Efficiency-Based Comparisons of One-Sided and Two-Sided Medicare Accountable Care Organizations (ACOs) and Their Potential Cost Savings

Abstract

Medicare ACOs represent the nation's largest initiative of Medicare alternative payment models toward value and health outcomes. Various ACO models have been experimented at differential risk structures, and the CMS has issued a final rule to accelerate the ACOs to assume greater downside financial risks. In response, this research conducts a comprehensive efficiency analysis of Medicare ACOs incorporating quality measures, investigates whether superiority exists among the various ACO models and determines their potential cost reductions. The results indicate that in minimizing expenditures given quality services, or maximizing quality services given health expenditures, one-sided ACOs are more efficient than two-sided ACOs, so it might not be advisable to mandate the transition of ACOs from one-sided to two-sided, as far as efficiency is concerned. This research also shows that the ACOs should be able to reduce expenditures significantly through efficiency improvement. Maintaining the same level of enrollment, utilization, and quality, without switching to two-sided ACO tracks, Track 1 ACOs are expected to reduce expenditures by 4.1% using the median efficiency target, and 1.5% using the 25th percentile efficiency target (compared to actual expenditures). Another finding is that the benchmark expenditures for one in four Medicare ACOs are below the efficient expenditures using the median efficiency target, and one in three using the 25th percentile efficiency target. The benchmark expenditures are probably too low for these ACOs, and should be adjusted upward.

Keywords: Medicare cost savings, accountable care organizations (ACOs), health expenditures, quality of care, economic efficiency

JEL classification: I13, I11, H51

Efficiency-Based Comparisons of One-Sided and Two-Sided Medicare Accountable Care Organizations (ACOs) and Their Potential Cost Savings

1. Introduction

The Affordable Care Act established the Medicare Accountable Care Organizations (ACOs) which represent the nation's largest initiative of Medicare alternative payment models toward value and health outcomes (Mechanic and Gaus, 2018). The majority of Medicare ACOs are in the Medicare Shared Savings Program (MSSP) (561 of all 649 Medicare ACOs in 2018) (Verma, 2018). According to the Centers for Medicare and Medicaid Services (CMS), the MSSP ACOs generated total savings of $1.1 billion, with $314 million net savings after accounting for shared savings earned by ACOs (Castellucci and Dickson, 2018). Multiple tracks with various financial risks currently exist among MSSP ACOs, most of which are upside-only sharing savings but not losses (460 of the 561 MSSP ACOs in 2018) (Verma, 2018). In contrast, the ongoing two-sided MSSP ACOs and Next Generation ACOs share both savings and losses. In December 2018, the CMS issued a final rule, referred to as "Pathways to Success", to advance accountability, competition, engagement, integrity, and quality for MSSP ACOs, and accelerate the path for MSSP ACOs to assume greater downside financial risks (CMS, 2018a). This research conducts a comprehensive evaluation of the performance of one-sided and two-sided Medicare ACOs incorporating health expenditures, utilization of medical services, and quality of health care, and aims to provide insights on the design of Medicare ACOs to increase quality for patients and drive towards cost savings.

Medicare ACOs are groups of doctors, hospitals, and other health care providers, who come together voluntarily to cooperate and share accountability in delivering better coordinated, higher quality, and more efficient care to Medicare beneficiaries (www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ACO/) . "Medicare ACOs are designed to provide financial incentives for fee-for-service (FFS) Medicare providers to reduce inefficiencies in care delivery for a population of beneficiaries under their care. ACOs are grounded in the theory that with the opportunity to share in financial rewards (or face penalties), ACOs will reduce fragmentation and duplication in medical care by facilitating improved communication and coordination across providers and between patients and their doctors, thereby improving quality and reducing spending" (L&M Policy Research, 2016). The CMS has created various ACO models, including MSSP ACOs, Pioneer ACOs, Next Generation ACOs, and the Comprehensive End Stage Renal Disease Care Model (CEC). The CEC Model was designed to identify, test, and evaluate new ways to improve care for Medicare beneficiaries with End-Stage Renal Disease (ESRD). The MSSP offers different participation options (tracks) that allow ACOs to assume various levels of risk/reward. The CMS selected a small subset of the MSSP ACOs to test the two pre-paid shared savings models: the Advanced Payment ACO model and the ACO Investment Model. The Advance Payment ACO Model was intended to help smaller physician-based organizations with less access to capital transition to MSSP ACOs. The ACO Investment Model builds on the experience with the Advance Payment Model, and tests the use of pre-paid shared savings to encourage new ACOs to form in rural and underserved areas and to encourage existing MSSP ACOs to progress to arrangements with greater financial risks. Consistent with, but separate from, MSSP ACOs, the Pioneer ACO Model and the Next Generation ACO Model were designed for health care organizations and providers that were already experienced in coordinating care for patients across care settings. The two models allow these provider groups to assume higher levels of financial risk and reward than are available under the Shared Savings Program. More details on the new payment and service delivery models are available from the CMS Innovation Center at https://innovation.cms.gov/initiatives/#views=models.

Evaluation analyses have been conducted to examine whether Medicare ACOs generate better results for medical costs and the quality of health care. L&M Policy Research (2015) evaluates the effects of the Pioneer ACOs on Medicare spending, utilization, and quality in their first two performance years (2012 and 2013). The results show that Pioneer ACOs saved a total of $384 million over the two years, $279.7 million in 2012 (a decrease of $35.62 per beneficiary per month) and $104.5 million in 2013 (a decrease of $11.18 per beneficiary per month). In May 2015, the CMS Office of the Actuary certified that expansion of the Pioneer ACO model as a permanent part of the Medicare program would reduce net program spending without any negative effects on quality of care (L&M Policy Research, 2016). NORC (2018) assesses the impact of the strong financial incentives of the Next Generation ACOs on Medicare spending, health services utilization, and quality of care in their first performance year (2016). The findings show that the 18 ACOs reduced Medicare spending by approximately $100 million, corresponding to a decrease of $18.2 per beneficiary per month, or 1.7% of the estimated average Medicare spending, absent the model. The net reduction in Medicare spending after adjusting for shared savings/losses payments totaled $62 million, a decrease of $11.2 per beneficiary per month, or 1.1%. Nonetheless, the reductions varied across ACOs. Only four of the 18 ACOs showed decreases that reached statistical significance (at the significance level of 0.10) and accounted for more than half of the total savings. Regarding utilization, the results indicate that there was a significant reduction in inpatient hospital days and evaluation and management visits. However, there was a significant increase in annual wellness visits (while no significant changes in other quality measures), indicating improved quality of care (NORC, 2018).

However, the CMS has not commissioned a formal evaluation of the MSSP ACOs, the majority of Medicare ACOs (Mechanic and Gaus, 2018). Furthermore the one-sided and two-sided ACO models have not been rigorously compared. This motivates the current research to investigate whether superiority exists among the various Medicare ACO models.

Most other discussions and analyses on ACOs focus on individual measures separately, such as health expenditures, inpatient hospital days, or unplanned readmissions. The individual measure is surely useful but only represents one performance aspect of ACOs. There is not even a financial ratio readily available for the evaluation of ACOs, such as the medical loss ratio for private health insurers, Medicare Advantage, and Medicaid managed care. Yang (2018) composes a comprehensive set of performance measures in evaluating Medicaid managed care, including individual measures, financial ratios, and efficiency measures. The efficiency measure combines more than two variables and provides a more complete evaluation of the insurer. Yang and Lin (2017) indicate that even financial ratios are not effective indicators of the efficiency of health insurers and suggest efficiency measures be incorporated to satisfy all stakeholders. Brockett, Golden and Yang (2018) apply Data Envelopment Analysis (DEA) to assess the efficiency and potential savings of Medicare, Medicare Advantage, and ACOs. The efficiency model of the ACOs integrates health expenditures and the utilization measures (inpatient admissions and ambulatory encounters) to generate an efficiency score for each ACO which is utilized for efficiency comparison and improvement. This current research incorporates quality measures into the efficiency analysis of Medicare ACOs, as has not been examined in any of the previous efficiency studies.

“One of the assumptions of the CMS’ MSSP final rule to accelerate risk is that MSSP one-sided ACOs have not saved money for Medicare, while other ACO models that require downside financial risk have generated savings" (Mechanic and Gaus, 2018). Verma (2018) indicates that, as of 2016, one-sided ACOs had increased Medicare spending. In 2016, the one-sided ACOs nominally increased Medicare spending by $49 million, while two-sided ACOs decreased Medicare spending by $33 million. However, in 2017, the MSSP ACOs generated a total of $1.1 billion in savings, and $313.7 million net savings (CMS, 2018). Additionally, Mechanic and Gaus (2018) argue that the CMS' benchmark expenditures produced underestimated savings of ACOs. McWilliams (2016) uses the difference-in-differences (DiD) regression models to compare changes in Medicare spending for ACO-attributed beneficiaries with concurrent changes for beneficiaries attributed to non-ACO providers, and records $867 total savings and $213 net savings for the MSSP ACOs during 2013 and 2014. McWillaims et al. (2018) find that the aggregate reduction in fee-for-service spending accrued by MSSP ACOs in 2015 was $583.4 million, 39% greater than the corresponding reduction reported by CMS ($419.3 million), and the net savings from these ACOs were nearly 2.8 times as great ($256.4 vs. $92.3 million). Similarly, Dobson et al. (2018) use the DiD statistical approach and document $541.7 million net MSSP savings in 2013-2015, in contrast to $344.2 million net losses reported by the CMS. During 2013-2015, the MSSP ACOs were almost exclusively one-sided (Track 1), and the research results provide evidence that one-sided ACOs should have the potential to achieve significant cost savings. Clifton Gaus, president and CEO of the National Association of ACOs, states that "ACOs are saving Medicare hundreds of millions of dollars, and given sufficient time, one-sided ACOs will return significant savings to the trust funds" (Castellucci and Dickson, 2018). In response, this research investigates the potential cost savings of Medicare ACOs. In the literature, Brockett, Golden and Yang (2018) examine the potential savings of ACOs and Medicare Advantage. This current research contributes to the literature by evaluating the cost reductions of the Medicare ACO models with different risk structures, using various efficiency models.

Specifically, this research utilizes two DEA models to analyze the performance of Medicare ACOs in two years (2016 and 2017). For efficiency comparison, firstly the cross-frontier DEA model is adopted to generate the efficiency scores of the ACOs. The Mann-Whitney test is then used to compare each pair of the Medicare ACO models. The collective frontier DEA model is selected to identify the "best practices" efficient frontier and determine the efficient health expenditures of the less efficient ACOs. The potential savings are obtained by comparing the efficient expenditures with benchmark and actual expenditures. Health expenditures are chosen as the input for all the efficiency analyses, with three different sets of outputs. The first set of outputs consists of enrollment (assigned beneficiaries or beneficiary-years), health services utilization, and quality measures, the second set enrollment and quality measures, and the third set enrollment and health services utilization. The results show that the MSSP one-sided ACOs are more efficient than two-sided models, as suggests that the Medicare ACOs should not be mandated to take on downside financial risks as far as efficiency is concerned. This research also documents significant cost savings by promoting efficient practices without switching to two-sided tracks.

This article proceeds as follows. An overview of Medicare ACOs and quality measures is presented in the next section. Section 3 discusses Medicare ACOs data, research desgin and DEA models. Section 4 presents descriptive statistics and univariate analyses of assigned beneficiaries, health expenditures, health services utilization, and quality measures. Sections 5 and 6 conduct the efficiency comparisons between the MSSP one-sided and two-sided tracks, and between MSSP ACOs and Next Generation ACOs. Section 7 investigates potential cost savings through efficiency improvement and discusses the setting of benchmark expenditures, and Section 8 concludes.

