

NBER WORKING PAPER SERIES

SPENDING REDUCTIONS IN THE MEDICARE SHARED SAVINGS PROGRAM:  
SELECTION OR SAVINGS?

J. Michael McWilliams  
Laura A. Hatfield  
Bruce E. Landon  
Michael E. Chernew

Working Paper 26403  
<http://www.nber.org/papers/w26403>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
October 2019

From the Department of Health Care Policy, Harvard Medical School (JMM, MEC, BEL, LAH, PH); Division of General Internal Medicine and Primary Care, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School (JMM); and Division of General Internal Medicine and Primary Care, Department of Medicine, Beth Israel Deaconess Medical Center (BEL), all in Boston, MA. Supported by grants from the National Institute on Aging of the National Institutes of Health (P01AG032952) and Arnold Ventures. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or Arnold Ventures. The authors thank Pasha Hamed, MA for statistical programming support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w26403.ack>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by J. Michael McWilliams, Laura A. Hatfield, Bruce E. Landon, and Michael E. Chernew. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Spending Reductions in the Medicare Shared Savings Program: Selection or Savings?  
J. Michael McWilliams, Laura A. Hatfield, Bruce E. Landon, and Michael E. Chernew  
NBER Working Paper No. 26403  
October 2019  
JEL No. I11,I13

### **ABSTRACT**

Evidence of patient and physician turnover in accountable care organizations (ACOs) has raised concerns that ACOs may be earning shared-savings bonuses by selecting for lower-risk patients or providers with lower-risk panels. We conducted three sets of analyses to examine risk selection in the Medicare Shared Savings Program. First, we estimated overall MSSP savings through 2015 using a difference-in-differences approach and methods that eliminated selection bias from ACO program exit or changes in the practices or physicians included in ACO contracts. We then checked for residual risk selection at the patient level. Second, we re-estimated savings with methods that address undetected risk selection but could introduce bias from other sources. These included patient fixed effects, baseline assignment, and area-level MSSP exposure to hold patient populations constant. Third, we tested for changes in provider composition or provider billing that may have contributed to bonuses, even if they were eliminated as sources of bias in the evaluation analyses. We find that MSSP participation was associated with modest and increasing annual gross savings in the 2012-2013 entry cohorts of ACOs that reached \$139-302/patient by 2015. Savings in the 2014 entry cohort were small and not statistically significant. Robustness checks revealed no evidence of residual risk selection. Alternative methods to address risk selection produced consistent results but were less robust than our primary analysis, suggesting the introduction of bias from within-patient changes in time-varying characteristics. We find no evidence of ACO manipulation of provider composition or billing to inflate savings. We further demonstrate that exit of high-risk patients or physicians with high-risk patients from ACOs is misleading without considering a counterfactual among non-ACO practices. We conclude that participation in the original MSSP program was associated with modest savings and not with favorable risk selection. These findings suggest an opportunity to build on early progress. Understanding the effect of new incentives and opportunities for risk selection in the revamped MSSP will be important for guiding future program reforms.

J. Michael McWilliams  
Harvard Medical School  
Department of Health Care Policy  
180 Longwood Avenue  
Boston, MA 02115  
mcwilliams@hcp.med.harvard.edu

Laura A. Hatfield  
Harvard Medical School  
hatfield@hcp.med.harvard.edu

Bruce E. Landon  
Harvard Medical School  
landon@hcp.med.harvard.edu

Michael E. Chernew  
Harvard Medical School Dept.  
of Health Care Policy 180  
Longwood Avenue Boston,  
MA 02115  
and NBER  
chernew@hcp.med.harvard.edu

## INTRODUCTION

In the voluntary Medicare Shared Savings Program (MSSP), participating accountable care organizations (ACOs) have incentives to reduce total Medicare spending for their attributed patient populations. Specifically, an ACO is eligible for a shared-savings bonus if total per-beneficiary Medicare spending is sufficiently below its spending target, or benchmark, and if its performance on a set of quality measures meets minimum standards. In Track 1 of the MSSP—the track in which almost all ACOs (99%) participated in for the first 4 years of the program<sup>1</sup>—the shared-savings bonus was 50% of the difference between an ACO's spending and its benchmark, less a small percentage for submaximal quality scores. The benchmark in a given performance year was based on an ACO's average historical spending before MSSP entry for analogously attributed patients, updated to the performance year based on average national Medicare spending growth.

Existing evidence indicates that participation in the MSSP produced modest reductions in Medicare spending for ACO patients,<sup>2-6</sup> even after accounting for bonus payments. However, churn in the patients attributed to ACOs and in the providers included in ACO contracts has raised concerns that some of the savings might be an artifact of risk selection—i.e., the result of ACOs encouraging high-cost patients to switch to a non-ACO provider or excluding clinicians with high-cost patients from ACO contracts. ACOs do have some incentives to engage in favorable risk selection, but the incentives were limited by several features of the MSSP's original design. Moreover, risk selection can be assessed and addressed by evaluation methods. Thus, evidence of risk selection does not imply that estimates of savings are biased; if there is favorable risk selection, net savings can still be determined by estimating gross savings using methods that address risk selection and subtracting the bonus payments (which include any

unearned gains from risk selection).<sup>7</sup> Below, we describe program features that shape incentives for risk selection in the MSSP before turning to the empirical contributions of this paper.

## **Incentives for Risk Selection in the MSSP**

### ***TIN Level***

The MSSP defines ACOs as collections of taxpayer identification numbers (TINs) identifying practices and Centers for Medicare & Medicaid Services (CMS) Certification Numbers (CCNs) identifying certain safety-net facilities (to simplify, we refer to both TINs and CCNs as TINs). An ACO's yearly attributed population includes all patients who receive more qualifying services from that ACO's TINs than any other ACO or non-ACO TIN. The attribution process assigns patients based on care received from primary care physicians (PCPs; defined by four specialties: internal medicine, family medicine, general practice, or geriatric medicine) as long as patients receive at least one qualifying service from a PCP. Since 88% of patients with a qualifying service have at least one visit with a PCP, attribution is largely based on the PCPs from which patients receive care. Patients without a qualifying service with a PCP are assigned based on receipt of the same services from non-PCP clinicians. Each year, ACOs in the MSSP can change the TINs in its contract but, unlike ACOs in the Pioneer model, cannot select among clinicians within TINs (i.e., all clinicians billing under an included TIN are in the contract).

Until 2017, the MSSP accounted for changes in TIN inclusion each year by adjusting benchmarks to reflect the baseline spending of the revised set of TINs. Thus, ACOs did not have clear incentives to favor lower-spending TINs because the reduced spending would be offset by reduced benchmarks. Excluding higher-spending TINs might improve performance on utilization-based quality measures such as readmission rates, thereby increasing their shared-savings rate (the percentage of savings they could keep if they qualified for a bonus), but this

would not artificially inflate the gross savings. Because the variance of medical spending is greater when spending is higher, an ACO with downside risk for spending in excess of benchmarks (i.e., in a two-sided contract) might avoid TINs with higher spending to minimize the probability of a large loss from random fluctuations in spending. Through 2015, however, almost all ACOs did not assume downside risk, and those that did had less risk for losses than savings.

Consequently, TIN-level selection incentives were minimal in the original MSSP and may actually have favored higher-spending TINs for two reasons. First, greater variability in spending at higher levels may present opportunities in one-sided contracts for larger bonuses (due to random fluctuations). Second, ACOs with higher spending might generate savings more easily, because the costs of lowering wasteful spending are likely lower when there is more wasteful spending to cut (i.e., more fat to trim). Indeed, ACOs with higher baseline spending have reduced spending more than other ACOs, on average.<sup>4</sup>

In 2017, the MSSP began to blend ACO benchmarks with average regional spending after 3 years of participation. The implementation of this regional blending will be accelerated by the recent overhaul of the MSSP, “Pathways for Success,” which also requires ACOs to assume more downside risk sooner after MSSP entry.<sup>8</sup> These changes create new incentives for ACOs in the MSSP to select TINs with spending below their regional average,<sup>9</sup> but these incentives were not in place during the period examined by MSSP evaluations to date.

### ***Clinician Level***

Given a set of TINs and the attendant benchmark, an MSSP ACO has incentives to exclude clinicians with spending in excess of that predicted by the Hierarchical Condition Categories (HCC) model used to adjust for case mix.<sup>10</sup> To selectively exclude such physicians

and their patients, an ACO would have to identify them and arrange for them to bill under a different TIN that is not included in the ACO's contract (because patient attribution is determined by the billing TINs, not clinicians). An ACO also could terminate clinicians' employment or move them to a different practice, though these behaviors seem implausible and might provoke legal action. The ability to alter the attributed patient population by changing the TIN under which clinicians bill does create some opportunities for risk selection, and there is some anecdotal evidence of ACOs exploiting this mechanism.<sup>11</sup> In particular, one strategy relates to the inclusion of encounters in post-acute or long-term nursing facilities among the qualifying services used by CMS to assign patients to ACOs. Consider an internist or geriatrician who sees patients in the office and rounds on patients in a skilled nursing facility (SNF), billing both types of services under the same TIN. After entry into the MSSP, the physician's organization could arrange for the SNF encounters to be billed under a separate excluded TIN, thereby causing patients who become acutely ill and receive more post-acute facility care than outpatient primary care to be assigned away from the ACO to the excluded TIN in a performance year but not in the baseline period used to calculate a benchmark. The resulting spending reduction would not be corrected by a benchmark reduction since the ACO did not change its constituent TINs.

To the extent that clinician-level selection results in lower risk-adjusted spending by favoring clinicians with patients who are lower risk, it should manifest as a change in case mix in the attributed population, assuming that unobservable patient factors not included in the risk adjustment are correlated with the observable factors that are included. It is possible that ACOs could select clinicians based on their efficiency (care patterns), independent of patient risk. This may not manifest as a change in case mix and therefore may not be testable. It is unlikely,

however, that ACOs possess the data and analytic capabilities to isolate physicians' efficiency from the case mix of their patients.

Selecting clinicians based on their efficiency has ambiguous normative implications. While selecting clinicians based on their patients' risk may be seen as wasteful gaming, selecting clinicians based on their efficiency could foster competition among PCPs to be more efficient as ACO programs expand and exert pressure on PCPs to participate. The associated spending reductions might offset any bonus payments to ACOs engaging in such selection.

### *Patient Level*

To mitigate ACO incentives to increase savings artificially through risk selection or up-coding, the MSSP uses an HCC-base risk adjustment model and only applies downward adjustments to benchmarks (if risk scores decrease). Thus, ACOs have incentives to avoid patients with spending in excess of what the HCC model predicts and to attract patients with below-predicted spending. Unlike Medicare Advantage (MA) plans, however, ACOs have no control over benefit or network design and therefore have fewer means to select favorable risks. In the absence of the provider-level selection strategies described above, an ACO would somehow have to induce high-cost patients in its attributed population to leave the ACO, for example by dropping them from the practice, successfully referring them to a different PCP, or otherwise limiting their access to the ACO (e.g., by capping appointments for high-risk patients).

In addition to supply-side selection efforts by ACOs, high-risk patients may exhibit stronger or weaker demand for care in ACOs. Many ACOs target high-risk patients for enhanced care management, and prior research has found that ACO efforts have been associated with improved overall care ratings among high-risk patients.<sup>12</sup> Thus, the tailored care ACOs offer may attract, rather than repel, high-risk patients.

Attracting low-risk patients may be more feasible than denying care to high-risk patients. For example, ACOs could reach out to healthy patients without qualifying services and schedule visits for them (e.g., annual wellness visits), thereby increasing the number of attributed patients with below-predicted spending. The proportion of Medicare beneficiaries without a qualifying service is low (8.5%), however, and they include a subgroup of very high-risk patients (e.g., those enrolled in hospice). Imperfect targeting of such efforts would also contribute to higher spending from the additional office visits for patients whose attribution is not altered.

### **Empirical Contributions of this Paper**

In this paper, we use evaluation methods to estimate savings in the MSSP that address selection bias, gauge the potential for residual selection, and test for risk selection that may have contributed to shared-savings bonuses but not to bias in our evaluation. First, we report new estimates of overall savings through 2015 from a difference-in-differences analysis that addresses bias from provider-level selection using an intention-to-treat approach.<sup>5</sup> Specifically, we hold ACOs' providers constant over time. These estimates reflect the combined results of earlier work that compared savings between physician-group and hospital-based ACOs. As in those stratified analyses,<sup>5</sup> we find no evidence of patient risk selection and estimate overall gross savings in excess of bonus payments.

Second, we implement alternative approaches to eliminate residual risk selection that may have gone undetected by tests of observable patient characteristics in our evaluation (summarized in Table 1). These include use of patient fixed effects in difference-in-differences models, an intention-to-treat analysis holding patients' baseline assignments to providers (prior to the start of MSSP incentives) constant, and an area-level analysis defining MSSP exposure based on program penetration in patients' hospital referral region (HRR). We provide empirical



evidence that these approaches—while ensuring no bias from selection of patients with fixed characteristics predictive of lower spending—introduce other sources of bias. Nevertheless, these approaches produce results that are generally consistent with our main findings. We also consider an alternative patient attribution approach (using data on referring PCPs) to address potential selection bias from ACO efforts to use annual wellness visits to boost attribution of healthy patients who might otherwise be unattributed. This approach increases savings estimates.

Third, we test for risk selection that may have been successfully eliminated in our evaluation but would have contributed to bonus payments (Table 2). We do so by allowing the provider composition of ACOs to change over time, as it did, and repeating our evaluation analysis. We also analyze patterns of patient and physician exit from ACOs over 3 performance years. We find that ACOs did not systematically favor providers with lower-risk patients or lower spending as they changed their provider composition. We also demonstrate that analyses of patient or physician exit can be misleading without considering the counterfactual—churn in the absence of MSSP incentives. Finally, we test for ACO manipulation of the TINs used by physicians for billing to achieve a lower-cost attributed population during the performance period. We find no evidence of this behavior.

