Evaluation of the Medicare Quality Bonus Payment Demonstration
Contract: HHSM-500-2011-00083C

FINAL REPORT
February 2016

Prepared for:
Sai Ma, COR
Centers for Medicare & Medicaid Services
7500 Security Boulevard, MS WB-06-05
Baltimore, MD 21244-1850

Prepared by:
L&M Policy Research, LLC
1743 Connecticut Ave NW
Suite 200
Washington, DC 20009
Attn: Lisa Green, Principal
LGreen@LMpolicyresearch.com
240-476-6663

www.LMpolicyresearch.com

PO Box 42026 Washington DC 20015 • tel 202-249-8945 • fax 202-249-1784

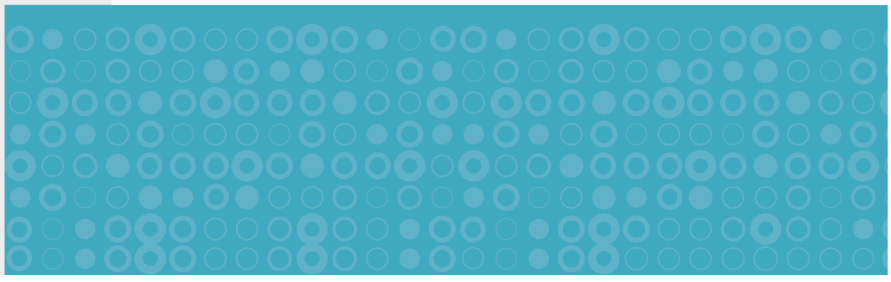


TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

The Centers for Medicare & Medicaid Services (CMS) launched the three-year Medicare Advantage (MA) Quality Bonus Payment (QBP) Demonstration ("QBP demo"), which extended quality bonus payments established in the Affordable Care Act of 2010 to 3- and 3.5-Star contracts in addition to contracts with Ratings of 4 Stars or higher and accelerated the phase-in of higher bonus payments to all levels of Star Ratings in payment years 2012 through 2014. The primary goal of this evaluation is to examine whether a causal relationship between the bonus payments and improved quality exists throughout the Ratings continuum. Towards this aim, we compare quality Ratings, enrollment, and benefit data for contracts in Star Ratings years 2009 through 2012 (derived from quality data collected prior to the announcement of the demonstration) to contracts in Star Ratings years 2013 through 2015 (derived from quality data collected after the announcement of the demonstration). Evaluating the causal impact of the QBP demo on quality is constrained by several factors, including the absence of an appropriate counterfactual, concurrent policy changes affecting MA plan payments and quality Ratings, and nonequivalence in Star Ratings measures over time. As a result, we provide descriptions of the payments made as a result of the QBP demo, contemporaneous changes in Star Ratings, enrollment, and benefits, but we cannot identify the unique contribution of the QBP demo from the effects of other factors the observed changes. We find the following in this evaluation:

QBP payments

Using payments made during the QBP demonstration to MA plans and estimated payments under the bonus structure set forth in the Affordable Care Act, provided by the CMS Office of the Actuary (OACT), we calculated that CMS paid nearly $10.9 billion in additional demonstration-related payments to MA contracts between 2012 and 2014. The majority of QBP payments above the ACA-level of bonuses went to 3 and 3.5 Star contracts, a finding consistent with the design of the demonstration and beneficiary enrollment. In later years, QBP demonstration payments above the ACA-level were smaller: $3.8 billion in 2012, $3.7 billion in 2013, and $3.4 billion in 2014. Again, the finding is consistent with the demonstration design, difference in quality bonus percentages under the QBP demonstration and the statutory ACA formulas, and the application of the bonus percentage to the entire blended benchmark. QBP-related payments are higher in 2012 than in 2014 across all Star Ratings categories.

Star Ratings changes

The distribution of MA contracts' Star Ratings improved between 2009 and 2015, with a higher proportion of 3.5 to 5 Star contracts in 2015 than in 2009. By 2014, 95 percent of U.S. counties had access to a 4-Star or higher rated contract. Enrollment-weighted Star Ratings, reflecting both improvements in contract Ratings and choices by beneficiaries, show an even more dramatic increase in Star Ratings. By 2015, more than half of MA enrollees enrolled in 4-Star or higher contracts. Although the number of enrollees in higher rated plans is greater in the QBP demo years, the trend started prior to the QBP demo. Conditional on contract characteristics, we found that:

- Ratings improved considerably for for-profit contracts during the QBP demonstration period, continuing a trend that began prior to the demonstration. The proportion of enrollees in 5-Star not-for-profit contracts also increased dramatically at the Start of the

QBP demonstration period; though, this was driven by one contract with high enrollment improving from 4 Stars in 2011 to 5 Stars in 2012.

- Enrollment-weighted Ratings also improved notably during the QBP demonstration period for contracts with local PPOs. In 2013, 36 percent of local PPO enrollees were in 4-Star or higher contracts; by 2015, that share increased to 79 percent. HMOs also showed an increase in enrollment-weighted Star Ratings, but it was less dramatic and appeared to continue a trend that began prior to the demonstration.

- Contracts that did not offer Special Needs Plans (SNP) had higher Star Ratings compared to contracts that had a mix of SNP and non-SNPs or contracts that had only SNPs.

Star Ratings changes were common for contracts throughout the periods examined, the pre-QBP demonstration period (2009 through 2012) and the QBP demonstration period (2012 through 2015). Year-to-year improvements in Star Ratings were more common than year-to-year declines, yielding overall improvement. Contracts with 2.5-, 3- and 3.5-Star Ratings were most likely to achieve a Ratings increase in both the pre- and QBP demonstration periods. Overall, 2.5 Star contracts were the most likely to improve in the next year and 4.5 Star contracts were the least likely to improve.

Measure-level analysis

To examine whether trends in MA contracts deviated from trends in other sectors during the QBP demonstration, we examined changes in national, enrollment-weighted means for several individual Part C and Part D measures that have adequate data over the period of interest (pre-demo and demo period) and compared those trends to comparison groups—commercial plans, Medicaid plans, Medicare FFS Consumer Assessment of Healthcare Providers and Systems (CAHPS), and stand-alone prescription drug plans (PDPs). Although these comparison groups suffer from shortcomings, we use these comparisons to evaluate trends in scores across MA and comparison groups in the pre-QBP and QBP demonstration periods. We found that, in general, MA contracts' enrollment-weighted average scores for Part C measures we examined were higher for measure data collected in the demo years than in the pre-demo period. Across the set of measures for which we have commercial, Medicaid, FFS CAHPS, and stand-alone PDP comparison group data, we observed no notable patterns across measures of deviation from the trend in enrollment-weighted score for MA contracts between the pre-QBP and QBP demonstration periods, nor do we observe deviations between the MA trends relative to the comparison group trends. Given that providers treat patients with different payers, it is possible that there are spillover effects from the Medicare program onto the quality scores for other payers and vice versa that influence these patterns.

Medicare Advantage organizations' quality improvement activities

To understand how the QBP demonstration may be affecting organizations' operations, we collected information regarding the quality improvement (QI) activities of MA plan sponsors through a contract-level survey and case studies with selected MA sponsors. The majority of survey respondents (88 percent) indicated that the budget for the contract's QI activities increased between 2010 (when the demonstration was announced) and 2013 (when the survey was

conducted). Overall, the survey results show that most respondents' self-reported perception was that the QBP demonstration helped to improve their contract's Star Rating. Among those respondents, the most commonly cited mechanism was the incentive it provided to senior management to focus on improving Star Ratings, closely followed by the funding QBP provided to implement or expand quality improvement. Linking survey results about organizations' QI activities to Star Ratings changes, we found that just one QI activity, provider incentive programs, was associated with changes in Star Ratings from 2013 to 2015.

Enrollment changes

Overall, analyses show enrollment changes are weakly related to changes in Star Ratings, with a higher average enrollment change for plans under 5-Star contracts and lower average enrollment change for those in 2/2.5-Star contracts. Furthermore, there is generally no statistically (or meaningfully) different relationship between increases in enrollment and changes in Star Ratings when comparing the demonstration period to previous years.

Premiums and cost sharing

Contracts with higher Star Ratings are provided a greater percentage of their rebate to lower premiums and enhance benefits. On average, beneficiaries in plans receiving bonus payments during the QBP demonstration period faced out-of-pocket (OOP) costs below what would be expected from a linear trend between 2011 and 2015. Plans in below 3-Star contracts, however, did not consistently have below-trend differences in OOP costs. As such, under the assumption that OOP costs would have grown linearly from 2011 to 2015 in the absence of the QBP demonstration, QBPs may have reduced Part C premiums, and expected non-premium and total OOP costs, from what they would have been in the absence of the QBP demonstration.

Across the QBP demonstration period (CY 2012-2014), average Star Ratings improved, more beneficiaries enrolled in higher rated plans, and more beneficiaries had access to higher rated plans. While there is no definitive way to attribute these changes (in whole or in part) to the QBP demonstration itself, evaluation analyses do show that the demonstration at least did not stall or reverse trends – Star Rating and plan enrollments increases that began prior to the demonstration continued throughout the demonstration period—and, in fact, QBP demo payments appear associated with reductions in OOP costs for beneficiaries.

BACKGROUND

The CMS Office of the Actuary's estimates indicate about 30 percent of 2013 Medicare beneficiaries elected to receive their Medicare benefits through private plans, the vast majority of which comprised Medicare Advantage (MA) plan enrollments.[1] To help monitor the quality and performance of MA contracts, CMS introduced the Five-Star Quality Rating System in 2008. Under the Star Ratings program, the Medicare program assigns a Star Rating of 1 to 5 Stars in half Star increments to MA contracts annually. Before 2012, Star Ratings were reported publicly but were not explicitly factored into payments to plans. The Affordable Care Act (ACA) of 2010 made a number of changes to the payments for MA plans, including requiring the Medicare program to incorporate Star Ratings into the payments for MA plans starting in calendar year 2012. Specifically, the ACA mandated that quality bonus payments (QBPs) be paid to all contracts earning 4 or 5 Stars in Medicare's Star Ratings program. The ACA also tied rebates, a portion of the amount when plan bids fall below the benchmarks paid back to the plan, to Star Ratings. Greater Star Ratings now increase the share of the difference between the bid and the benchmark the plan receives as a rebate.

Rather than implement the quality bonus payments as structured in the ACA, CMS launched the three-year MA QBP Demonstration ("QBP demo"), which accelerated the phase-in of bonus percentages to 4-, 4.5-, and 5-Star contracts and by extending bonus eligibility to 3- and 3.5-Star contracts.[2] In effect, the demonstration increased payments above what they would have been under the ACA payment formula. The QBP demonstration applied to payments in 2012 through 2014 (based on overall Star Ratings from one year prior to the payment, 2011 to 2013). Through evaluation of the demonstration, CMS seeks to understand how incentive payments affect plan quality across a broader spectrum of plans.

