SUPPLEMENTARY APPENDIX

INTRODUCTION

This supplementary appendix expands the discussion of several technical details mentioned in the text.

DETAILS ON SAMPLE CONSTRUCTION AND ATTRIBUTION

As noted in the text, due to CMS redaction of 2013 and 2014 files under the DUA for this paper, claims were redacted in all years for consistency. Our own work suggests that redaction does not affect patient attribution to an ACO nor estimates of total spending in cohorts, like ours, that are unrestricted by illness or age of Medicare beneficiary [1, 2].

In regression (4), we address the group of beneficiaries that were unattributed to an ACO due to a lack of qualifying claims used in attribution. Excluding non-attributed beneficiaries from longitudinal analyses could overstate average spending by excluding beneficiaries with zero spending in the post-ACO period. We reversed this exclusion by attributing patients enrolled in continuous fee-for-service Medicare to the last provider to which they could be attributed following CMS rules. Thus, we assigned zero spending to unattributed beneficiaries if three conditions were met:

- the beneficiary was enrolled in fee-for-service Medicare in that year;
- the beneficiary was not attributed to a provider in that year; and
- spending was missing for that beneficiary.

Appendix Figure A1 shows that only a small percentage of beneficiaries, 0.6%, were affected by this process of imputing zero spending when no claims with positive spending were present in a year.

Appendix Figure A1 shows that the vast majority of beneficiaries (86.4%) had no gaps in their time-series records for any reason. The gaps appear to be mostly due to the beneficiaries' enrollment being only partial in Medicare Parts A and B in the relevant year. As Appendix Figure A2 shows, switches from fee for service into Medicare Advantage was quite rare in our sample (only 0.7% of beneficiaries making this switch at some point in our sample period).

To identify these continuing beneficiaries, we fixed an anchor year, 2011, and required positive spending for those beneficiaries prior to that anchor year, i.e. in 2009 or 2010. Restricting the sample to beneficiaries with positive 2009 or 2010 spending excluded 14% of beneficiary-year observations compared to the baseline sample. To avoid bias from selection on positive spending in that first year, we further restricted the sample by dropping the first observation year (2009 or 2010) for each beneficiary, excluding 21% of the remaining beneficiary-year observations. These procedures restored beneficiary-year observations among those with positive spending in 2009 or 2010 for any year missing spending through 2014.

We attributed beneficiaries to provider organizations in each year following CMS MSSP rules [29-32], whenever the beneficiaries' plurality of allowed charges from qualifying primary care services were delivered by an eligible healthcare professional affiliated with that organization. While Pioneer's attribution rules were slightly different from MSSP's [31], the difference is immaterial for 90% of beneficiaries attributed to Pioneer ACOs. The organization to which a beneficiary was attributed could change from year to year, as in these contracts.

DETAILS ON ORDINARY LEAST SQUARES REGRESSIONS

Our baseline regression implements a difference-in-differences design following prior studies [3-5]. Letting i represent beneficiaries, t years, j the organization to which i is attributed in year t, and h the HRR in which i resides in year t, the regression can be written as

$$\begin{aligned} E(Y_{it}) = {} & \beta_1(\mathit{2012\,Pioneer}_{jt} \times \mathit{Participation\,Year\ 1}_{jt}) \\ & + \beta_2(\mathit{2012\,Pioneer}_{jt} \times \mathit{Participation\,Year\ 2}_{jt}) \\ & + \beta_3(\mathit{2012\,Pioneer}_{jt} \times \mathit{Participation\,Year\ 3}_{jt}) \\ & + \beta_4(\mathit{2012\,MSSP}_{jt} \times \mathit{Participation\,Year\ 1}_{jt}) \\ & + \beta_5(\mathit{2012\,MSSP}_{jt} \times \mathit{Participation\,Year\ 2}_{jt}) \\ & + \beta_6(\mathit{2012\,MSSP}_{jt} \times \mathit{Participation\,Year\ 3}_{jt}) \\ & + \beta_7(\mathit{2013\,MSSP}_{jt} \times \mathit{Participation\,Year\ 1}_{jt}) \\ & + \beta_8(\mathit{2013\,MSSP}_{jt} \times \mathit{Participation\,Year\ 2}_{jt}) \\ & + \beta_9(\mathit{2014\,MSSP}_{jt} \times \mathit{Participation\,Year\ 1}_{jt}) \\ & + \boldsymbol{x}'_{it}\boldsymbol{\delta} + \theta_{ht} + \alpha_{j(\mathrm{ACO})}. \end{aligned} \tag{A1}$$

On the left-hand side, E denotes an expected value, and Y_{it} denotes total Medicare spending for beneficiary i in year t. On the right-hand side, $\mathit{2012\,Pioneer}_{jt}$, $\mathit{2012\,MSSP}_{jt}$, $\mathit{2013\,MSSP}_{jt}$, and $\mathit{2014\,MSSP}_{jt}$ are indicators (functions equaling 1 in the stated event and 0 otherwise) for beneficiary i being served in year t by a provider organization j that joins the indicated ACO cohort when it signs an ACO contract. The ACO cohort variables allow cohort status to be prospective: for example, $\mathit{2012\,Pioneer}_{jt}$ is set to 1 if the provider organization j signs up for the Pioneer program in 2012 even for observation year $t = 2009$, $t = 2010$, or $t = 2011$. The variables $\mathit{Participation\,Year\ 1}_{jt}$, $\mathit{Participation\,Year\ 2}_{jt}$, and $\mathit{Participation\,Year\ 3}_{jt}$ are indicators for t being the specified participation year in an ACO contract, where participation year

1 represents the year that j enters an ACO contract, participation year 2 represents the next year, and participation year 3 the next. The main object of our study are estimates of the coefficients β_1 through β_9 on the interactions between the indicators; these coefficients capture savings due to ACO participation. The flexible specification, having a different coefficient on each interaction term, allows estimated savings effect to differ by program, cohort, and participation year.

Additional covariates on the right-hand side include a suite of beneficiary-level characteristics represented by the vector $\boldsymbol{x}_{it}$. This vector includes a set of indicators for age interval (64 and below serving as the omitted reference category, as well as the intervals 65–69, 70–74, 75–79, 80–89, 90–94, 95 and above), an indicator for sex (female), indicators for race/ethnicity (white serves as the omitted reference category as well as black, Hispanic, Asian and Pacific Islander, and Other), an indicator for living in a high-poverty census tract, an indicator for dual eligibility for Medicaid benefits (either full or partial), an indicator for disability status as a basis of original entitlement. The vector also includes a set of indicators for the number of hierarchical condition category (HCC) diagnoses (1, 2, 3, 4 or more, or no information due to no visits and the omitted reference category being 0). Vector $\boldsymbol{\delta}$ represents coefficients on these variables to be estimated.

The right-hand side of (A1) also includes θ_{ht}, representing a suite of fixed effects for HRR-year interactions, and $\alpha_{j(\text{ACO})}$, representing a suite of fixed effects for each ACO (these are 352 individual ACOs, as distinct from a cruder indicator for belonging to an ACO cohort or to any ACO).

The only notable departure of the baseline regression from prior studies is that we imposed the same size restriction on our sample of non-ACOs as the programs place on ACOs (serving at least 5,000 beneficiaries). While we sought to improve the comparability of the reference group to

ACOs, the restriction turned out to be inconsequential since, despite the restriction, baseline estimates have overlapping confidence intervals with previous research taking this difference in difference approach, as will be seen in Appendix Figures A7 and A8. Other minor departures from individual prior studies also turn out to be inconsequential. In results not reported, we find that estimates from equation (A1) match those from a regression restricting the sample to MSSP participants and match results from another regression limiting the sample to beneficiaries attributed exclusively through primary care.

Regression (2) augments the baseline by including individual fixed effects $\alpha_{j(\text{non-ACO})}$ for 474 non-ACO organizations. Because beneficiaries who follow their physicians tend to have higher predicted spending [6], if beneficiaries switching out of ACOs tend to move to the higher spending (as determined by the organization's attributed beneficiaries' service use, risk profile, and spending) among non-ACOs, this could bias the results toward finding ACO efficiencies. Even if there was no relative change in provider efficiency, the compositional effect of an increase in the number of beneficiaries selecting the higher spending among non-ACOs could be measured as a relative decline in the efficiency of non-ACOs. To control for such compositional effects, we added fixed effects for each non-ACO provider organization.