## **DATA AND METHODS**

### **Evaluation of the MSSP through 2015**

Using Medicare claims for 20% random annual samples of fee-for-service Medicare beneficiaries from 2009-2015, we conducted a difference-in-differences analysis comparing beneficiaries attributed to providers that entered the MSSP in 2012, 2013, or 2014 with local beneficiaries attributed to non-participating providers (control group), before and after program entry by participating providers. In each year, we attributed beneficiaries to the ACO or non-

ACO TIN that accounted for the plurality of allowed charges for their office visits with PCPs (Current Procedural Terminology [CPT] codes 99201-15, 99241-45, G0402, and G0438-39 in Carrier claims and corresponding codes in Outpatient claims for specific safety net settings).<sup>13</sup> In the pre-entry period, attribution to an ACO meant attribution to a group of providers who would subsequently enter the MSSP. We limited the qualifying services used for assignment to office visits with PCPs to achieve comparability between the ACO and control group. Use of all qualifying services in the CMS assignment rules, which include evaluation and management services from outpatient specialists and physicians in nursing facilities, introduces systematic differences between ACO-attributed patients and the control group.<sup>5</sup> This occurs because many ACOs do not provide specialty care or post-acute or long-term care in nursing facilities.<sup>4,14,15</sup> Consequently, beneficiaries using nursing facility care or only specialty care would be disproportionately assigned to the control group. When comparison groups in a difference-in-differences analysis systematically differ, a stronger common shocks assumption is required; in this context, drivers of spending growth other than the MSSP would more likely affect the attributed populations of ACOs and non-ACO providers differently if the populations differ.

Our modifications to beneficiary assignment also minimized bias from gaming strategies that involve changes in the TINs used for billing (described above). For example, a patient that requires more post-acute care than primary care would be assigned by the CMS algorithm to the TIN billing for the post-acute care but would remain assigned to the ACO or non-ACO TIN providing the primary care in our analysis.

As expected from the dominant role of primary care in the CMS attribution algorithm, our assignments and CMS assignments overlapped substantially. For example, 89% of beneficiaries we attributed in 2013 to ACOs entering the MSSP in 2012-2013 were found in the

2013 MSSP Beneficiary-level attribution file;<sup>16</sup> of those, the assigned ACO matched in over 99% of cases. Of beneficiaries in the 2013 MSSP Beneficiary-level attribution file, we attributed 84% to ACOs; of those, the assigned ACO matched in 96% of cases.

After assigning beneficiaries to providers, we fit the following linear regression model:

$$Y_{itkh} = \beta_0 + \beta_1 ACO_{itk} + \beta_2 HRR\_Year_{ith} + \beta_3 ACOcohort\_Post_{itk} + \beta_4 Covariates_{it} + \varepsilon_{itkh}$$

where  $Y$  is the annual Medicare spending for beneficiary  $i$  in year  $t$  attributed to ACO  $k$  or a non-ACO TIN and residing in HRR  $h$ ;  $ACO$  is a vector of indicators for each ACO with the non-ACO control group as the omitted reference group;  $HRR\_Year$  is a vector of indicators for each HRR-year combination;  $ACOcohort\_Post$  is a vector of indicators of attribution to a specific entry cohort of ACOs (2012, 2013, or 2014 cohort) in a specific post-entry year;  $Covariates$  is a vector of the patient characteristics listed in Table 4; and  $\beta_1$ - $\beta_4$  are vectors of coefficients corresponding to each term. The ACO fixed effects adjust for pre-entry differences between ACOs and the control group and for any changes in the distribution of ACO-attributed beneficiaries across ACOs. The HRR-year fixed effects adjust for geographic differences between ACOs and the control group and for regional changes in spending for the control group. Thus, the estimated effect of MSSP participation ( $\beta_3$ ) is the difference between spending for ACO-attributed patients in a post-entry year and spending that would be expected for ACO patients if the change from the pre-entry period to that year was equal to the change observed for patients in the same HRR served by non-ACO providers (the differential change in spending for ACO patients, or *gross savings*). We used a robust variance estimator, specifying clusters as ACOs (for ACO-attributed beneficiaries) or HRRs (for the control group). Specifying HRRs as the clusters for all beneficiaries yielded similar results. Additional details of the methods have been published elsewhere,<sup>5</sup> including exclusion of patients attributed to Pioneer ACOs.

To eliminate bias from selective dropout of ACOs by 2015, our analysis followed an intention-to-treat approach in which we retained all ACOs through 2015 regardless of participation status. To eliminate bias arising from compositional changes in the TINs or physicians composing ACOs, we held constant from 2009-2015 the definition of ACOs as collections of TINs or physician National Provider Identifiers (NPIs), in the latter case modifying attribution to assign patients to a group of ACO NPIs or a non-ACO NPI. We conducted additional analyses to assess potential violations of the identifying assumption of our difference-in-differences analysis (that the ACO-control group difference would have remained constant in the absence of MSSP participation). We estimated differential changes in patient characteristics and compared savings estimates with and without adjustment for fixed and time-varying patient covariates. For time-varying covariates potentially affected by the MSSP (e.g., HCC scores via upcoding), we checked the sensitivity of results to adjusting for values derived from data several years prior to a given study year, as opposed to the year prior. In addition to regression adjustments, we also implemented a propensity score weighting technique to balance covariates between ACOs and the control group in each year.<sup>5,17</sup>

We estimated differences in pre-entry trends between ACOs and the control group and conducted falsification tests treating pre-entry years as hypothetical entry years. We also conducted falsification tests treating both non-ACO TINs that were large enough after the start of the MSSP to participate (an expected 5000+ assigned beneficiaries in the full Medicare population) and the 2015 MSSP entrants (which we did not analyze in the main analyses) as hypothetical entrants in various years. In addition to testing whether large provider organizations or groups that eventually joined the MSSP had slower spending growth when *not* participating, these falsification tests also explored whether our intention-to-treat approach, which categorized

TINs by their ACO status at the outset of MSSP entry and held the treatment group of TINs constant, could produce differential reductions in spending in the absence of the MSSP.

## **Approaches to Assess and Address Residual Risk Selection**

### ***1. Patient Fixed Effects***

#### *Conceptual considerations*

One approach to ensure that differential changes in population composition do not contribute to savings estimated by difference-in-differences analysis is to use patient fixed effects in the model to control for all characteristics of patients that are fixed. Replacing ACO fixed effects with patient fixed effects in the model above produces a difference-in-difference estimate based on within-patient changes. Specifically, for a given performance year, the estimate becomes the mean within-patient difference between spending for a patient attributed to an ACO in the performance year and spending in years when the patient is not attributed to an ACO in a performance year, minus the concurrent within-patient spending difference for patients not attributed to an ACO in the performance year.

There are two major drawbacks to this approach. First, basing the estimation on within-patient changes converts the analysis from one of annual cross-sectional samples, each representative of the fee-for-service Medicare population, to a longitudinal cohort of patients who were alive, enrolled in fee-for-service Medicare, and eligible for attribution in both the pre- and post-entry periods. As illustrated in Figure 1, the spending trends over the study period for these two samples differ dramatically. Adjusted annual Medicare spending of the serial cross-sectional samples analyzed in our main evaluation approach increased by \$374/patient (4.2%) from 2009-2015, demonstrating that, despite substantial turnover in the sample over time, spending growth reflected modest secular trends. In contrast, adjusted spending increased by

\$1740/patient (26.4%) for a longitudinal cohort of continuously enrolled and attributable patients that would serve as the basis for estimation of savings in a model with patient fixed effects. The spending increase is most rapid at the end of the study period. This reflects the fact that patients must remain alive from the pre-period to 2015 to contribute to estimation of 2015 savings in a model with patient fixed effects, but they may then die or enter a long-term care facility, for example, and no longer be alive or attributable on the basis of outpatient primary care in 2016. Thus, the cohort becomes more acutely ill (in ways not accounted for by the adjustments) as they near the end of the cohort inclusion period, unlike patients in consistently defined annual cross-sectional samples. Figure 1 demonstrates how the rapid increase in spending occurs near the end of the inclusion period, no matter when that inclusion period ends.

Thus, an analysis with patient fixed effects requires the strong assumption that within-patient spending changes would be the same for ACO and non-ACO patients in the absence of the MSSP. This implies both similar health declines and similar treatment of patients with declining health by ACO and non-ACO providers, yet the rapid acceleration in spending for the longitudinal cohorts in Figure 1 is likely to vary across different patient populations and providers. In contrast, an analysis of serial cross-sectional samples allows patients to die or experience health declines consistently across time. The analysis can therefore net out differences in health care needs or treatment patterns between severely ill ACO and non-ACO patients (because severely ill patients are consistently present in the pre- and post-period).

Second, because patient attribution to ACOs is time-varying, the difference-in-difference estimator in a model with patient fixed effects reflects not only pre- to post-period changes in spending associated with a patient's provider *entering* the MSSP but also changes in spending associated with changes in patient attribution from a non-ACO to ACO provider, or vice-versa,

*during the post-period.* If patients are assigned to ACOs vs. non-ACO providers based on their time-varying health care needs, the estimates from a model with patient fixed effects would be biased because differences in spending caused by endogenous assignment to ACO or non-ACO providers would not be differenced out, as they would be in a model without patient fixed effects. This second source of bias may interact with the first (e.g., if sorting based on time-varying needs is prominent among patients experiencing severe health declines).

To ameliorate the bias due to shifts in attribution between ACO and non-ACO providers in the post-period, ACO attribution in the post-period could be treated as an absorbing state (turned on indefinitely after the first post-period year of ACO attribution). However, this would not remove bias from endogenous sorting into ACOs in an initial post-period year and would tend to bias estimates away from savings because attribution of high-risk patients is less stable (as described below); thus, treating ACO attribution as an absorbing state would selectively retain high-risk patients in the ACO group selectively in the post-period.

Recognizing these conceptual concerns, results from a model with patient fixed effects must be interpreted with caution. Although patient fixed effects eliminate bias from differential compositional changes in the fixed characteristics of patients exposed to the MSSP, their deployment can exacerbate bias from differential changes in time-varying characteristics within patients, effectively reversing the bias corrections achieved by a difference-in-difference comparison of serial cross-sectional samples that are stably different (at the population level) in their fixed and time-varying characteristics.

### *Empirical Analysis*

To understand the impact of using patient fixed effects, first we limited our study sample to a longitudinal cohort of continuously attributable patients that supports estimation of a

difference-in-difference from within-patient changes and re-estimated our main difference-in-difference model. The resulting estimate might differ from our main estimates for several reasons, including the concerns described above and the much lower mean spending for this cohort (Figure 1). Second, we substituted patient fixed effects for the ACO fixed effects in the model to isolate the incremental effect of holding the patient constant. Third, to gauge selection bias introduced by this approach, we compared estimates with and without adjustment for patients' time-varying characteristics.

We did not estimate a model with PCP fixed effects because we hold the PCPs in each ACO constant in our main approach to eliminate bias from changes in ACO PCP composition. Using physician fixed effects instead could introduce bias if ACOs shift high-risk patients to more cost-effective clinicians, for example, and we do not wish to remove the effects of such strategic shifting from our evaluation of savings.

## ***2. Holding Baseline Assignments Fixed***

### *Conceptual Considerations*

Another approach to eliminating bias from risk selection is to hold patients' attribution to providers at baseline constant. This type of intention-to-treat approach was implemented by the Medicare Payment Advisory Commission, for example.<sup>18</sup> In addition to removing the contribution of differential changes in patient attribution from the difference-in-differences estimate (by disallowing changes in attribution), this approach also does not require utilization of qualifying services to categorize patients into ACO and non-ACO groups after the initial year. This latter advantage may address bias from differential changes in the attributed patient population caused by provision of qualifying services (such as annual wellness visits). More generally, ACO effects on primary care use and patient attribution are endogenous, though in



prior work, we found no evidence of differential changes in ACO provision of PCP office visits that would substantiate this concern.<sup>5</sup>

Like the use of patient fixed effects, however, use of baseline patient assignments can introduce other sources of bias. If ACO and non-ACO providers differ in their reimbursement rates or practice patterns in the absence of MSSP exposure, or if patient attribution to ACO vs. non-ACO providers (in the absence of MSSP exposure) is influenced by their time-varying health needs, we should expect spending differences between groups of patients defined by their baseline assignments to change over time, even if the MSSP has no effect on spending. In the framework of an instrumental variable analysis, the exclusion restriction is unlikely to hold when using baseline assignment as an instrument for MSSP exposure in the post period. That is, baseline assignment to ACO vs non-ACO providers likely predicts changes in spending that are not solely reflective of greater exposure to the MSSP.

The bias arises because a constraint is applied asymmetrically in time. It is therefore similar to the problem noted above of requiring a cohort to be alive and continuously enrolled for some period and also to the problem of regression to the mean when matching on time-varying variables.<sup>19,20</sup> For example, outpatient Medicare spending for patients of independent physician groups is likely to be lower than for other patients, on average, because they are likely to receive less outpatient care at more generously reimbursed hospital-owned facilities. Consequently, spending for patients initially attributed to independent primary care groups is likely to increase over time relative to a local control group served by a mix of PCPs in independent and hospital-based practices. As patients switch practices, the proportion of patients attributed to hospital-based practices can only increase among those initially attributed to independent groups, whereas switching would be bidirectional in the control group, leading to a smaller net shift to hospital-

based practices. Thus, an evaluation holding baseline assignments constant would tend to underestimate savings by independent physician group ACOs, all else equal. More generally, use of baseline assignments could bias overall MSSP savings estimates if the mix of independent and hospital-based practices participating in the MSSP differs from the surrounding delivery system.