2. Medicare ACOs and quality measures

The Medicare ACO program was launched in 2012 and keeps proliferating. As of 2018, Medicare ACOs serve approximately 12.3 million beneficiaries, representing over 20 percent of Medicare beneficiaries (NAACOS, 2018). The CEC, Pioneer ACOs and Next Generation ACOs are all two-sided models which share both savings and losses (the CMS Innovation Center: www.innovation.cms.gov/initiatives/#views=models). The Pioneer ACOs started in 2012 and concluded in the end of 2016, while the ongoing CEC and the Next Generation Model were initiated in 2015 and 2016 respectively. The MSSP offers both one-sided and two-sided tracks, with the one-sided tracks being upside-only and sharing only savings but not losses. Tracks 1&2 are available since the beginning in 2012, and Track 3 and Track 1+ were established in 2016 and 2018 respectively (KFF, 2018). For agreement periods beginning on July 1, 2019, ACOs may participate in the Shared Savings Program under one of two tracks: the BASIC track (which includes a glide path for eligible ACOs), or the ENHANCED track, which offers the highest level of risk and potential reward. ACOs participating in the BASIC track's glide path may begin under a one-sided model and progress through incremental levels of increasing risk and potential reward (CMS, 2018a). The BASIC track comprises five levels: Levels A and B correspond to Track 1, and Levels C, D, and E are similar to Track 1+. The ENHANCED track is the same as Track 3 and Track 2 will not be available anymore. Among all the tracks/levels, Track 1 (and the BASIC A and B levels) are one-sided, while the others are all two-sided. The number of ACOs in each track/model (as of 2018) and the risk structures of the ACO models are presented in Tables 1 and 2.

Table 1. Medicare ACO annual counts in each ACO model and track

ACO Model		2012	2013	2014	2015	2016	2017	2018
MSSP ACOs	Track 1	110	217	330	389	410	433	460
	Track 1+							55
	Track 2	4	5	3	3	6	6	8
	Track 3					16	33	38
	Advanced Payment ACOs*	20	36	35	33			
	ACO Investment Model*				4	45	45	45
Pioneer ACOs		32	23	20	19	9		
Next Generation ACOs						18	45	51
CEC						13	37	37
All Medicare ACOs		146	245	353	411	472	554	649

Sources: NAACOS (2018), the CMS Innovation Center (www.innovation.cms.gov/initiatives/#views=models).
*The ACOs participating in the Advanced Payment Model and the Investment Model are included in their respective tracks.

Table 2. Risk structures of Medicare ACO models

ACO model			Status	Start year	Risk structure	Savings share rate	Loss share rate
Pioneer ACOs			Ended (2016)	2012	Two-sided	Up to 60-75%	Up to 60-75%
Next Generation ACOs			Active (ending in 2020)	2016	Two-sided	Up to 80% or 100%	Up to 80% or 100%
Medicare Shared Savings Program (MSSP)	Track 1		Active*	2012	One-sided	Up to 50%	Not applicable
	Track 1+		Active*	2018	Two-sided	Up to 50%	30%
	Track 2		Active*	2012	Two-sided	Up to 60%	40-60%
	Track 3		Active*	2016	Two-sided	Up to 75%	40-75%
	Basic track	Level A& B	Not yet active	2019	One-sided	Up to 40%	Not applicable
		level C	Not yet active	2019	Two-sided	Up to 50%	Up to 30%
		Level D	Not yet active	2019	Two-sided	Up to 50%	Up to 30%
		Level E	Not yet active	2019	Two-sided	Up to 50%	Up to 30%
	Enhanced track		Not yet active	2019	Two-sided	Up to 75%	40-75%

Sources: KFF (2018), CMS (2017a), CMS (2018a), NAACOS (2018).
*Tracks 1, 1+, 2, and 3 are phasing out starting in 2019.

Population-based payments are the underlying payment model applied to Medicare ACOs (Champagne and McEwen, 2018). Firstly, the CMS assigns a group of Medicare fee-for-service beneficiaries to the ACO. After CMS completes ACO beneficiary assignment, it establishes the ACO's benchmark expenditures for each performance year. At the end of each year, the CMS compares the benchmark expenditures to the ACO's actual expenditures during the performance year to see whether the ACO may share in savings or losses. To qualify for shared savings, an ACO must meet or exceed a prescribed Minimum Savings Rate (MSR) and fulfill the minimum quality performance standards. ACOs that meet these requirements may share in savings at a rate determined by their quality performance up to a performance payment limit. To be liable for shared losses, an ACO must meet or exceed a prescribed Minimum Loss Rate (MLR). Once this MLR is met or exceeded, the ACO will share in losses at a rate determined by its quality performance up to a loss recoupment limit (also referred to as a loss sharing limit). If the ACO does not meet the quality performance standard for the performance year, it will not be eligible for any shared savings and will be accountable for shared losses based on the highest shared loss rate. The details of the shared savings and losses and assignment methodology are provided in CMS (2017b).

The Medicare ACOs are required to completely and accurately report quality data that are used to calculate and assess their quality performance (CMS, 2016). In the first performance year of their first agreement period, ACOs satisfy the quality performance standard when they completely and accurately report on all quality measures (pay-for-reporting). Complete and accurate reporting in the ACO's first performance year qualifies the ACO for the maximum quality score and sharing rate. In subsequent performance years, quality measures are phased in to pay-for-performance and national performance benchmarks are used to calculate the ACO's quality score and final sharing rate. CMS measures quality of care using a set of nationally recognized quality measures in four key domains: patient/caregiver experience, care coordination/patient safety, preventive health, and clinical care for at-risk population. There are 31 quality measures in 2018, 31 in 2017, 34 in 2016, 33 in 2015, 33 in 2014, and 33 in 2013. The 31 quality measures of 2018 are provided in Appendix I (CMS, 2018b). The two diabetes measures (ACO-27 and ACO-41) are scored together as a composite measure (Diabetes Composite). The quality measures for the Next Generation ACOs are the same as the quality measure set of the MSSP ACOs, except that ACO-11 is exempted for the Next Generation ACOs (CMS, 2015).

For most of the measures, the higher the rate of the quality measure, the higher the level of performance. However, a lower rate is indicative of better performance on the following seven measures (CMS, 2016):

ACO 8: Risk Standardized, all condition readmissions,

ACO 27: Diabetes Mellitus: Hemoglobin A1c poor control,

ACO 35: Skilled Nursing Facility 30-Day All-Cause Readmission Measure (SNFRM),

ACO 36: All-Cause Unplanned Admissions for Patients with Diabetes,

ACO 37: All-Cause Unplanned Admissions for Patients with Heart Failure,

ACO 38: All-Cause Unplanned Admissions for Patients with Multiple Chronic Conditions,

ACO 43: Ambulatory Sensitive Condition Acute Composite (AHRQ PQI #91).

Currently, four quality measures are pay-for-reporting in all years (CMS, 2016):

ACO 7: CAHPS: Health Status/Functional Status,

ACO 40: Depression Remission at Twelve Months, ACO 42: Statin Therapy for the Prevention and Treatment of Cardiovascular Disease,

ACO 44: Use of Imaging Studies for Low Back Pain.

3. Data and research design

The CMS publishes the ACO-level public-use file (PUF) that contains ACO-specific metrics for some performance years of Medicare ACOs (MSSP: www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SSPACO/index.html, Pioneer and Next Generation ACOs: www.innovation.cms.gov/initiatives/#views=models). Currently the data are available for MSSP ACOs from 2013 to 2017, Next Generation ACOs 2016 and 2017. This research is designed to compare the one-sided and two-sided ACOs incorporating the quality measures. However, the quality measures are not published for the three years of 2013-2015 for the MSSP ACOs, and there were very few two-sided ACOs during this period of time (5 in 2013, 3 in 2014, and 3 in 2015). Therefore, only the data of 2016 and 2017 are used for the analyses of this research. The Pioneer ACO model had ended in 2016 with only nine ACOs so they are not included in this research, nor the CEC ACOs which serve a special population of Medicare beneficiaries with ESRD. Originally, the CMS PUF contains 904 MSSP and 62 Next Generation ACO-years (ACOs hereafter) in 2016 and 2017. After excluding the ACOs with incomplete data and the outliers with extreme values, the final sample consists of 60 Next Generation ACOs and 873 MSSP ACOs (Table 3).

Table 3. The sample of Medicare ACOs

ACO model		2016	2017	Total
MSSP ACOs	Track 1	401	412	813
	Track 2	6	6	12
	Track 3	16	32	48
Next Generation ACOs		17	43	60
Total		440	493	933

There are 34 quality measures in 2016 and 31 in 2017 for MSSP ACOs (CMS, 2016). Six of the 2016 quality measures (ACO-9, ACO-10, ACO-39, ACO-21, ACO-31, and ACO-33) were retired and three new quality measures (ACO-12 and ACO-43, and ACO-44) were introduced in 2017, which are excluded from the analysis because they are not applicable to both 2016 and 2017.Furthermore, three of the 28 remaining quality measures (ACO-7, ACO-40, and ACO-42) are pay for reporting (not for performance) for all the years. Therefore, these three quality measures are also excluded. The 25 quality measures of this research are presented in Table 4 (the Diabetes Composite measure includes two individual component measures). ACO-11 is exempted for Next Generation ACOs.

Table 4. 25 quality measures of this research (applicable to both 2016 and 2017, and not pay-for-reporting for all the years)

Domain	ACO measure	Measure title
Patient/ caregiver experience	ACO-1	CAHPS: Getting Timely care, Appointments, and Information
	ACO-2	CAHPS: How Well Your Providers Communicate
	ACO-3	CAHPS: Patients' Rating of Provider
	ACO-4	CAHPS: Access to Specialists
	ACO-5	CAHPS: Health Promotion and Education
	ACO-6	CAHPS: Shared Decision Making
	ACO-34	CAHPS: Stewardship of Patient Resources
Care coordination/ patient safety	ACO-8	Risk-Standardized, All Condition Readmission
	ACO-35	Skilled Nursing Facility 30-Day Readmission (SNFRM)
	ACO-36	All-Cause Unplanned Admissions for Patients with Diabetes
	ACO-37	All-Cause Unplanned Admissions for Patients with Heart Failure
	ACO-38	All-Cause Unplanned Admissions for Patients with Multiple Chronic Conditions
	ACO-11*	Use of Certified EHR Technology
	ACO-13 (CARE-2)	Falls: Screening for Future Fall Risk
Preventive health	ACO-14 (PREV-7)	Preventive Care and Screening: Influenza Immunization
	ACO-15 (PREV-8)	Pneumonia Vaccination Status for Older Adults
	ACO-16 (PREV-9)	Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up
	ACO-17 (PREV-10)	Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention
	ACO-18 (PREV-12)	Preventive Care and Screening: Screening for Clinical Depression and Follow- up Plan
	ACO-19 (PREV-6)	Colorectal Cancer Screening
	ACO-20 (PREV-5)	Breast Cancer Screening
At-risk population	Diabetes Composite**	ACO-27: Diabetes Mellitus: Hemoglobin A1c Poor Control
		ACO-41: Diabetes: Eye Exam
	ACO-28 (HTN-2)	Hypertension (HTN): Controlling High Blood Pressure
	ACO-30 (IVD-2)	Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet

Source: CMS (2016).
*ACO-11 is exempted for Next Generation ACOs.
**The Diabetes Composite measure includes two individual component measures.

All the quality measures in three of the four quality domains (patient/caregiver experience, preventive health, and at-risk population) are reported on the 0-100 scale, as well as the two quality measures (ACO-11 and ACO-13) in the other quality domain (care coordination/patient safety) (CMS, 2017c). For the other five quality measures (in the domain of care coordination/patient safety), ACO-8 and ACO-35 represent the predicted readmission rate divided by the expected readmission rate and then multiplied by an average readmission rate; while ACO-36, ACO-37, and ACO-38 represent the predicted acute admission rate divided by the expected acute admission rate and then multiplied by an average acute admission rate. Therefore, it is possible that the scores of these measures may be more than 100%. Also, the five measures are all inverted rates, that is, a higher rate is indicative of worse performance. For the consistency among all the quality measures, the scores of the five measures are converted to the 0-100 scale and then subtracted from 100. As a result, a higher value of the transformed rate is now indicative of better performance, while it is still on the 0-100 scale. In this research, the quality measures are included in the outputs of the efficiency analysis. The re-inversion of the inverted rate is required for the efficiency analysis because a higher value of the output is supposed to be indicative of higher efficiency given the input level.