Regression (3) adds a suite of beneficiary fixed effects, γ_i. This straightforward addition does not require further explanation except to note that the proper procedure for including fixed effects at this level entails dropping any singleton observations (beneficiaries having only one year of data in the sample) before running the regression. The coefficient estimates are the same whether or not the singleton observations are dropped but dropping them ensures standard errors are computed correctly. Adding beneficiary fixed effects narrows the sources of identifying variation down to two: first, spending changes (compared to controls) for a given beneficiary who remains with a provider organization before and after that organization enters an ACO and, second, spending changes for a given beneficiary whose provider organization switches from a control to an organization in the ACO program or from an ACO to a control. The method ignores changes in a provider organization's spending due to changes in its beneficiary mix.

Regression (4) takes measures to address truncation of zero spending beneficiaries. For beneficiary-years in which a beneficiary was enrolled in fee-for-service Medicare, but unattributed because of a lack of qualified Medicare spending, we attribute the beneficiary to the last provider organization to which CMS rules allow definitive attribution. Beneficiary-year observations gaining attribution via this procedure that have no spending are assigned $Y_{it} = 0$. Beneficiary-year observations gaining attribution via this procedure that have positive recorded spending keep that value for Y_{it}.

This method allows us to impute zero spending only for continuing patients who have positive spending in some previous year; otherwise there is no link to a previous provider to impute attribution. In anticipation of our IV procedure which requires beneficiaries to be observed in 2011, the year used to construct instrumental variables, we restrict the sample just to beneficiaries who are continuing patients in 2011. This allows the possibility of imputing $Y_{i\,2011} = 0$ for any beneficiary in the restricted sample, which would avoid bias from selecting on positive spending in the anchor year. For us to know that the patient is continuing in 2011, we must observe positive spending for them in some earlier year in the sample, 2009 or 2010. Thus, our sample restriction amounts to the requirement that beneficiaries have positive spending in either 2009 or 2010, leading us to drop 12,472,274 beneficiary-year observations.

By construction, the restricted sample includes beneficiaries having positive spending the first year they appear in the sample. To avoid the bias from selection on positive spending in that first year, we further restrict the sample by dropping the first observation year (2009 or 2010) for each beneficiary. This additional restriction resulted in 11,945,951 beneficiary-year observations being dropped.

All regressions use two-way clustering for standard errors, clustering at both the beneficiary and HRR level.

Details on Fixed-Attribution and Instrumental-Variables Methods

A subtler beneficiary-selection problem arises if beneficiaries change providers in anticipation of changing healthcare needs. Beneficiary fixed effects do not account for within-beneficiary changes in spending over time, and may even exacerbate the bias from this form of selection [5]. If changing healthcare needs are correlated with changes in observed factors, our beneficiary controls help address this selection problem. However, spending changes may relate to patient preferences or more subtle clinical differences not observed in claims.

To address beneficiary selection on spending changes, we took two related approaches. The paper reports a fixed-attribution (FA) method, which involves picking an anchor year prior to the formation of ACOs and fixing beneficiaries' attributed organization to the one in that anchor year regardless of whether the beneficiary later switches. Regression (5), reported in Table 4 in the paper, implements this method using 2011, the year before participation started in ACOs, as the anchor year. The FA method can be thought of as an intent-to-treat (ITT) estimator because not everyone who receives the "treatment" of being attributed in the anchor year to an organization that will become an ACO will remain to receive that treatment in the participation year.

A related instrumental-variables method effectively converts the ITT estimator to a treatment-on-the-treated (TT) one. IV uses the beneficiary's attribution in the anchor year as an instrument for later attribution. It can be shown that the IV estimator scales the FA estimator by an estimate of the reciprocal of the probability of remaining as initially attributed in the later participation year. Regression (5'), reported in Appendix Table A1, implements the IV approach using 2011 as the anchor year.

To better understand our IV approach used in regression (5'), consider the example of instrumenting for the following representative interaction term:

$$\mathit{2013\,MSSP}_{jt} \times \mathit{Participation\,Year}\ 2_{jt}. \tag{A2}$$

The second factor (and thus the whole expression) equals 0 unless $t = 2014$ because 2014 is the second participation year for the 2013 MSSP cohort. Call 2014 the *target year*. Rather than using the whole dataset, the first stage can equivalently just use target-year data, regressing $\mathit{2013\,MSSP}_{j\,2014}$, which is the indicator for the event that i is attributed in the target year to a provider in the 2013 MSSP cohort, on $\mathit{2013\,MSSP}_{j\,2011}$, which is the indicator for the event that i is attributed in 2011 to a provider that will join the 2013 MSSP cohort. Call 2011 the *anchor year*; it has special significance as the last year in the pre-ACO period, used to construct instruments for all the interaction terms analogous to (2). In effect, the IV procedure uses the beneficiary's attribution in the anchor year to predict attribution in the target year. Our IV approach narrows the sources of variation used to identify savings, in effect ignoring variation arising from beneficiaries' switching from provider organizations in the control group into ACOs, as such switches may be caused by spending changes that would bias estimated ACO savings. Our IV approach ignores this source of variation by mapping beneficiaries back to their provider organizations before ACO contracts began. The first-stage regression also includes other target-year covariates: a vector of 2014 beneficiary level characteristics ($\boldsymbol{x}_{i\,2014}$) and HRR fixed effects ($\theta_{h\,2014}$). Following a two-stage least squares procedure, the fitted value $\widehat{\mathit{2013\,MSSP}}_{j\,2014}$ of the dependent variable from the first-stage regression just described is entered in a second-stage regression in place of the interaction term in equation (2).

As discussed in the text, dynamics in beneficiary spending (including mean reversion, whereby beneficiaries with high spending needs one year tend to revert to more typical spending levels the next) can bias FA estimates. This factor can likewise bias IV estimates for similar reasons. To test the robustness of our FA and IV approaches, we shift the anchor year one year prior to our initial choice of 2011, thus instrumenting for beneficiary attribution to an ACO cohort using 2010 attribution. Regressions (5) and (5') use the 2011 anchor year (for FA and IV methods, respectively) and (6) and (6') use the 2010 anchor year.

To properly implement this approach, slightly different sample restrictions are needed in (6) than (5). Regression (5) requires definitive evidence that the beneficiary was continuing in the 2011 anchor year, requiring positive spending for the beneficiary in some previous year in our sample, either 2009 or 2010. With the anchor year moved back to 2010 in regression (6), definitive evidence that the beneficiary was continuing is required for 2010, which requires positive spending in the only prior year left in our sample, 2009. In the presence of spending dynamics that could bias our IV estimates, the different lags should interact with different points in the dynamic spending process, producing different estimates in (5) and (6); absent such spending dynamics, the estimates should be relatively similar. In sum, (5) includes beneficiary with positive spending in at least one of 2009 or 2010, (6) drops those that only have positive spending in 2010 but not 2009. We impute $Y_{i\,2010} = 0$ for any beneficiary in the restricted sample with missing spending used for regression (6) and drop the first sample year (2009) for them. All these sample restrictions leave us with 11,624,345 fewer beneficiary-year observations in (6) than (5). The comparison of the samples behind (5’) and (6’) is identical because the IV estimator imposes the same sample restrictions as the FA estimator.

SCENARIOS ILLUSTRATING REGRESSION SAMPLES

Moving from regression (1) to (6) involves increasingly restricted samples discussed in the text. Several of the sample restrictions are complex by themselves, and layering them on top of other restrictions is yet more complex. As an aid to the reader, Appendix Table A2 illustrates which observations are included in which regression through several hypothetical scenarios.

PLACEBO AND PRE-TRENDS TESTS

Appendix Table A3 returns to the baseline specification but instead of just estimating post-ACO treatment effects, estimates treatment effects for organizations that will later become ACOs in the pre-period before organizations in that cohort could form ACOs. These function as placebo tests. We see that only one of the 11 estimates in the pre-period is significant, that in 2010, so four years prior to ACO formation, for 2014 MSSPs. This seems to be somewhat of an anomaly, preceding the formation of ACOs by that cohort by a considerable span, and not showing a consistent trend in placebo coefficients that could generate the later estimates.

Note that the coefficients in the post-ACO period closely match those estimated in regression (1) of Table 4. This is not a foregone conclusion as Appendix Table A3 uses just one reference year, the year prior to ACO formation, where Table 4 uses that an all preceding years as the reference period. This adds evidence that secular trends do not drive instability in the results.

In regressions not reported, we tested for linear trends in the pre-period and obtained fairly precise zeros. After absorbing these pre-trends, our coefficients in the post-ACO period remained about the same size as estimated in baseline regression (1). We thus have no evidence that our results are driven by underlying pre-trends.