Similarly, practice patterns might differ systematically between ACOs and non-ACO providers. The substantial patient churn in provider patient populations<sup>21-23</sup> could therefore introduce bias in an evaluation using time-invariant baseline assignments to define comparison groups that would not be present in an evaluation using time-varying assignments.

In addition, changes in health care needs may cause changes in attribution of patients to ACO or non-ACO providers, whether because of true change in providers or the attribution algorithm. If patients are disproportionately assigned to ACOs when they become ill and to non-ACO providers when they are healthy, or vice-versa, one would expect differences in spending between patients initially assigned to ACOs and non-ACO providers to converge as their health status reverts to the population mean. Use of the CMS attribution algorithm could exacerbate this source of bias. Its inclusion of services in post-acute facilities would cause acutely ill patients to be disproportionately assigned away from ACOs at baseline,<sup>24</sup> inducing a subsequent differential increase in spending for patients assigned to ACOs at baseline as the control group's acute care needs subside and the ACO group's needs emerge.

We do not attempt to assess or address these sources of potential bias introduced by using baseline assignments. Rather, we note that the bias is difficult to predict and could be substantial, interpret savings estimates produced by this approach with caution, and conduct falsification tests to determine whether this approach might estimate an erroneous differential change in spending in the absence of MSSP participation.

### *Empirical Analysis*

First, we assigned patients to ACOs or non-ACO TINs in 2009 based on office visits with PCPs. We then fit the model above, limiting the sample to beneficiaries with a 2009 assignment, replacing the time-varying indicators for the ACO or cohort to which a patient is assigned with time-invariant 2009 assignments. We dropped the 2009 data from our analysis to minimize bias from regression to the mean that would arise because we require a qualifying service in 2009 but not after that.

Assuming absence of the biases described above, the differential change in spending for patients assigned to ACOs at baseline estimated by this reduced form model is interpretable as attributable to the MSSP. Because only 66.6% of patients assigned to an ACO in 2009 were assigned to an ACO in 2015 (among those eligible for assignment in both years), we inflate the differential change estimate to recover the MSSP effect as if all patients assigned at baseline to ACOs and none assigned to non-ACO providers were exposed to the MSSP in the performance years. To do so, we estimated the difference in the probability of being assigned to an ACO in a performance year between patients assigned to ACOs and non-ACO providers at baseline, among those with an assignment in 2009 and the performance year. We use the inverse of this difference, which averaged roughly 2 for performance year 2015, as the inflation factor. We use this approximation in lieu of a formal two-state estimation procedure to avoid limiting the analysis to a cohort of continuously attributable patients, which would negate one of the advantages of holding the baseline assignment constant and require a stronger common shocks assumption (as described above). In falsification tests, we applied the same estimation procedure in hypothetical entry years to large non-ACO TINs and ACOs that entered the MSSP in 2015.

### ***3. Area-level Analysis***

### *Conceptual Considerations*

Another approach to eliminate bias from strategic selection of lower-risk patients by ACOs is to compare spending changes between areas with higher vs lower exposure to the MSSP. Basing exposure on an area-level measure of MSSP penetration (an ecologic instrument) rather than patient-level attribution to an ACO ensures that systematic re-sorting of lower-risk patients to ACOs after program entry would not contribute to savings estimates, assuming that the mechanisms for risk selection do not change patients' location of residence. This approach also captures spillover effects of ACO efforts to lower spending on patients served by, but not attributed to, ACOs, as well as any spillover effects on practice patterns of other providers.

This strategy, too, is not without its disadvantages. First, the counterfactual (spending in the absence of MSSP participation) is no longer based on local trends in spending for an unexposed group but rather based on average national spending growth in HRRs with no (or less) MSSP participation. Greater MSSP participation in low-growth regions (selection relative to benchmarks based on national spending growth) would therefore contribute to savings in an area-level analysis but not in our main analysis. As described below, we take an intention-to-treat approach to remove bias from selective ACO continuation or expansion in areas determined by ACOs to be low-growth ex post (e.g., based on their bonuses), but this does not remove bias from selective entry based on ex ante knowledge of spending growth. Because spending growth is challenging to predict—e.g., regional growth in one period does not predict growth in the next<sup>25,26</sup>—we would not expect bias from selective entry but cannot exclude this possibility.

Second, because few HRRs had no MSSP penetration, and no HRRs had 100% penetration, an area-level analysis requires strong parametric assumptions about the relationship between MSSP penetration and spending growth and extrapolation to estimate an effect of 100%

vs. 0% participation that is analogous to effects estimated by our main evaluation approach. Third, like any area-level analysis, inferences about lower-level units are subject to ecological fallacy. For example, ACOs that most effectively reduce spending could be in low-penetration areas. Fourth, differences in fixed or time-varying characteristics of the Medicare fee-for-service population between areas may be less stable than differences between providers within areas, on average. For example, growing MSSP penetration may be correlated with faster or slower growth in regional Medicare Advantage enrollment, potentially causing differential changes in the study population that would be minimized in a within-area analysis. Finally, and perhaps most importantly, an area-level analysis does not hold constant market-level changes in unobserved determinants of spending growth, and spending growth is known to vary widely across regions.

### *Empirical Analysis*

For each performance year, we calculated MSSP penetration in each HRR as the proportion of attribution-eligible beneficiaries attributed to an ACO in a given program year, using our main method of attribution and an intention-to-treat approach that holds constant ACO definitions as the sets of TINs included at the outset of program participation and retains exiting ACOs as continuing in the program. MSSP penetration in 2014, for example, is the proportion of beneficiaries in an HRR attributed in 2014 to an ACO in the 2012, 2013, or 2014 entry cohorts. We then fit the following model for Medicare spending ( $Y$ ) for beneficiary  $i$  in year  $t$  and HRR  $h$ :

$$Y_{ith} = \beta_0 + \beta_1 HRR_{ith} + \beta_2 Year_{it} + \beta_3 ACOPenetration_{ith} \times ProgramYr_{it} + \beta_4 Covariates_{it} + \varepsilon_{ith}$$

where  $HRR$  and  $Year$  are vectors of HRR and year fixed effects, respectively, and

$ACOPenetration \times ProgramYr$  is an interaction between ACO penetration and indicators of each program year from 2012-2015, allowing the effect of ACO penetration to differ in each program

year as more ACOs enter and continuing ACOs gain experience (the interaction creates four variables equal to the MSSP penetration in HRR  $h$  in program year  $t$  when  $Year$  is program year  $t$ , and zero otherwise). To gauge whether this area-level approach was more or less immune to bias from changes in population characteristics than our main within-area approach, we compared the differential changes in patient characteristics estimated in our main approach with analogous differential changes associated with 100% increases in area-level MSSP participation.

#### ***4. Attribution Based on Referring PCPs***

##### *Conceptual Considerations*

While attributing patients based on PCP office visits only minimizes some forms of bias, it leaves an average of 23% of beneficiaries unassigned in each year. To reduce this and to address potential selection bias from ACO efforts to boost attribution of low-cost patients without altering patients' actual PCP, we modified the attribution procedure to use information about the referring PCP for other services. Thus, in a year in which a patient sees a specialist or has an imaging procedure or laboratory test but does not have an office visit with a PCP, we can attribute the patient to the PCP listed as the referring physician for those other services. This approach should reduce bias from a differential increase in the assigned share of low-risk ACO patients after MSSP entry, whether because of strategic annual wellness visits or other ACO efforts to enhance primary care access (though we did not find evidence of this in prior work).<sup>5</sup>

##### *Empirical Analysis*

Specifically, we used Medicare Carrier claims to determine the most common NPI with a PCP specialty appearing in the referring NPI field of a beneficiary's claims. For a given year, we then attributed the beneficiary to an ACO if that NPI was listed in the ACO's participant list in the first year of MSSP participation. We implemented these alternate assignments if the patient

had no office visits with a PCP and re-estimated savings using our main evaluation approach. Doing so increased the proportion of beneficiaries with an assignment in a given year from 77% to 87%, on average. Among beneficiaries for whom assignments could be made using either approach, 88.8% were assigned to the same ACO or to the control group in both cases, indicating that the most common referring PCP is usually the PCP providing the most office visits.

## **Assessing Risk Selection Potentially Contributing to Bonuses but Removed in Evaluation**

### ***1. Reconfiguration of ACOs to Favor Lower-cost Primary Care Providers***

To assess the extent to which ACOs reconfigured their provider composition over performance years to favor primary care practices or PCPs with lower per-patient spending, we modified our difference-in-difference analysis to allow the sets of TINs or PCPs (NPIs) constituting each ACO to change over the performance years per the annual MSSP Provider-level ACO participation files.<sup>27</sup> The changes in ACO PCPs reflected both changes in TINs and changes in the PCPs billing under the included TINs. Because the CMS participation files are available only for ACOs participating in the MSSP, we limited this analysis to ACOs participating through 2015 to eliminate effects of ACO dropout.

We then compared savings estimates when holding the set of TINs or NPIs constant, as in our main approach, with estimates when allowing them to change. Greater savings produced by the compositional changes would be a necessary but not sufficient condition for concluding that ACOs favored providers with lower spending as they evolved. Such a finding would not be sufficient because it might be expected from attenuation bias in our intention-to-treat analysis, which treated TINs or PCPs no longer exposed to ACO incentives as still part of an ACO. In addition, ACOs may have successfully identified providers who were more responsive to MSSP incentives, as opposed to providers with lower baseline spending.

Moreover, greater spending reductions produced by changes in ACO TIN inclusion would be negated by benchmark adjustments in the MSSP's calculation of shared savings, as noted above. Thus, compositional changes favoring lower-cost providers would only contribute to bonus payments if the changes in PCP composition of ACOs produced greater spending reductions than the changes in TIN composition of ACOs.

## ***2. Gaming of CMS Attribution Algorithm via Manipulation of TINs Used for Billing***

As noted above, our modifications to the attribution rules would act to minimize bias from ACO manipulation of the TINs used for billing to shift the attributed population toward lower-cost patients. To assess the potential for this selection strategy, among others, we assessed the effect of patient covariate adjustment on savings estimates when employing the original CMS attribution algorithm, which included additional qualifying services (CPT codes 99304-99310, 99315-99316, 99318, 99324-99328, 99334-99337, 99339-99340, 99341-99345, 99347-99350) and an additional step to attribute beneficiaries with no services from PCPs on the basis of services from non-PCPs (specialists, nurse practitioners, and physician assistants).<sup>13</sup> Specifically, we implemented the CMS algorithm to attribute beneficiaries to providers and repeated our evaluation analyses, holding constant the sets of TINs composing ACOs over the study period. We compared gross savings estimates with vs. without adjustment for observable patient characteristics. Substantial attenuation of savings estimates by patient covariate adjustment in analyses using the CMS attribution algorithm but not in our main approach (using only office visits with PCPs for attribution) would suggest risk selection that was removed in our evaluation but may have contributed to bonus payments. This assumes that risk selection is based on observables or that unobservable factors used to select are correlated with the observables.

Employing the CMS attribution algorithm, we also compared savings estimates from



evaluation analyses holding constant the composition of ACOs as fixed sets of TINs vs. fixed sets of clinician NPIs (the NPIs billing primarily under TINs included in ACOs in their first year of participation). If ACOs strategically changed the TINs used for billing by member clinicians to cause selective attribution of lower-cost patients (e.g., by shifting billing for nursing facility services to excluded TINs), then savings estimates should be attenuated by holding ACOs constant as sets of NPIs. For example, if ACOs shifted billing for nursing facility services, but not office visits by the same clinician, to an excluded TIN, or if ACOs shifted billing by clinicians with high-cost patients to a excluded TIN, the billing changes would increase savings when ACOs are defined as sets of TINs but not when they are defined as sets of NPIs. In the latter case, patients would remain assigned to an ACO even if their assigned clinician changed the TIN used to bill for all or some of their services.

### ***3. Patient and Physician Exit from ACOs***

We also examined whether higher-risk patients or PCPs with higher-risk patients were more likely to exit from ACOs. We categorized beneficiaries attributed to ACOs in 2013 or 2014 (year  $t$ ) into deciles based on their concurrent HCC score (i.e., using diagnoses year  $t$ ). We then compared the proportion who were no longer attributed to the same ACO in the subsequent year ( $t+1$ ) across deciles. We used the MSSP Beneficiary-level attribution files to determine actual beneficiary assignments in years  $t$  and  $t+1$  and limited the sample to beneficiaries who were attributed to ACOs that remained in the MSSP in 2015, so that patient exit could be interpreted as the patient, the patient's physician, or the physician's practice leaving an ACO, rather than an ACO leaving the MSSP. We additionally limited the sample to beneficiaries continuously eligible for attribution from 2013-2015 so that exit did not reflect lack of a qualifying service.

In an alternate analysis, we used our attribution approach (based on office visits with

PCPs) and held ACO composition of TINs constant (using ACO composition upon MSSP entry) so that patient exit could be interpreted as the patient or the patient's PCP leaving a set of ACO TINs (the more relevant quantity since ACO benchmarks adjust for TIN inclusion). In each version, we calculated the difference in HCC scores between leavers and stayers and fit a model of HCC scores as a function of ACO fixed effects and an indicator of leaving to estimate the mean *within-ACO* difference in HCC scores between leavers and stayers, thereby controlling for any relationship between organizational case mix and patient churn.