QBP Demo evaluation study design and limitations

CMS contracted with L&M Policy Research, LLC (L&M), to conduct an evaluation of the demonstration, guided by the following aims:

- Document changes in quality improvement (QI) under the demonstration;
- Document MA contract, enrollee, and market characteristics associated with quality Ratings and improvements; and
- Document the relationship across QBPs, quality initiatives, and enrollment.

The primary goal of the evaluation is to examine the relationship between the bonus payments and improved quality, as measured by the 5-Star Rating system. Because the MA QBP demonstration allowed nearly all MA plans to participate, the evaluation relies on a quasi-experimental pre-post design—quality Ratings and associated contract, market, and enrollment characteristics observed

[1] http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/ReportsTrustFunds/downloads/tr2014.pdf

[2] Quality Bonus Payment Demonstration policy was outlined in a December 2012 memo from Cynthia G. Tudor, Director, Medicare Drug Benefit and C & D Data Group to Medicare Advantage Compliance Officers. See https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/MemoQBPAppeals_121610.pdf.

for Star Ratings years 2009 through 2012, prior to the Start of the demonstration, are compared to Ratings and characteristics under the demonstration period for Star Ratings years 2013 through 2015.

Several threats to internal validity complicate establishing a causal relationship between the QBP demonstration and any observed changes in the quality Ratings. We cannot rule out the possibility that factors other than the demonstration payments are affecting observed improvements in plan quality. Key threats to validity and approaches to address them, where possible, include:

Instrumentation. Observed changes in quality Ratings may result from changes in the measurement process over time rather than from the demonstration. The underlying measures and algorithm used to calculate the Star Ratings and cut points evolved over the period under consideration. That is, though the Star Ratings are used as a measure of quality comparing multiple years, the Rating varied from one year to the next in which measures are included in the Ratings and how the Ratings are calculated.

CMS did not begin calculating the overall Star Rating, which is the basis for the QBPs, until 2011. In 2011, all measures comprising the overall Star Rating were weighted equally. Starting with the 2012 Star Ratings, CMS assigned a weight of 1 for process measures, 1.5 for CAHPS, complaints, and access measures, and 3 for outcome and intermediate outcome measures.

Furthermore, the underlying measures of overall Ratings changed over time. Of 80 quality measures that have been included at least once in the overall Star Rating between the 2011 and 2014 Star Ratings years, only 26 were included in all years; 12 were included as composites in some years and as separate metrics in others; and, 42 appeared in only one to three years. Even among the 26 consistently included measures, differences in the year-to-year specifications exist.

Not only were the components of the overall Ratings changing over time, the thresholds to achieve Ratings on individual measures also changed—overall Ratings are calculated as an average of 1-5 Star Ratings achieved on the underlying measures. We assessed whether changing thresholds would drive observed changes in overall Ratings by examining 36 Part C measures that were fairly consistently collected for the 2011 to 2015 Ratings. Thresholds for achieving given Star Ratings on these measures moved to make Star levels both easier and more difficult to achieve. However, it was more common for a given Rating to become more difficult to obtain from one year to the next than easier. As a result, it should be more difficult to achieve Ratings increases rather than easier due to changes in thresholds (See Appendix 2. Star Ratings Cut Point Analysis).

Non-equivalence of comparison groups. All MA contracts are eligible for the treatment, in this case bonus payments, under the QBP demonstration. Therefore, equivalent entities are not available during the same time to serve as appropriate comparisons. When available, we used quality data on measures from comparisons that were not directly exposed to the QBP demonstration—Medicaid, commercial plans, Medicare FFS, and stand-alone PDPs. These comparison groups have only a fraction of quality Ratings data available for inclusion. Furthermore, the beneficiaries of other payers are exposed to differing circumstances than beneficiaries of MA plans and are not robust comparison groups for the purposes of observing trends in plan quality outside of the demonstration. In addition, performance changes may spillover

in provider practices as a result of incentives from one type of payer (i.e. MA) to beneficiaries of other payers, biasing the comparison outcomes.

Complementary quality initiatives in MA program. A number of concurrent initiatives beyond the demonstration likely impact a managed care organization's behavior and decisions involving quality improvement activities. These include: introduction of the low performer icon on the Medicare Plan Finder Web tool, the threat of contract termination after multiple years of low Star Rating performance, CMS letters to enrollees of low-performing plans, introduction of a five-Star plan icon on Plan Finder, allowing five-Star plans year-round enrollment, and allowing higher quality plans to use a higher percentage of their rebate to lower premiums or enhance benefits. In addition, the QBP demonstration coincided with the phasing in of MA payment changes passed in the ACA. As a result, we are not able to disentangle the effects of the QBP demonstration from potential effects of other concurrent changes.

Data collection lags provide limited time to observe changes in quality. Time elapses between the data collection periods for the measures, their inclusion in the Star Ratings, and their use for determining payment. The data collection period for the measures composing the 2011 overall Star Rating, which determined MA contracts' 2012 QBP, extends from January 2009 to July 2010. Since the demonstration was announced several months later (November 2010), the incentives of the demonstration methodology could not have impacted the 2011 Star Ratings.

Potential impact of the demonstration on 2012 overall Star Ratings is limited by the fact that data collection occurred after the November 2010 announcement for only 16 of 50 measures included in the Star Ratings. The remaining 34 measures are based on data collected prior to the announcement and could not have been impacted by the demonstration. It was not until the 2013 overall Star Rating, the basis for the third and final QBP payment, when quality measure collection occurred entirely after the demonstration announcement.

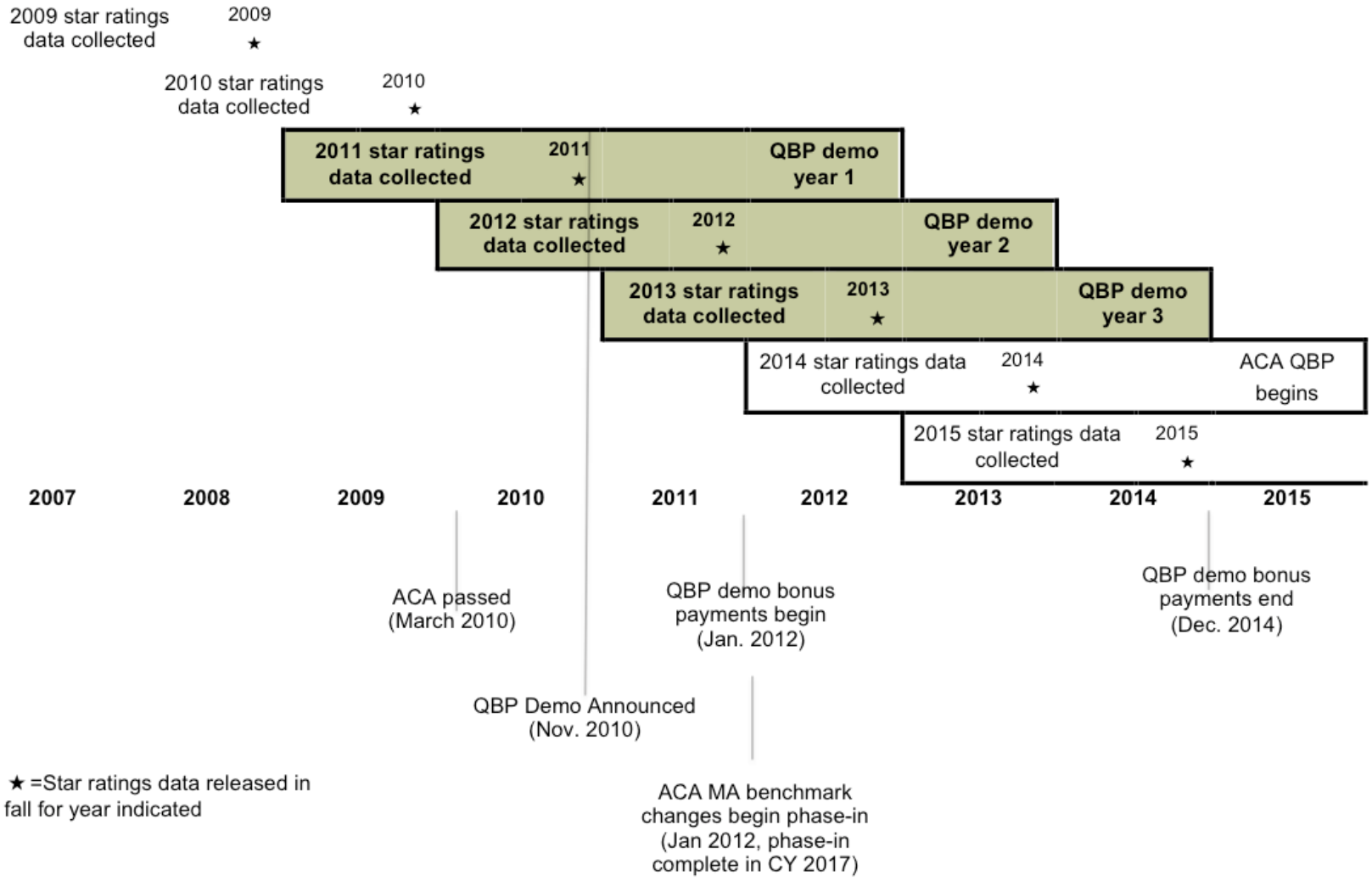
Table 1 illustrates the relationship between the data collection periods and the announcement of the demonstration. Shading in the table highlights when data collection occurred; darker shading indicates collection prior to the QBP demonstration period. Because the 2012 Star Ratings are a transition year in which some measures have pre-announcement and others have post-announcement data collection periods, we have displayed the data collection periods for specific types of measures.

Figure 1 is an accompanying timeline that shows Star Ratings data collection and reporting years in the context of dates and events that are important to the QBP demonstration.