Pooled Results

Robustness to Different ACO Fixed Effect Definitions

Following previous research, the baseline regression included an indicator for each of the 352 ACO provider organizations (without an analogous indicator for each non-ACO organization). We tested robustness of this specification against two separate variants: one replacing the 352 ACO indicators with a single aggregate ACO indicator, and another with an indicator for each of the 4 ACO cohorts, pooling the separate cohort and participation year estimates in into a single ACO effect to provide a summary measure. Figure A3 shows very stable estimates across specifications.

In the appendix, all tables of regression results report two panels, one with separate cohort and participation year estimates, and one pooling those estimates into a single ACO effect. For space considerations, the main table of regression results in the paper, Table 4 just reports the separate cohort by participation year results. To complement those results, Appendix Table A4 reports the analogous pooled results. These are the estimates graphed in Figure 1 in the paper.

Robustness to Sample Restrictions

The text claimed that the sample restrictions were not driving changes in results. To support this claim, Appendix Table A5 reports regressions maintaining the specification from the baseline regression (1) but run on the smaller samples used for the other regressions. The first column of results repeats of the baseline (1) results from Table 4. Note that the second column of results in

Appendix Table A5 is the same as the first, reinforcing the point that the samples used in regressions (1) and (2) are the same.

Looking across each row, the results do not change appreciably with different samples, suggesting the sample restrictions are not driving the changes in results seen in Table 4. The point is even clearer in Appendix Figure A4, which displays the pooled ACO effects from Appendix Table A5. The figure shows that the estimates and confidence intervals hardly budge across samples.

Alternative Sequencing of Methods

The results in Table 4 and Figure 2 show how introducing a sequence of statistical methods changes the results. The order in which the methods were introduced is somewhat arbitrary. A natural question is whether the effect of the method would be different in a different order. Order may matter if the methods have a complex interaction, say being substitutes or complements for each other.

To investigate order effects, Appendix Table A6 reports the results from adding the methods in regressions (2)–(6) to the baseline rather than cumulatively. Appendix Figure A5 displays the pooled ACO effects. For reference, both table and figure repeat baseline specification (1), which remains unchanged from the text.

Comparing Figure 2 and Appendix Figure A5, we see that they are different. The FA and IV methods in regressions (5) through (6’) tend to increase estimated savings when added to the baseline but have the reverse effect when layered on later, after the other methods. This phenomenon would be observed if adding beneficiary fixed effects magnifies the bias due to selection on spending changes, and bias is reversed by FA and IV estimators, which control for this type of dynamic selection

Overall, we see that the alternate order of imposing the methods on top of the baseline does not eliminate swings in the results. The swings show a different pattern across regressions but are apparent nonetheless. The analysis in this section shows that the sequence in which the methods were introduced in the text was not chosen to exaggerate the instability of the estimates.

Heterogeneity of Estimates by Hospital Integration

Although our baseline regression (1) was modeled on McWilliams et al.[5], one might worry that the results are not directly comparable because of several remaining specification differences. First, they omit the Pioneer ACO cohort. Second, they break the ACO effect further into that for ACOs integrated with a hospital and not, whereas we estimate an ACO effect averaged across these categories. Third, they use spending at the primary-care physician for beneficiary attribution, where we attribute based on all Medicare spending.

Appendix Table A7 shows the results for a new baseline that moves further toward McWilliams et al.[5] in these three dimensions. The baseline results in (1) are quite close to their main results in their Figure 1. The remaining columns add the statistical method controlling for selection on top of this new baseline. The effect of introducing these methods is perhaps better seen in Appendix Figure A6, which displays the estimates from the panel in Appendix Table A7 that pools the estimates into a single ACO effect (although still broken out by hospital-integrated ACOs and not in the two panels). The significant savings they found for ACOs formed by physician-group practices not integrated with hospitals, which we replicate in the new baseline (1), disappear in all subsequent specifications (2)–(6'). The zero savings they found for hospital-integrated ACOs, which we replicate in our new baseline (1), become spending increases in specifications (2)–(4), swinging back to significant savings in FA and IV specifications (5)–(6’).

It turns out that which type of spending is used for attribution is not consequential for the results. To demonstrate this, we reproduced Appendix Table A4 but, as in the text, used all Medicare spending rather than just primary care for beneficiary attribution. The results, not reported, are quite similar.

Appendix Figures A7 and A8 return to the baseline specification in Appendix Table A7 but plot the estimates (both broken out by ACO cohort and participation year and, at the bottom of the figures, pooled estimates) alongside the analogous ones from McWilliams et al. (2018) for comparison. With few exceptions across the 12 different cases, the estimates and confidence intervals line up closely.

References

1. Austin AM, Bynum JPW, Maust DT, Gottlieb DJ, Meara E. Long-Term Implications Of A Short-Term Policy: Redacting Substance Abuse Data. Health Aff (Millwood). 2018;37(6):975-9. doi:10.1377/hlthaff.2017.1524.
2. Ouayogodé MH, Meara ER, Chang C-H, Raymond SR, Bynum JP, Lewis VA et al. Forgotten patients: ACO attribution omits low-service users and the dying. American Journal of Managed Care. 2018;24(7):e207-e15.
3. Colla CH, Lewis VA, Kao L-S, O’Malley AJ, Chang C-H, Fisher ES. Association between Medicare accountable care organization implementation and spending among clinically vulnerable beneficiaries. JAMA Internal Medicine. 2016;176(8):1167-75.
4. McWilliams JM, Hatfield LA, Chernew ME, Landon BE, Schwartz AL. Early performance of accountable care organizations in Medicare. New England Journal of Medicine. 2016;374(24):2357-66.
5. McWilliams JM, Hatfield LA, Landon BE, Hamed P, Chernew ME. Medicare spending after 3 years of the Medicare Shared Savings Program. New England Journal of Medicine. 2018;379(12):1139-49.
6. Hsu J, Vogeli C, Price M, Brand R, Chernew ME, Mohta N et al. Substantial physician turnover and beneficiary ‘churn’in a large Medicare Pioneer ACO. Health Affairs. 2017;36(4):640-8.

Table A1
Instrumental variables (IV) regressions of effect of ACO participation on Medicare spending

ACO Cohort	Regression (5')	Regression (6')
2012 Pioneer		
Participation year 1	-62.7	-14.3
	(51.3)	(64.9)
Participation year 2	-2.5	-25.2
	(80.2)	(106.6)
Participation year 3	-242.7 *	-328.0 **
	(115.1)	(125.3)
2012 MSSP		
Participation year 1	-4.3	-55.5
	(46.0)	(46.5)
Participation year 2	-16.7	-192.3 **
	(53.1)	(62.7)
Participation year 3	-191.8 **	-337.4 **
	(72.3)	(77.9)
2013 MSSP		
Participation year 1	-57.4	-161.3 **
	(46.0)	(55.5)
Participation year 1	-235.5 **	-420.0 **
	(63.9)	(79.2)
2014 MSSP		
Participation year 1	-36.5	-102.8
	(46.9)	(52.7)
Pooled ACO Estimate		
ACO × Post period	-65.3 *	-142.7 **
	(28.3)	(34.9)
Specification		
Beneficiary-level covariates	Yes	Yes
HRR-year fixed effects	Yes	Yes
ACO fixed effects	Yes	Yes
Non-ACO fixed effects	Yes	Yes
Beneficiary fixed effects	Yes	Yes
Address zero-spending truncation	Yes	Yes
IV using attribution to anchor year	Yes (2011)	Yes (2010)
Observations		
Beneficiary level (# of clusters)	11,656,202	10,094,293
Beneficiary level	13,581,898	11,999,258
Beneficiary-year level	54,039,875	47,846,537

Notes: Instrumental variables (IV) regressions estimated using two-stage least squares. Regression (5') instruments for attribution to an ACO cohort using 2011 attribution and (6') using 2010 attribution.Specificiations are otherwise the same as regression (4) from Table 4; see the notes from that table for further information. Regressions require different sample restrictions to address truncation of zero spending: (5') requires beneficiaries to have positive spending in either 2009 or 2010, while (6') requires positive spending in 2009, resulting in fewer observations. The F-statistic for the test of joint significance of instruments in the first stage for regression (5') is F=3.38, p<.001. The panel labled "pooled ACO estimate" re-runs the same regressions as in the panel above but estimates a pooled ACO effect across all ACO cohorts and participation years. Standard errors reported in parentheses adjust for two-way clustering on beneficiaries and HRRs. Significantly different from 0 in a two-sided test at the *5%, **1% level.