Prior research demonstrates that attribution in the MSSP is less stable over time for higher-risk patients because attribution is based on utilization.<sup>18,23,24</sup> Higher-risk patients use more qualifying services provided by more TINs (Appendix Table 1) and have a higher risk of health declines that may prompt a change in provider. Hence, they should be more likely to have changes in attribution due to changes in health care needs that cause them to favor different established providers in different years or switch to new providers. Differential exit from ACOs of high-risk patients is therefore not necessarily the consequence of risk selection. Moreover, it may not lead to a differential change in the average risk of ACO-attributed patients relative to non-ACO patients, because the risk of continuously assigned patients changes over time and new patients enter the ACO-assigned population.

To characterize the relationship between assignment churn and patient risk in the absence of MSSP incentives, we conducted a falsification test in which we applied the above analyses of patient exit to large non-participating TINs (those meeting the MSSP eligibility criterion of 5000+ beneficiaries). For consistency with the analysis of exit determined from the MSSP Beneficiary-level attribution file, we used the CMS attribution algorithm. This comparison remained inconsistent, however, because changes in TIN inclusion contributed to patient exit

from ACOs and we could not simulate such compositional changes among non-ACO providers. To achieve a more consistent comparison, we employed our attribution approach in an alternate version that held ACO or non-ACO composition constant over time.

We conducted an analogous analysis at the PCP level to characterize the relationship between the average health risk of a PCP's patients and the probability of PCP exit from the ACO. Specifically, we modified our attribution method to attribute beneficiaries to a PCP NPI, rather than to an ACO or non-ACO TIN, based on qualifying office visits. We focused on PCPs actively billing for visits from 2012-2015 so that exit from an ACO or non-ACO TIN by 2015 would reflect a switch to a different practice or different TIN for billing purposes, as opposed to exit from the workforce. We also limited the analysis to PCPs with at least 20 attributed patients per year (accounting for 85.7% of patient-years) to reduce sampling error in estimation of PCPs' average patient risk and to avoid giving undue weight to exiting PCPs with very few patients.

Using 2012-2013 data, we estimated the average HCC risk score of each PCP's attributed patients by fitting a linear regression model of patients' HCC scores as a function of PCP fixed effects and an indicator for year. We categorized PCPs into deciles based on their patients' mean HCC score. We determined the primary TIN under which PCPs billed in 2013 from the Medicare Data on Provider Practice and Specialty file.<sup>28</sup> Among PCPs billing under TINs included by the 2012 or 2013 entry cohorts of ACOs upon program entry (per the MSSP Provider-level RIF), we then determined the proportion of PCPs in each decile who were no longer billing under any of those TINs in 2015. Similarly, among PCPs billing under large non-ACO TINs in 2012 or 2013, we determined the proportion in each decile no longer billing under any of those TINs by 2015.

Because ACOs and large non-ACO TINs differ, our falsification analyses could not reliably establish a counterfactual (the extent to which higher-risk patients, or PCPs with higher-

risk patients, would exit ACOs in the absence of MSSP incentives). Nevertheless, a relationship between patient risk and patient or PCP exit that is similar for ACOs and non-ACO TINs would reject an interpretation of a strong relationship for ACOs as prima facie evidence of strategic risk selection—including manipulation of TINs used for billing—in response to MSSP incentives.

## **RESULTS**

### **Main Evaluation of the MSSP through 2015**

Table 3 summarizes the overall results of our main evaluation approach. In the pre-entry period, ACO spending levels and trends did not differ from those for local controls. Estimates of annual gross savings grew over performance years to \$302/patient by 2015 in the 2012 entry cohort and \$139/patient in the 2013 cohort. Overall gross savings did not grow in the 2014 cohort over two performance years and were not significant in 2015. Aggregating these gross savings across all ACO-attributed patients from 2013-2015, multiplying by 5 to correct for the 20% sampling, and subtracting bonus payments yielded a total program-wide estimate of net savings to Medicare from 2013-2015 of \$358 million.<sup>5</sup>

Differential changes from the pre-entry period to 2015 in ACO-attributed patients' sociodemographic and clinical characteristics, relative to local control patients attributed to non-ACO providers, were consistently minimal (Table 4). These findings included minimal differential changes in patients' history of hip fracture or acute myocardial infarction, conditions that have been used as exogenous markers of health risk (though could be affected by efforts to improve quality).<sup>29</sup> Not only were all differential changes in observable patient characteristics small, there is no suggestion in the Table 4 estimates of consistently greater imbalance in entry cohorts with greater savings or of growing imbalance within cohorts as savings grew.

Estimates were nearly identical for the 2012 and 2013 cohorts with and without

adjustment for patient covariates and with and without propensity-score weighting (Figure 2). Holding ACO definitions constant as sets of PCPs instead of TINs increased gross savings slightly in the 2012 cohort and appreciably in the 2013 cohort. Thus, we can reject changes in the PCPs billing under ACO TINs as contributing to the main estimates of savings. Falsification tests of pre-entry years for ACOs and hypothetical entry years for 439 large non-ACO TINs revealed no evidence of differential reductions in spending in the absence of MSSP participation (Appendix Figure 1 and Appendix Table 2).

### **Approaches to Assess and Address Residual Risk Selection**

#### ***Patient Fixed Effects***

After limiting the study population to a longitudinal cohort of continuously enrolled beneficiaries who were attributable to an ACO or non-ACO provider in at least one pre-MSSP year and in 2015, gross savings estimates were attenuated and less precise (Table 5), as expected from the substantially lower spending for this cohort (Figure 1) and its smaller size (35% of beneficiaries and 55% of beneficiary-year observations in the full study population). Within this cohort, replacing ACO fixed effects with patient fixed effects increased savings by \$1/patient in the 2012 entry cohort, decreased savings by \$33/patient in the 2013 cohort, and increased savings by \$52/patient in the 2014 cohort (Table 5), providing no consistent evidence that turnover in ACO-attributed populations differentially favored patients with fixed characteristics predictive of lower spending. In models with patient fixed effects, estimates of gross savings adjusted for time-varying patient factors were consistently greater (larger savings) than unadjusted estimates (Table 5), suggesting that restricting to a cohort of continuously attributable patients and implementing patient fixed effects introduced differential changes in time-varying characteristics that biased savings toward zero and were not present in our main analysis.

### ***Holding Baseline Assignments Fixed***

Estimates of gross savings from analyses holding patients' baseline assignments fixed across the study period were generally similar to estimates from our main approach (Figure 3). Falsification tests applying the same approach to large non-ACO TINs or the 2015 entry cohort of MSSP ACOs as hypothetical entrants in 2013 yielded significant differential spending *increases* in hypothetical performance years, despite pre-period spending differences from the control group that were similar to those for ACOs (Appendix Table 3). While the results of these falsification tests suggest that we may have underestimated savings in our main intention-to-treat approach (e.g., from attenuation bias due to retaining ACOs and ACO TINs no longer participating), they suggest more generally that an approach using baseline assignments may introduce substantial and bias in unknown direction.

### ***Area-level Analysis***

Adjusted gross savings associated with a 100% increase in MSSP penetration in the area-level analysis were larger than gross savings estimated in our main approach and grew from \$299/patient in 2013 to \$463/patient in 2015. While the larger savings may be indicative of spillovers, we also found that populations in HRRs with greater growth in MSSP participation became differentially lower risk relative to HRRs with lower growth in participation (Table 6). Unlike in our main analysis, adjusted gross savings in 2015 (\$463/patient) was substantially smaller than unadjusted gross savings (\$788/patient) because of the growing imbalance in patient characteristics described in Table 6.

### ***Attribution Based on Referring PCPs***

Attribution based on the dominant referring PCP when beneficiaries had no PCP office visits to support attribution increased gross savings estimates by \$50-112/patient (Appendix

Table 4). In previous work, we also found that modifying attribution to include office visits with nurse practitioners and physician assistants (who conduct annual wellness visits in many practices) did not substantively change estimates.<sup>5</sup>

**Assessing Risk Selection Potentially Contributing to Bonuses but Removed in Evaluation**  
*Reconfiguration of ACOs to Favor Lower-cost Primary Care Providers*

Allowing the TINs composing ACOs to change after the first performance year did not appreciably affect adjusted gross savings in the 2012 entry cohort, decreased savings in the 2013 entry cohort, and increased savings in the 2014 cohort (Table 7). The changes in adjusted savings due to compositional changes in the 2013 and 2014 cohorts were at least partly mediated by shifts to providers with sicker (2013 cohort) or healthier (2014 cohort) patients, as opposed to providers with different levels of efficiency, based on comparisons of adjusted and unadjusted estimates when allowing ACO TIN composition to change (Table 7),

Compared with changes produced by allowing the composition of TINs to change, allowing ACOs' composition of PCPs to change caused lesser changes in adjusted savings (Table 7), suggesting no systematic selection of lower-cost PCPs within ACO TINs. These findings provide no consistent evidence of favorable risk selection mediated by changes in ACO provider inclusion to increase bonuses, but they do support our main intention-to-treat approach to eliminate bias from changes in ACO provider composition.

***Gaming of CMS Attribution Algorithm via Manipulation of TINs Used for Billing***

In analyses employing the CMS attribution algorithm and holding ACO TIN composition constant over performance years, adjustment for patient covariates had modest and inconsistent effects, providing no consistent evidence of patient-level risk selection within ACO TINs via gaming of the attribution algorithm. Adjusted gross savings in 2015 was 94% of unadjusted

gross savings in the 2012 cohort, 133% of unadjusted gross savings in the 2013 cohort, and 74% of the (smaller) unadjusted gross savings in the 2014 cohort (Appendix Table 5).

Use of the CMS attribution algorithm caused trends in the pre-entry period to differ between ACOs and the control group in a direction that would exaggerate savings estimates if the trend difference continued over performance years. The trend difference was due to the inclusion of visits in nursing facilities, which were largely dropped from the attribution algorithm in a 2017 rule change,<sup>30</sup> and not due to the inclusion of outpatient visits with specialists, which increased differences in pre-period levels but not trends (Appendix Table 6). These findings support our a priori decision to base attribution only on PCP office visits in our main approach and suggest that savings estimated with use of the CMS attribution algorithm may be biased.

In analyses using the CMS attribution algorithm, adjusted gross savings were substantively similar when treating ACOs as fixed groups of initially participating TINs or as fixed groups of clinicians billing under those TINs in the ACOs' first year of participation (Appendix Table 5). Savings were somewhat smaller when defining ACOs as fixed groups of clinicians, but so were differences in pre-period trends (data not shown), suggesting that redefining ACOs as groups of clinicians attenuated estimates by correcting for bias related to pre-existing trends as opposed to bias from gaming related to changes in the TINs used for billing. (The latter should have manifested as larger and more consistent effects of adjustment for patient characteristics on savings estimates in Appendix Table 5). Taken together, the results of these analyses do not provide clear evidence of ACOs shifting the billing of some clinicians to excluded or included TINs in order to attain a lower-risk attributed patient population.

### ***Patient and Physician Exit from ACOs***

Among patients attributed to an ACO in 2013 or 2014 per the CMS Beneficiary-level



attribution file, the proportion no longer attributed to the same ACO in the subsequent year per the attribution file was higher among patients with higher HCC scores, for example 22.4% in the highest decile of HCC scores vs. 15.3% in second to lowest decile (Figure 4A). As has been previously described,<sup>23</sup> the exit rate was also higher among patients in the lowest decile of risk scores, consistent with their low use of qualifying services providing a less reliable basis for assessing a patient's regular source of care. We observed a nearly identical pattern when using the CMS attribution algorithm to assign patients to large non-ACO TINs, with exit rates rising from 15.2% among patients in the second lowest decile of HCC scores to 22.3% in the highest decile (Figure 4A). These results are consistent with the lack of differential changes detected in our analysis of measurable patient characteristics (Table 4), which quantify the net effect of non-random patient exit and entry and suggest churn is similar in both ACO and non-ACO groups.

Results were similar when attribution was based on office visits with PCPs; patient exit rates were higher for patients with higher HCC scores in both ACOs and large non-ACO TINs (Figure 4A). Within ACOs, the mean difference in HCC scores between patients exiting and staying was nearly identical for ACOs (0.095;  $P < 0.001$ ) and large non-ACO TINs (0.096;  $P < 0.001$ ). Thus, in the absence of MSSP incentives, higher-risk patients had less stable assignments over time, as might be expected because they receive more services from more physicians in PCP specialties (Appendix Table 1). When attribution was based on office visits with PCPs only, exit rates were substantially lower (Figure 4A), as expected by the exclusion of other qualifying services in the CMS attribution algorithm (e.g., post-acute visits) that directly reflect changes in patients' health care needs and thus introduce instability in attribution.<sup>24</sup>

In analogous analyses at the PCP level, PCPs with higher-risk patients also had higher rates of exit from both ACO TINs and large non-ACO TINs by 2015 (Figure 4B). The mean

difference in PCPs' mean patient HCC score between exiting and staying PCPs was 0.029 (P=0.009) for ACOs and 0.026 (P=0.005) for non-ACO TINs. These differences were reduced to 0.019 (P=0.09) and 0.018 (P=0.05), respectively, after adjustment for ACO or TIN fixed effects to estimate within-ACO or within-TIN differences, suggesting that physician turnover is higher in organizations serving higher risk patients. We could not explain the residual differences in patient risk between exiting and staying PCPs. One possibility is that physicians who have been at a practice longer may be more likely to switch practices and also may have sicker patients due to aging of their patient panel. Regardless of the explanation, our analysis demonstrates that PCP turnover is greater for PCPs with higher-risk patients in the absence of MSSP incentives. Thus, greater exit from ACOs by physicians with higher-risk patients should not be interpreted as evidence of risk selection by ACOs.<sup>31</sup>

## **DISCUSSION**

Through 2015, we estimate that the MSSP lowered Medicare spending modestly for ACO patients. We implemented several measures to minimize selection bias and found no evidence that residual risk selection drove the estimated savings. The larger savings and greater growth in savings reported for physician group ACOs were similarly found to be robust in previous work.<sup>5</sup> In addition, we detected no evidence of risk selection that may have contributed to ACO bonuses but was eliminated by our evaluation approach. Specifically, we did not find evidence that ACOs consistently manipulated their composition of providers to favor practices or physicians with lower-risk patients (or more efficient practice patterns). The lack of detectable risk selection is consistent with ACOs' limited incentives to favor practices (TINs) with low spending under the benchmarking rules during the first phase of the MSSP (through 2016) and a limited ability to select specific physicians or patients within practices for inclusion or exclusion.