Table 1. Data collection periods for Star Ratings measures and their significance to the QBP demonstration evaluation

Star Ratings Year	Measure Type	Number of Measures	Approx. Data Collection Period	Relationship between Demo Announcement and Data Collection Period	QBP Payment Year
2009	Overall Star Rating (All measures)	55	Jan. 2007 – Jul. 2008	Pre-demo announcement	NA
2010	Overall Star Rating (All measures)	52	Jan. 2008 – Jul. 2009	Pre-demo announcement	NA
2011	Overall Star Rating (All measures)	51	Jan. 2009 – Jul. 2010	Pre-demo announcement (However, some measures could have been affected by the announcement of the ACA methodology.)	QBP Year 1 (2012) - Demo
2012	Overall Star Rating (All measures)	50	Jan. 2010 – Jun. 2011	Mixed, depending on measure	QBP Year 2 (2013) - Demo
	HEDIS	*18*	*Jan. 2010 – Dec. 2010*	*Pre-demo announcement*	
	CAHPS	*10*	*Feb. 2011 – June 2011*	*Post-demo announcement*	
	Health Outcomes Survey (HOS)	*5*	*Apr. 2010 – Aug. 2010*	*Pre-demo announcement*	
	Prescription Drug Event (PDE)	*5*	*Jan. 2010 – Dec. 2010*	*Pre-demo announcement*	
	Appeals	*4*	*Jan. 2010 – Jun. 2011*	*3 measures are pre-demo announcement; 1 is post-demo*	
	Call Center	*3*	*Jan. 2011 – May 2011*	*Post-demo announcement*	
	Other	*5*	*Jan. 2010 – Jun. 2011*	*3 measures are pre-demo announcement; 2 are post-demo*	
2013	Overall Star Rating (All measures)	51	Jan. 2011 – Jun. 2012	Post-demo announcement	QBP Year 3 (2014) - Demo
2014	Overall Star Rating (All measures)		Jan. 2012 – Jun. 2013	Post-demo announcement	QBP Year 4 (2015) - ACA

Figure 1. Data collection, reporting, and payment timeline for 2009 through 2015 Star Ratings



Given the lag between data collection, reporting, and payments based on the Star Ratings, the QBP demonstration is hypothesized to potentially induce plans to improve their Ratings through two mechanisms, each of which took effect in different years:

1. The announcement of the demonstration could have induced Medicare Advantage organizations to improve their quality to achieve bonus payments. CMS announced in November 2010 that the demonstration would be in effect in calendar years 2012, 2013, and 2014. As discussed earlier, the quality data on which 2012 quality bonus payments would be based (the 2011 Star Ratings) had already been collected between 2009 and 2010. Similarly, most, though not all, of the quality data to determine the 2013 quality bonus payments had been collected prior to the demonstration announcement as shown in Table 1. The only payment year for which the data collection period fell entirely after the demo announcement was 2014, which used 2013 Star Ratings (collected January 2011 to June 2012) to determine bonus payments. We hypothesize that the announcement effect of the QBP demonstration could have had an effect on the data used to calculate Star Ratings beginning in 2013.

2. The payments that Medicare Advantage organizations with three or more Stars received under the QBP demonstration in calendar years 2012, 2013, and 2014 could have provided them with additional funds to invest in improving their quality scores. We hypothesize that the payment effect on Star Ratings would have occurred in Star Ratings years 2014 and partial 2013 (data collected in 2012 and 2013) and Star Ratings years 2015 (data collected in 2013 and 2014).

Combining the announcement effect and payment effect and taking the data collection and payment lag into effect, we consider Star Ratings years 2013, 2014, and 2015 to be the "demonstration years" in the analysis of the quality bonus payments. Unless otherwise indicated, data in this report are presented by Star Ratings year.

Total enrollment in MA increased and the total number of MA contracts declined over the demonstration period

There was much speculation during the debate over the ACA that MA plans would drop out of the MA program and enrollment would decline as it did after the payment reductions for Medicare+Choice plans enacted in the Balanced Budget Act of 1997 (BBA '97).[3] However, the number of beneficiaries enrolled in MA steadily increased between 2009 and 2015, as shown in Table 2. The number of MA contracts fluctuated between a high in 2009 and a low in 2015, with contract counts increasing during the QBP period.

[3] Marsha Gold, Gretchen Jacobson, Anthony Damico, and Tricia Neuman, "Medicare Advantage 2014 Spotlight: Enrollment Market Update," Kaiser Family Foundation Issue Brief, April 2014, p. 1. Available at https://kaiserfamilyfoundation.files.wordpress.com/2014/04/8588-medicare-advantage-2014-spotlight-enrollment-market-update.pdf

Table 2. MA contracts and enrolled beneficiaries, 2009 to 2015

Calendar Year	Count of contracts	Count of enrolled beneficiaries
2009	634	10,649,601
2010	583	11,217,454
2011	531	11,378,805
2012	546	13,047,876
2013	553	14,248,403
2014	566	15,526,277
2015	525	16,352,957

Source: MA contracts of interest were identified using data from HPMS. Enrollment data are from CMS monthly enrollment files.

Note: This table includes all MA contracts, including those not eligible for a Star Rating or Quality Bonus Payment and those with zero beneficiary enrollment.

FINDINGS

Findings of the evaluation are organized into five major topic areas and addressed in the following order: (1) payments due to the QBP demonstration, (2) Star Ratings changes in the pre-demonstration and demonstration periods, (3) QI activities and their relationship to Star Ratings, (4) the relationship between enrollment and Star Ratings changes, and (5) the relationship between bonus payments, premiums, and beneficiaries' out-of-pocket costs.

Finding 1. The Medicare program paid Medicare Advantage organizations an estimated $10.96 billion under the QBP Demonstration

The CMS Office of the Actuary (OACT) provided the evaluation team with (1) data on Medicare program payments made to each plan and the number of beneficiary months by county for each payment year, 2012 through 2014 and (2) an estimate of what payments would have been under the statutory ACA payment formula assuming that (a) risk-adjustment and enrollment did not differ under the alternative payment and (b) the bid made under the QBP would have been the same under the ACA, regardless of the differences in benchmarks. To calculate the amount paid as a result of the QBP demonstration, we took the difference in payments made under the QBP and estimated payments under the ACA. Using these data, we found that CMS paid $10.96 billion in demonstration-related payments to MA contracts from 2012 through 2014, as shown in Table 3. The $10.96 billion represents about 2.6 percent of total Medicare spending on MA for the same period.

Table 3. Total QBP demonstration payments by overall Star Rating, 2012 through 2014 (payment years)

	2012	2013	2014	Total
3 Stars	$1,018,910,289	$846,706,563	$527,948,586	$2,393,565,438
3.5 Stars	$1,180,660,119	$1,678,383,576	$1,836,673,072	$4,695,716,767
4 Stars	$432,929,016	$325,879,952	$340,018,780	$1,098,827,748
4.5 Stars	$642,247,320	$274,516,440	$388,863,004	$1,305,626,764
5 Stars	$56,590,782	$435,129,623	$223,393,135	$715,113,540
New contracts	$394,585,905	$106,097,635	$91,007,088	$591,690,627
Low enrollment	$87,210,405	$50,314,764	$24,264,384	$161,789,553
Total QBP demo payments	$3,813,133,835	$3,717,028,553	$3,432,168,048	$10,962,330,436
QBP demo payments as a share of total MA payments	2.9%	2.6%	2.2%	2.6%

Sources: Payment and enrollment data are from the CMS Office of the Actuary. Star Ratings data are from HPMS.

Notes: Star Ratings are the overall Star Ratings from HPMS for the year (e.g., overall Star Ratings for 2011 were used to determine payments in 2012). For purposes of payment, new contracts and low enrollment contracts that did not have a Star Rating in HPMS were assigned a Star Rating according to program rules as described above to determine their payment.

Table 3 shows that the majority of QBP payments went to 3 and 3.5 Star contracts, which is consistent with enrollment distribution and the design of the QBP demonstration. As shown in Table 4, under the ACA 3- and 3.5-Star contracts would not have received bonus payments, whereas under the QBP demo these contracts received bonus percentages of 3 and 3.5 percent, respectively.

In addition to extending bonuses to 3 and 3.5 Star contracts, QBP payments above ACA expectations shown in Table 3 reflect the following changes, which effectively increased the benchmarks under the QBP compared to what they would have been under the formula in the ACA:

- Increasing the quality bonus percentage for contracts with 4 or more Stars in 2012 and 2013, relative to what it would have been under the ACA, as shown in Table 4.
- Applying bonus payments to the entire blended benchmark amount, rather than just the ACA portion of the blended benchmark. Under the ACA, bonuses will only be applied to the ACA portion of the blended benchmark.
- Eliminating the cap on blended county benchmarks. Under the ACA, the blended county benchmarks could not exceed the pre-ACA benchmark amount.

For simplified examples of how the bonus payments are applied in the MA payment formula, see Appendix 1.

Table 4. Quality bonus percentages under the Affordable Care Act and the QBP demo

	ACA			QBP demo		
Overall Star Rating	2012	2013	2014	2012	2013	2014
5 Stars	1.5	3	5	5	5	5
4 or 4.5 Stars	1.5	3	5	4	4	5
3.5 Stars	0	0	0	3.5	3.5	3.5
3 Stars	0	0	0	3	3	3
Fewer than 3 Stars	0	0	0	0	0	0
New plan	1.5	2.5	3.5	3*	3*	3.5*
Low enrollment plan	**	**	**	3	3	3

Source: Adapted from United States General Accounting Office, "Medicare Advantage: Quality Bonus Payment Demonstration Undermined by High Estimated Costs and Design Shortcomings, GAO-12-409R, March 21, 2012, p. 17. Table was supplemented with information from Call Letters published after the GAO report.

*Notes: * A new MA contract offered by a parent organization that has not had any MA contract(s) with CMS in the previous three years is treated as a qualifying contract, per statute, and is assigned three Stars for QBP purposes for 2012 and 2013, and 3.5 Stars in 2014. These contracts are treated as new MA contracts during the demonstration until the contract has enough data to calculate a Star Rating. For a parent organization that has had MA contract(s) with CMS in the previous three years, any new MA contract under that parent organization will receive a weighted average of the Star Ratings earned by the parent organization's existing MA contracts, per the 2012 Call Letter.*

***Bonus payment percentages for low enrollment plans were not specified in law. According to the Health Care and Education Reconciliation Act of 2010 (P.L. 111-152 March 20, 2010) low enrollment plans should be considered qualifying plans in 2012; the law also specifies that for 2013 and beyond the Secretary shall establish a method to determine whether an MA plans with low enrollment is a qualifying plan for purposes of calculating the quality rating.*

Table 5 shows the average per-beneficiary-per-month (PBPM) QBP demonstration-related payment by Star Rating. Consistent with the difference in quality bonus percentages under the QBP demonstration and the statutory ACA formulas in Table 4 and the application of the bonus percentage to the entire blended benchmark, the relative-to-ACA payments are higher in 2012 than in 2014 across all the Star Ratings categories. The relatively higher QBP-related payments to 3.5 Star contracts are likely a function of their relatively high average PBPM total payments compared to the average total PBPM payments for contracts across the Star Ratings continuum. (See Table 6.) These higher payments may be a function of these contracts' service areas (and thus their benchmarks) and differences in the risk scores of their enrolled beneficiaries.