The DEA models are adopted to generate the efficiency scores of ACOs, which are compared among the three tracks (Track 1, Track 2, and Track 3) and Next Generation ACOs to show the superiority of any specific ACO model. Data Envelopment Analysis (DEA) is a mathematical programming frontier approach to estimating the relative efficiency of a homogeneous set of peer entities called Decision Making Units (DMUs). DEA is a multi-input, multi-output efficiency measurement technique that generalizes the classical single input, single output approach (Charnes, Cooper, and Rhodes, 1978). It is non-parametric, which requires no distributional assumptions and avoids the need to specify a functional form between inputs and outputs. The relative efficiency of a DMU is measured by comparing this DMU to "best practice" efficient frontiers formed by the most efficient DMUs. The "collective frontier" DEA model pools all the DMUs (from different groups) together to perform a joint DEA analysis, which means that each unit is compared against members of its own group in addition to members of other groups. Consequently, a characterization of "inefficient" may result from "within group" effects instead of the "between group" effects (Brockett et al., 2004). Nonetheless, the "cross-frontier" DEA model overcomes this shortcoming by comparing the test unit exclusively against members outside its own group, which is also referred to as the Variable-Benchmark DEA model (Zhu, 2009) or the Game-Theoretic DEA model (Brockett et al, 2004). In the current research, appropriately, the cross-frontier DEA model is utilized to generate the efficiency scores for the purpose of comparing any track of ACOs exclusively against other track(s), for example, Track 1 versus Track 2 or Track 3. Afterwards, the efficiency scores of the ACOs in both groups are pooled together, ranked, and then the Mann-Whitney test is conducted to detect the efficiency difference between the two groups. The rank statistical method is outlined in Brockett and Golany (1994).

The value-based Medicare ACOs are an innovative component of the national quality strategy to fulfill the triple aims of health care: better care, better health and lower costs. The efficiency of ACOs can be evaluated from the perspectives of beneficiaries, the ACO itself, or the society. To be consistent with the three quality aims of the health care reform, this current research adopts the societal perspective to analyze the efficiency of ACOs (Brockett et al., 2004). From a societal perspective, the policymakers care about providing necessary quality health services at reasonable expenses and expanding the coverage for more beneficiaries. Obviously, the ACO's health expenditures, the costs incurred by the ACO and beneficiaries, should be the inputs. In the literature, the outputs of some health insurer efficiency studies comprise enrollment and utilization measures (Brockett et al., 2004; Yang, 2014, 2018; Yang and Wen, 2017; Yang and Lin, 2018; and Brockett, Golden and Yang, 2018). Following Brockett, Golden and Yang (2018), inpatient admissions and ambulatory encounters are included in the outputs of the DEA model, as well as enrollment which is selected as an output because ACOs are intended to minimize medical expenditures given a certain amount of covered beneficiaries.

In evaluating the efficiency of ACOs (a group of medical professionals) from the societal perspective, one emphasis should be the value of health care services: the quality of care and health outcomes of beneficiaries (Brockett, Golden and Yang, 2018). Therefore, the outputs of the ACO efficiency analysis should also contain "better care and better health", which are represented by the CMS quality measures. Specifically, four quality domain composite measures are incorporated: patient/caregiver experience composite, care coordination/patient safety composite, preventive health composite, and at-risk population composite. The at-risk population composite measure is the average of the Diabetes Composite and the two individual measures (ACO-28 and ACO-30). The other three domain composite measures are the average measure scores of all the individual quality measures of their respective quality domain. The quality domain composite measures of this current research are consistent with other health plan quality reporting. For example, the National Committee for Quality Assurance (NCQA) publishes health plan ratings at the composite, subcomposite and individual measure level (NCQA, 2018).

Some utilization measures might be deemed as intermediate outputs. Therefore, for the comparison purpose and a robustness check, this current research conducts the efficiency analysis using several different DEA models. DEA model 1 adopts all the three sets of outputs as discussed above: enrollment, utilization measures, and quality measures. DEA model 2 only uses enrollment and quality measures as outputs, and DEA model 3 enrollment and utilization measures.

The CMS PUF of Next Generation ACOs does not include the utilization data. One quality measure (ACO-11) is exempted for the Next Generation ACOs. Additionally, regarding enrollment, the PUF only reports total assigned beneficiaries, not total beneficiary-years. Therefore, another DEA model (DEA model 4) is adopted for the efficiency comparison between Next Generation ACOs and MSSP ACOs. The input of DEA model 4 is still health expenditures. The outputs consist of total assigned beneficiaries (instead of total beneficiary-years) and the four quality domain composite measures, except that the quality measure ACO-11 is not included in the care coordination/patient safety composite. The total beneficiary-years are the total assigned beneficiaries adjusted downwards for beneficiaries with less than a full 12 months of eligibility. The inputs and outputs of the four DEA models are presented in Table 5.

Table 5. Inputs and outputs of DEA models

Model	Input variable	Output variable
Model 1	Health expenditures	Enrollment (total beneficiary-years)
		Inpatient admissions
		Ambulatory encounters
		Patient/caregiver experience composite
		Care coordination/patient safety composite (a)
		Preventive health composite
		At-risk population composite
Model 2	Health expenditures	Enrollment (total beneficiary-years)
		Patient/caregiver experience composite
		Care coordination/patient safety composite (a)
		Preventive health composite
		At-risk population composite
Model 3	Health expenditures	Enrollment (total beneficiary-years)
		Inpatient admissions
		Ambulatory encounters
Model 4	Health expenditures	Enrollment (total assigned beneficiaries)
		Patient/caregiver experience composite
		Care coordination/patient safety composite (b)
		Preventive health composite
		At-risk population composite

*ACO-11 is included in care coordination/patient safety composite (a) but not Care coordination/patient safety composite (b).

The collective frontier DEA model is utilized to obtain the potential cost savings of ACOs. The DEA's "efficient frontier" is formed by the most efficient DMUs, and this efficient frontier can provide a benchmark for the inefficient DMUs to follow and improve performance. Specifically, for this research, "efficient frontiers" or efficient ACOs should provide other less efficient ACOs with "best efficient practices" in health care delivery and management to reduce costs and improve efficiency. In identifying the efficient frontier for an ACO, it does not have to be constrained to the ACOs of the other group alone. Any other ACOs, even in its own group, if more efficient, should be selected as its efficiency benchmark. Therefore, the collective frontier DEA model to pool together all the ACOs is appropriate for the purpose of efficiency improvement and cost reductions. Other than efficiency scores, the DEA model also generates efficient inputs and outputs for each ACO. In this research, the potential cost savings are derived by subtracting the efficient health expenditures of the ACO from its actual or benchmark health expenditures.

Two types of returns to scale are generally employed for the DEA model: constant returns-to-scale (CRS) and variable returns to scale (VRS). The CRS model generates "overall technical efficiency" which can be decomposed into two components: pure technical efficiency and scale efficiency (Cummins and Weiss, 2013). The "pure technical efficiency" is generated by using the VRS model (Banker, Charnes and Cooper, 1984). Similar to Brockett et al. (2004) and Yang (2014), CRS is selected to analyze the "overall" efficiency, not just one of its components. Furthermore, two alternative model orientations are available in DEA to determine the relative efficiency of DMUs: input-oriented where the inputs are minimized conditional on the level of outputs and output-oriented where the outputs are maximized conditional on input usage. This current research adopts the input-oriented approach to examine the efficiency of

Medicare ACOs in minimizing health expenditures given enrollment, utilization and quality (DEA model 1), enrollment and quality (DEA model 2 and 4), and enrollment and utilization (DEA model 3). However, with CRS, the efficiency scores are reciprocals to each other using the input-oriented and output-oriented model. Therefore, it doesn't affect the efficiency analysis whether the input-oriented or output-oriented model is used. In the current research, the DEA optimization problems are solved by using the DEA software developed by Joe Zhu (http://www.deafrontier.net/software.html). The interested reader is referred to Cooper, Seiford and Tone (2007) and Zhu (2009) for details and references.

4. Descriptive and univariate analyses of Medicare ACOs

This research conducts efficiency analysis to identify the advantageous design of Medicare ACOs. The sample of this research consists of 60 Next Generation ACOs and 873 MSSP ACOs, 813 of which are in Track 1, 12 in Track 2, and 48 in Track3. Four different DEA models are adopted to compare the efficiency of the Medicare ACOs and determine their potential cost savings. The efficiency analysis integrates various variables to deliver a holistic evaluation of the ACOs, but firstly, this section presents some informative descriptive and univariate analyses of the input and output variables of the DEA models separately: assigned beneficiaries and beneficiary-years, health expenditures, medical services utilization, and quality measures.

Total beneficiary-years and total assigned beneficiaries are one of the outputs. Total assigned beneficiaries are available for all the ACOs of the sample. Total beneficiary-years are available for MSSP ACOs, but they are not published for Next Generation ACOs. Some summary statistics of the assigned beneficiaries and beneficiary-years are presented in Table 6.

On average the assigned beneficiary-years are 97.3% (standard deviation 0.5%) of the assigned beneficiaries in the sample of the 873 MSSP ACOs. However, the assigned beneficiaries and beneficiary-years are perfectly correlated (the correlation coefficient is 1.0). Therefore, it is equivalent to select either the total beneficiaries or the total beneficiary-years for the DEA analysis. In the sample of this research, generally the two-sided ACOs of Tracks 2&3 and the Next Generation model are assigned more beneficiaries than Track 1 ACOs. It shows that the average total assigned beneficiary-years are 24,413 for Track 3 and 27,752 for Next Generation ACOs, 29.4% and 47.1% more than that of the one-sided Track 1 ACOs (18,871) respectively. Nonetheless, the two-sided Track 2 ACOs are 43.0% smaller than the one-sided Track 1 ACOs in assigned beneficiaries.

Table 6. Summary statistics of total assigned beneficiaries and beneficiary-years of Medicare ACOs*

ACO model	Number of ACOs	Total assigned beneficiaries		Total beneficiary-years		
		Mean	StDev	Mean	StDev	% of total assigned beneficiaries
Track 1	813	18,871	18,835	18,371	18,341	97.3%
Track 2	12	10,759	3,983	10,452	3,888	97.1%
Track 3	48	24,413	17,573	23,917	17,224	97.9%
Tracks 2&3	60	21,682	16,712	21,224	16,390	97.8%
Next Generation ACOs	60	27,725	18,757	n/a	n/a	n/a
Total	933	19,621	18,815	18,567	18,219	97.3%

* The total beneficiary-years are the total assigned beneficiaries adjusted downwards for beneficiaries with less than a full 12 months of eligibility.

Actual expenditures are the inputs of all the DEA analyses of this research, which minimize the expenditures given the outputs. In other words, higher expenditures result in lower efficiency given equivalent outputs. The adjusted and unadjusted actual expenditures per beneficiary and per beneficiary-year are presented in Table 7. The adjusted and unadjusted total actual expenditures are highly correlated (the correlation coefficient is 0.98). Therefore, there is no significant difference on the results of this research using the adjusted or unadjusted expenditures. On average, the actual expenditures per beneficiary or beneficiary-year of one-sided Track 1 ACOs is lower than that of two-sided ACOs. Specifically, the average adjusted actual expenditures per beneficiary-year is $11,810 for MSSP Tracks 2&3 ACOs, 2.4% higher than that of Track 1 ACOs. Similarly, the average adjusted actual expenditures per beneficiary is $11,167 for Next Generation ACOs, 5.1% higher than that of Track 1 ACOs. Among two-sided ACO models, the actual expenditures are not significantly different between Track 2 and Track 3 ($11,858 vs. $11,798 per beneficiary-year, and $11,505 vs. $11,553 per beneficiary. However, the average expenditures per beneficiary of Next Generation ACOs is 2.2% higher than that of Tracks 2&3 ($11,792 vs. 11,543).

Table 7. Summary statistics of adjusted and unadjusted actual expenditures of Medicare ACOs (per beneficiary and per beneficiary year)*

Actual expenditures ($)		Track 1	Track 2	Track 3	Tracks 2&3	Next Generation ACOs	Total
Unadjusted actual expenditures per beneficiary year ($)	Mean	10,676	11,066	11,088	11,083	n/a	10,704
	StDev	1,990	1,758	1,183	1,301	n/a	1,953
Adjusted actual expenditures per beneficiary year ($)	Mean	11,535	11,858	11,798	11,810	n/a	11,553
	StDev	2,356	2,692	1,696	1,909	n/a	2,328
Unadjusted actual expenditures per beneficiary ($)	Mean	10,384	10,741	10,857	10,834	11,167	10,464
	StDev	1,892	1,653	1,155	1,254	1,924	1,871
Adjusted actual expenditures per beneficiary ($)	Mean	11,218	11,505	11,553	11,543	11,792	11,276
	StDev	2,242	2,539	1,657	1,841	2,046	2,210

*The adjusted expenditures are in 2017 Texas dollars (Yang, 2014).