As the basis for benchmarks increasingly transitions to ACOs' regional spending average, as opposed to their own historical spending, ACOs have new incentives to favor practices with spending below the regional average. Identifying practices with lower predicted spending levels is easier than identifying those with slower predicted spending growth, because spending levels are strongly correlated over time, whereas growth rates are not.<sup>25,26</sup> Wide variation in risk-adjusted spending levels between providers within regions<sup>32,33</sup> suggests opportunities for ACOs in the revamped MSSP, Pathways for Success, to earn bonuses by selectively including or excluding practices. Such practice-level selection may be easier than physician- or patient-level selection because it requires only a change in ACO participant lists and because spending can be more reliably profiled at the practice level than at the physician or patient levels.

In the short run under Pathways, selective inclusion of low-spending practices by ACOs (or selective participation of low-spending ACOs) will be costly to the Medicare program, effectively increasing subsidies to providers that have lower-risk patients or are already more efficient.<sup>9</sup> In the long run, selectively attracting more efficient practices could conceivably enhance social welfare if demand for efficiency in the MSSP applies sufficient pressure on other providers to become more efficient, and thus attractive to ACOs, as the program expands. Such a scenario, however, would require the fee-for-service alternative to be less attractive to inefficient providers than the MSSP and better risk adjustment to mitigate transfers from practices serving sicker patients to those serving healthier patients. As new incentives under Pathways play out, evaluations that judge ACO spending against valid counterfactuals, rather than benchmarks, will continue to be important to quantify savings from changes in care delivery. Unbiased estimates, however, may be challenging to produce as new payment models proliferate and expand.

Our analyses also have implications for the application of quasi-experimental research

methods. First, our findings exemplify the importance of establishing a plausible counterfactual when attempting to draw causal conclusions. Our estimates of savings are greater than those generated by comparisons with ACO benchmarks, which systematically underestimated ACO savings in the first phase of the program.<sup>34</sup> In addition, our tests of risk selection reveal that analysis of patient or physician exit from ACOs can be misleading without considering exit under a counterfactual scenario in which providers do not face MSSP incentives.<sup>31</sup> When comparing against churn among non-ACO providers, we found no evidence of risk selection at the clinician or patient level; turnover is higher for higher-risk patients and their clinicians, regardless of MSSP incentives.

Second, our analyses illustrate common tradeoffs between approaches to address different sources of bias; strategies that ensure elimination of one source of bias can exacerbate bias from other sources. In particular, our findings demonstrate that analytic steps to eliminate bias from changes in fixed characteristics of patients due to changes in ACO population composition can exacerbate bias from within-patient changes in time-varying characteristics. While we found no suggestion of residual selection in robustness checks of our main approach—which allowed patient turnover within ACOs—the same checks suggested introduction of bias by use of patient fixed effects, baseline assignments, or area-level comparisons to hold patients constant. The estimates produced by these approaches were generally consistent with those from our main approach but less robust in sensitivity or falsification analyses. Thus, an overemphasized conceptual concern about one source of potential bias (in this case, selection on unobserved fixed traits of patients) can lead to a suboptimal approach, if not erroneous conclusions. Our findings suggest that assessing assumptions with observable information can help guide choice of an approach. Since identifying assumptions in quasi-experimental studies

cannot be tested directly (counterfactuals cannot be observed) one must ultimately rely on tests of observable quantities to gauge the extent of residual bias. By that standard, we have most confidence in the estimates produced by our primary approach and would advise caution in interpreting estimates produced by the alternative approaches.

### **Limitations**

Because providers were not randomized to the MSSP, our analysis was subject to forms of selection bias other than risk selection conditional on participation. One concern is that providers selected into the MSSP based on anticipated changes in spending growth. Spending trends in the pre-entry period, however, were nearly identical for ACOs and local control groups, on average, suggesting that ACOs did not select into the MSSP based on established trajectories. Although we could not rule out selective entry as contributing to gross savings estimates, we do not find it plausible that ACOs could accurately predict their future risk-adjusted spending growth relative to their region. We further note that ACOs had incentives to enter if their anticipated spending growth was slower than national spending growth (the basis for updating benchmarks), not local spending growth (the basis for counterfactuals in our evaluation). Thus, even if ACOs managed to selectively enter in a way that resulted in bonus payments, that would not necessarily bias our estimates of gross savings, and the unearned bonuses would be accounted for in our calculation of net savings. However, because regional spending growth in one period is not predictive of regional spending growth in the next, and because regional spending levels are not correlated with regional spending growth, ACOs' basis for predicting their region's growth relative to the nation is limited.

Clearly, ACOs were likely to have greater capacity to respond to MSSP incentives than non-participating providers. But in the absence of selection on future changes in spending

growth, non-equivalence between ACO and non-ACO providers would compromise only the external validity (generalizability), not the internal validity, of our findings. As long as the differential reductions in spending we estimated were due to provider responses to new incentives, they would be valid estimates of the causal effects of MSSP participation on participants. We would not expect ACOs to slow fee-for-service spending, counter to their financial self-interest, in the absence of an incentive to do so.

Another concern is that other time-varying determinants of spending growth affected ACOs and non-ACO providers differently because they are different. However, spending levels were similar for ACOs and non-ACO providers, and differential changes in spending for ACO patients in pre-entry years and for patients of large non-ACO providers were small. Thus, violations of the common shocks assumption would have had to coincide with the staggered entry of ACOs into the MSSP and grow with longer participation.<sup>5</sup>

Nevertheless, in the absence of randomization to the MSSP, we cannot entirely exclude the possibility of selection bias contributing to our estimates of gross savings. The source of bias, however, would have to evade detection by the many robustness checks we conducted. We also cannot reject the possibility of some gaming behavior undetected by our many tests. The costs of such gaming, however, would be reflected in bonus payments. Thus, our estimates of net savings would still be valid as long as our approach generated unbiased estimates of gross savings.

## **Conclusion**

Through its first 3 full years of operation, we found that participation in the MSSP was associated with modest savings and not with favorable risk selection. These findings suggest an opportunity to build on early progress. Understanding the effect of stronger selection incentives on savings in the revamped MSSP will be important to guide future program reforms.

## REFERENCES

1. Centers for Medicare and Medicaid Services. Fast Facts. All Medicare Shared Savings Program ACOs and Pioneer ACOs. 2015. (Accessed October 1, 2019, at <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/PioneersMSSPCombinedFastFacts.pdf>.)
2. Colla CH, Lewis VA, Kao LS, O'Malley AJ, Chang CH, Fisher ES. Association Between Medicare Accountable Care Organization Implementation and Spending Among Clinically Vulnerable Beneficiaries. *JAMA Intern Med* 2016;176:1167-75.
3. McWilliams JM. Changes in Medicare Shared Savings Program Savings from 2013 to 2014. *JAMA* 2016;316:1711-13.
4. McWilliams JM, Hatfield LA, Chernew ME, Landon BE, Schwartz AL. Early Performance of Accountable Care Organizations in Medicare. *N Engl J Med* 2016;374:2357-66.
5. McWilliams JM, Hatfield LA, Landon BE, Hamed P, Chernew ME. Medicare Spending after 3 Years of the Medicare Shared Savings Program. *N Engl J Med* 2018;379:1139-49.
6. Trombley MJ, Fout B, Brodsky S, McWilliams JM, Nyweide DJ, Morefield B. Early Effects of an Accountable Care Organization Model for Underserved Areas. *N Engl J Med* 2019;381:543-51.
7. McWilliams JM, Zaslavsky AM, Landon BE, Chernew ME. Spending reductions in the Medicare Shared Savings Program: selection or savings? (Accessed October 1, 2019, at <https://theincidentaleconomist.com/wordpress/mssp-selection-savings/>.)
8. Department of Health and Human Services. Centers for Medicare and Medicaid Services. 42 CFR Part 425. Medicare Program; Medicare Shared Savings Program; Accountable Care Organizations--Pathways to Success and Extreme and Uncontrollable Circumstances Policies for Performance Year 2017. Final rules. (Accessed October 1, 2019, at <https://www.govinfo.gov/content/pkg/FR-2018-12-31/pdf/2018-27981.pdf>.)
9. McWilliams JM, Landon BE, Rathi VK, Chernew ME. Getting More Savings from ACOs - Can the Pace Be Pushed? *N Engl J Med* 2019;380:2190-2.
10. Medicare program; Medicare Shared Savings Program: accountable care organizations. Final rule. 2011. (Accessed October 1, 2019, at <http://www.gpo.gov/fdsys/pkg/FR-2011-11-02/pdf/2011-27461.pdf>.)
11. Friedberg MW, Chen PG, Simmons M., Sherry T., Mendel P, et al. Effects of Health Care Payment Models on Physician Practice in the United States. Follow-up Study. 2018. (Accessed October 1, 2019, at [https://www.rand.org/pubs/research\\_reports/RR2667.html](https://www.rand.org/pubs/research_reports/RR2667.html).)

12. McWilliams JM, Landon BE, Chernew ME, Zaslavsky AM. Changes in patients' experiences in Medicare Accountable Care Organizations. *N Engl J Med* 2014;371:1715-24.
13. Centers for Medicare and Medicaid Services. Medicare Shared Savings Program. Shared savings and losses and assignment methodology. Specifications. 2014. (Accessed October 1, 2019, at <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/Shared-Savings-Losses-Assignment-Spec.pdf>.)
14. Colla CH, Lewis VA, Bergquist SL, Shortell SM. Accountability across the Continuum: The Participation of Postacute Care Providers in Accountable Care Organizations. *Health Serv Res* 2016;51:1595-611.
15. McWilliams JM, Chernew ME, Landon BE, Schwartz AL. Performance differences in year 1 of Pioneer accountable care organizations. *N Engl J Med* 2015;372:1927-36.
16. Research Data Assistance Center. Shared Savings Program Accountable Care Organizations Beneficiary-level RIF. (Accessed October 1, 2019, at <https://www.resdac.org/cms-data/files/ssp-aco-beneficiary-level-rif>.)
17. Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association* 2018;113:390-400.
18. Medicare Payment Advisory Commission. Report to Congress: Medicare and the Health Care Delivery System. Chapter 6: Assessing the Medicare Shared Savings Program's Effects on Spending. (Accessed October 1, 2019, at [http://www.medpac.gov/docs/default-source/reports/jun19\\_ch6\\_medpac\\_reporttocongress\\_sec.pdf?sfvrsn=0](http://www.medpac.gov/docs/default-source/reports/jun19_ch6_medpac_reporttocongress_sec.pdf?sfvrsn=0) )
19. Daw JR, Hatfield LA. Matching and Regression to the Mean in Difference-in-Differences Analysis. *Health Serv Res* 2018;53:4138-56.
20. Chabe-Ferret S. Should we combine difference in differences with conditioning on pre-treatment outcomes? (Accessed October 1, 2019, at <https://www.tse-fr.eu/publications/should-we-combine-difference-differences-conditioning-pre-treatment-outcomes>.)
21. Pham HH, Schrag D, O'Malley AS, Wu B, Bach PB. Care patterns in Medicare and their implications for pay for performance. *N Engl J Med* 2007;356:1130-9.
22. Lewis VA, McClurg AB, Smith J, Fisher ES, Bynum JP. Attributing patients to accountable care organizations: performance year approach aligns stakeholders' interests. *Health Aff (Millwood)* 2013;32:587-95.
23. McWilliams JM, Chernew ME, Dalton JB, Landon BE. Outpatient care patterns and organizational accountability in Medicare. *JAMA Intern Med* 2014;174:938-45.
24. McWilliams JM, Chernew ME, Zaslavsky AM, Landon BE. Post-acute care and ACOs - who will be accountable? *Health Serv Res* 2013;48:1526-38.



25. McWilliams JM, Song Z. Implications for ACOs of variations in spending growth. *N Engl J Med* 2012;366:e29.
26. Chernew M, Decicca P, Town R. Managed care and medical expenditures of Medicare beneficiaries. *J Health Econ* 2008;27:1451-61.
27. Research Data Assistance Center. Shared Savings Program Accountable Care Organizations Provider-level RIF. (Accessed October 1, 2019, at <https://www.resdac.org/cms-data/files/ssp-aco-provider-level-rif>.)
28. Research Data Assistance Center (ResDAC). Medicare Data on Provider Practice and Specialty (MD-PPAS). (Accessed October 1, 2019, at <https://www.resdac.org/cms-data/files/md-ppas>.)
29. Colla CH, Wennberg DE, Meara E, et al. Spending differences associated with the Medicare Physician Group Practice Demonstration. *JAMA* 2012;308:1015-23.
30. Sawhney TG, Fitch K, Gusland C. The exclusion of some nursing facility visits from MSSP assignment has potential unintended consequences. Milliman. (Accessed October, 2019, at <https://www.milliman.com/uploadedFiles/insight/2018/exclusion-nursing-facility-visits-unintended-consequences.pdf>.)
31. Markovitz AA, Hollingsworth JM, Ayanian JZ, et al. Risk Adjustment In Medicare ACO Program Deters Coding Increases But May Lead ACOs To Drop High-Risk Beneficiaries. *Health Aff (Millwood)* 2019;38:253-61.
32. Rose S, Zaslavsky AM, McWilliams JM. Variation In Accountable Care Organization Spending And Sensitivity To Risk Adjustment: Implications For Benchmarking. *Health Aff (Millwood)* 2016;35:440-8.
33. Schwartz AL, Zaslavsky AM, Landon BE, Chernew ME, McWilliams JM. Low-Value Service Use in Provider Organizations. *Health Serv Res* 2018;53:87-119.
34. Chernew ME, Barbey C, McWilliams JM. Savings Reported by CMS Do Not Measure True ACO Savings. *Health Affairs Blog*, 2017. (Accessed October 1, 2019, at <https://www.healthaffairs.org/doi/10.1377/hblog20170619.060649/full/>.)