Table 5. PBPM differences in payment between QBP and simulated ACA by overall Star Rating, 2012-2014 (payment years)

	2012	2013	2014	Total
3 Stars	$20	$18	$15	$18
3.5 Stars	$30	$30	$26	$28
4 Stars	$31	$19	$14	$20
4.5 Stars	$29	$21	$14	$21
5 Stars	$37	$29	$15	$23
New contracts	$25	$22	$23	$24
Low enrollment	$25	$19	$17	$21
All contract average	$26	$24	$19	$23

Sources: Payment and enrollment data are from the CMS Office of the Actuary. Star Ratings data are from HPMS.

Notes: Star Ratings are the overall Star Ratings from HPMS for the year (e.g., overall Star Ratings for 2011 were used to determine payments in 2012). For purposes of payment, new contracts and low enrollment contracts that did not have a Star Rating in HPMS were assigned a Star Rating according to program rules as described above to determine their payment.

Table 6. Average actual PBPM total payments by overall Star Rating, 2012-2014 (payment years)

	2012	2013	2014	Total
3 Stars	$824	$819	$844	$827
3.5 Stars	$916	$873	$833	$866
4 Stars	$815	$828	$807	$815
4.5 Stars	$791	$768	$821	$800
5 Stars	$750	$816	$799	$805
New contracts	$729	$791	$766	$747
Low enrollment	$826	$811	$795	$815
All contract average	$832	$834	$825	$830

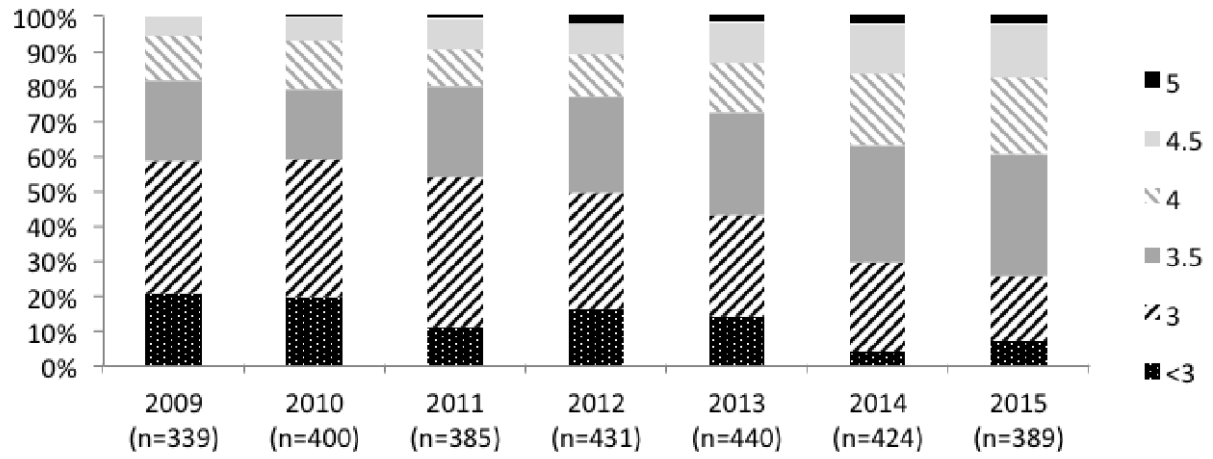
Sources: Payment and enrollment data are from the CMS Office of the Actuary. Star Ratings data are from HPMS.

Notes: Star Ratings are the overall Star Ratings from HPMS for the year (e.g., overall Star Ratings for 2011 were used to determine payments in 2012). New contracts and low enrollment contracts did not have a Star Rating in HPMS, but were assigned a Star Rating according to program rules as described above to determine their payment. All contract averages do not include payments to contracts with fewer than 3 Stars.

Finding 2. MA contract Ratings continued a pre-demonstration trend of improved overall Star Ratings and increasing beneficiary enrollment for higher rated contracts

Overall, MA contracts' Star Ratings improved between 2009 and 2015, as shown in Figure 2, with higher rated contracts making up a larger proportion of the contracts available to beneficiaries over time. The proportion of contracts with Star Ratings of 3.5, 4, 4.5, and 5 was higher in 2015 than in 2009. Notably, the proportion of active contracts with an overall Rating of 3.5 Stars or higher moved from fewer than half, in 2011, to over 70 percent of all contracts in 2014 and 2015.

Figure 2. Contract-weighted overall Star Ratings, 2009 to 2015

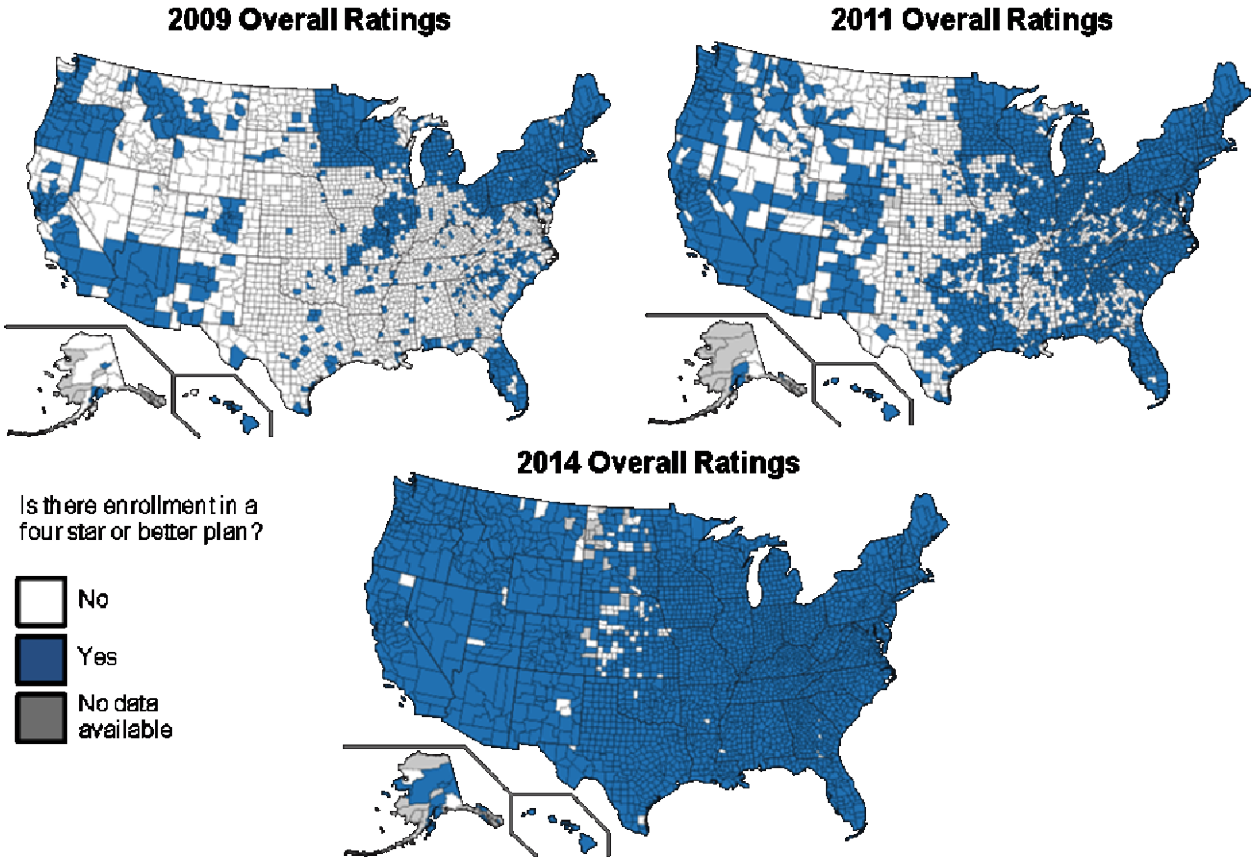


Source: Publicly available Star Ratings are from http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html. Contracts of interest and other characteristics were identified using data from HPMS.

Note: L&M estimated the contract-weighted overall Star Ratings for 2009 and 2010. Figures exclude contracts that were too new or low enrollment and thus, no Star Rating.

As a result of these increases in Star Ratings, by 2014, 95 percent of the counties in the country had access to a 4-Star or higher-rated contract. This increase is in contrast to 2009, the earliest year examined in the pre-QBP demonstration period, when access to a 4-Star or higher-rated contract varied significantly across the country, as shown in Figure 3. (For additional data showing the share of counties with 4-Star or higher contracts in each year, 2009 to 2015, and additional data about contract characteristics and overall Star Ratings, see Appendix 3. Additional Overall Star Ratings Data.)

Figure 3. County-level access to a 4-Star (or higher) rated contract, 2009, 2011, and 2014

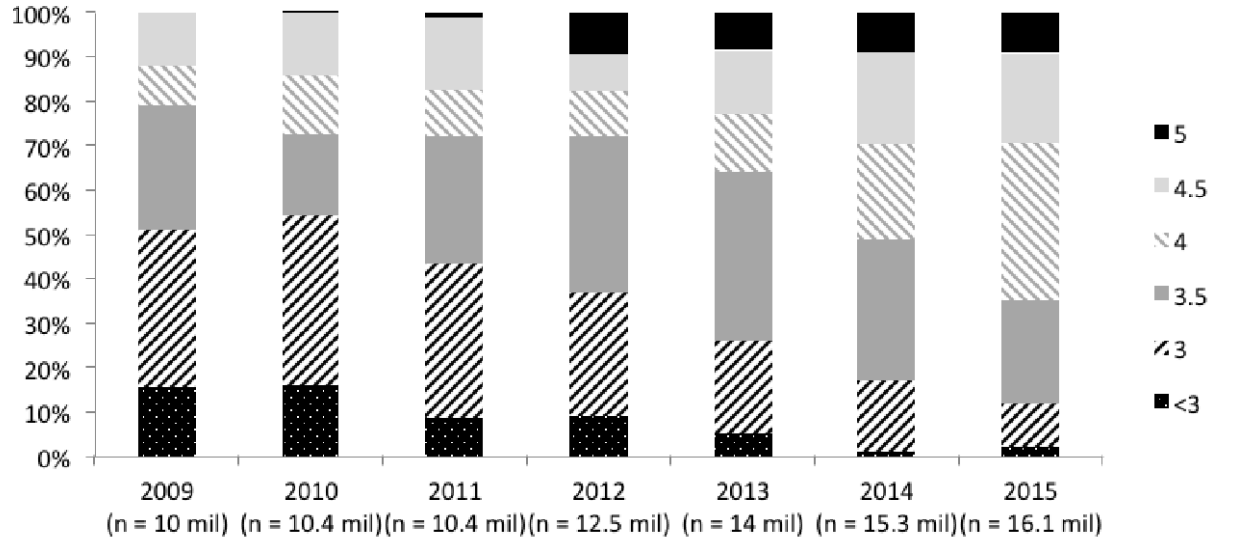


Source: Publicly available Star Ratings are from http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html. Enrollment data are from CMS monthly enrollment files.