The key underlying rationale of the ACOs is the reduction of duplicated and wasteful medical care through better improved coordination. However, given health expenditures and enrollment, more necessary quality medical services reflect more benefits for beneficiaries and higher efficiency of the providers. The summary statistics of the two utilization measures (inpatient admissions and ambulatory encounters) are presented in Table 8. In the CMS PUF, the utilization measures are not published for Next Generation ACOs. On average, one-sided MSSP Track 1 ACOs incur lower utilization of medical services. Specifically, the average inpatient admissions per beneficiary-year of Track 1 is 0.32, 3.0% less than that of the two-sided Tracks 2&3 (0.33). The average ambulatory encounters per beneficiary-year of Track 1 is 11.08, 1.5% less than that of the two-sided Track 3 (11.25). However, the number of ambulatory encounters is similar for Track 1 and Track 2 (11.08 vs. 11.06). It is shown that two-sided ACOs also incur more health expenditures. The analysis of individual variables does not readily distinguish which ACO is more efficient when expenditures and medical services move in the same direction. In contrast, the efficiency analysis is necessitated and better equipped to combine expenditures, utilization and other variables for an overall evaluation of the ACO's performance.

Table 8. Summary statistics of utilization measures (inpatient admissions and ambulatory encounters) of Medicare ACOs

ACO model	Inpatient admissions per beneficiary year		Ambulatory encounters per beneficiary year	
	Mean	StDev	Mean	StDev
Track 1	0.32	0.07	11.08	1.63
Track 2	0.33	0.07	11.06	1.25
Track 3	0.33	0.05	11.25	1.87
Tracks 2&3	0.33	0.05	11.21	1.75
Next Generation ACOs	n/a	n/a	n/a	n/a
Total	0.32	0.07	11.09	1.63

Quality measures are among the outputs of three of the four DEA efficiency models. Some summary statistics of the quality domain composite measures of Medicare ACOs are presented in Table 9. As stated, there are four quality domain composite measures:

patient/caregiver experience composite (quality domain composite 1), care coordination/patient safety composite (quality domain composite 2), preventive health composite (quality domain composite 3), and at-risk population composite (quality domain composite 4). The quality measure ACO-11 is exempted for Next Generation ACOs. Both the care coordination/patient safety composite measures with and without ACO-11 are presented. However, they are highly correlated (the correlation coefficient is 0.88).

It is understandable that health may not be improved no matter what medical services are provided or how much is spent. Accordingly, it is appropriate to designate patient experience or satisfaction as one (if not the only one) of the ultimate goals. Interestingly, the rating of patient/caregiver experience is almost the same for all the different ACO models, one-sided or two-sided (73.1 for Track 1, 73.5 for Track 2, 73.1 for Track 3, and 73.3 for Next Generation ACOs). This implies that, as far as patient experience is concerned, the extra services and expenditures of some ACOs might not be warranted.

As to care coordination/patient safety, two-sided Track 3 ACOs score 2.7% higher than one-sided Track 1 ACOs (65.6 vs. 63.9). However, two-sided Track 2 ACOs are 3.1% lower than Track 1 (61.9 vs. 63.9). The two-sided Next Generation ACOs are almost the same as Track 1 and Track 2, but 2.0% lower than Track 3.

Track 1 and Track 2 are equivalent in preventive health (71.1 vs. 71.2), and they are around 5.5% and 2.5% lower than Track 3 (75.2) and Next Generation ACOs (72.9) respectively. Regarding care for at-risk population, Track 1 is 5.8% lower than Track 3 (66.7 vs. 70.8), 2.9% lower than Next Generation ACOs (66.7 vs. 68.7), but 2.3% higher than Track 2 (66.7 vs. 65.2).

Overall, two-sided Track 3 ACOs rank the highest in care coordination/patient safety, preventive health, and care for at-risk population; two-sided Next Generation ACOs are higher than one-sided Track 1 in preventive health and care for at-risk population; and two-sided Track 2 ACOs are lower than Track 1 in care coordination/patient safety and care for at-risk population.

As stated, the triple aims of health care are better care, better health, and lower costs. A higher quality of care indicates higher efficiency given health expenditures and other variables. The efficiency analysis of this research is designed to determine the degree of efficiency of ACOs with differential levels of quality, expenditures, and medical services utilization.

Table 9. Summary statistics of quality domain composite measures of Medicare ACOs*

ACO model		Quality domain composite 1	Quality domain composite 2 (with ACO-11)	Quality domain composite 2 (without ACO-11)	Quality domain composite 3	Quality domain composite 4
Track 1	Mean	73.1	63.9	60.0	71.1	66.7
	StDev	1.7	5.7	5.2	9.6	8.0
Track 2	Mean	73.5	61.9	60.1	71.2	65.2
	StDev	1.5	5.8	5.3	5.7	5.9
Track 3	Mean	73.1	65.6	61.3	75.2	70.8
	StDev	1.3	4.4	4.7	10.2	7.7
Tracks 2&3	Mean	73.2	64.9	61.1	74.4	69.7
	StDev	1.3	4.9	4.8	9.6	7.7
Next Generation ACOs	Mean	73.3	n/a	60.1	72.9	68.7
	StDev	1.4	n/a	5.4	11.4	8.5
Total	Mean	73.1	64.0	60.1	71.4	67.0
	StDev	1.6	5.7	5.2	9.7	8.0

*Quality domain 1: patient/caregiver experience; quality domain 2: care coordination/patient safety; quality domain 3: preventive health; quality domain 4: at-risk population.

5. Efficiency-based comparisons of MSSP ACOs

The Medicare ACO models at various risk levels are designed to award differential financial incentives to reduce expenditures in providing an optimal mix of quality medical services to an assigned Medicare beneficiary population. In this section, the performance efficiency of different MSSP ACOs is compared using several sets of inputs and outputs. Firstly, the 873 MSSP ACOs are temporarily pooled together for the purpose of scaling. Each input and output is divided by its average to ensure that the results are unit invariant (Brockett et al., 2004).

The sample of this research consists of 813 Track 1 ACOs, 12 Track 2 ACOs, and 48 Track 3 ACOs. Equal sample sizes for each pair of comparison are used (Brockett et al., 2004). For example, in comparing Track 1 with Track 3, 48 Track 1 ACOs are randomly drawn with replacement from the 813 Track 1 ACOs. The 48 random Track 1 ACOs are then compared to the 48 Track 3 ACOs. The process is repeated 30 times to check the robustness of the results. Therefore, 30 different Track 1 samples (each of size 48) are run against the 48 (fixed) Track 2 ACOs. The same procedure is applied to all the other pairs of comparison: Track 1 vs. Track 2, Track 1 vs. Tracks 2&3, and Track 2 vs. Track 3.

As discussed, the cross-frontier DEA model is adopted to generate the efficiency scores of ACOs, with the ACOs of one track (for example, Track 1) being run exclusively against the efficient frontier of the alternative track (for example, Track 3). After obtaining the efficiency scores, the analysis to detect efficiency differences between two tracks is performed by the Mann-Whitney test (Brockett and Golany, 1994). The alternative hypothesis is accepted if the p-value is <10%.

Track 2 and Track 3 are both two-sided ACO models. The first comparison of this research groups Track 2 and Track 3 together to compare them with the one-sided Track 1 ACOs. The results of the efficiency comparison between Track 1 and Tracks 2&3 are presented in Table 10. Under DEA model 1 (with actual expenditures as inputs, and enrollment, utilization and quality as outputs), the alternative hypothesis that Track 1 is more efficient than Tracks 2&3

is supported in 26 (86.7%) of the 30 random runs. For the other 4 runs, the null hypothesis that Track 1 is equally efficient as Tracks 2&3 is accepted.

The efficiency comparison is also conducted using the other two alternative DEA models: DEA model 2 (with actual expenditures as inputs, and enrollment and quality as outputs), and DEA model 3 (with actual expenditures as inputs, and enrollment and utilization as outputs). Under DEA model 2, Track 1 is more efficient than Tracks 2&3 in 28 (93.3%) of the 30 random runs, and Track 1 is equally efficient as Tracks 2&3 in the other two runs. Under DEA model 3, Track 1 is more efficient than Tracks 2&3 in 23 (76.7%) of the 30 random runs, Track 1 is equally efficient as Tracks 2&3 in five runs, and Track 1 is less efficient than Tracks 2&3 in only one run.

Overall, the result indicates that two-sided ACOs are actually less efficient than one-sided ACOs, in minimizing health expenditures given enrollment (total beneficiary-years), utilization of medical services, and quality of care (input-oriented), or in maximizing the covered beneficiaries, utilization, and quality given health expenditures (output-oriented). Therefore, it might not be advisable to mandate the transition of ACOs from one-sided to two-sided, as far as efficiency is concerned.

Table 10. Efficiency comparison between Track 1 and Tracks 2&3 of MSSP ACOs

Run No.	DEA model 1		DEA model 2		DEA model 3	
	Accepted hypothesis	p-value	Accepted hypothesis	p-value	Accepted hypothesis	p-value
1	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
2	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
3	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.002	$AH_{1vs2\&3}$-1	0.000
4	$AH_{1vs2\&3}$-1	0.034	$AH_{1vs2\&3}$-1	0.005	$AH_{1vs2\&3}$-2	0.408
5	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.005	$AH_{1vs2\&3}$-1	0.000
6	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
7	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
8	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
9	$AH_{1vs2\&3}$-1	0.028	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-2	0.211
10	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
11	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
12	$AH_{1vs2\&3}$-2	0.405	$AH_{1vs2\&3}$-1	0.023	$AH_{1vs2\&3}$-3	0.032
13	$AH_{1vs2\&3}$-1	0.018	$AH_{1vs2\&3}$-1	0.017	$AH_{1vs2\&3}$-2	0.301
14	$AH_{1vs2\&3}$-1	0.004	$AH_{1vs2\&3}$-1	0.078	$AH_{1vs2\&3}$-1	0.072
15	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
16	$AH_{1vs2\&3}$-2	0.374	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-3	0.044
17	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
18	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.002	$AH_{1vs2\&3}$-1	0.000
19	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
20	$AH_{1vs2\&3}$-2	0.796	$AH_{1vs2\&3}$-2	0.720	$AH_{1vs2\&3}$-2	0.585
21	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
22	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
23	$AH_{1vs2\&3}$-1	0.044	$AH_{1vs2\&3}$-1	0.039	$AH_{1vs2\&3}$-1	0.053
24	$AH_{1vs2\&3}$-1	0.040	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.066
25	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
26	$AH_{1vs2\&3}$-1	0.007	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.001
27	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000
28	$AH_{1vs2\&3}$-2	0.408	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-2	0.527
29	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.000	$AH_{1vs2\&3}$-1	0.001
30	$AH_{1vs2\&3}$-1	0.007	$AH_{1vs2\&3}$-2	0.254	$AH_{1vs2\&3}$-1	0.014

DEA model 1: inputs – health expenditures; outputs – enrollment, utilization, and quality.
DEA model 2: inputs – health expenditures; outputs – enrollment and quality.
DEA model 3: inputs – health expenditures; outputs – enrollment and utilization.
$AH_{1vs2\&3}$-1: Track 1 is more efficient than Tracks 2&3.
$AH_{1vs2\&3}$-2: Track 1 is equally efficient as Tracks 2&3.
$AH_{1vs2\&3}$-3: Track 1 is less efficient than Tracks 2&3.
$AH_{1vs2\&3}$-1 and $AH_{1vs2\&3}$-3 are the alternative hypothesis, and $AH_{1vs2\&3}$-2 is the null hypothesis.
The significance level is 10%.

Among two-sided MSSP ACOs, there are very few in Track 2 (only 12 out of 60 in the sample), and Track 3 ACOs represent the majority (48 out of 60). Therefore, Track 1 ACOs are also compared with Track 3 ACOs alone (to eliminate the confounding effect, if any, of Track 2 ACOs). The efficiency comparison between Track 1 and Track 3 is presented in Table 11.

Under DEA model 1, Track 1 is more efficient than Track 3 in 21 (70%) of the 30 random runs; Track 1 is equally efficient as Track 3 in eight runs; and Track 1 is less efficient than Track 3 in only one run. Under DEA model 2, Track 1 is more efficient than Track 3 in 25 of the 30 runs; Track 1 is equally efficient as Track 3 in five runs; and no run supports that hypothesis that Track 3 is more efficient than Track 1. Under DEA model 3, relatively more runs support the hypothesis that Track 3 is more efficient than Track 1 (five of the 30 runs); however, there is still an overwhelming majority of the 30 runs (25, or 83.3%) which support the hypothesis that Track 1 is more efficient than or equally efficient as Track 3.