**Table 1. Approaches to Assess and Address Residual Risk Selection in Evaluation**

<b>Analytic strategy</b>	<b>How strategy intends to address risk selection</b>	<b>Potential problems with strategy</b>
Patient fixed effects	Eliminate differential changes in fixed characteristics of patients attributed to ACOs vs. non-ACO providers by comparing <i>within-patient</i> changes in spending	Limiting to a continuous cohort introduces large within-patient health declines that could differ between ACO and non-ACO providers in absence of MSSP if: attribution to an ACO is affected by time-varying health needs ACO and non-ACO providers differ in treatment of declining patients Conflates within-patient changes from before to after MSSP entry with within-patient changes in attribution from ACO to non-ACO providers during post-period
Intention-to-treat with fixed baseline assignments	Eliminate differential changes in patient attribution to ACOs vs. non-ACO providers by holding baseline assignments constant over the study period	In absence of MSSP, baseline attribution to ACOs associated with subsequent differential changes in spending as patients switch providers if: ACO and non-ACO providers differ in practice patterns or reimbursement rates attribution to an ACO is affected by time-varying health needs, causing differential regression to the mean
Area-level analysis	Eliminate re-sorting of patients to ACO vs. non-ACO providers after MSSP entry by basing exposure on area-level MSSP penetration instead of patient-level attribution to an ACO	Allows bias from selective program participation in low-spending-growth areas Does not control for market-level determinants of spending growth, including differential changes in population health Ecological fallacy (e.g., if most effective ACOs are in low-growth areas)
Attribution based on referring PCP	Augment attribution using the referring PCP for services used by unattributed patients to mitigate bias from ACO use of annual wellness visits or other strategies to attract low-cost patients who would not otherwise be attributed	Narrowly addresses only one potential selection behavior 13% of beneficiaries remain unassigned

**Table 2. Approaches to Assess Risk Selection Contributing to Bonuses but Removed in Evaluation**

<b>Selection strategy</b>	<b>Analytic approach to assess extent of selection</b>	<b>Interpretation</b>
Reconfiguration of ACOs to favor lower-cost providers	Compare savings estimates from main approach holding ACO TINs or clinician NPIs constant vs. approach allowing changes in provider composition after first performance year	Greater savings estimated when allowing provider turnover would suggest selective inclusion of lower-cost providers. Since ACO benchmarks adjust for changes in TINs, only increases in savings from within-TIN changes in clinician composition could contribute to bonuses and constitute risk selection.
Gaming of CMS attribution algorithm via manipulation of TINs used for billing	<ol style="list-style-type: none"> <li>1. Compare savings with and without adjustment for patient characteristics when employing the CMS attribution algorithm</li> <li>2. Using CMS attribution algorithm, compare savings when holding ACO composition fixed as sets of TINs vs. NPIs</li> </ol>	<ol style="list-style-type: none"> <li>1. Attenuation of savings by patient covariate adjustment in analysis using the CMS attribution rules but not in our main approach (using only PCP office visits for attribution) would suggest risk selection enabled by the CMS rules that was removed in our evaluation but could have contributed to bonuses.</li> <li>2. If ACOs strategically changed the TINs used for billing by member clinicians to induce attribution of lower-cost patients, savings estimates should be attenuated by holding ACOs constant as sets of NPIs.</li> </ol>
Exclude high-risk patients or physicians with high-risk patients	Compare differences in exit from ACO-attributed populations between higher- vs. lower-risk patients (or differences in exit from ACO TINs between PCPs with higher- vs. lower-risk patients) with analogous differences in patient or PCP exit among non-ACO TINs	Disproportionate exit of higher-risk patients or PCPs with higher-risk patients that is greater for ACOs than non-ACO providers would suggest potential risk selection (a necessary but not sufficient observation).

TIN = taxpayer identification number; NPI = national provider identifier

Table 3. Estimated gross savings estimates by MSSP entry cohort and performance year

Entry cohort	Unadjusted pre-entry sample mean, <sup>a</sup> \$/patient	Adjusted pre-entry difference in annual spending level between ACOs and control group, \$/patient (95% CI)	Adjusted pre-entry difference in annual spending trend between ACOs and control group, \$/patient (95% CI)	Estimated gross savings (adjusted differential change in spending from pre-entry period to performance year for ACOs vs. control group), <sup>b</sup> \$/patient (95% CI)		
				2013	2014	2015
<b>2012 cohort</b> (N=114 ACOs)	9649	139 (-79, 357)	-3 (-58, 53)	-129 (-261, 2)	-291 (-425, -157)	-302 (-437, -166)
<b>2013 cohort</b> (N=106 ACOs)	9649	31 (-84, 146)	-5 (-39, 29)	-15 (-112, 82)	-114 (-214, -14)	-139 (-243, -35)
<b>2014 cohort</b> (N=115 ACOs)	9649	33 (-90, 155)	8 (-18, 34)	-	-72 (-150, 7)	-36 (-122, 50)

<sup>a</sup>In the analyses, the pre-entry period differed for each entry cohort, but for the purpose of describing the study sample in this table, years 2009-2011 were used to calculate a single mean for each characteristic.

<sup>b</sup>A negative differential change in spending indicates savings

**Table 4. Differential changes from the pre-entry period to each performance year in the characteristics of patients served by ACOs, as compared with the control group, by entry cohort of ACOs<sup>a</sup>**

Patient Characteristic	Unadjusted pre-entry sample mean, <sup>b</sup> \$/patient	Differential change from pre-entry period to each performance year for ACOs vs. control group							
		2012 entry cohort			2013 entry cohort			2014 entry cohort	
		2013	2014	2015	2013	2014	2015	2014	2015
Age (yr)	72.2	0.0	0.0	0.0	0.1	0.2	0.1	0.1	0.1
Female sex (%)	58.5	-0.2	-0.3	-0.2	-0.1	-0.2	-0.2	0.0	-0.1
Race or ethnic group <sup>c</sup> (%)									
Non-Hispanic white	83.5	0.0	-0.1	-0.2	0.0	-0.1	-0.2	0.0	0.1
Non-Hispanic black	8.5	-0.1	0.0	0.2	0.0	0.0	0.2	-0.1	-0.2
Hispanic	4.7	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0
Other	3.2	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.1
Medicaid recipient (%)	15.3	-0.2	-0.2	-0.2	-0.1	-0.3	-0.1	-0.1	0.1
Disability was original reason for Medicare eligibility (%)	21.9	-0.1	-0.1	0.0	-0.3	-0.5	-0.3	-0.2	-0.3
End-stage renal disease (%)	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CCW conditions, <sup>d</sup> no.									
Through prior year	5.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Through 3 years prior <sup>e</sup>	4.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
History of hip fracture (%)	2.9	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.0	-0.1
History of myocardial infarction (%)	4.2	0.0	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0
HCC risk score <sup>f</sup>									
Based on claims in prior year	1.23	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.00
Based on claims 3 years prior <sup>e</sup>	1.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ZCTA-level characteristic									
% Below federal policy level	9.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% With high school diploma	75.4	-0.1	-0.2	-0.2	0.0	0.0	-0.1	0.0	0.0
% With college degree	19.4	-0.2	-0.2	-0.2	0.0	0.0	0.0	-0.1	0.0

<sup>a</sup>Means and percentages were adjusted for geography to reflect comparisons within hospital referral regions. ZCTA denotes ZIP Code tabulation area.

<sup>b</sup>In the analyses, the pre-entry period differed for each entry cohort, but for the purpose of describing the study sample in this table, years 2009-2011 were used to calculate a single mean for each characteristic.

<sup>c</sup>Race or ethnic group was determined from Medicare Master Beneficiary Summary Files.

<sup>d</sup>Chronic conditions from the Chronic Conditions Data Warehouse (CCW) included 27 conditions: acute myocardial infarction, Alzheimer's disease, Alzheimer's disease and related disorders or senile dementia, anemia, asthma, atrial fibrillation, benign prostatic hyperplasia, chronic kidney disease, chronic obstructive pulmonary disease, depression, diabetes, heart failure, hip or pelvic fracture, hyperlipidemia, hypertension, hypothyroidism, ischemic heart disease, osteoporosis, rheumatoid arthritis or osteoarthritis, stroke or transient ischemic attack, breast cancer, colorectal cancer, endometrial cancer, lung cancer, prostate cancer, cataracts, and glaucoma. Analytic models included indicators for all 27 conditions and indicators for the presence of multiple conditions ranging from 2 to 9 or more conditions. Counts of conditions included all conditions except cataracts and glaucoma.

<sup>e</sup>For analyses of CCW condition indicators and Hierarchical Condition Categories (HCC) scores derived from earlier claims, we limited the sample to a subgroup of beneficiaries who were also continuously enrolled in fee-for-service Medicare 3 years prior to the study year. The purpose of this was to assess the extent to which any differential changes may have been due to differential changes in coding practices in response to MSSP incentives.

<sup>f</sup>HCC risk scores are derived from demographic and diagnostic data in Medicare enrollment and claims files, with higher scores indicating higher predicted spending in the subsequent year. For each beneficiary in each study year, we assessed the HCC score based on enrollment and claims data in the prior year, two years prior, and three years prior, in each case requiring continuous enrollment in fee-for-service Medicare in the study year and the year of claims used to calculate HCC scores.

**Table 5. Impact of patient fixed effects on estimated savings**

Entry cohort	Estimated gross savings in 2015 <sup>a</sup> (differential change in spending from pre-entry period to 2015 for ACOs vs. control group), \$/patient (95% CI)			
	Primary sample and approach, no patient fixed effects in model, adjusted <sup>b</sup>	Continuously attributable sample, <sup>c</sup> no patient fixed effects in model, adjusted <sup>b</sup>	Continuously attributable sample, <sup>c</sup> patient fixed effects added to model	
			Adjusted <sup>b</sup>	Unadjusted
<b>2012 cohort</b>	-302 (-437, -166)	-203 (-299, -107)	-204 (-281, -126)	-161 (-240, -83)
<b>2013 cohort</b>	-139 (-243, -35)	-109 (-225, 8)	-76 (-154, 1)	-72 (-150, 6)
<b>2014 cohort</b>	-36 (-122, 50)	-18 (-134, 99)	-70 (-144, 4)	-43 (-118, 32)

<sup>a</sup>A negative differential change in spending indicates savings

<sup>b</sup>Adjusted for patient characteristics

<sup>c</sup>Sample limited to beneficiaries continuously enrolled in fee-for-service Medicare from 2009-2015 who were attributable to an ACO or non-ACO TIN in 2015 and at least one year from 2009-2011 (and thus could contribute to estimation of a differential change in a model with patient fixed effects). We further excluded decedents, long-term nursing home residents, and beneficiaries enrolled in hospice so that the large increase in spending in 2015 displayed in Figure 1 would not be even larger (i.e., to limit potential bias from allowing these health declines in at the end of the study period but not earlier).