Note: "No data available" means that there was no enrollment in any plan in the county for the year. "No" means that there was enrollment in the county but not in a four star or better plan.

Enrollment-weighted Star Ratings reflect the contract Ratings distributions of MA plans that beneficiaries have chosen, rather than the menu of contracts available. The enrollment-weighted Star Ratings show an even more dramatic shift toward higher Ratings than the contract-weighted data. As shown in Figure 4, by 2014, more than half of MA enrollees are in contracts with 4 or more Stars. Although the number of enrollees in these higher-rated plans is greater in the QBP demonstration years, this increase appears to be a continuation of a trend that started prior to the QBP demonstration period.

Figure 4. Enrollment-weighted overall Star Ratings, 2009 to 2015



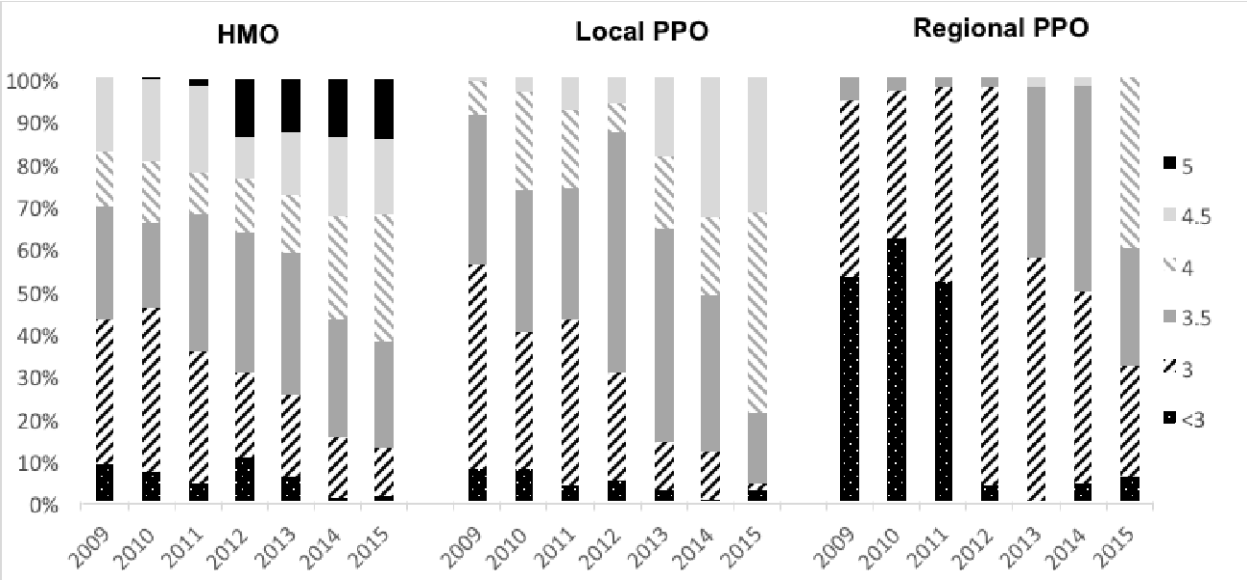
Sources: Publicly available Star Ratings are from http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html. Contracts of interest and other characteristics were identified using data from HPMS. Enrollment data are from CMS monthly enrollment files.

Note: L&M estimated the contract-weighted overall Star Ratings for 2009 and 2010. Figures exclude contracts that were too new or low enrollment and thus did not have a Star Rating. Contract enrollment for the Star Ratings year was used to weight the Ratings data, e.g., the 2013 Star Ratings were weighted using 2013 enrollment. The (n)s at the bottom of the figure are the totals for contracts in the figure and do not represent the total MA population in the given year.

Enrollment-weighted overall Ratings for HMO and local PPOs show increases, with local PPOs showing dramatic increases in 2015

As shown in Figure 5, HMOs and local and regional PPO contracts' enrollment-weighted Ratings improved over the demonstration period. In 2013, the median enrollment-weighted Star Rating for HMO and local PPO contracts was 3.5; by 2015, that had improved to 4 Stars for both HMO and local PPOs. While the improvement for HMOs, which represent about 65 percent of enrollment in 2015, follows a steady trend that began in the pre-demonstration period, the improvement for local PPOs, which represent about 22 percent of enrollment in 2015, was more dramatic in 2015 as enrollment shifted to higher-rated contracts. In 2013, 41 percent of beneficiaries in HMOs were in 4-Star or higher contracts; by 2015, that share increased to 62 percent. In 2013, 36 percent of local PPO enrollees were in 4-Star or higher contracts; by 2015, that share increased to 79 percent. The median for regional PPOs, which represented about 8 percent of enrollment in 2015, remained lower than HMOs and local PPOs, though also improved over the period. By 2015, 40 percent of regional PPO enrollees were in a single regional PPO that had a 4-Star Rating.

Figure 5. Overall Star Ratings for HMOs, local PPOs, and regional PPOs, 2009 to 2015



Sources: Publicly available Star Ratings are from http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html. Contracts of interest and other characteristics were identified using data from HPMS. Enrollment data are from CMS monthly enrollment files.

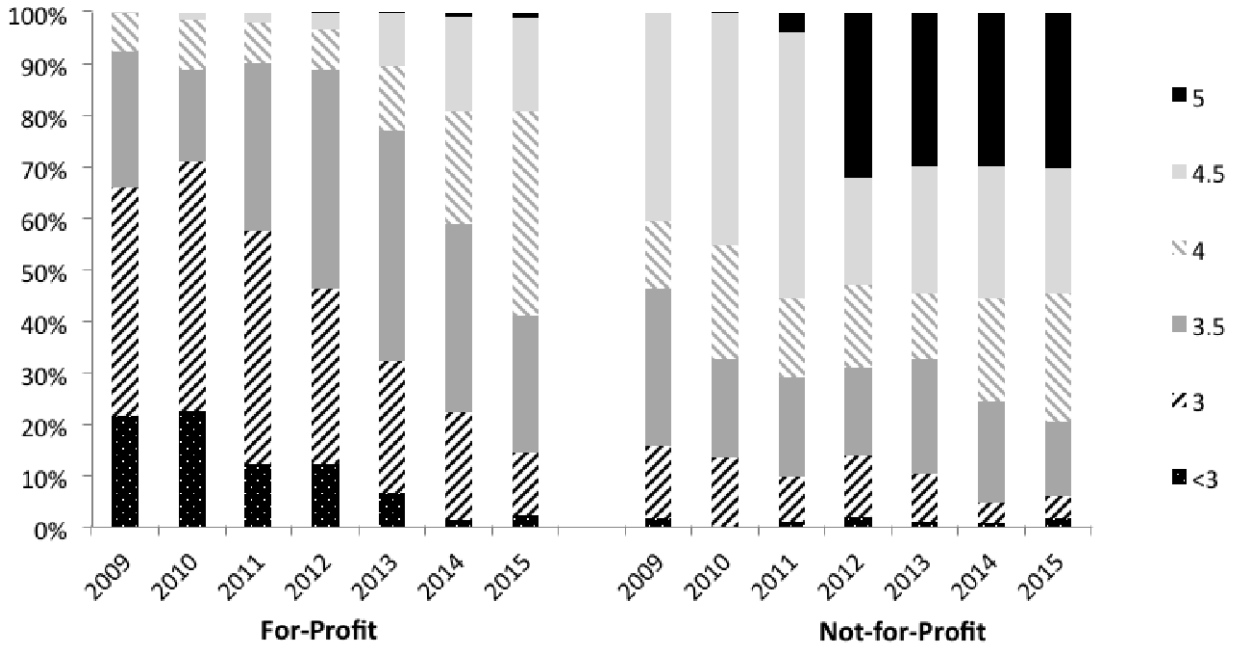
Note: L&M estimated the contract-weighted overall Star Ratings for 2009 and 2010. Figures exclude contracts that were too new or low enrollment and thus did not have a Star Rating. Contract enrollment for the Star Ratings year was used to weight the Ratings data, e.g., the 2013 Star Ratings were weighted using 2013 enrollment. "HMO" includes contracts with plans labeled "HMO" and "HMOPOS" in HPMS.

Trends in overall Star Ratings reflect changes in for-profit contract Ratings

The shift in the overall Star Ratings distribution generally reflects movements in the overall Star Ratings distribution of for-profit contracts, which had lower star ratings the start of our period of interest. Enrollment-weighted Star Ratings improved meaningfully between 2010 and 2015 for for-profit contracts—see Figure 6. These contracts make up roughly three-quarters of contracts with Ratings and, thus, are reflected in the overall Ratings distribution.

Not-for-profit contracts experienced a substantial shift in the enrollment of beneficiaries in plans of 5-Star contracts. Between 2011 and 2012, not-for-profit contracts moved from having four percent of their beneficiaries enrolled in a 5-Star contract to thirty-two percent. For the remainder of the period, not-for-profit contracts had thirty percent of their beneficiaries enrolled in a 5-Star contract. This increase is the result of one large contract moving from a 4.5-Star to a 5-Star Rating prior to the demonstration. Meanwhile, for-profit contracts improved in-line with general trends.

Figure 6. Overall Star Ratings by tax status, 2009 to 2015



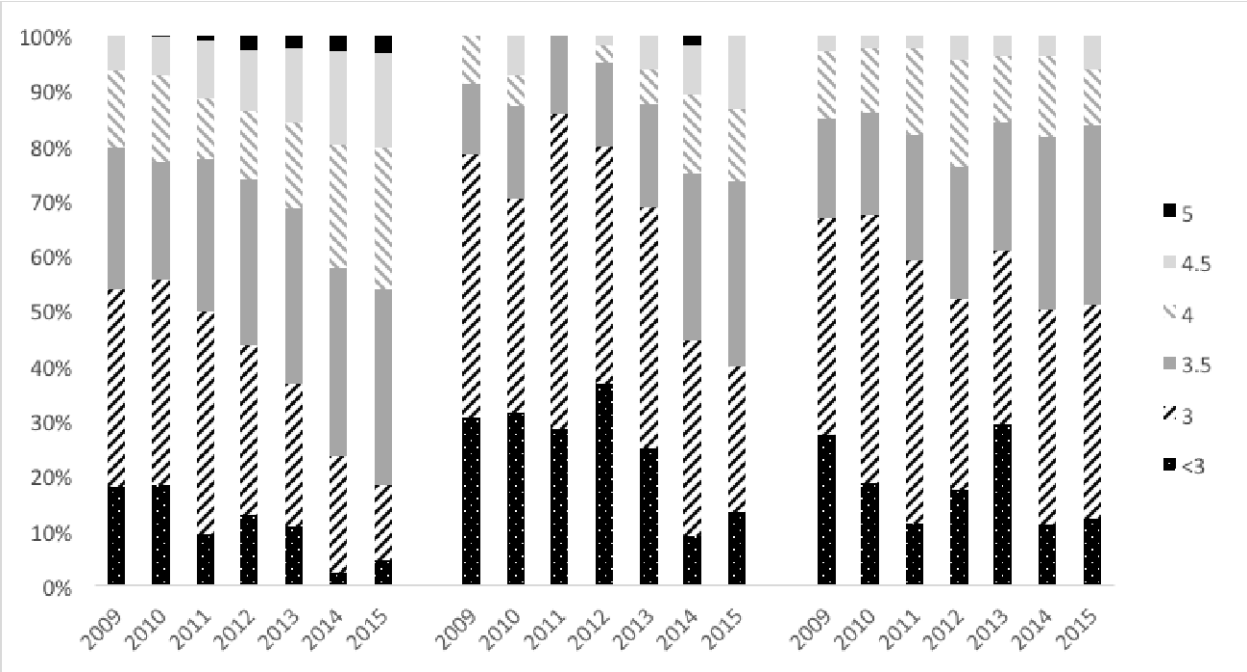
Source: Publicly available Star Ratings are from http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html. Contracts of interest and other characteristics were identified using data from HPMS.