In summary, the result shows that generally one-sided Track 1 ACOs are more efficient than the two-sided Track 3 ACOs in minimizing health expenditures or maximizing covered beneficiaries, utilization, and quality.

Table 11. Efficiency comparison between Track 1 and Track 3 of MSSP ACOs

Run No.	DEA model 1		DEA model 2		DEA model 3	
	Accepted hypothesis	p-value	Accepted hypothesis	p-value	Accepted hypothesis	p-value
1	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.001	AH_{1vs3}-1	0.004
2	AH_{1vs3}-1	0.005	AH_{1vs3}-1	0.013	AH_{1vs3}-1	0.000
3	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000
4	AH_{1vs3}-2	0.875	AH_{1vs3}-1	0.000	AH_{1vs3}-2	0.580
5	AH_{1vs3}-3	0.095	AH_{1vs3}-1	0.000	AH_{1vs3}-3	0.023
6	AH_{1vs3}-2	0.406	AH_{1vs3}-1	0.034	AH_{1vs3}-3	0.006
7	AH_{1vs3}-1	0.001	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.014
8	AH_{1vs3}-3	0.000	AH_{1vs3}-2	0.991	AH_{1vs3}-3	0.000
9	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000
10	AH_{1vs3}-1	0.040	AH_{1vs3}-2	0.221	AH_{1vs3}-2	0.956
11	AH_{1vs3}-1	0.001	AH_{1vs3}-2	0.956	AH_{1vs3}-1	0.000
12	AH_{1vs3}-2	0.240	AH_{1vs3}-1	0.000	AH_{1vs3}-2	0.679
13	AH_{1vs3}-2	0.590	AH_{1vs3}-1	0.010	AH_{1vs3}-3	0.004
14	AH_{1vs3}-1	0.004	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.002
15	AH_{1vs3}-1	0.052	AH_{1vs3}-2	0.545	AH_{1vs3}-1	0.005
16	AH_{1vs3}-1	0.007	AH_{1vs3}-1	0.000	AH_{1vs3}-2	0.262
17	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000
18	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000
19	AH_{1vs3}-2	0.317	AH_{1vs3}-1	0.053	AH_{1vs3}-2	0.755
20	AH_{1vs3}-2	0.711	AH_{1vs3}-1	0.015	AH_{1vs3}-3	0.005
21	AH_{1vs3}-1	0.001	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000
22	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000
23	AH_{1vs3}-2	0.294	AH_{1vs3}-1	0.028	AH_{1vs3}-3	0.020
24	AH_{1vs3}-2	0.772	AH_{1vs3}-1	0.004	AH_{1vs3}-3	0.056
25	AH_{1vs3}-1	0.048	AH_{1vs3}-2	0.761	AH_{1vs3}-1	0.017
26	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.004	AH_{1vs3}-1	0.000
27	AH_{1vs3}-1	0.016	AH_{1vs3}-1	0.000	AH_{1vs3}-2	0.348
28	AH_{1vs3}-1	0.088	AH_{1vs3}-1	0.048	AH_{1vs3}-2	0.406
29	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000
30	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000	AH_{1vs3}-1	0.000

DEA model 1: inputs – health expenditures; outputs – enrollment, utilization, and quality.
DEA model 2: inputs – health expenditures; outputs – enrollment and quality.
DEA model 3: inputs – health expenditures; outputs – enrollment and utilization.
AH_{1vs3}-1: Track 1 is more efficient than Track 3.
AH_{1vs3}-2: Track 1 is equally efficient as Track 3.
AH_{1vs3}-3: Track 1 is less efficient than Track 3.
AH_{1vs3}-1 and AH_{1vs3}-3 are the alternative hypothesis, and AH_{1vs3}-2 is the null hypothesis.
The significance level is 10%.

Even though Track 2 ACOs only accounts for a very small share of the two-sided MSSP ACOs, they are still compared with one-sided Track 1 ACOs and two-sided Track 3 ACOs. The efficiency comparison of Track 2 with Track 1 and Track 3 is presented in Appendix II and III. Under DEA model 1, Track 1 is more efficient than Track 2 in 12 of the 30 runs, and Track 2 is more efficient than Track 1 in four runs. Under DEA model 2, Track 1 is more efficient than Track 2 in seven runs, and Track 2 is more efficient than Track 1 also in seven runs. Under DEA model 3, Track 1 is more efficient than Track 2 in 17 runs, and Track 2 is more efficient than Track 1 in only one run. Overall, it is safe to claim that one-sided Track 1 ACOs are more efficient than two-sided Track 2 ACOs using DEA models 1&3, and at least Track 2 is not more efficient than Track 1 using DEA model 2. This implies that, efficiency-wise, there is no justification to switch an ACO from Track 1 to Track 2.

So far, none of the results of the efficiency comparison have documented any advantages of two-sided MSSP ACOs, either Track 2 or Track 3, compared to one-sided Track 1 ACOs. Therefore, the one-sided Track 1 ACO model should be the desirable design of Medicare ACOs from the perspective of efficiency. However, the discussion of the efficiency difference between Track 2 and Track 3 is still presented to complete the “comparison” analysis of every pairs of the ACO models. Under DEA model 1, Track 2 is more efficient than Track 3 in eight runs, and less efficient than Track 3 in four runs. Under DEA model 2, Track 2 is more efficient than Track 3 in 20 runs, and less efficient than Track 3 in none of the 30 runs. Under DEA model 3, Track 2 is more efficient than Track 3 in four runs and less efficient than Track 3 in 10 runs. This indicates that Track 2 is more efficient than Track 3 using DEA models 1&2, but less efficient than Track 3 using DEA model 3.

6. Efficiency-based comparisons between MSSP ACOs and Next Generation ACOs

Among Medicare ACOs, the two-sided Next Generation ACOs assume the highest level of financial risks, higher than MSSP Track 3 ACOs. They are an experimental innovation of Medicare payment reform moving more closely towards the capitation payment model. This section evaluates the performance of Next Generation ACOs from the efficiency perspective.

As stated, for Next Generation ACOs, the total beneficiary-years and the utilization measures are not available, and the quality measure ACO-11 is excluded from the care coordination/patient safety domain. Therefore, instead of DEA models 1, 2, and 3, DEA model 4 is utilized for the comparison of Next Generation ACOs with MSSP ACOs. Same as the other DEA models, DEA model 4 picks actual expenditures as the input. However, the outputs of DEA model 4 comprise total assigned beneficiaries and quality measures. It is shown that total assigned beneficiaries are highly correlated with total beneficiary-years, so are the care coordination/patient safety domain composite measures with and without ACO-11. Subsequently, the results using DEA model 4 are not expected to be significantly different from those of DEA model 2.

There are 60 Next Generation ACOs in the sample, which are compared with MSSP ACOs. As in section 5 of this article, 60 Next Generation ACOs and the 873 MSSP ACOs are pooled together temporarily to scale the input and output variables. Afterwards, the 60 Next Generation ACOs are firstly compared with the 60 MSSP two-sided ACOs (both Track 2 and Track 3 as a single group). Because they have the same sample size (60 ACOs for each), random samples are not necessary for this comparison. The cross-frontier DEA model is run on each group exclusively against the other group to obtain the efficiency scores of the ACOs. Then the Mann-Whitney test is applied to compare the two groups. The results show that the MSSP two-sided ACOs are more efficient than Next Generation ACOs (the p-value is 0.024) in minimizing health expenditures or maximizing covered beneficiaries and quality of care.

Next Generation ACOs are also compared to Track 1, Track 2, and Track 3 separately. Similarly, 30 random samples are drawn for each pair of comparison. For example, to compare the 60 Next Generation ACOs with Track 1, 30 random samples are selected from the 813 Track 1 ACOs, each of which comprises 60 Track 1 ACOs. The efficiency comparison results between Next Generation ACOs and MSSP ACOs are presented in Table 12.

The results show that all the 30 random runs support the hypothesis that MSSP Track 1 ACOs are more efficient than Next Generation ACOs. Track 2 is more efficient than Next Generation ACOs in 25 of the 30 runs, and they are equally efficient in the other five runs. As to the efficiency difference between Track 3 and Next Generation ACOs, 18 of the 30 runs support the hypothesis that Track 3 is more efficient, and they are equally efficient in the other 12 runs. None of 90 random runs show that Next Generation ACOs are more efficient than any track of MSSP ACOs.

In summary, given the number of assigned beneficiaries and the quality of medical services, Next Generation ACOs are the least efficient in minimizing health expenditures. Alternatively, given health expenditures, Next Generation ACOs are the least efficient in maximizing covered beneficiaries and quality of medical services. Based on the similar results between DEA model 1 and DEA model 2, it would be expected that Next Generation ACOs should still be the least efficient if utilization measures were incorporated. Therefore, the higher the financial risk that the ACO assumes (from Track 1 to Tracks 2&3, then to Next Generation ACOs), the less efficient the ACO becomes in minimizing health expenditures or maximizing enrollment (total assigned beneficiaries or beneficiary-years), utilization, and quality.

Table 12. Efficiency comparison between Next Generation ACOs and MSSP ACOs

Run No.	Track 1 vs. Next Generation ACOs		Track 2 vs. Next Generation ACOs		Track 3 vs. Next Generation ACOs	
	Accepted hypothesis	p-value	Accepted hypothesis	p-value	Accepted hypothesis	p-value
1	AH_{1vsNG}-1	0.000	AH_{2vsNG}-2	0.932	AH_{3vsNG}-1	0.088
2	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.022	AH_{3vsNG}-1	0.040
3	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.003	AH_{3vsNG}-2	0.296
4	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.001	AH_{3vsNG}-2	0.444
5	AH_{1vsNG}-1	0.004	AH_{2vsNG}-2	0.469	AH_{3vsNG}-2	0.690
6	AH_{1vsNG}-1	0.001	AH_{2vsNG}-1	0.022	AH_{3vsNG}-1	0.000
7	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-1	0.000
8	AH_{1vsNG}-1	0.000	AH_{2vsNG}-2	0.284	AH_{3vsNG}-1	0.056
9	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-1	0.050
10	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.014	AH_{3vsNG}-1	0.000
11	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-2	0.453
12	AH_{1vsNG}-1	0.000	AH_{2vsNG}-2	0.285	AH_{3vsNG}-1	0.042
13	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.001	AH_{3vsNG}-1	0.064
14	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.002	AH_{3vsNG}-1	0.000
15	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-2	0.631
16	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-2	0.307
17	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.034	AH_{3vsNG}-1	0.000
18	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-2	0.921
19	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-1	0.000
20	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.002	AH_{3vsNG}-1	0.000
21	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-2	0.248
22	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.001	AH_{3vsNG}-2	0.365
23	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.022	AH_{3vsNG}-2	0.823
24	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.016	AH_{3vsNG}-1	0.027
25	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-1	0.009
26	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-2	0.700
27	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-2	0.267
28	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.000	AH_{3vsNG}-1	0.000
29	AH_{1vsNG}-1	0.000	AH_{2vsNG}-1	0.016	AH_{3vsNG}-1	0.000
30	AH_{1vsNG}-1	0.000	AH_{2vsNG}-2	0.744	AH_{3vsNG}-1	0.068

The efficiency scores are generated using DEA model 4: inputs – health expenditures; outputs – enrollment and quality (without ACO-11).

AH_{1vsNG}-1: Track 1 is more efficient than Next Generation ACOs.

AH_{2vsNG}-1: Track 2 is more efficient than Next Generation ACOs.

AH_{2vsNG}-2: Track 2 is equally efficient as Next Generation ACOs.

AH_{3vsNG}-1: Track 3 is more efficient than Next Generation ACOs.

AH_{3vsNG}-2: Track 3 is equally efficient as Next Generation ACOs.

AH_{1vsNG}-1, AH_{2vsNG}-1 and AH_{3vsNG}-1 are the alternative hypothesis, and AH_{2vsNG}-2 and AH_{3vsNG}-2 are the null hypothesis.

The significance level is 10%.

7. Efficiency-based potential cost savings of Medicare ACOs

The DEA efficiency score is obtained as the optimal ratio of the weighted sum of outputs over the weighted sum of inputs. Less efficient ACOs can resort to the reduction of inputs (health expenditures) to improve their efficiency. The potential cost savings are determined by the difference between the actual (or benchmark) expenditures and the efficient expenditures, the latter of which are generated by the DEA analysis. This section investigates the potential cost savings of the MSSP and Next Generation ACOs, by comparing efficient expenditures to both actual and benchmark expenditures, based on different efficiency goals.