Table A2
Scenarios illustrating inclusion in various regression samples

Scenario	Included Beneficiary-Year Observations			
	Regressions (1)–(3)	**Regression (4)**	**Regression (5)**	**Regression (6)**
Beneficiary A has qualifying spending allowing attribution to a provider organization throughout whole sample period.	2009-2014.	2010-2014. First observation year (2009) dropped.	Same as previous column: 2010-2014.	Same as previous column: 2010-2014.
Beneficiary B ages into Medicare in 2010. Has qualifying spending allowing attribution to a provider organization	2010-2014.	2011-2014. First observation year (2010) dropped.	Same as previous column: 2011-2014.	Beneficiary entirely dropped because lacks required spending in 2009
Beneficiary C has qualifying spending allowing attribution to a provider organization in 2009 and 2010 but dies in 2010.	2009-2010.	2010. First observation year (20009) dropped.	Beneficiary entirely dropped from this restricted sample because not observed in anchor year, 2011.	After dropping initial year (2009), left with singleton observation (2010), which is dropped as appropriate when using beneficiary fixed effects.
Beneficiary D has qualifying spending allowing attribution to a provider organization in 2009 and from 2011 to 2013, exiting the panel before 2014.	2009, 2011-2013.	2010-2013. First observation year (2009) dropped. Zero spending assigned for 2010 and attributed to 2009 provider organization.	Same as previous column: 2010-2013.	Same as previous column: 2010-2013.

Table A3
Placebo tests of ACO effects in pre-ACO period

	ACO Cohort			
Participation year	**2012 Pioneer**	**2012 MSSP**	**2013 MSSP**	**2014 MSSP**
Pre-ACO Period				
-5				14.3
				(71.1)
-4			99.1	-138.2 **
			(77.4)	(52.3)
-3	30.2	106.5	-33.3	22.9
	(91.3)	(73.8)	(63.5)	(46.7)
-2	-21.7	33.4	-17.3	-19.7
	(66.3)	(45.3)	(49.7)	(40.7)
Reference Year				
-1	—	—	—	—
Post-ACO Period				
1	74.3	-91.7 *	-113.7 **	-51.2
	(59.8)	(42.5)	(41.6)	(35.5)
2	-150.2 *	-152.4 *	-186.6 **	
	(72.7)	(60.8)	(51.7)	
3	-447.4 **	-295.8 **		
	(86.0)	(51.0)		
Joint Test of Pre-ACO Period Effects				
Test distribution	$F(2, 305)$	$F(2, 305)$	$F(3, 305)$	$F(4, 305)$
Test statistic	0.45	1.04	1.34	9.23**

Notes: Re-runs baseline regression (1) from Table 4, adding interactions between an indicator for beneficiary attribution to provider organization that will join ACO program and each year preceeding participation. Year before ACO participation (-1) treated as reference year for each ACO cohort (2011 for 2012 Pioneer and 2012 MSSP, 2012 for 2013 MSSP, and 2013 for 2014 MSSP). Although results presented in different columns to aid visualization, they are jointly estimated in one regression. See notes to Table 4 for details on specification and number of observations in baseline regression (1), which are identical here. Standard errors reported in parentheses adjust for two-way clustering on beneficiaries and HRRs. Significantly different from 0 in a two-sided test at the *5%, **1% level.

Table A4
Pooled estimate of effect of ACO participation on Medicare spending

Pooled ACO Estimate	**Regression (1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
ACO × Post period	-142.1 **	106.3 **	31.1	17.5	-45.6 *	-85.5 **
	(35.4)	(31.8)	(25.3)	(22.6)	(19.8)	(21.0)
Specification						
Beneficiary-level covariates	Yes	Yes	Yes	Yes	Yes	Yes
HRR-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
ACO fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Non-ACO fixed effects	No	Yes	Yes	Yes	Yes	Yes
Beneficiary fixed effects	No	No	Yes	Yes	Yes	Yes
Address zero-spending truncation	No	No	No	Yes	Yes	Yes
Fixed attribution to anchor year	No	No	No	No	Yes (2011)	Yes (2010)
Number of Observations						
Beneficiary level (# of clusters)	21,776,132	21,776,132	17,297,018	13,265,011	11,656,202	10,094,293
Beneficiary level	27,609,638	27,609,638	23,130,524	16,484,142	13,581,898	11,999,258
Beneficiary-year level	88,965,381	88,965,381	84,486,267	60,215,727	54,039,875	47,846,537

Notes: Regressions analogous to those in Table 4 but estimate single ACO effect pooled across all ACO cohorts and participation years. See Table 4 for additional notes. Standard errors reported in parentheses adjust for two-way clustering on beneficiaries and HRRs. Significantly different from 0 in a two-sided test at the *5%, **1% level.

Table A5

Robustness of baseline estimates of ACO effect to sample changes

	Sample on which Baseline Specification Estimated							
ACO Cohort	**Same as (1)**	**Same as (2)**	**Same as (3)**	**Same as (4)**	**Same as (5)**	**Same as (5')**	**Same as (6)**	**Same as (6')**
2012 Pioneer								
Participation year 1	74.1	74.1	42.9	45.7	74.7	74.7	62.8	62.8
	(65.9)	(65.9)	(63.0)	(59.6)	(53.4)	(53.4)	(58.9)	(58.9)
Participation year 2	-155.7 *	-155.7 *	-168.5 *	-145.4 *	-46.7	-46.7	-126.5	-126.5
	(75.4)	(75.4)	(72.8)	(72.1)	(64.4)	(64.4)	(68.9)	(68.9)
Participation year 3	-449.7 **	-449.7 **	-438.3 **	-447.7 **	-373.2 **	-373.2 **	-458.2 **	-458.2 **
	(84.8)	(84.8)	(84.3)	(85.8)	(83.8)	(83.8)	(90.2)	(90.2)
2012 MSSP								
Participation year 1	-136.9 *	-136.9 *	-154.3 **	-152.4 **	-81.0	-81.0	-85.5	-85.5
	(54.9)	(54.9)	(55.9)	(50.0)	(45.3)	(45.3)	(47.7)	(47.7)
Participation year 2	-202.1 **	-202.1 **	-226.6 **	-227.3 **	-114.2	-114.2	-170.7 *	-170.7 *
	(71.9)	(71.9)	(72.1)	(65.3)	(60.5)	(60.5)	(68.4)	(68.4)
Participation year 3	-342.8 **	-342.8 **	-345.9 **	-350.1 **	-283.4 **	-283.4 **	-356.1 **	-356.1 **
	(64.8)	(64.8)	(64.7)	(60.7)	(60.5)	(60.5)	(67.4)	(67.4)
2013 MSSP								
Participation year 1	-130.3 *	-130.3 *	-150.6 **	-119.6 *	-32.8	-32.8	-86.0	-86.0
	(50.3)	(50.3)	(49.9)	(48.2)	(43.0)	(43.0)	(49.7)	(49.7)
Participation year 2	-199.2 **	-199.2 **	-203.2 **	-195.8 **	-161.6 **	-161.6 **	-233.3 **	-233.3 **
	(61.2)	(61.2)	(60.4)	(61.6)	(61.7)	(61.7)	(66.4)	(66.4)
2014 MSSP								
Participation year 1	-27.1	-27.1	-37.6	-17.8	18.9	18.9	-38.7	-38.7
	(47.7)	(47.7)	(43.8)	(45.6)	(45.2)	(45.2)	(46.1)	(46.1)
Pooled ACO Estimate								
ACO × Post period	-142.1 **	-142.1 **	-155.9 **	-137.8 **	-72.7 *	-72.7 *	-120.7 **	-120.7 **
	(35.4)	(35.4)	(35.3)	(30.6)	(29.0)	(29.0)	(33.0)	(33.0)
Specification								
Beneficiary-level covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HRR-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ACO fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Non-ACO fixed effects	No	No	No	No	No	No	No	No
Beneficiary fixed effects	No	No	No	No	No	No	No	No
Address zero-spending truncation	No	No	No	No	No	No	No	No
Fixed attribution to anchor year	No	No	No	No	Yes (2011)	No	Yes (2010)	No
IV using prior attribution to anchor year	No	No	No	No	No	Yes (2011)	No	Yes (2010)
Number of Observations								
Beneficiary level (# of clusters)	21,776,132	21,776,132	17,297,018	13,183,134	11,582,873	11,582,873	10,042,717	10,042,717
Beneficiary level	27,609,638	27,609,638	23,130,524	16,324,197	13,647,674	13,647,674	11,897,083	11,897,083
Beneficiary-year level	88,965,381	88,965,381	84,486,267	57,961,912	52,176,040	52,176,040	46,254,744	46,254,744

Notes: Re-runs specification from baseline regression (1) on restricted samples used for other regressions in Tables 4 and A3. First column repeats baseline results for reference. Results are the same in first two columns because regression (2) used same sample as (1). Results in column (5') same as (5) because they use the same sample, and likewise for columns (6) and (6'). See notes to Table 4 for details on baseline specification and notes to Tables 4 and A3 for details on sample restrictions. The panel labled "pooled ACO estimate" re-runs the same regressions as in the panel above but estimates a pooled ACO effect across all ACO cohorts and participation years. Standard errors reported in parentheses adjust for two-way clustering on beneficiaries and HRRs. Significantly different from 0 in a two-sided test at the *5%, **1% level.