Note: L&M estimated the contract-weighted overall Star Ratings for 2009 and 2010. Figures exclude contracts that were too new or low enrollment and thus did not have a Star Rating. Contract enrollment for the Star Ratings year was used to weight the Ratings data, e.g., the 2013 Star Ratings were weighted using 2013 enrollment.

Contracts with no SNP enrollment had the highest Ratings compared to contracts with SNP enrollment

Contracts that did not have any SNP plans (see the left-most set of bars in Figure 7) followed the trends in overall Star Ratings observed across all contracts. Contracts with a mix of SNP and non-SNP plans (middle set of bars) or SNP plans (right-most set of bars) showed less improvement in Star Ratings over the period.

Figure 7. Overall Star Ratings by SNP enrollment, 2009 to 2015



Source: Publicly available Ratings are from http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html. Contracts of interest and other characteristics were identified using data from HPMS.

Note: L&M estimated the contract-weighted overall Star Ratings for 2009 and 2010. Figures exclude contracts that were too new or low enrollment and thus did not have a Star Rating. Contract enrollment for the Star Ratings year was used to weight the Ratings data, e.g., the 2013 Star Ratings were weighted using 2013 enrollment.

In the next year, 2.5-Star contracts were the most likely to improve, 4.5-Star contracts were the least likely

To examine improvement throughout the Ratings continuum, we calculated the share of contracts at each half-Star level with higher and lower overall Star Ratings in the next year in the pre-QBP demonstration period (2009 through 2012) and the QBP demonstration period (2012 through 2015). Figure 8 displays the proportion of contracts with a higher Star Rating in the next year, conditional on the current year Star Rating. Specifically, the y-axis shows the share of contracts that improved from one year to the next (i.e., a higher level on the y-axis means a higher proportion improved). The slope of the line shows if this proportion changed over time. Figure 8 shows that 2.5-, 3-, and 3.5-Star contracts generally showed the highest likelihood of improvement over the pre- and QBP demonstration periods. Overall, 2.5-Star contracts were the most likely to improve in the next year and 4.5-Star contracts were the least likely to improve; this pattern was fairly consistent in the pre-QBP and QBP demonstration period. Across the Ratings continuum, fewer contracts improved Star Ratings in the final year, 2014 to 2015, than in the prior year, 2013 to 2014.

As shown in Figure 9, the probability of a contract declining in the next year remained relatively stable, with a slight uptick between 2014 and 2015. Data for the 2016 Star Ratings, which have not yet been released, were collected partially during the demonstration (CY2014) and partially

after (CY2015). The Star Ratings for 2016 and 2017 will provide more information about how Ratings change once the QBP demonstration bonus payments end and the ACA bonus payment system is implemented. Transition matrices for each year-to-year change by overall Star Rating can be found in Appendix 4. Transition Matrices.

Figure 8. Share of contracts with a higher overall Star Rating in the next year

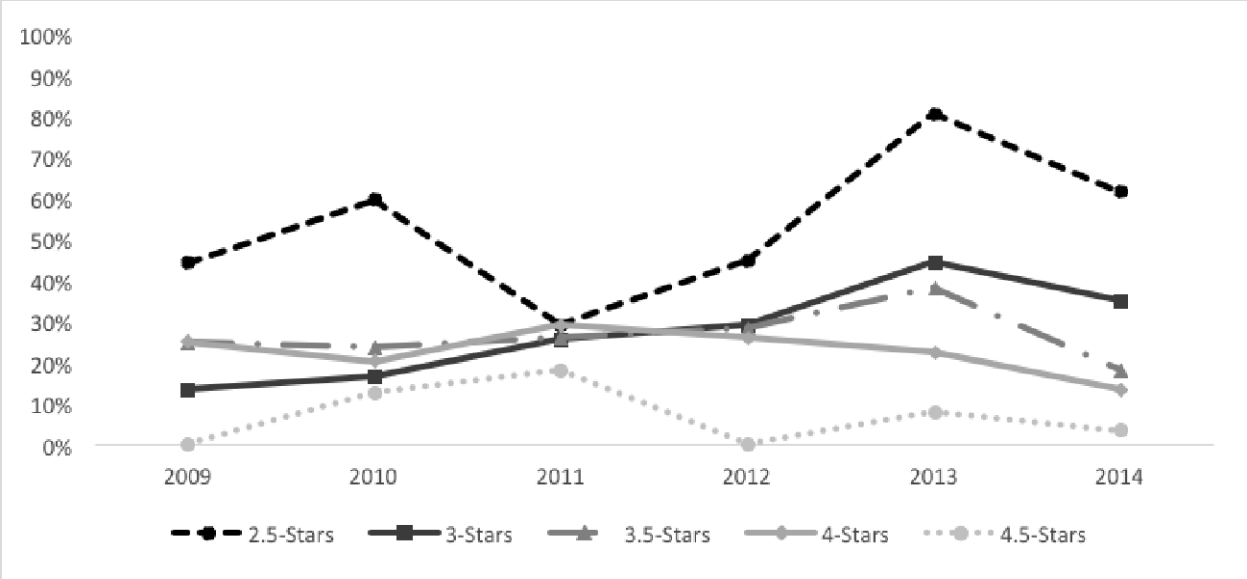
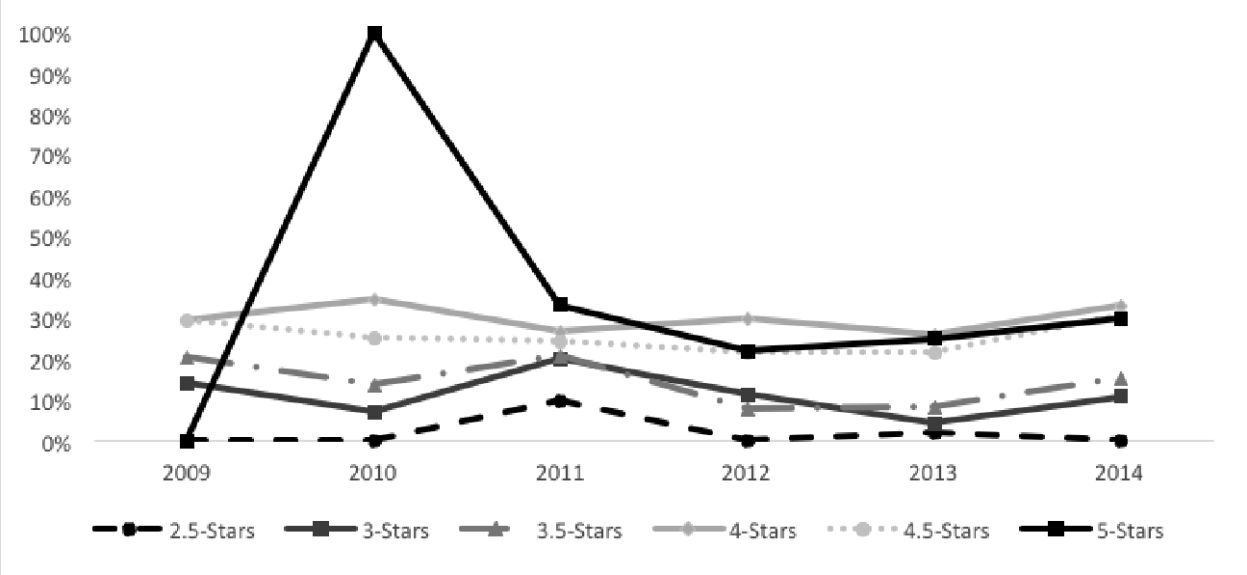


Figure 9. Share of contracts with a lower overall Star Rating in the next year



Source: Publicly available Star Ratings are from http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html. Contracts of interest and other characteristics were identified using data from HPMS.

Finding 3. Compared to other coverage types, MA contracts show similar trends in average enrollment-weighted scores for selected measures

As noted in the background, the absence of an appropriate comparison group makes it difficult to identify which, if any, observed changes in overall Star Ratings during the QBP demonstration period are attributable to the QBP demonstration. To examine whether trends in MA contracts deviated from trends in other sectors during the QBP demonstration, we examined changes in national, enrollment-weighted means for several individual Part C and Part D measures over the period of interest (pre-demo and demo period) and compared those trends to enrollment-weighted means for comparison groups—commercial plans, Medicaid plans, Medicare FFS, and stand-alone PDPs. Although these comparison groups suffer from shortcomings, we use these comparisons to evaluate trends in scores across MA and comparison groups in the pre-QBP and QBP demonstration periods.

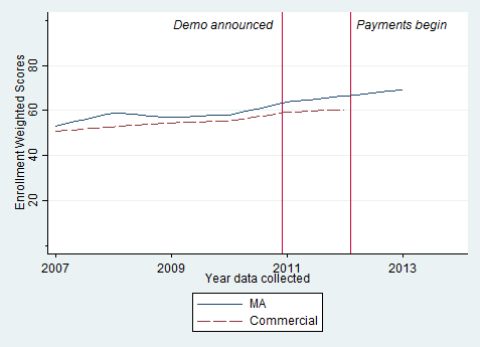
In general, MA contracts' enrollment-weighted average scores for Part C measures were higher in 2013 (data collection year) than in 2007 (data collection year).[4] Across the set of measures for which we have commercial and Medicaid comparison group data, we observed no notable patterns of deviation from the trend in enrollment-weighted score for MA contracts between the pre-QBP and QBP demonstration periods, nor do we observe deviations between the MA trends relative to the comparison group trends. Figure 10 shows largely parallel trends across multiple measures for which we have Medicaid and commercial plan data.