Any efficient ACO(s), no matter which group it belongs to, should be qualified to serve as the efficiency benchmark for the less efficient ACOs. Accordingly, instead of the cross-frontier DEA model, the collective frontier DEA model is applied to calculate the efficient expenditures of the ACOs. It is unrealistic to expect all the less efficient ACOs to reach the highest possible efficiency level. Therefore, two conservative efficiency goals are selected for the efficiency improvement purpose: the median efficiency and the 25th percentile efficiency (Brockett, Golden and Yang, 2018). Firstly the potential savings are calculated for all the MSSP ACOs using DEA models 1, 2, and 3 (which are not applicable to Next Generation ACOs). The collective frontier DEA model is run on all the 873 MSSP ACOs to obtain their efficiency scores. With the median efficiency goal, the ACOs with an efficiency score above the median efficiency of the 873 ACOs are not required to reduce their health expenditures; in other words, their actual expenditures are the efficient expenditures. The collective frontier DEA model is run again on the 437 ACOs with an efficiency score at or below the median efficiency of the 873 ACOs, to derive their efficient expenditures. Similarly, with the 25th percentile efficiency goal, the efficient expenditures are obtained using the DEA analysis only for the ACOs with an efficiency score at or below the 25th percentile efficiency of the 873 ACOs; and the actual expenditures of other ACOs are deemed as efficient.

The potential cost savings of the MSSP ACOs are presented in Table 13 (median efficiency target) and Table 14 (25th percentile efficiency target), with respect to benchmark and actual expenditures. The potential cost reductions are similar under DEA model 1 and DEA model 2. Specifically, using the median efficiency target, the potential savings (compared to actual expenditures) are 4.0% (Track 1), 4.1% (Track 2), and 4.7% (Track 3) under DEA model 1; 4.3% (Track 1), 4.2% (Track 2), and 4.5% (Track 3) under DEA model 2. Using the 25th percentile efficiency target, the potential savings (compared to actual expenditures) are 1.5% (Track 1), 1.2% (Track 2), and 1.5% (Track 3) under DEA model 1; 1.7% (Track 1), 1.2% (Track 2), and 1.5% (Track 3) under DEA model 2.

The potential cost reductions are significantly higher using DEA model 2 than using DEA model 1 or 3. Under DEA model 2, the potential savings (compared to actual expenditures) are 6.2% (Track 1), 9.7% (Track 2), and 6.5% (Track 3) using the median efficiency target; 2.7% (Track 1), 5.6% (Track 2), and 2.2% (Track 3) using the 25th percentile efficiency target.

This current research suggests the incorporation of utilization and quality in the DEA analysis (DEA model 1). In conclusion, without any sacrifice in enrollment (total beneficiary-years), utilization or quality, the MSSP Track 1, Track 2, and Track 3 are expected to reduce expenditures by 4.0%, 4.1%, and 4.7% using the median efficiency target, and 1.7%, 1.2%, and 1.5% using the 25th percentile efficiency target (compared to actual expenditures).

Table 13. Potential cost savings of MSSP ACOs using the median efficiency of all the three tracks as the target

DEA model	ACO model	Savings per member year (compared to benchmark expenditures)				Savings per member year (compared to actual expenditures)			
		$		%		$		%	
		Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Model 1	Track 1	696	1,084	5.1%	7.5%	538	893	4.0%	6.2%
	Track 2	1,123	1,148	7.7%	6.9%	608	825	4.1%	5.5%
	Track 3	900	957	6.7%	6.8%	614	814	4.7%	6.0%
	All tracks	713	1080	5.2%	7.5%	543	887	4.1%	6.2%
Model 2	Track 1	1,081	1,895	7.2%	10.5%	922	1,731	6.2%	9.7%
	Track 2	1,964	2,364	12.9%	13.4%	1,450	2,043	9.7%	12.6%
	Track 3	1,174	1,390	8.4%	9.2%	889	1,198	6.5%	8.2%
	All tracks	1,098	1,879	7.4%	10.4%	928	1,710	6.3%	9.6%
Model 3	Track 1	739	1,103	5.4%	7.6%	581	923	4.3%	6.4%
	Track 2	1,137	1,172	7.8%	7.1%	623	848	4.2%	5.6%
	Track 3	875	933	6.5%	6.6%	589	783	4.5%	5.7%
	All tracks	752	1,095	5.5%	7.5%	582	914	4.4%	6.4%

Table 14. Potential cost savings of MSSP ACOs using the 25th percentile efficiency of all the three tracks as the target

DEA model	ACO model	Savings per member year (compared to benchmark expenditures)				Savings per member year (compared to actual expenditures)			
		$		%		$		%	
		Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Model 1	Track 1	370	853	2.6%	6.0%	212	566	1.5%	3.8%
	Track 2	692	723	4.8%	4.5%	178	303	1.2%	1.9%
	Track 3	491	678	3.6%	5.0%	205	470	1.5%	3.4%
	All tracks	381	843	2.7%	5.9%	211	558	1.5%	3.8%
Model 2	Track 1	603	1,545	3.7%	8.2%	444	1,342	2.7%	6.9%
	Track 2	1,380	1,846	9.0%	10.0%	865	1,507	5.6%	8.9%
	Track 3	603	1,005	4.1%	6.7%	317	728	2.2%	4.9%
	All tracks	614	1,526	3.8%	8.2%	443	1,318	2.7%	6.8%
Model 3	Track 1	391	865	2.7%	6.1%	232	596	1.7%	4.1%
	Track 2	694	710	4.9%	4.6%	180	302	1.2%	2.0%
	Track 3	482	681	3.5%	5.0%	196	465	1.5%	3.4%
	All tracks	400	854	2.8%	6.0%	230	586	1.6%	4.0%

The efficiency comparison of this research suggests that the MSSP Track 1 ACOs should be the most efficient, which currently also accounts for the majority of all MSSP ACOs (813 out of 873 in the sample of this research). Therefore, the potential savings are also calculated for Track 1 ACOs without pooling Track 2 and Track 3 ACOs in the reference set. That is, the efficient frontier for the less efficient Track 1 ACOs only comprises Track 1 ACOs. By comparing to the efficient ACOs in the same track, this analysis examines the potential savings of Track 1 ACOs without switching to other two-sided tracks. However, due to the small share of Track 2 and Track 3 ACOs (only 60 out of 873), the potential savings of Track 1 ACOs are not expected to be significantly different. The potential savings of Track 1 are presented in Table 15 using the median and 25th percentile efficiency targets.

The same pattern of potential savings of Track 1 ACOs is documented. The potential savings are similar under DEA model 1 and DEA model 3, but significantly higher under DEA model 2. Specifically, the potential savings of Track 1 are 4.1% (DEA model 1), 6.4% (DEA model 2), and 4.3% (DEA model 3) using the median efficiency target, and 1.5% (DEA model 1), 2.8% (DEA model 2), and 1.7% (DEA model 3) using the 25th percentile efficiency target.

In summary, maintaining the same level of enrollment (total beneficiary-years), utilization, and quality, without switching to the two-sided ACO tracks, MSSP Track 1 ACOs are expected to reduce expenditures by 4.1% using the median efficiency target, and 1.5% using the 25th percentile efficiency target (compared to actual expenditures). The CMS might set its savings objective anywhere in the middle, for example, 2.8% (the average of 4.1% and 1.5%).

Table 15. Potential savings of MSSP Track 1 ACOs using the median efficiency and the 25th percentile efficiency of Track 1 as the targets

Efficiency target	DEA model	Savings per member year (compared to benchmark expenditures)				Savings per member year (compared to actual expenditures)			
		$		%		$		%	
		Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Median efficiency of Track 1	Model 1	703	1,088	5.1%	7.5%	544	897	4.1%	6.3%
	Model 2	1,098	1,904	7.4%	10.5%	939	1,740	6.4%	9.7%
	Model 3	739	1,102	5.4%	7.6%	580	922	4.3%	6.4%
25th percentile efficiency of Track 1	Model 1	375	857	2.6%	6.0%	217	575	1.5%	3.9%
	Model 2	619	1558	3.9%	8.3%	461	1,358	2.8%	7.0%
	Model 3	391	865	2.7%	6.1%	233	597	1.7%	4.1%

As stated, DEA models 1, 2, and 3 are not applicable to Next Generation ACOs. The potential savings of Next Generation ACOs are calculated using DEA model 4. The outputs of DEA model 4 consists of enrollment (total assigned beneficiaries) and quality domain composite measures (the care coordination/patient safety domain composite does not include ACO-11). The potential savings of the 60 Next Generation ACOs are computed by pooling all the 933 ACOs together (873 MSSP ACOs and 60 Next Generation ACOs) (Table 16). The potential savings of the MSSP tracks are also presented to show their sensitivity to the different enrollment variable and the quality measure ACO-11.

The result shows that, without any sacrifice in enrollment (total assigned beneficiaries), or quality, the Next Generation ACOs are expected to reduce expenditures by 9.0% using the median efficiency of the 933 ACOs as the efficiency target, and 4.1% using the 25th percentile efficiency target.

DEA model 4 is a modified version of DEA model 2 with two minor changes. The potential savings of MSSP ACOs are similar under these two DEA models, using total assigned beneficiaries or total beneficiary-years, with the quality measure ACO-11 or without.

Specifically, with regard to actual expenditures, the potential savings of MSSP ACOs are 6.2% (Track 1), 9.7% (Track 2), and 6.5% (Track 3) under DEA model 2, versus 5.9% (Track 1), 9.2% (Track 2), and 6.4% (Track 3) under DEA model 4, using the median efficiency target; 2.7% (Track 1), 5.6% (Track 2), and 2.2% (Track 3) under DEA model 2, versus 2.5% (Track 1), 5.2% (Track 2), and 2.1% (Track 3) under DEA model 4, using the 25th percentile efficiency target.

It is shown that DEA model 2 (and DEA model 4) tends to overestimate the potential savings compared to DEA model 1 or 3. For example, the potential savings of MSSP Track 3 are 6.5% (DEA model 2) and 4.7% (DEA model 1) using the median efficiency target (compared to actual expenditures); 2.2% (DEA model 2) and 1.5% (DEA model 1) using the 25th percentile efficiency target.

Therefore, roughly, it is expected that Next Generation ACOs would reduce 7.2% (using the median efficiency target) and 3.4% (using the 25th percentile efficiency target) of actual expenditures under DEA model 1, without any sacrifice in enrollment, utilization, or quality.

Table 16. Potential savings of Next Generation and MSSP ACOs using DEA model 4

Efficiency target	ACO model	Savings per member year (compared to benchmark expenditures)				Savings per member year (compared to actual expenditures)			
		$		%		$		%	
		Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Median efficiency of MSSP and Next Generation ACOs	Track 1	1,004	1,782	6.9%	10.2%	850	1,622	5.9%	9.4%
	Track 2	1,834	2,175	12.5%	12.9%	1,337	1,872	9.2%	12.1%
	Track 3	1,139	1,352	8.3%	9.1%	859	1,161	6.4%	8.1%
	Next Generation ACOs	1,470	1,843	10.4%	10.8%	1,268	1,664	9.0%	10.5%
25th percentile efficiency of MSSP and Next Generation ACOs	Track 1	553	1,440	3.5%	8.0%	399	1,242	2.5%	6.6%
	Track 2	1,274	1,681	8.7%	9.6%	777	1,361	5.2%	8.4%
	Track 3	584	979	4.1%	6.7%	304	702	2.1%	4.8%
	Next Generation ACOs	826	1,508	5.6%	8.6%	624	1,291	4.1%	7.9%

Chernew, Barbey and McWilliams (2017) argue that benchmark expenditures are not true counterfactuals and the use of benchmark expenditures likely underestimates ACO savings. In the current research, the potential savings compared to benchmark expenditures are presented mainly for the informative and comparison purpose but not discussed much. However, the efficiency analysis of this current research also provides some implications for setting benchmark expenditures for Medicare ACOs.