Table A6
Results adding methods to baseline separately rather than cumulatively

	Regression Specification							
ACO Cohort	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(5')**	**(6)**	**(6')**
2012 Pioneer								
Participation year 1	74.1	244.4 **	46.5	-50.0	-57.3	-122.0	61.5	164.2
	(65.9)	(74.6)	(51.0)	(55.1)	(59.0)	(141.9)	(50.2)	(152.2)
Participation year 2	-155.7 *	176.9 *	-118.2 *	-186.2 *	10.5	-8.3	24.4	90.6
	(75.4)	(82.6)	(55.5)	(75.8)	(67.6)	(189.2)	(58.5)	(209.8)
Participation year 3	-449.7 **	-30.7	-287.4 **	-499.7 **	-35.8	-123.9	-12.3	-36.1
	(84.8)	(85.1)	(55.5)	(84.1)	(70.2)	(229.6)	(60.2)	(253.3)
2012 MSSP								
Participation year 1	-136.9 *	40.7	-105.2 **	-124.7 **	76.9	135.0	142.1 **	266.4 *
	(54.9)	(48.1)	(36.4)	(45.0)	(42.2)	(75.4)	(54.1)	(110.2)
Participation year 2	-202.1 **	83.3	-178.6 **	-187.0 **	132.0 **	224.3 *	79.6	178.0
	(71.9)	(60.7)	(55.1)	(62.9)	(45.6)	(87.3)	(56.3)	(124.2)
Participation year 3	-342.8 **	21.0	-289.2 **	-333.5 **	65.9	141.3	30.7	100.1
	(64.8)	(61.5)	(58.9)	(57.5)	(52.9)	(107.6)	(58.1)	(142.0)
2013 MSSP								
Participation year 1	-130.3 *	91.6	-104.2 **	-96.1 *	94.4	154.3	-11.9	-35.2
	(50.3)	(49.1)	(36.8)	(47.5)	(49.2)	(86.5)	(49.5)	(110.8)
Participation year 2	-199.2 **	111.8	-170.8 **	-191.4 **	35.6	72.1	-88.6	-192.7
	(61.2)	(58.3)	(47.9)	(59.7)	(48.9)	(94.0)	(52.4)	(129.3)
2014 MSSP								
Participation year 1	-27.1	152.1 **	-30.7	-12.3	154.4 **	229.2 **	113.2 **	179.3 *
	(47.7)	(44.0)	(35.6)	(50.3)	(42.7)	(63.7)	(40.6)	(69.3)
Pooled ACO Estimate								
ACO × Post period	-142.1 **	106.3 **	-108.5 **	-139.8 **	51.9	104.0	45.5	116.0
	(35.4)	(31.8)	(25.6)	(29.3)	(34.2)	(67.9)	(33.6)	(85.1)
Specification								
Beneficiary-level covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HRR-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ACO fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Non-ACO fixed effects	No	No	No	No	No	No	No	No
Beneficiary fixed effects	No	No	No	No	No	No	No	No
Address zero-spending truncation	No	No	No	No	No	No	No	No
Fixed attribution to anchor year	No	No	No	No	Yes (2011)	No	Yes (2010)	No
IV using prior attribution to anchor yeaı	No	No	No	No	No	Yes (2011)	No	Yes (2010)
Number of Observations								
Beneficiary level (# of clusters)	21,776,132	21,776,132	17,297,018	15,978,776	13,445,555	13,445,555	12,904,330	12,904,330
Beneficiary level	27,609,638	27,609,638	23,130,524	19,197,907	15,580,525	15,580,525	15,076,751	15,076,751
Beneficiary-year level	88,965,381	88,965,381	84,486,267	62,929,492	68,455,718	68,455,718	65,071,745	65,071,745

Notes: Takes methods from Tables 4 and A3 and adds them to baseline rather than cumulating them in sequence. See notes to those tables for further information. Standard errors reported in parentheses adjust for two-way clustering on beneficiaries and HRRs. The panel labled "pooled ACO estimate" re-runs the same regressions as in the panel above but estimates a pooled ACO effect across all ACO cohorts and participation years. Significantly different from 0 in a two-sided test at the *5%, **1% level.

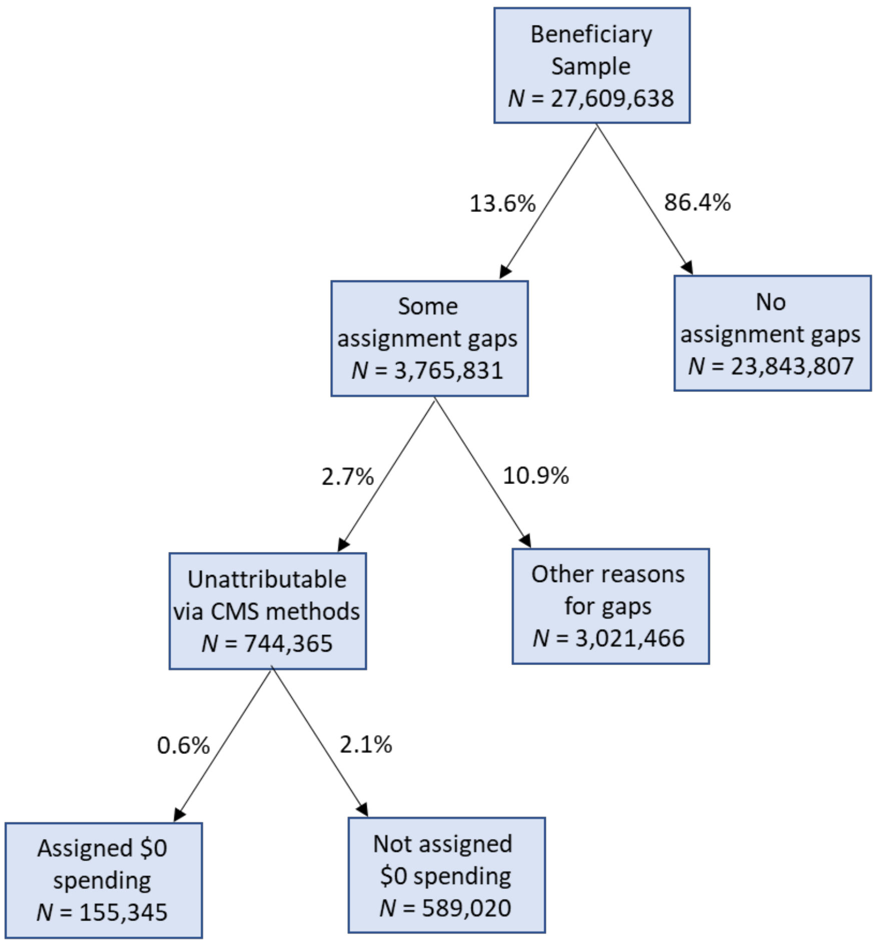
Table A7
Results attributing beneficiaries based on primary-care spending and broken out by hospital integration