We also compared a set of consistently measured Part D measure scores for MA plans and stand-alone PDPs, which do not receive quality bonus payments, to assess the effect of the QBP demonstration on Part D scores. Similar to the results in Figure 10, Figure 11 shows similar trends in the national average enrollment-weighted scores on these Part D measures for MA-PD plans and stand-alone PDPs. See Appendix 5 for MA contracts' national average enrollment-weighted scores for the measures presented in Figure 10 and Figure 11, as well as additional measures for which we do not have comparison groups from 2009 to 2015.

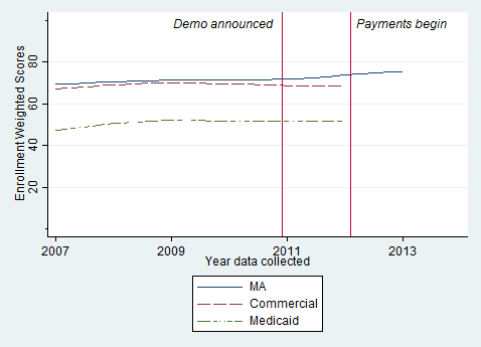
[4] Note that in the measure-level analysis in this section, scores are labeled with the year the data were collected to make them comparable across sectors.

Figure 10. Enrollment-weighted quality measure scores for select Part C measures: MA contracts vs. comparison

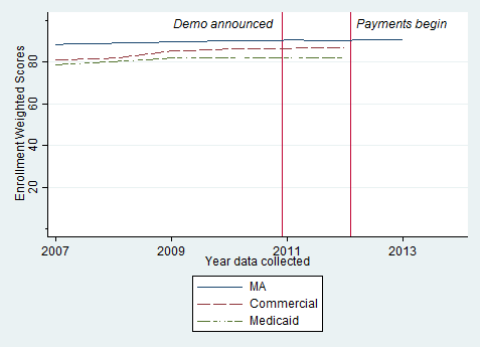
Colorectal cancer screening



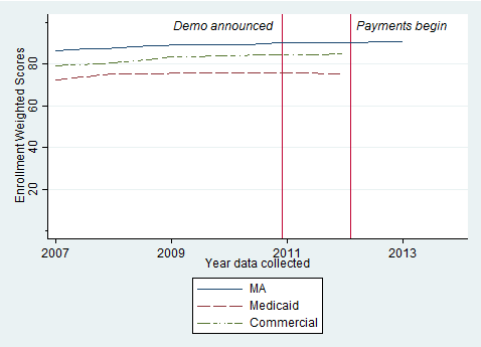
Breast cancer screening



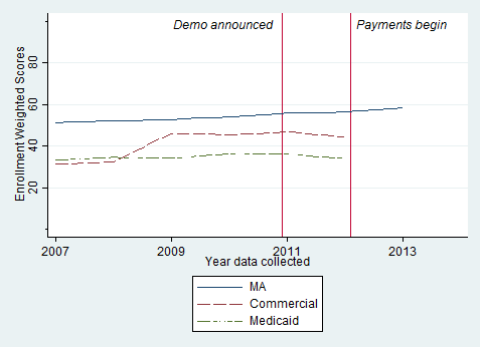
Cholesterol screen for members with heart disease



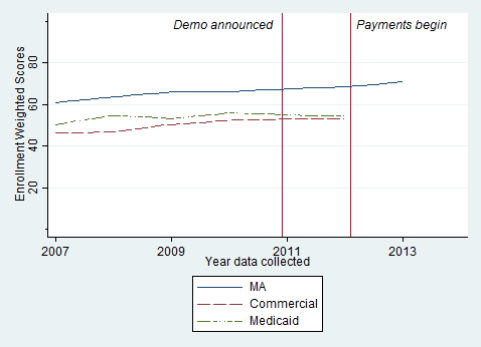
Cholesterol screen for members with diabetes



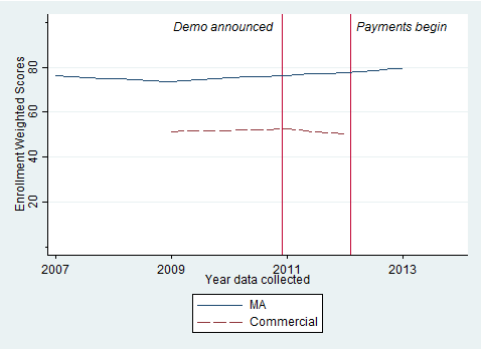
Diabetes care—cholesterol controlled



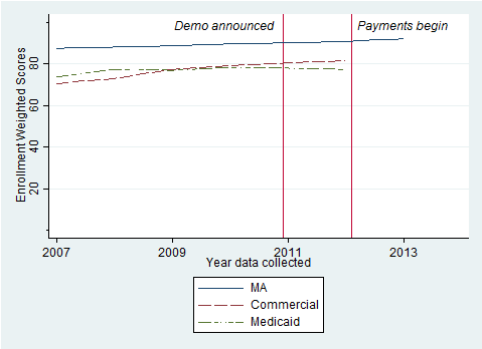
Diabetes care—eye exam



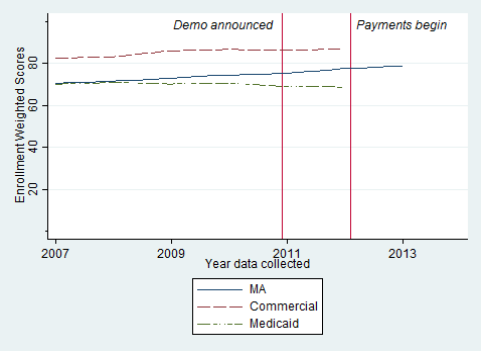
Diabetes care—blood sugar controlled



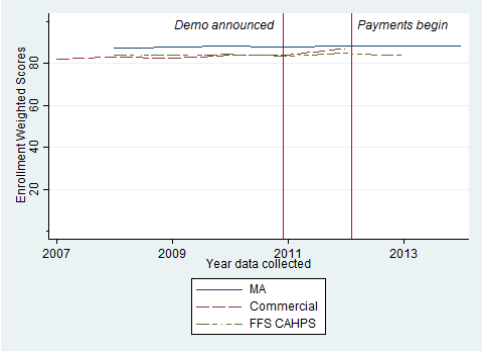
Diabetes care—kidney disease screening



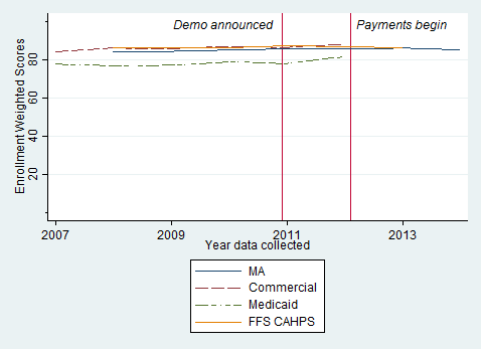
Rheumatoid arthritis management



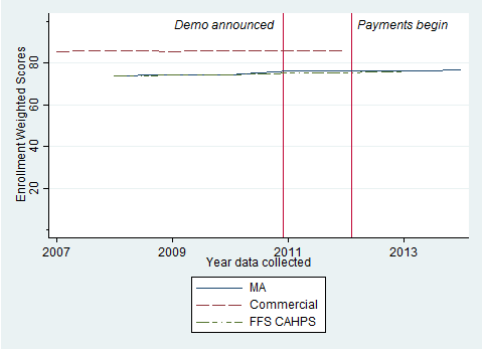
Customer service



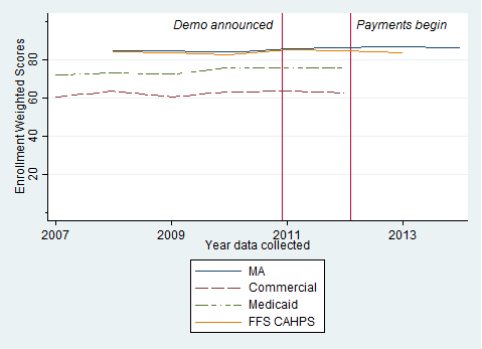
Ease of getting needed care and seeing specialists



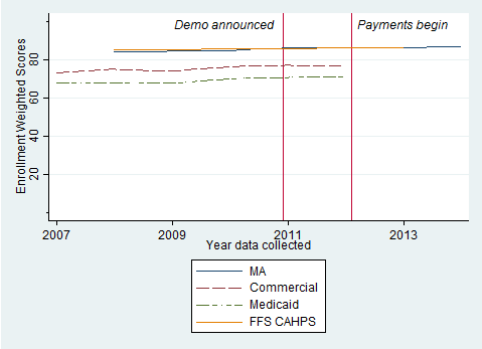
Getting appointments and care quickly



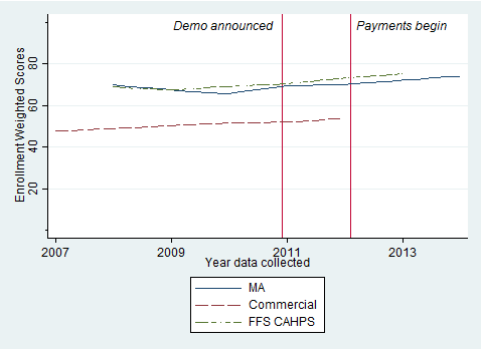
Members' overall rating of health plan



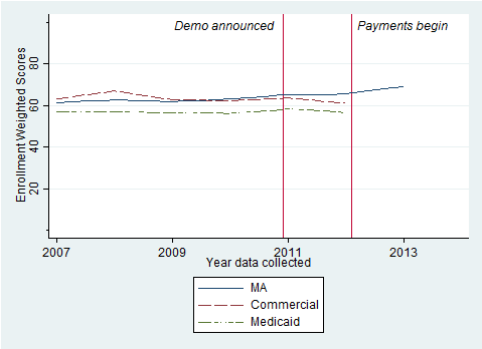
Overall rating of health care quality



Flu vaccine



Controlling blood pressure

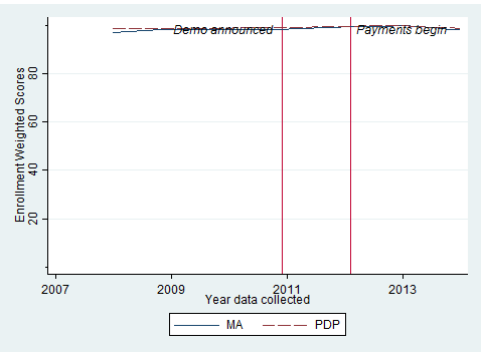


Sources: MA data sources: Star Ratings from http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html, contracts of interest and other characteristics were identified using data from HPMS, enrollment data are from CMS monthly enrollment files. FFS CAHPS data were provided to the evaluation team by CMS. Medicaid. Commercial and Medicaid plan data, including measure scores and enrollment were obtained from NCQA.