The results show that the benchmark expenditures of a significant number of ACOs are actually below the efficient expenditures using the median or the 25th percentile efficiency goal (Table 17). Using DEA model 1, the benchmark expenditures for 229 ACOs (26.2%) are below the efficient expenditures using the median efficiency target, and 34.4% using the 25th percentile efficiency target. Under the other three DEA models, 24.4% (DEA model 2), 24.5% (DEA model 3), and 24.5% (DEA model 4) are below the efficient expenditures using the median efficiency target, and 32.5% (DEA model 2), 33.4% (DEA model 3), and 32.6% (DEA model 4) using the 25th percentile efficiency target.

Generally, the benchmark expenditures for one in four Medicare ACOs are below the efficient expenditures using the median efficiency target, and one in three using the 25th percentile efficiency target. The benchmark expenditures are probably too low for these ACOs. Therefore, it might be advisable that these benchmark expenditures be adjusted upward.

Table 17. Number of ACOs with benchmark expenditures below efficient expenditures

DEA model	ACO model	Total number of ACOs	Median efficiency as the target		25th percentile efficiency as the target	
			Number of ACOs with benchmark expenditures below efficient expenditures	%	Number of ACOs with benchmark expenditures below efficient expenditures	%
Model 1	Track 1	813	216	26.6%	282	34.7%
	Track 2	12	2	16.7%	2	16.7%
	Track 3	48	11	22.9%	16	33.3%
	Total	873	229	26.2%	300	34.4%
Model 2	Track 1	813	201	24.7%	268	33.0%
	Track 2	12	2	16.7%	2	16.7%
	Track 3	48	10	20.8%	14	29.2%
	Total	873	213	24.4%	284	32.5%
Model 3	Track 1	813	202	24.8%	274	33.7%
	Track 2	12	2	16.7%	2	16.7%
	Track 3	48	10	20.8%	16	33.3%
	Total	873	214	24.5%	292	33.4%
Model 4	Track 1	813	209	25.7%	273	33.6%
	Track 2	12	2	16.7%	2	16.7%
	Track 3	48	10	20.8%	15	31.3%
	Next Generation ACOs	60	8	13.3%	14	23.3%
	Total	933	229	24.5%	304	32.6%

8. Conclusions

Medicare ACOs represent the nation's largest initiative of Medicare alternative payment models toward value and health outcomes, which are designed to provide financial incentives for Medicare providers to reduce inefficiencies in care delivery by facilitating improved communication and care coordination. Various ACOs models have been experimented at differential risk structures, with one-sided ACOs sharing only savings and two-sided ACOs sharing both savings and losses. It is concerned that one-sided ACOs might not generate significant savings due to insufficient financial responsibilities. Therefore, the CMS has issued a final rule to advance accountability, competition, engagement, integrity, and quality for Medicare ACOs, and accelerate the path to assume greater downside financial risks. In response, this research conducts a comprehensive efficiency analysis of Medicare ACOs incorporating health expenditures, utilization of medical services, and quality of health care, investigates whether superiority exists among the various Medicare ACO models and determines their potential cost reductions, and aims to provide insights on the effective design of Medicare ACOs to increase quality for patients and drive towards cost savings.

Most other discussions and analyses on Medicare ACOs focus on individual measures separately, such as health expenditures, inpatient hospital days, or unplanned readmissions. The efficiency analysis of this research integrates various input and output variables to deliver a holistic evaluation of the ACOs. The value-based Medicare ACOs are an innovative component of the national quality strategy to fulfill the triple aims of health care: better care, better health and lower costs. In evaluating the efficiency of ACOs from the societal perspective, one emphasis should be the value of health care services: the quality of care and health outcomes of beneficiaries. Therefore, besides enrollment (assigned beneficiaries) and utilization of medical services, the outputs of the ACO efficiency analysis also contain "better care and better health", which are represented by the CMS quality measures. Specifically, four quality domain composite measures are incorporated: patient/caregiver experience composite, care coordination/patient safety composite, preventive health composite, and at-risk population composite. Analyses are also conducted using several alternative sets of outputs for robustness check and comparison purpose.

The results indicate that generally the Medicare ACOs with higher financial risks are associated with better quality of care but more health expenditures and utilization of medical services. In minimizing expenditures given assigned beneficiaries, utilization and quality levels, or maximizing the outputs given health expenditures, one-sided ACOs are more efficient than two-sided ACOs. Specifically the MSSP Track 1 is the most efficient and the Next Generation ACO model is the least efficient. The efficiency of ACOs deteriorates with the assumption of more financial risks. None of the results of the efficiency comparison document any advantages of two-sided MSSP ACOs, either Track 2 or Track 3, compared to one-sided Track 1 ACOs. Therefore, the one-sided Track 1 ACO model should be the desirable design of the Medicare ACOs from the perspective of efficiency. Correspondingly, it might not be advisable to mandate the transition of ACOs from one-sided to two-sided, as far as efficiency is concerned.

This research investigates the potential cost savings of the MSSP and Next Generation ACOs based on different efficiency goals. Less efficient ACOs can resort to the reduction of health expenditures to improve their efficiency. The results indicate that, without any sacrifice in enrollment (assigned beneficiaries), utilization or quality, the MSSP Track 2, Track 3, and Next Generation ACOs are expected to reduce expenditures by 4.1%, 4.7%, and 7.2% using the median efficiency target, and 1.2%, 1.5%, and 3.4% using the 25^{th} percentile efficiency target (compared to actual expenditures). Maintaining the same level of enrollment, utilization, and quality, without switching to the two-sided ACO tracks, MSSP Track 1 ACOs are expected to reduce expenditures by 4.1% using the median efficiency target, and 1.5% using the 25^{th} percentile efficiency target (compared to actual expenditures).

The efficiency analysis of this current research also provides some implications for setting benchmark expenditures for Medicare ACOs. The results show that the benchmark expenditures of a significant number of ACOs are actually below the efficient expenditures using the median or the 25th percentile efficiency goal. Generally, the benchmark expenditures for one in four Medicare ACOs are below the efficient expenditures using the median efficiency target, and one in three using the 25th percentile efficiency target. The benchmark expenditures are probably too low for these ACOs. Therefore, it might be advisable that these benchmark expenditures be adjusted upward.

References

Banker, R., A. Charnes and W. Cooper, 1984, Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science* 30(9): 1078-1092.

Brockett, P., R. Chang, J. Rousseau, J. Semple, and C. Yang, 2004, A Comparison of HMO Efficiencies as a Function of Provider Autonomy, *Journal of Risk and Insurance* 71: 1-19.

Brockett, P., and B. Golany, 1994, Determining Programmatic Efficiency Differences in DEA, *Management Science* 42: 466 – 472.

Brockett, P., L. Golden, and C. Yang, 2018, "Potential "Savings" of Medicare: the Analysis of Medicare Advantage and Accountable Care Organizations (ACOs)," *North American Actuarial Journal* 22:3, 458-472.

Castellucci, M., V. Dickson, 2018, Medicare ACOs Saved CMS $314 Million in 2017, *Modern Healthcare*, https://www.modernhealthcare.com/article/20180830/NEWS/180839987.

Centers for Medicare & Medicaid Services (CMS), 2015, Pioneer ACO Model and Next Generation ACO Model: Comparison across Key Design Elements, https://innovation.cms.gov/Files/fact-sheet/nextgenaco-comparefactsheet.pdf.

Centers for Medicare & Medicaid Services (CMS), 2016, Medicare Shared Savings Program Quality Measure Benchmarks for the 2016 and 2017 Reporting Years, https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/MSSP-QM-Benchmarks-2016.pdf.

Centers for Medicare & Medicaid Services (CMS), 2017a, New Accountable Care Organization Model Opportunity: Medicare ACO Track 1+ Model, https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/New-Accountable-Care-Organization-Model-Opportunity-Fact-Sheet.pdf.

Centers for Medicare & Medicaid Services (CMS), 2017b, Shared Savings and Losses and Assignment Methodology, https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/Shared-Savings-Losses-Assignment-Spec-V5.pdf.

Centers for Medicare & Medicaid Services (CMS), 2017c, Accountable Care Organization 2017 Quality Measure Narrative Specifications, https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/2017-Reporting-Year-Narrative-Specifications.pdf.

Centers for Medicare & Medicaid Services (CMS), 2018a, Medicare Program, Medicare Shared Savings Program, Accountable Care Organizations—Pathways to Success and Extreme and Uncontrollable Circumstances Policies for Performance Year 2017, https://www.govinfo.gov/content/pkg/FR-2018-12-31/pdf/2018-27981.pdf.

Centers for Medicare & Medicaid Services (CMS), 2018b, Accountable Care Organization (ACO) 2018 Quality Measures, https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/2018-reporting-year-narrative-specifications.pdf.

Champagne, N., J. McEwen, 2018, Population-based payments (PBPs) open door for ACOs to more effectively manage care, *Milliman Issue Brief*, http://www.milliman.com/uploadedFiles/insight/2018/PBPs-open-door-for-ACOs.pdf.

Charnes, A., W. Cooper, and E. Rhodes, 1978, Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research* 2: 429-444.

Chernew M., C. Barbey and J. McWilliams, 2017, Savings Reported by CMS Do Not Measure True ACO Savings, *Health Affairs Blog*, https://www.healthaffairs.org/do/10.1377/hblog20170619.060649/full/.

Cooper, W., L. Seiford, and K. Tone. 2007, Data Envelopment Analysis: A Comprehensive Text with Models, Applications, and References and DEA Solver Software, 2nd Ed., Kluwer Academic Publishers, Norwell, MA.

Cummins, J. and M. Weiss, 2011, Analyzing Firm Performance in the Insurance Industry Using Frontier Efficiency and Productivity Methods, available at SSRN: https://ssrn.com/abstract=1997468 or http://dx.doi.org/10.2139/ssrn.1997468).

Dobson, A., S. Pal, A. Hartzman, L. Arzaluz, K. Rhodes, and J. DaVanzo, 2018, Estimates of Savings by Medicare Shared Savings Program Accountable Care Organizations, *Dobson DaVanzo & Associates, LLC*, https://www.naacos.com/assets/docs/pdf/Study_of_MSSP_Savings_2012-2015.pdf.

Kaiser Family Foundation (KFF), 2018, Side-by-Side Comparison: Medicare Accountable Care Organization (ACO) Models, http://files.kff.org/attachment/Evidence-Link-Side-by-Side-ACOs.

L&M Policy Research, 2015, Pioneer ACO Evaluation Findings from Performance Years One and Two, Centers for Medicare & Medicaid Services (CMS), https://innovation.cms.gov/Files/reports/PioneerACOEvalRpt2.pdf.

L&M Policy Research, 2016, Pioneer ACO Final Report, Centers for Medicare & Medicaid Services (CMS), https://innovation.cms.gov/files/reports/pioneeraco-finalevalrpt.pdf.

McWilliams, M., 2016, Changes in Medicare Shared Savings Program Savings from 2013 to 2014, Journal of the American Medical Association 36(16): 1711-1713.

McWilliams, M., L. Hatfield, B. Landon, P. Hamed, and M. Chernew, 2018, Medicare Spending after 3 Years of the Medicare Shared Savings Program, New England Journal of Medicine 379(12): 1139-1149.

Mechanic, R., C. Gaus, 2018, Medicare Shared Savings Program Produces Substantial Savings: New Policies Should Promote ACO Growth, *Health Affairs Blog*, https://www.healthaffairs.org/do/10.1377/hblog20180906.711463/full/.

National Association of ACOs (NNACOS), 2018, NAACOS Overview of the 2018 Medicare ACO Class, https://www.naacos.com/assets/docs/pdf/Overivew2018MedicareACOCohortFinal043018.pdf.

National Committee for Quality Assurance (NCQA), 2018, NCQA Health Insurance Plan Ratings Methodology July 2018, https://www.ncqa.org/wpcontent/uploads/2018/09/201808013_Health_Plan_Ratings_Methodology.pdf.

NORC, 2018, Next Generation Accountable Care Organization (NGACO) Model Evaluation, Centers for Medicare & Medicaid Services (CMS), https://innovation.cms.gov/files/reports/nextgenaco-firstannrpt.pdf.

Verma S., 2018, Pathways to Success: A New Start for Medicare's Accountable Care Organizations, *Health Affairs Blog*, https://www.healthaffairs.org/do/10.1377/hblog20180809.12285/full/.

Yang, C., 2014, Health Care Reform, Efficiency of Health Insurers, and Optimal Health Insurance Markets, *North American Actuarial Journal* 18(4): 478-500.

Yang, C., 2018, The Impact of Medicaid Expansion, Diversity and the ACA Primary Care Fee Bump on the Performance of Medicaid Managed Care, *Journal of Insurance Regulation* 37(6): 1-34.