ACO Cohort	Hospital Integrated	(1)	(2)	(3)	(4)	(5)	(5')	(6)	(6')
		Regression Specification							
2012 MSSP									
Participation year 1	No	-154.7 *	-9.5	-38.9	-22.5	181.0 **	244.6 **	132.0 *	212.8 *
		(78.0)	(76.2)	(56.0)	(48.6)	(65.9)	(90.8)	(54.9)	(87.9)
	Yes	-76.1	74.3	32.6	-9.2	-210.7 **	-236.5 **	-129.8 *	-160.5 *
		(54.7)	(50.6)	(36.3)	(36.2)	(55.5)	(62.7)	(51.2)	(61.8)
Participation year 2	No	-310.4 **	-86.2	-114.1	-73.9	210.7 **	306.0 **	93.6	153.3
		(102.8)	(95.4)	(81.7)	(69.9)	(69.0)	(106.2)	(66.7)	(115.8)
	Yes	-53.9	180.5 **	114.4 *	69.7	-130.4 *	-152.8 *	-121.9 *	-147.6 *
		(80.0)	(65.7)	(49.1)	(47.8)	(56.1)	(65.4)	(58.4)	(71.5)
Participation year 3	No	-406.8 **	-66.2	-132.7	-109.3	148.2	221.5	72.5	101.4
		(121.5)	(158.9)	(126.7)	(106.3)	(85.4)	(146.1)	(79.2)	(153.4)
	Yes	-186.3 **	122.1 *	30.4	-25.6	-210.7 **	-247.0 **	-210.4 **	-262.5 **
		(61.0)	(50.6)	(51.4)	(55.6)	(61.1)	(72.7)	(68.2)	(85.6)
2013 MSSP									
Participation year 1	No	-259.0 **	-122.5	-67.4	-45.1	41.3	57.0	12.6	21.2
		(88.4)	(89.6)	(63.0)	(60.5)	(53.2)	(81.3)	(44.0)	(88.1)
	Yes	13.0	193.7 **	114.9 **	65.4	-121.4 **	-132.3 **	-96.9 *	-111.1 *
		(55.6)	(53.4)	(35.5)	(35.7)	(35.3)	(40.9)	(45.0)	(54.9)
Participation year 2	No	-269.2 **	-70.6	-140.9	-182.8 *	-92.0	-173.2	-113.7	-280.3 *
		(96.9)	(101.2)	(77.2)	(72.5)	(68.0)	(112.3)	(62.9)	(140.2)
	Yes	-17.9	253.3 **	182.8 **	116.4 *	-75.3	-76.6	-98.0	-109.7
		(64.9)	(61.1)	(43.8)	(47.3)	(46.5)	(55.9)	(55.9)	(71.2)
2014 MSSP									
Participation year 1	No	-28.4	105.3	82.6	121.4 *	163.3 **	221.3 **	147.5 **	218.7 *
		(66.0)	(68.3)	(46.5)	(51.8)	(45.5)	(67.1)	(54.9)	(93.7)
	Yes	54.4	214.3 **	151.3 **	76.1	-41.8	-46.3	-51.9	-64.7
		(54.1)	(53.1)	(43.3)	(47.4)	(43.2)	(48.9)	(52.6)	(62.6)
Pooled ACO Estimate									
Participation year 1	No	-215.0 **	-37.9	-52.6	-35.1	105.4 **	159.9 **	24.4	41.4
		(56.9)	(61.6)	(45.7)	(40.0)	(35.4)	(56.4)	(35.0)	(69.5)
	Yes	-34.0	170.2 **	106.9 **	52.1 *	-112.2 **	-123.1 **	-87.5 **	-103.0 **
		(38.7)	(31.4)	(23.5)	(23.3)	(31.1)	(35.3)	(30.1)	(36.4)
Specification									
Beneficiary-level covariates		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HRR-year fixed effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ACO fixed effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Non-ACO fixed effects		No	No	No	No	No	No	No	No
Beneficiary fixed effects		No	No	No	No	No	No	No	No
Address zero-spending truncation		No	No	No	No	No	No	No	No
Fixed attribution to anchor year		No	No	No	No	Yes (2011)	No	Yes (2010)	No
IV using prior attribution to anchor year		No	No	No	No	No	Yes (2011)	No	Yes (2010)
Observations									
Beneficiary level (# of clusters)		16,905,530	16,905,530	12,903,774	8,495,736	8,099,593	8,099,593	6,625,964	6,625,964
Beneficiary level		20,424,103	20,424,103	16,422,347	9,847,080	9,130,402	9,130,402	7,500,781	7,500,781
Beneficiary-year level		61,762,702	61,762,702	57,760,946	36,518,283	34,989,784	34,989,784	28,964,294	28,964,294

Notes: Numbers refer to regression specifications in Tables 4 and A2. See those tables for further notes. Regression here are identical to those except for three changes, to more closely follows the specification in McWilliams et al. (2018). First, the Pioneer ACO cohort is not included. Second, beneficiaries are attributed based not on all spending but on spending at primary-care physician. Third, results broken out by ACOs that are physician group practices versus integrated with a hospital. The panel labled "pooled ACO estimate" re-runs the same regressions as in the panel above but estimates a pooled ACO effect across all ACO cohorts and participation years. Standard errors reported in parentheses adjust for two-way clustering on beneficiaries and HRRs. Significantly different from 0 in a two-sided test at the *5%, **1% level.

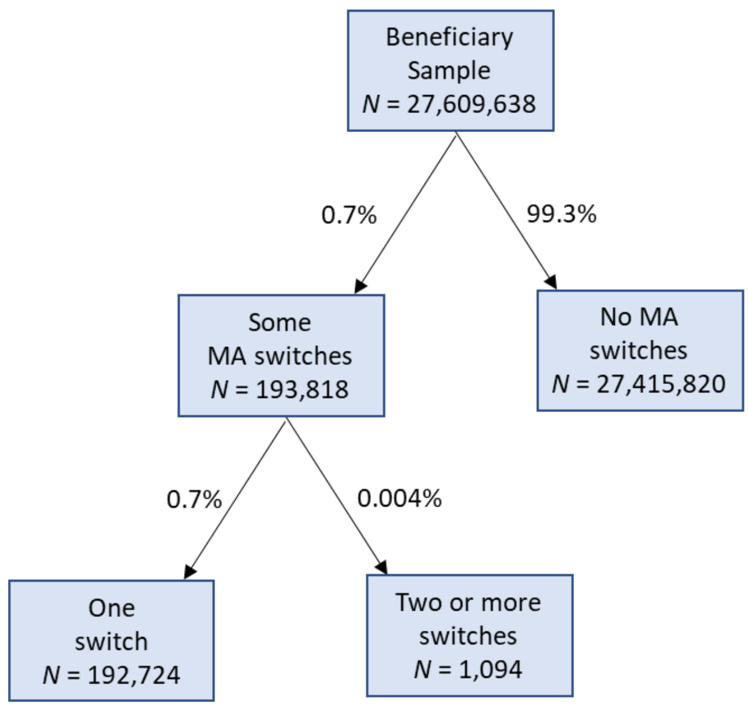
Figure A1

Breakdown of beneficiary sample by time-series gaps, attribution, and zero-spending allocation



Notes: Percentages along paths are computed by dividing by total beneficiaries in sample (*N*=27,609,638). Beneficiaries with gap in assignment in some year (*N*=3,765,831) may due to several reasons: being unattributed, switching to Medicare Advantage, having partial enrollment in Medicare parts A or B.

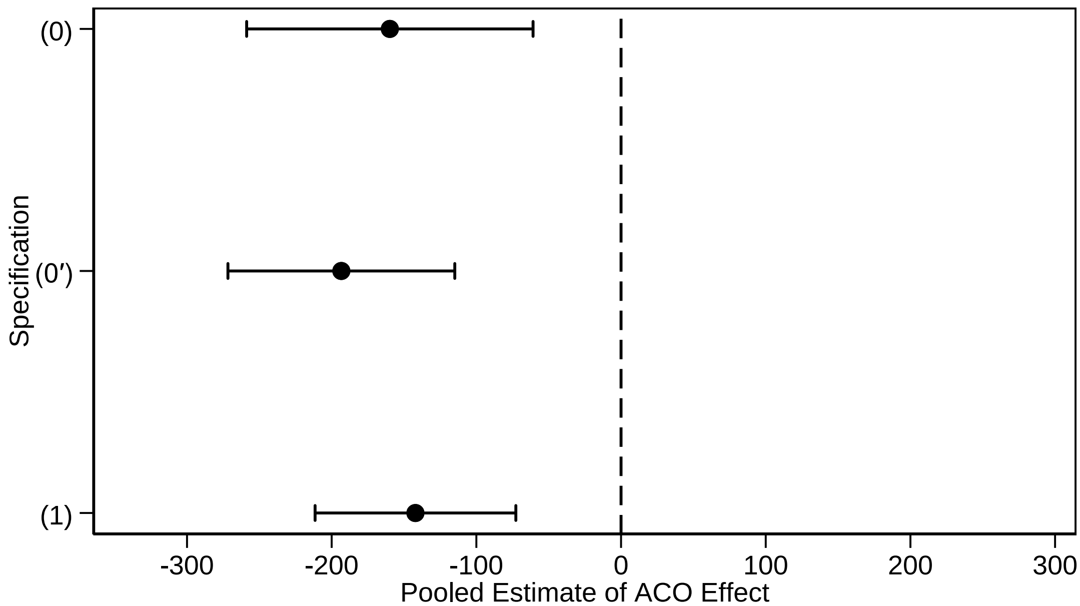
Figure A2
Transitions from fee for service into Medicare Advantage



Notes: Percentages along paths are computed by dividing by total beneficiaries in sample (N=27,609,638). MA stands for Medicare Advantage.