**Table 6. Comparison of balance on patient characteristics in within-area patient-level analysis vs. area-level analysis**

Patient Characteristic	Differential change from pre-entry period to 2015 associated with MSSP exposure	
	Primary approach (exposure = patient attribution to MSSP ACO in 2015) <sup>a</sup>	Area-level approach (exposure = 100% MSSP penetration in HRR in 2015)
Age (yr)	0.1	0.1
Female sex (%)	-0.2	-0.6
Race or ethnic group <sup>c</sup> (%)		
Non-Hispanic white	-0.1	-1.5
Non-Hispanic black	0.0	0.0
Hispanic	0.0	0.1
Other	0.0	1.3
Medicaid recipient (%)	-0.1	0.4
Disability was original reason for Medicare eligibility (%)	-0.2	-0.7
End-stage renal disease (%)	0.0	-0.2
CCW conditions, <sup>d</sup> no.		
Through prior year	0.0	-0.1
Through 3 years prior <sup>e</sup>	0.0	0.0
HCC risk score <sup>f</sup>		
Based on claims in prior year	0.01	-0.04
Based on claims 3 years prior <sup>e</sup>	0.00	-0.03
ZCTA-level characteristic		
% Below federal policy level	0.0	0.0
% With high school diploma	-0.1	0.0
% With college degree	-0.1	0.2

<sup>a</sup>Estimates from Table 2, pooled across 2012-2014 entry cohorts



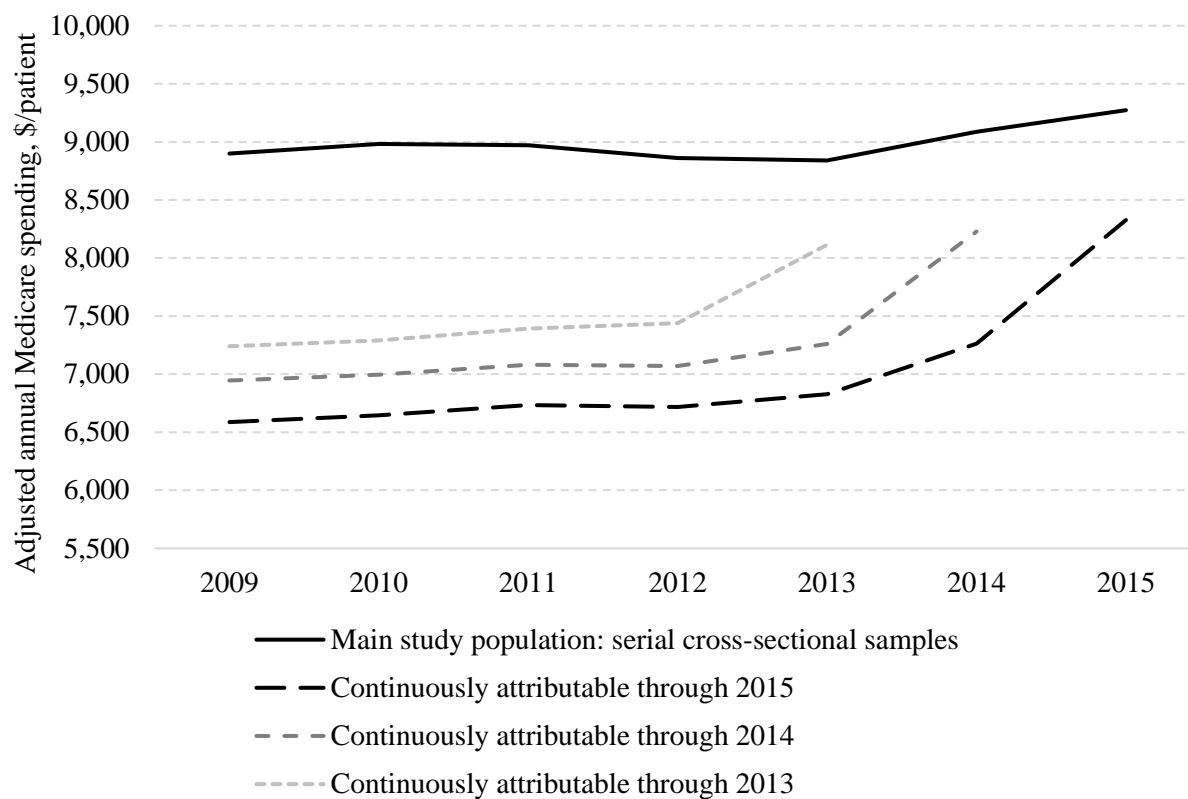
**Table 7. Impact on estimated savings of allowing provider composition of ACOs to change over time<sup>a</sup>**

Entry cohort	Estimated gross savings in 2015 <sup>a</sup> (differential change in spending from pre-entry period to 2015 for ACOs vs. control group), \$/patient (95% CI)					
	ACOs defined as sets of TINs			ACOs defined as sets of PCPs		
	Composition fixed over time, adjusted <sup>b</sup>	Composition allowed to change, adjusted <sup>b</sup>	Composition allowed to change, unadjusted	Composition fixed over time, adjusted <sup>b</sup>	Composition allowed to change, adjusted <sup>b</sup>	Composition allowed to change, unadjusted
<b>2012 cohort</b>	-286 (-431, -142)	-296 (-455, -136)	-307 (-563, -50)	-303 (-408, -199)	-257 (-367, -146)	-243 (-385, -102)
<b>2013 cohort</b>	-115 (-223, -7)	-62 (-185, 61)	44 (-143, 230)	-167 (-274, -60)	-146 (-251, -41)	-120 (-263, 23)
<b>2014 cohort</b>	-33 (-122, 56)	-116 (-308, 76)	-215 (-456, 26)	-52 (-143, 39)	-12 (-205, 182)	-107 (-334, 120)

<sup>a</sup>Limited to ACOs still participating in 2015 for which time-varying ACO definitions are available in the MSSP Provider-level RIF

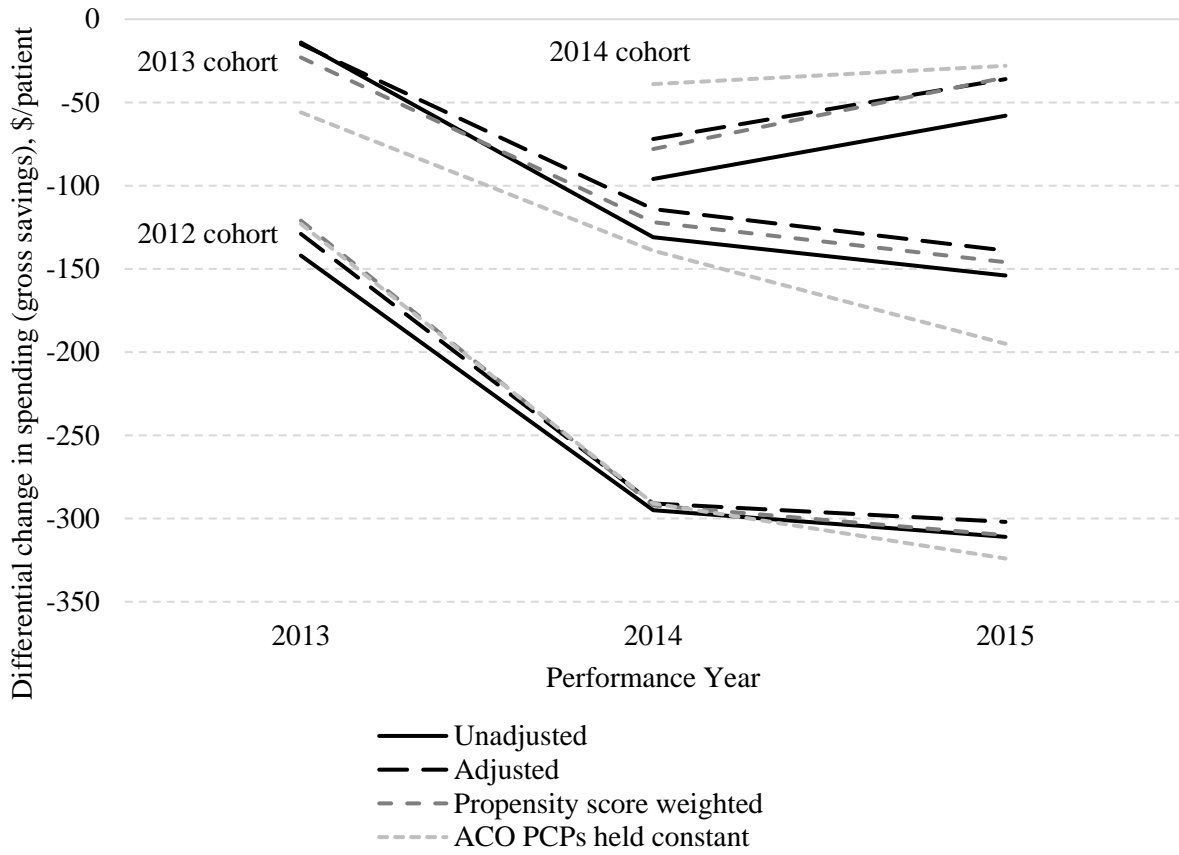
<sup>b</sup>Adjusted for patient characteristics

**Figure 1. Adjusted annual per-patient Medicare spending from 2009-2015 for serial cross-sectional samples vs. longitudinal cohorts of continuously attributable patients<sup>a</sup>**

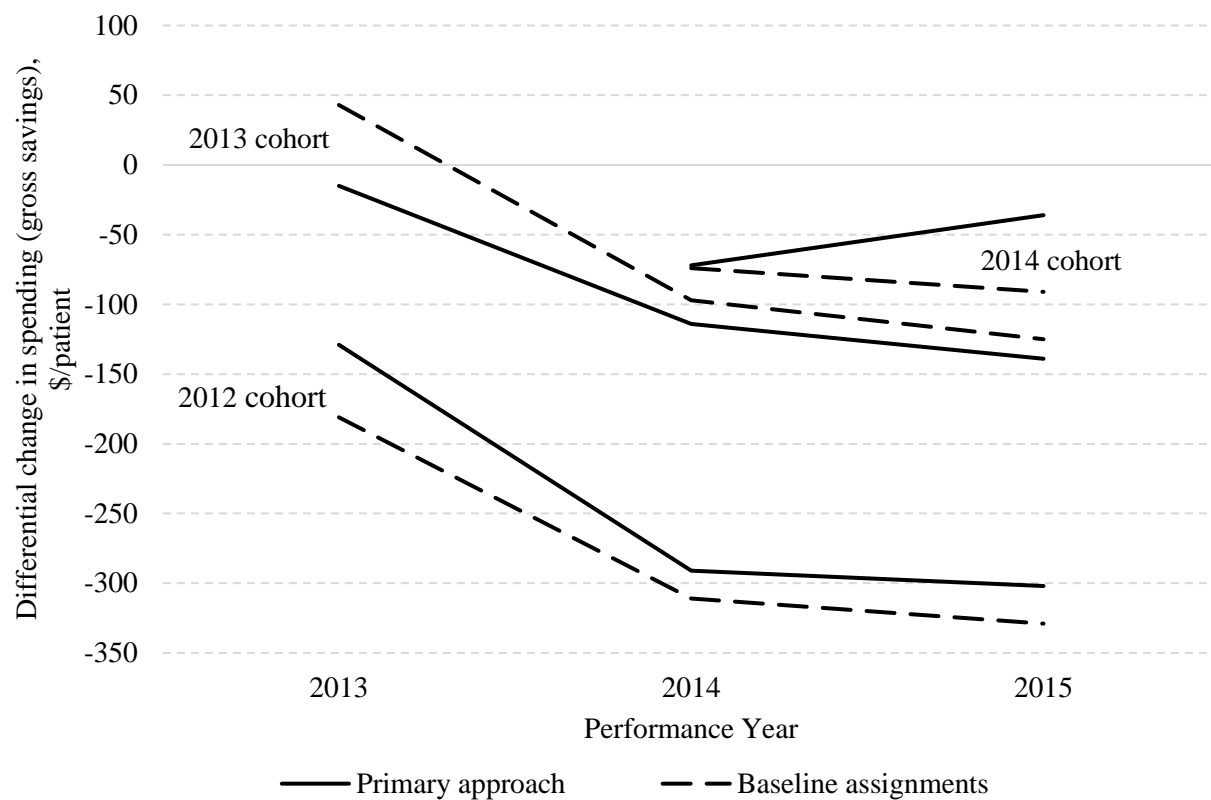


<sup>a</sup>Decedents excluded throughout study period in all groups. Continuously attributable cohorts are subgroups of the main study population who have at least one office visit with a PCP in each year to support attribution.

**Figure 2. Effects of adjusting patient characteristics and holding PCP composition of ACOs constant on savings estimates in main analysis**



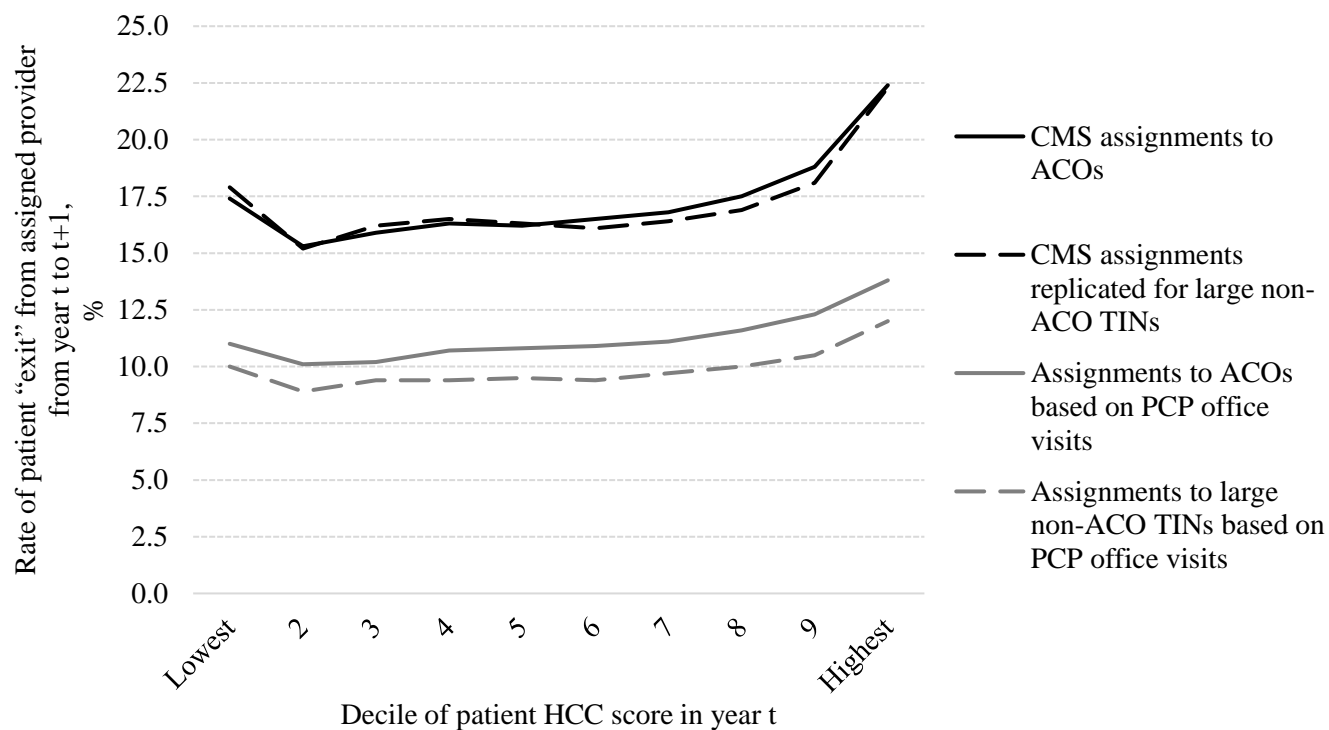
**Figure 3. Comparison of savings estimates from primary approach vs. approach holding baseline assignments constant**