Yang, C., and H. Lin, 2017, The (Mis)alignment of Health Insurers' Efficiency Measures from Different Perspectives and Their (Un)linkage with Financial Ratios and Asset Allocation, *Journal of Insurance Regulation* 36(8): 1–21.

Yang, C., M. Wen. 2017, An Efficiency-Based Approach to Determining Potential Cost Savings and Profit Targets for Health Insurers: the Case of Obamacare Health Insurance CO-OPs, *North American Actuarial Journal* 21(2): 305-321.

Zhu J., 2009, Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets, 2nd Edition, Springer, New York.

Appendix I. Measures for use in establishing the quality performance standard that ACOs must meet for shared savings, 2018

Domain	ACO measure	Measure title
Patient/ caregiver experience	ACO-1	CAHPS*: Getting Timely care, Appointments, and Information
	ACO-2	CAHPS: How Well Your Providers Communicate
	ACO-3	CAHPS: Patients' Rating of Provider
	ACO-4	CAHPS: Access to Specialists
	ACO-5	CAHPS: Health Promotion and Education
	ACO-6	CAHPS: Shared Decision Making
	ACO-7	CAHPS: Health Status/Functional status
	ACO-34	CAHPS: Stewardship of Patient Resources
Care coordination/ patient safety	ACO-8	Risk-Standardized, All Condition Readmission
	ACO-35	Skilled Nursing Facility 30-Day Readmission (SNFRM)
	ACO-36	All-Cause Unplanned Admissions for Patients with Diabetes
	ACO-37	All-Cause Unplanned Admissions for Patients with Heart Failure
	ACO-38	All-Cause Unplanned Admissions for Patients with Multiple Chronic Conditions
	ACO-43	Ambulatory Sensitive Condition Acute Composite (AHRQ Prevention Quality Indicator (PQI) #91)
	ACO-11	Use of Certified EHR Technology
	ACO-12 (CARE-1)	Medication Reconciliation Post- Discharge
	ACO-13 (CARE-2)	Falls: Screening for Future Fall Risk
	ACO-44	Use of Imaging Studies for Low Back Pain
Preventive health	ACO-14 (PREV-7)	Preventive Care and Screening: Influenza Immunization
	ACO-15 (PREV-8)	Pneumonia Vaccination Status for Older Adults
	ACO-16 (PREV-9)	Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up
	ACO-17 (PREV-10)	Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention
	ACO-18 (PREV-12)	Preventive Care and Screening: Screening for Clinical Depression and Follow- up Plan
	ACO-19 (PREV-6)	Colorectal Cancer Screening
	ACO-20 (PREV-5)	Breast Cancer Screening
	ACO-42 (PREV-13)	Statin Therapy for the Prevention and Treatment of Cardiovascular Disease
Clinical care for at-risk population	ACO-40 (MH-1)	Depression Remission at Twelve Months
	ACO-27 (DM-2)	Diabetes Mellitus: Hemoglobin A1c Poor Control
	ACO-41 (DM-7)	Diabetes: Eye Exam
	ACO-28 (HTN-2)	Hypertension (HTN): Controlling High Blood Pressure
	ACO-30 (IVD-2)	Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet

Source: CMS (2018b).
*CAHPS = Consumer Assessment of Healthcare Providers and Systems.

Appendix II. Efficiency comparison between Track 1 and Track 2 of MSSP ACOs

Run No.	DEA model 1		DEA model 2		DEA model 3	
	Accepted hypothesis	p-value	Accepted hypothesis	p-value	Accepted hypothesis	p-value
1	AH_{1vs2}-1	0.096	AH_{1vs2}-2	0.541	AH_{1vs2}-1	0.033
2	AH_{1vs2}-2	0.378	AH_{1vs2}-2	0.977	AH_{1vs2}-2	0.478
3	AH_{1vs2}-2	0.370	AH_{1vs2}-2	0.744	AH_{1vs2}-1	0.022
4	AH_{1vs2}-1	0.022	AH_{1vs2}-2	0.551	AH_{1vs2}-1	0.000
5	AH_{1vs2}-2	0.514	AH_{1vs2}-1	0.003	AH_{1vs2}-2	0.799
6	AH_{1vs2}-3	0.057	AH_{1vs2}-3	0.006	AH_{1vs2}-2	0.319
7	AH_{1vs2}-2	0.551	AH_{1vs2}-2	0.291	AH_{1vs2}-1	0.099
8	AH_{1vs2}-3	0.039	AH_{1vs2}-3	0.022	AH_{1vs2}-3	0.080
9	AH_{1vs2}-1	0.000	AH_{1vs2}-1	0.000	AH_{1vs2}-1	0.000
10	AH_{1vs2}-1	0.001	AH_{1vs2}-1	0.017	AH_{1vs2}-1	0.000
11	AH_{1vs2}-2	0.755	AH_{1vs2}-3	0.089	AH_{1vs2}-2	0.291
12	AH_{1vs2}-2	0.378	AH_{1vs2}-3	0.057	AH_{1vs2}-2	0.478
13	AH_{1vs2}-2	0.755	AH_{1vs2}-2	0.347	AH_{1vs2}-1	0.014
14	AH_{1vs2}-1	0.001	AH_{1vs2}-1	0.014	AH_{1vs2}-1	0.000
15	AH_{1vs2}-1	0.000	AH_{1vs2}-1	0.000	AH_{1vs2}-1	0.000
16	AH_{1vs2}-3	0.055	AH_{1vs2}-3	0.002	AH_{1vs2}-2	0.619
17	AH_{1vs2}-3	0.026	AH_{1vs2}-3	0.003	AH_{1vs2}-2	0.887
18	AH_{1vs2}-2	0.347	AH_{1vs2}-2	0.671	AH_{1vs2}-2	0.551
19	AH_{1vs2}-2	0.590	AH_{1vs2}-2	0.347	AH_{1vs2}-1	0.039
20	AH_{1vs2}-2	0.713	AH_{1vs2}-2	0.887	AH_{1vs2}-2	0.378
21	AH_{1vs2}-2	0.630	AH_{1vs2}-2	0.630	AH_{1vs2}-2	0.551
22	AH_{1vs2}-1	0.064	AH_{1vs2}-2	0.843	AH_{1vs2}-1	0.001
23	AH_{1vs2}-2	0.347	AH_{1vs2}-2	0.590	AH_{1vs2}-1	0.002
24	AH_{1vs2}-1	0.000	AH_{1vs2}-1	0.001	AH_{1vs2}-1	0.000
25	AH_{1vs2}-1	0.089	AH_{1vs2}-2	0.887	AH_{1vs2}-1	0.072
26	AH_{1vs2}-1	0.000	AH_{1vs2}-1	0.000	AH_{1vs2}-1	0.000
27	AH_{1vs2}-1	0.022	AH_{1vs2}-2	0.410	AH_{1vs2}-1	0.000
28	AH_{1vs2}-1	0.026	AH_{1vs2}-2	0.378	AH_{1vs2}-1	0.007
29	AH_{1vs2}-3	0.072	AH_{1vs2}-3	0.019	AH_{1vs2}-2	0.932
30	AH_{1vs2}-2	0.629	AH_{1vs2}-2	0.977	AH_{1vs2}-2	0.712

DEA model 1: inputs – health expenditures; outputs – enrollment, utilization, and quality.
DEA model 2: inputs – health expenditures; outputs – enrollment and quality.
DEA model 3: inputs – health expenditures; outputs – enrollment and utilization.
AH_{1vs2}-1: Track 1 is more efficient than Track 2.
AH_{1vs2}-2: Track 1 is equally efficient as Track 2.
AH_{1vs2}-3: Track 1 is less efficient than Track 2.
AH_{1vs2}-1 and AH_{1vs2}-3 are the alternative hypothesis, and AH_{1vs2}-2 is the null hypothesis.
The significance level is 10%.

Appendix III. Efficiency comparison between Track 2 and Track 3 of MSSP ACOs

Run No.	DEA model 1		DEA model 2		DEA model 3	
	Accepted hypothesis	p-value	Accepted hypothesis	p-value	Accepted hypothesis	p-value
1	AH_{2vs3}-2	0.920	AH_{2vs3}-1	0.049	AH_{2vs3}-3	0.025
2	AH_{2vs3}-3	0.070	AH_{2vs3}-2	0.788	AH_{2vs3}-3	0.000
3	AH_{2vs3}-2	0.378	AH_{2vs3}-1	0.002	AH_{2vs3}-2	0.887
4	AH_{2vs3}-1	0.001	AH_{2vs3}-1	0.000	AH_{2vs3}-1	0.019
5	AH_{2vs3}-2	0.443	AH_{2vs3}-2	0.843	AH_{2vs3}-3	0.006
6	AH_{2vs3}-1	0.000	AH_{2vs3}-1	0.000	AH_{2vs3}-1	0.003
7	AH_{2vs3}-2	0.421	AH_{2vs3}-1	0.019	AH_{2vs3}-2	0.891
8	AH_{2vs3}-1	0.087	AH_{2vs3}-1	0.001	AH_{2vs3}-2	0.469
9	AH_{2vs3}-3	0.012	AH_{2vs3}-2	0.340	AH_{2vs3}-3	0.000
10	AH_{2vs3}-3	0.062	AH_{2vs3}-2	0.659	AH_{2vs3}-3	0.016
11	AH_{2vs3}-2	0.541	AH_{2vs3}-1	0.014	AH_{2vs3}-2	0.579
12	AH_{2vs3}-1	0.005	AH_{2vs3}-1	0.002	AH_{2vs3}-1	0.010
13	AH_{2vs3}-2	0.504	AH_{2vs3}-2	0.659	AH_{2vs3}-2	0.504
14	AH_{2vs3}-2	0.377	AH_{2vs3}-2	0.590	AH_{2vs3}-3	0.012
15	AH_{2vs3}-1	0.008	AH_{2vs3}-1	0.005	AH_{2vs3}-2	0.434
16	AH_{2vs3}-2	0.477	AH_{2vs3}-1	0.057	AH_{2vs3}-2	0.551
17	AH_{2vs3}-2	0.876	AH_{2vs3}-1	0.007	AH_{2vs3}-2	0.340
18	AH_{2vs3}-2	0.843	AH_{2vs3}-2	0.514	AH_{2vs3}-2	0.410
19	AH_{2vs3}-2	0.319	AH_{2vs3}-1	0.004	AH_{2vs3}-2	0.887
20	AH_{2vs3}-2	0.434	AH_{2vs3}-1	0.016	AH_{2vs3}-2	0.312
21	AH_{2vs3}-2	0.340	AH_{2vs3}-2	0.619	AH_{2vs3}-2	0.369
22	AH_{2vs3}-2	0.370	AH_{2vs3}-1	0.043	AH_{2vs3}-3	0.038
23	AH_{2vs3}-2	0.213	AH_{2vs3}-1	0.010	AH_{2vs3}-2	0.620
24	AH_{2vs3}-3	0.017	AH_{2vs3}-2	0.932	AH_{2vs3}-3	0.001
25	AH_{2vs3}-1	0.078	AH_{2vs3}-1	0.003	AH_{2vs3}-2	0.702
26	AH_{2vs3}-1	0.038	AH_{2vs3}-1	0.004	AH_{2vs3}-3	0.012
27	AH_{2vs3}-2	0.417	AH_{2vs3}-1	0.030	AH_{2vs3}-2	0.887
28	AH_{2vs3}-2	0.235	AH_{2vs3}-2	0.235	AH_{2vs3}-3	0.000
29	AH_{2vs3}-2	0.311	AH_{2vs3}-1	0.014	AH_{2vs3}-2	0.876
30	AH_{2vs3}-1	0.000	AH_{2vs3}-1	0.000	AH_{2vs3}-1	0.086

DEA model 1: inputs – health expenditures; outputs – enrollment, utilization, and quality.
DEA model 2: inputs – health expenditures; outputs – enrollment and quality.
DEA model 3: inputs – health expenditures; outputs – enrollment and utilization.
AH_{2vs3}-1: Track 2 is more efficient than Track 3.
AH_{2vs3}-2: Track 2 is equally efficient as Track 3.
AH_{2vs3}-3: Track 2 is less efficient than Track 3.
AH_{2vs3}-1 and AH_{2vs3}-3 are the alternative hypothesis, and AH_{2vs3}-2 is the null hypothesis.
The significance level is 10%.