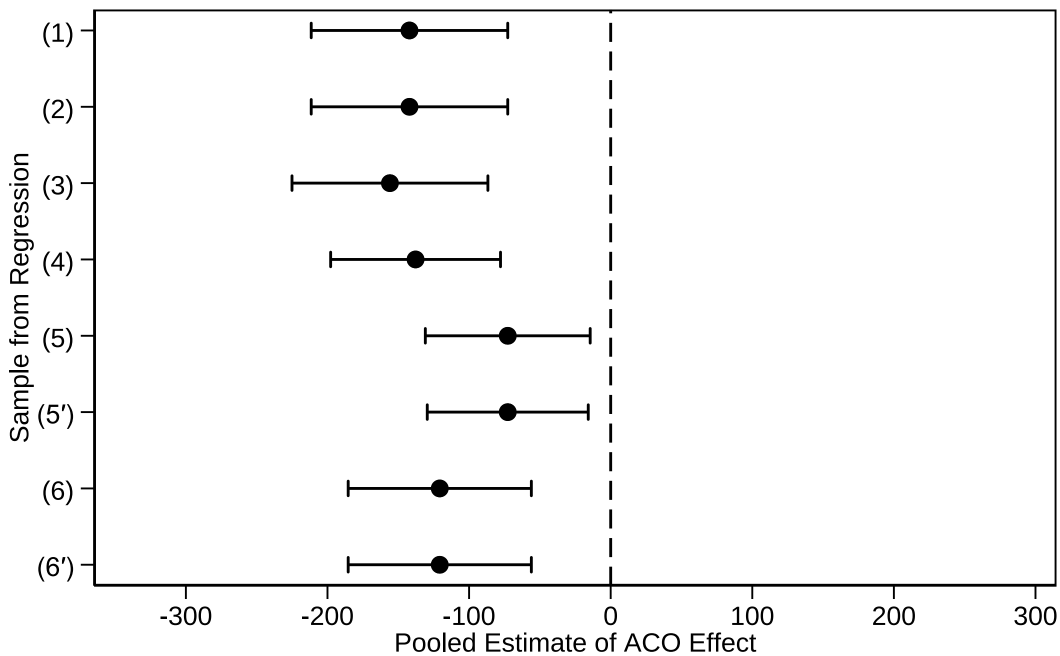
Figure A3

Comparing pooled estimates of ACO effect in baseline regression with different ACO fixed effects



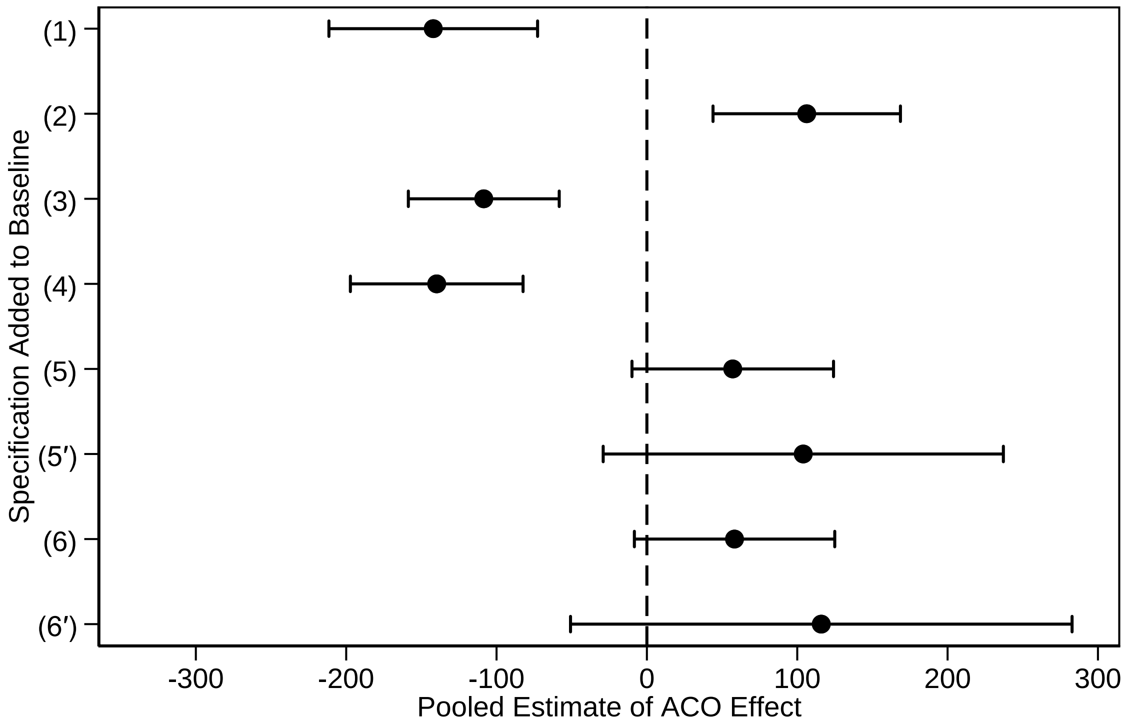
Notes: Plot of results from variants of baseline regression in Table 3 column 1 with a pooled ACO effect estimated rather than a separate effect by cohort and participation year. We replaced the individual ACO fixed effects in the baseline (1) with aggregate ACO (0) and aggregate ACO cohort fixed effects (0′).Estimates shown as dots and 95% confidence intervals as bars.

Figure A4
Pooled Estimates Running Baseline Specification on Restricted Samples



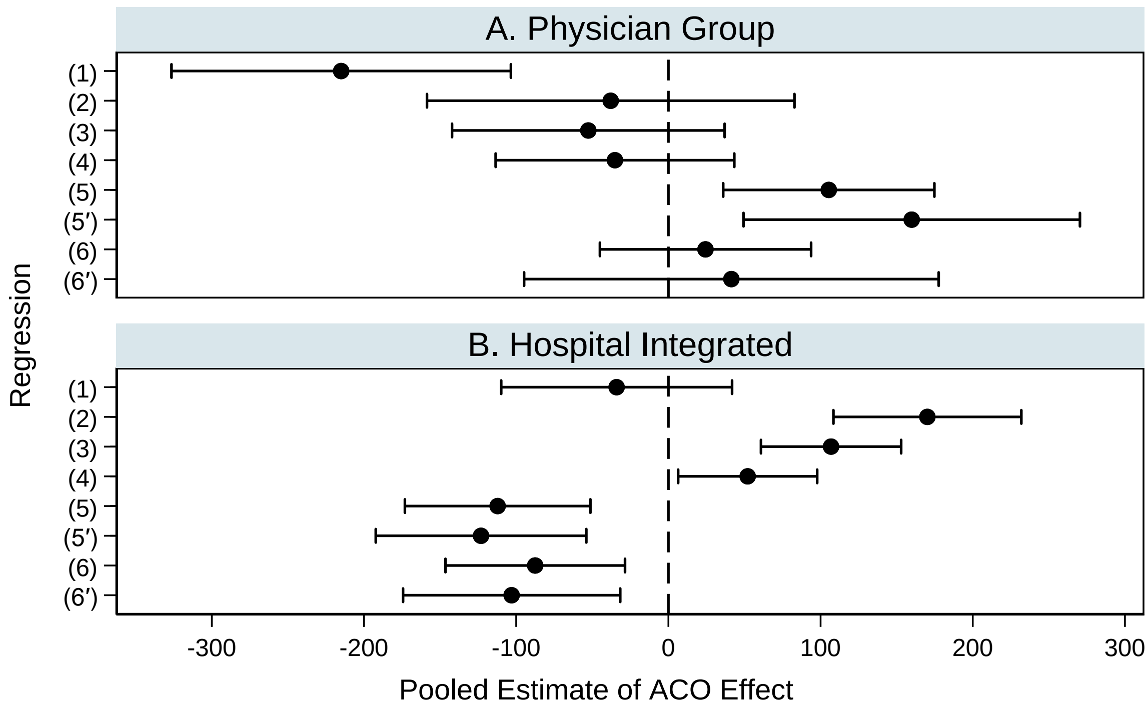
Notes: Plot of results from Table A4 panel presenting pooled ACO estimates. Estimates shown as dots and 95% confidence intervals as bars. Results for (5) and (5′) are same since regressions use same samples; similarly for (6) and (6′).

Figure A5
Pooled estimates adding methods separately to baseline regression



Notes: Plot of results from Table A5 panel presenting pooled ACO estimates. Estimates shown as dots and 95% confidence intervalus as bars.