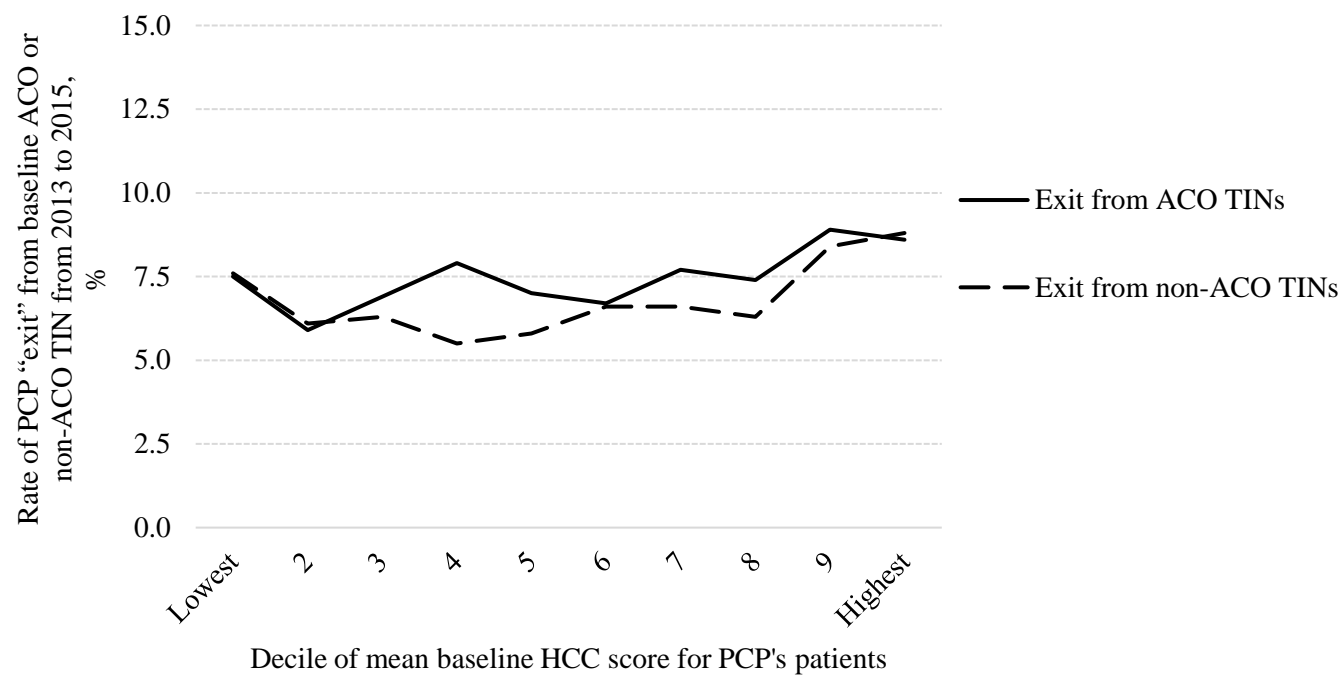
<sup>a</sup>Estimates from analysis holding baseline assignments fixed over study period were inflated to correct for attenuation bias as described in the methods section.

**Figure 4. Patient and PCP exit from ACOs and large non-ACO TINs by decile of patient risk**

**Panel A. Patient exit**



**Panel B. PCP exit**



**Appendix Table 1. Number of qualifying services received and number of TINs “competing” for assignment, by decile of per-beneficiary spending<sup>a</sup>**

<b>Decile of spending</b>	<b>CMS attribution algorithm</b>		<b>Attribution based only on office visits with PCPs</b>	
	<b>Mean number of qualifying services received by beneficiary</b>	<b>Mean number of TINs providing qualifying services</b>	<b>Mean number of qualifying services received by beneficiary</b>	<b>Mean number of TINs providing qualifying services</b>
1 (lowest)	2.24	1.07	2.21	1.06
2	3.39	1.12	3.30	1.12
3	4.05	1.16	3.92	1.16
4	4.58	1.20	4.40	1.19
5	5.08	1.24	4.86	1.23
6	5.57	1.27	5.33	1.26
7	6.16	1.31	5.83	1.30
8	6.76	1.36	6.41	1.33
9	7.23	1.44	6.59	1.36
10 (highest)	9.20	1.68	6.53	1.37

<sup>a</sup>Based on 2011 data

**Appendix Table 2. Falsification tests of 439 non-ACO TINs large enough to participate in the MSSP and of 2015 MSSP entrants prior to entry**

Hypothetical entry year	Differential change in spending in hypothetical performance year, \$/patient		
	2013	2014	2015
<b>Large non-ACO TINs</b>			
2012	14	3	12
2013	13	2	11
2014	-	-1	8
<b>2015 MSSP entrants</b>			
2012	143**	93	-
2013	115**	65	-
2014	-	39	-

\*P<0.05 \*\*P<0.01 \*\*\*P<0.001

**Appendix Table 3. Application of baseline assignment approach in falsification tests treating large non-ACO TINs or 2015 MSSP entrants as hypothetical entrants in 2013**

	Pre-period (2009-2012) difference in spending vs. control group, \$/patient	Differential change in spending after hypothetical entry in 2013, \$/patient (inflated estimate)		
		2013	2014	2015
<b>Large non-ACO TINs</b>	39*	-3 (-5)	99** (174)	80** (150)
<b>2015 MSSP entry cohort of ACOs</b>	35	114** (186)	23 (40)	-

\*P<0.05 \*\*P<0.01 \*\*\*P<0.001



Appendix Table 4. Attribution based on referring PCP when beneficiary has no office visits with a PCP

Entry cohort	Unadjusted pre-entry sample mean, <sup>a</sup> \$/patient	Adjusted pre-entry difference in annual spending level between ACOs and control group, \$/patient (95% CI)	Adjusted pre-entry difference in annual spending trend between ACOs and control group, \$/patient (95% CI)	Estimated gross savings (adjusted differential change in spending from pre-entry period to performance year for ACOs vs. control group), <sup>b</sup> \$/patient (95% CI)		
				2013	2014	2015
<b>2012 cohort</b>	10,390	289 (63, 515)	-25 (-78, 28)	-191 (-329, -53)	-356 (-492, -219)	-408 (-561, -254)
<b>2013 cohort</b>	10,390	217 (58, 377)	-23 (-65, 18)	-37 (-141, 68)	-147 (-251, -44)	-251 (-356, -145)
<b>2014 cohort</b>	10,390	160 (10, 310)	8 (-21, 37)	-	-94 (-171, -16)	-86 (-179, 7)

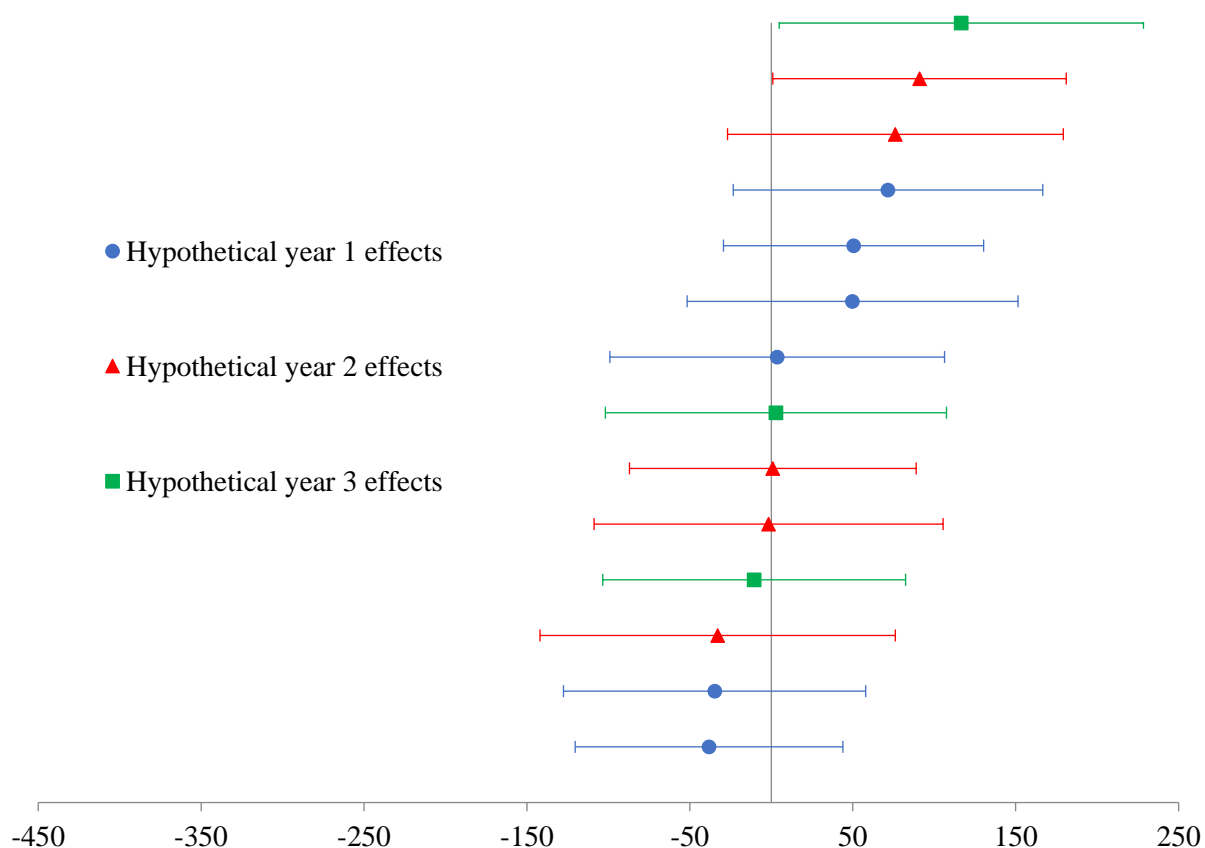
**Appendix Table 5. Tests of ACO gaming of CMS attribution algorithm via manipulation of TINs used for billing**

Entry cohort	Estimated gross savings in 2015 using CMS attribution algorithm (differential change in spending from pre-entry period to 2015 for ACOs vs. control group), \$/patient (95% CI)		
	ACOs defined as fixed sets of TINs over time		ACOs defined as fixed sets of NPIs over time
	Unadjusted for patient characteristics	Adjusted for patient characteristics	Adjusted for patient characteristics
<b>2012 cohort</b>	-465 (-665, -266)	-438 (-588, -289)	-364 (-495, -233)
<b>2013 cohort</b>	-179 (-383, 26)	-238 (-385, -91)	-130 (-263, 2)
<b>2014 cohort</b>	-57 (-234, 121)	-42 (-170, 87)	21 (-96, 138)

**Appendix Table 6. Comparison of pre-period differences in spending levels and trends by different methods of patient attribution**

	<b>Unadjusted pre-entry sample mean \$/patient</b>	<b>Adjusted pre-entry difference in annual spending level between ACOs and control group, \$/patient (95% CI)</b>	<b>Adjusted pre-entry difference in annual spending trend between ACOs and control group, \$/patient (95% CI)</b>
<b>Attribution based on PCP office visits</b>			
2012 cohort	9649	139 (-79, 357)	-3 (-58, 53)
2013 cohort	9649	31 (-84, 146)	-5 (-39, 29)
2014 cohort	9649	33 (-90, 155)	8 (-18, 34)
<b>Attribution based on PCP office visits or office visits with specialists if no visits with a PCP</b>			
2012 cohort	9778	188 (-20, 396)	-6 (-60, 47)
2013 cohort	9778	92 (-23, 206)	-1 (-35, 33)
2014 cohort	9778	95 (-38, 228)	6 (-19, 30)
<b>CMS attribution algorithm</b>			
2012 cohort	10,038	98 (-51, 247)	-74 (-138, -10)
2013 cohort	10,038	78 (-127, 282)	-64 (-112, -16)
2014 cohort	10,038	-65 (-227, 97)	14 (-27, 55)

**Appendix Figure 1. Falsification tests estimating differential changes in spending for ACOs hypothetically entering in pre-entry period years<sup>a</sup>**



<sup>a</sup>For each entry cohort of ACOs, we estimated all possible differential changes in pre-entry spending generated by treating each pre-entry year as a hypothetical entry year. For the 2013 entry cohort, for example, this allowed procedure generated 3 hypothetical effects of 1 year of participation (differential changes from 2009 to 2010, from 2009-2010 to 2011, and from 2009-2011 to 2012), 2 hypothetical year 2 effects (differential changes from 2009 to 2011 and from 2009-2010 to 2012) and 1 hypothetical year 3 effect (differential change from 2009 to 2012). Plotted are all the possible hypothetical effects across the 2012-2014 entry cohorts. The estimates were distributed evenly around zero, with a slight skew toward positive differential changes rather than negative differential changes. Moreover, hypothetical year 2 effects were not systematically more negative than hypothetical year 1 effects, and hypothetical year 3 effects were not systematically more negative than hypothetical year 1 or 2 effects.