACO Assigned Beneficiaries by County PUF. The area factors used in this analysis reflected only differences in cost per service, not utilization. Separate reimbursement factors were developed for each baseline year.
- Local FFS trends for each ACO were estimated using the publicly available CMS FFS Data 2015 in conjunction with the Number of ACO Assigned Beneficiaries by County PUF.
- Physician-led ACOs were identified using information from the Leavitt Partners ACO Database.

We used these features to predict gross savings percentage with a machine learning algorithm known as a random forest. A random forest averages the predictions from a large number of decision tree models developed from bootstrapped samples of the data. In our case, we used 10,000 decision trees.

For purposes of this analysis, the biggest advantage of the random forest algorithm is that it is able to handle a large number of features, including features that are highly collinear (for instance, BY1 costs and BY2 costs). Although it does not produce coefficients or p-values, as we typically see in linear models, the random forest algorithm provides a useful measure of feature importance, which we utilized in this paper. The feature importance roughly measures how each feature contributed to the predicted accuracy.

Limitations and qualifications

The information in this paper is intended to identify and rank ACO characteristics most associated with gross savings under the MSSP. It may not be appropriate, and should not be used, for other purposes.

In performing the analysis for this paper, we relied on data made available by CMS and Leavitt Partners. We have not audited or verified this data and other information. If the underlying data or information is inaccurate or incomplete, the results of our analysis may likewise be inaccurate or incomplete.

We performed a limited review of the data used directly in our analysis for reasonableness and consistency and have not found material defects in the data. If there are material defects in the data, it is possible that they would be uncovered by a detailed, systematic review and comparison of the data to search for data values that are questionable or for relationships that are materially inconsistent.

Differences between our projections and actual amounts depend on the extent to which future experience conforms to the assumptions made for this analysis. It is certain that actual experience will not conform exactly to the assumptions used in this analysis. Actual amounts will differ from projected amounts to the extent that actual experience deviates from expected experience.

Guidelines issued by the American Academy of Actuaries require actuaries to include their professional qualifications in all actuarial communications. Jill S. Herbold, Anders Larson, and Cory Gusland are members of the American Academy of Actuaries and meet the qualification standards for performing the analyses presented in this report.

FOR MORE ON MILLIMAN'S PERSPECTIVE:

Visit our reform library at milliman.com/hcr

Visit milliman.com/medicare-insight

Visit our blog at healthcaretownhall.com

Follow us at twitter.com/millimanhealth



Milliman is among the world's largest providers of actuarial and related products and services. The firm has consulting practices in life insurance and financial services, property & casualty insurance, healthcare, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.

milliman.com

CONTACT

Jill S. Herbold
jill.herbold@milliman.com

Anders Larson
anders.larson@milliman.com

Cory Gusland
cory.gusland@milliman.com