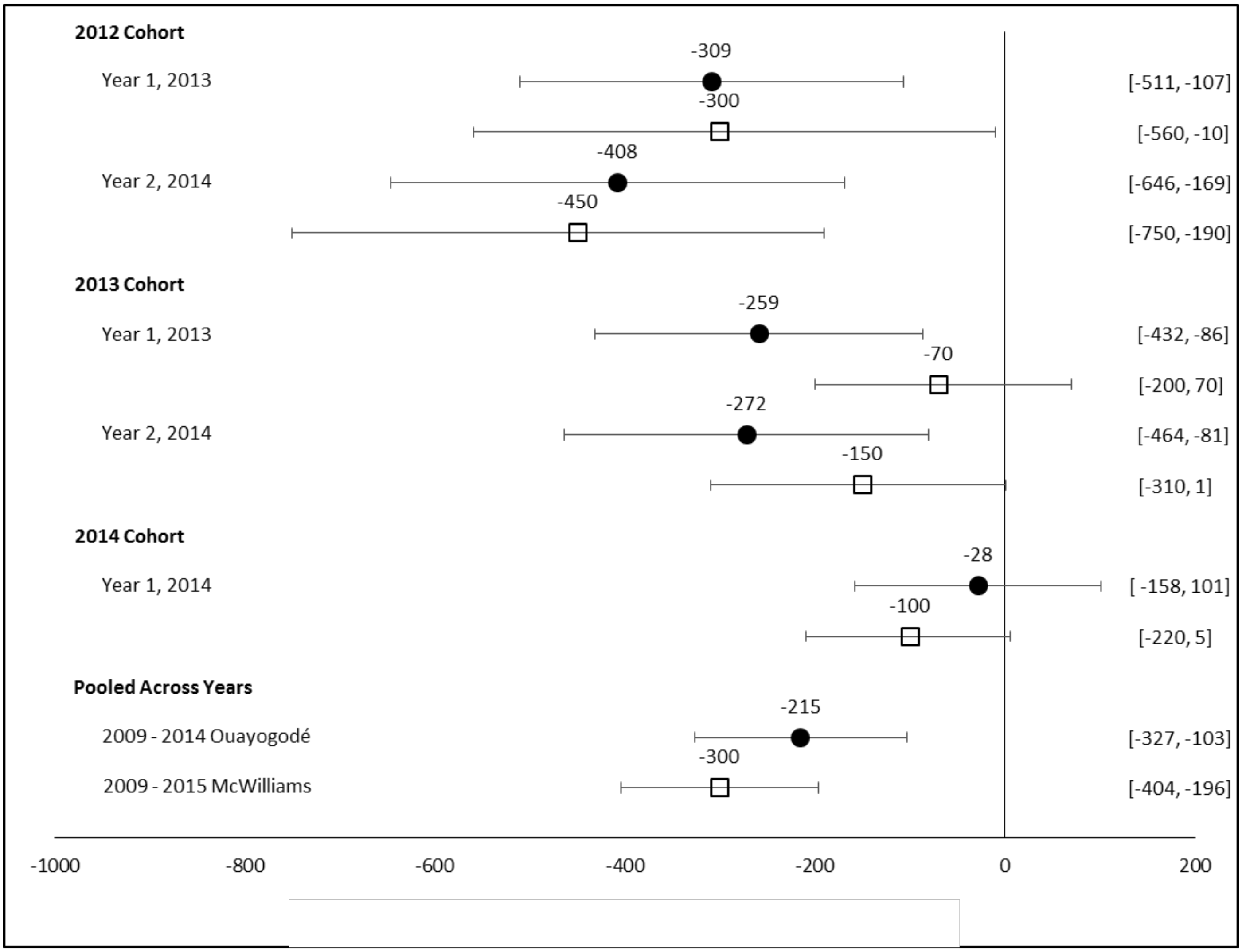
Figure A6
Pooled estimates attributing beneficiaries based on primary-care spending and broken out by hospital integration



Notes: Plot of results from Table A6 panel presenting pooled ACO estimates. Estimates shown as dots and 95% confidence intervals as bars.

Figure A7

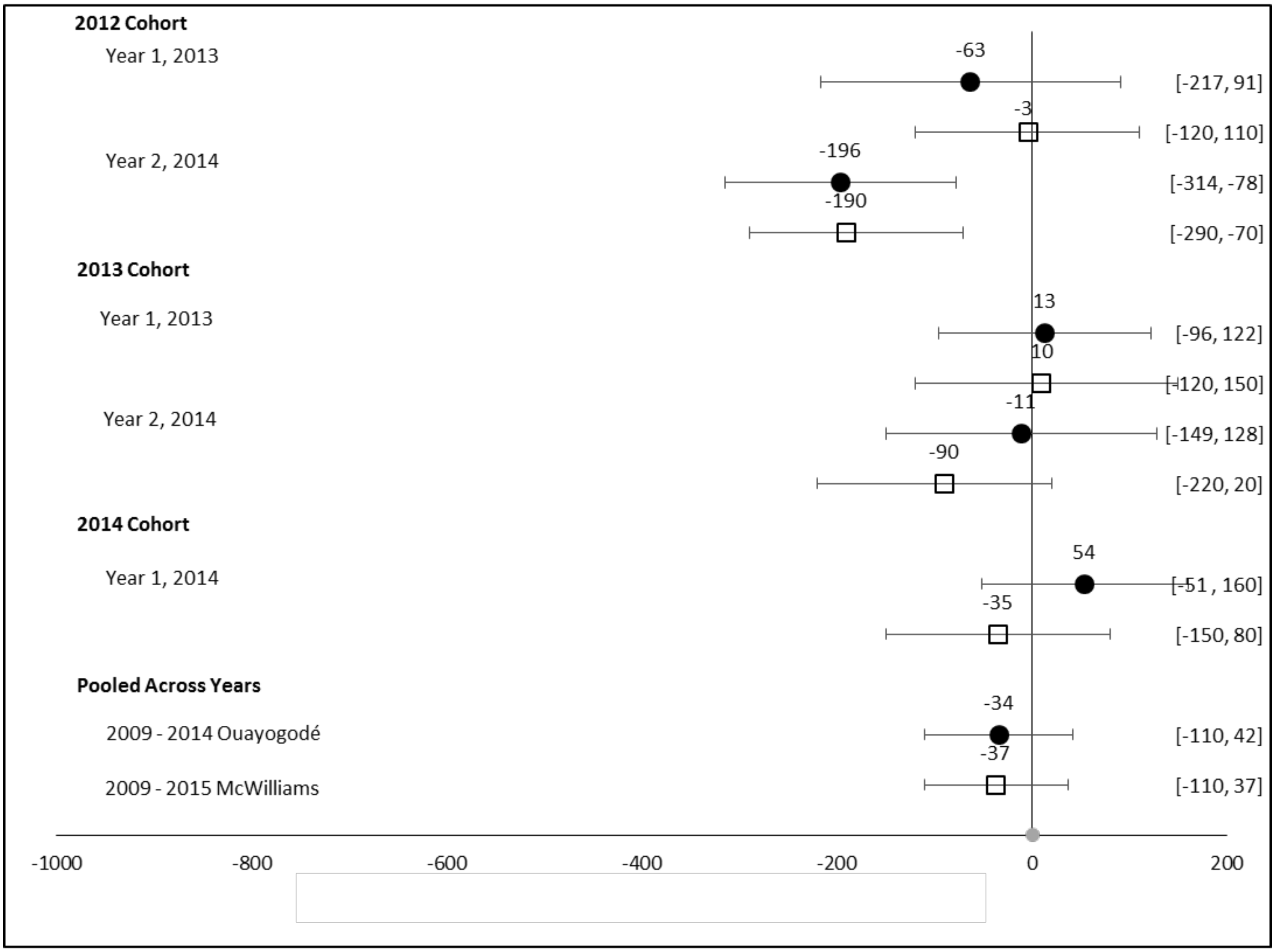
Comparison to McWilliams et al. (2018) of results for physician-group ACOs



Notes: Plot of results from Table A6 panel presenting pooled ACO estimates to analogous results from McWilliams et al. (2018). As in that paper, beneficiaries are attributed based on primary-care spending. Ouayogodé et al. estimate is a black circle and McWilliams et al. is a hollow square.

Figure A8

Comparison to McWilliams et al. (2018) of results for hospital-integrated ACOs



Notes: Plot of results from Table A6 panel presenting pooled ACO estimates to analogous results from McWilliams et al. (2018). As in that paper, beneficiaries are attributed based on primary-care spending. Ouayogodé et al. estimate is a black circle and McWilliams et al. is a hollow square.