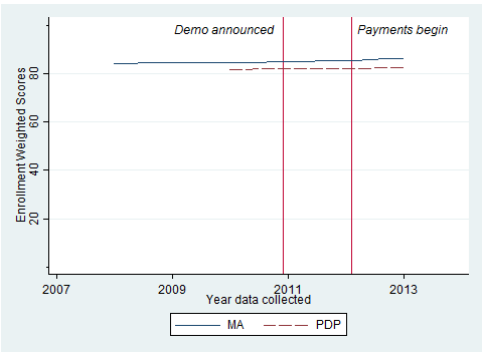
Notes: Data are shown by collection year to compare across sectors. The first vertical line indicates the announcement of the QBP demonstration; the second vertical line indicates the beginning of QBP payments. Data were weighted by enrollment in the year the data were collected, except for the NCQA data, which were weighted by enrollment in the year the data were released.

Figure 11. Enrollment-weighted quality measure scores for select Part D measures: MA contracts vs. stand-alone PDPs

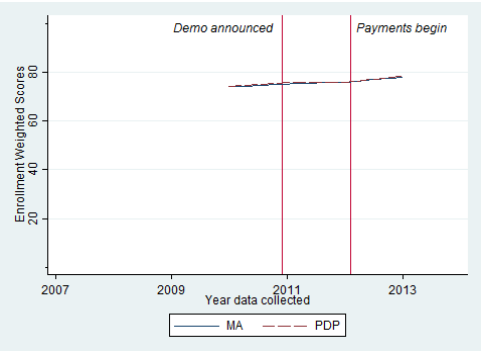
Availability of drug coverage and cost information



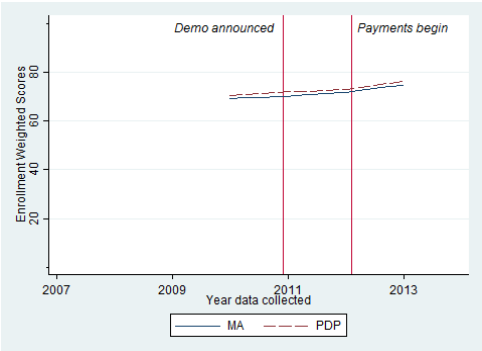
Diabetes treatment



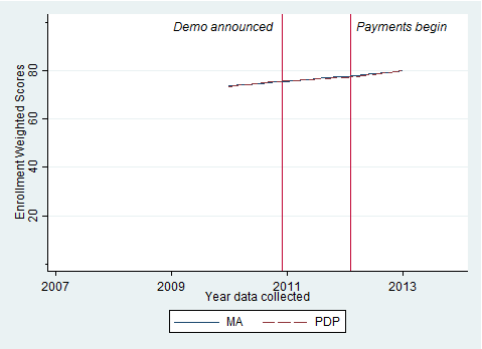
Medication adherence—oral diabetes medications



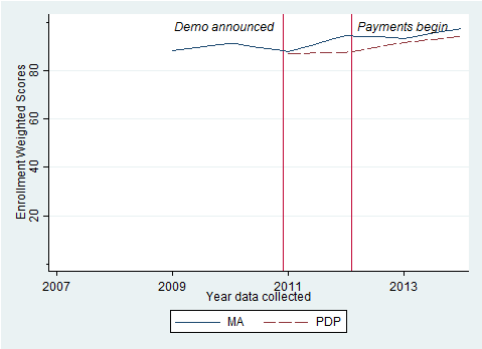
Medication adherence—statins



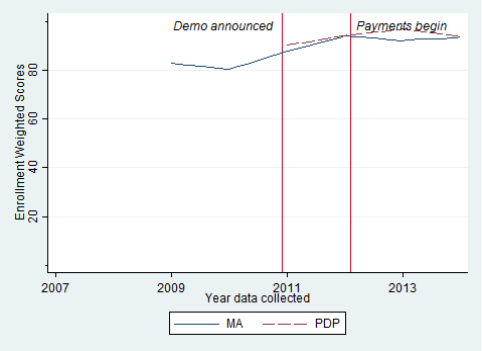
Medication adherence—ACEIs/ARBs



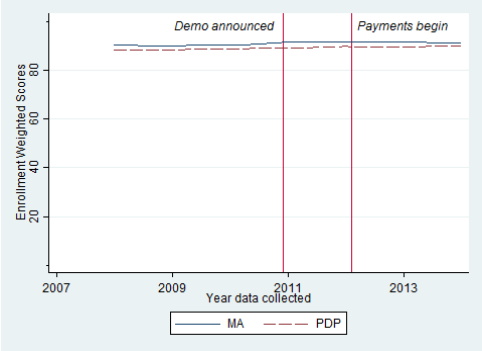
Timely decisions about appeals



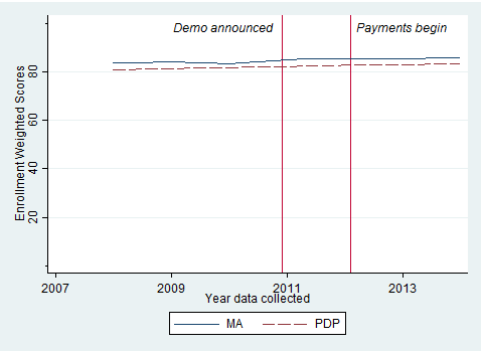
Appeals decisions upheld



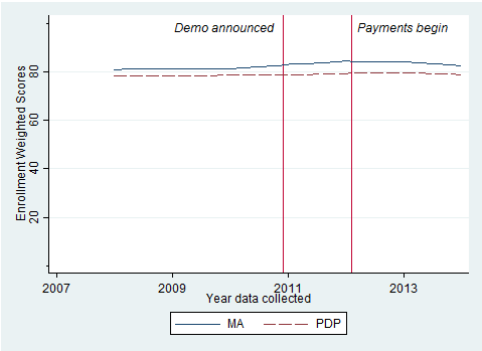
Getting needed prescriptions



Members' overall rating of drug plan



Getting information from the drug plan



Sources: MA data sources: Star Ratings for MA and PDPs from http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html, contracts of interest and other characteristics were identified using data from HPMS, enrollment data are from CMS monthly enrollment files.

Notes: Data are shown by collection year and weighted using enrollment data for the year that the data were collected. The first vertical line indicates the announcement of the QBP demonstration; the second vertical line indicates the beginning of QBP payments.

Finding 4. MA organizations reported that the demonstration encouraged a focus on QI efforts

To understand how the QBP demonstration may be affecting organizations' operations and affecting their measure scores and Star Ratings, we collected information regarding the quality improvement (QI) activities of MA plan sponsors through a contract-level survey. The survey, conducted in the summer of 2013, was mailed to 533 contracts (165 parent organizations). We received completed questionnaires for 442 contracts, a response rate of 83 percent. See the Methods section of this report for additional details about development of the sampling frame, data collection procedures, response rate, and data processing.

The majority of survey respondents (88 percent) indicated that the budget for the contract's QI activities increased from 2010 levels (when the demonstration was announced) by 2013 (when the survey was conducted). Of those reporting a budget increase, most indicated that the increase was, to "a large extent" (70 percent) or "moderate extent" (23 percent), due to expansion of QI activities specifically related to MA's Star Ratings.

Overall, the survey results show that most respondents (281/442, 64 percent) think that the QBP demonstration helped to improve their contract's Star Rating. The remaining 36 percent were evenly split among those who said that it was too soon to tell and those who said the QBP demonstration has not helped improve the contract's Star Rating. Among those who said that the QBP demonstration helped to improve the contract's Star Rating, the most commonly cited mechanism was the incentive it provided to senior management to focus on improving Star Ratings, closely followed by the funding QBP provided to implement or expand quality improvement. (See Table 7.)

Table 7. Ways in which the demonstration supported quality improvement

How did the QBP demo affect quality improvement?	Share of contracts (N=281)
Funded implementation/expansion of QI activities	83.3%
Incentivized senior management to focus on Star Ratings	96.8%
Incentivized staff to focus on Star Ratings	53.7%
Other	3.6%

Source: L&M/MPR MAO contract survey

Case study discussions supported and elucidated the results from the survey. Several of the case study organizations indicated that the demonstration garnered focus from leadership on Star Ratings and their improvement because it tied funding to QI in the wake of MA payment reductions in the ACA. One case study organization explained that Star Ratings are an important part of its business strategy because the Ratings are a revenue stream for contracts that are performing well as MA reimbursement declines. Another case study organization noted that QI had historically been a cost center, but the opportunity to make up some of the ACA's MA payment reductions through QBPs is changing that. (For additional information, see the complete case study report,

"Findings from Medicare Advantage Organization Case Studies," available at http://innovation.cms.gov/Files/reports/MAQBP-FirstEvalRpt.pdf.)

Provider incentive programs are the only QI activity associated with changes in Star Ratings from 2013 to 2015

Survey respondents were asked to indicate in which of 19 types of QI activities (plus an "other" category) the contract engaged. Excluding the "other" category, the survey results indicated that all 19 types of activities were common, with at least 60 percent of respondents reporting the contract engaged in any given QI activity.

To determine whether any specific QI activities are associated with improvement or decline in the Star Ratings, the team estimated an ordinary least squares regression, using whether the contract's overall Rating improved or declined/stayed the same between Star Ratings year 2013 and 2015 as the dependent variable.[5] The team used the following independent variables for the model:

- The number of activities in which the plan engaged (which ranged from zero to twelve).
- The contract's overall Star Rating in 2013.
- An indicator variable for parent organization Aetna, the parent organization of 23 of the 416 contracts that completed the survey.[6]

In estimating this model, we found the QI activity 'Provider incentive programs' was the only statistically significant QI activity and associated with contracts that improved their overall Rating between 2013 and 2015. Table 8 provides the number of contracts engaging in each type of QI activity and the results of the regression analysis.

[5] The linear model is descriptive of how the proportion of surveyed plans improving between Star Ratings years 2013 and 2015 varied with engagement in the activities. A linear model was used, rather than logistic regression, to describe how the means/proportions in the sample varied with the QI activities, rather than inferring how the likelihood of improvement changes with an activity.

[6] Of the 442 contracts answering the survey in 2013, 26 were no longer operating in 2015, leaving a total N of 416 for this analysis. Comprising a high proportion of contracts engaging in "other" activities (23 of 48), a relationship between the parent organization, Aetna, and Star Ratings outcomes may drive a spurious relationship between "other" activities and Star Ratings outcomes. Therefore, we included an indicator variable for the parent organization to measure the effect of "other" activities independent of a potential relationship between the parent organization and Star Ratings outcomes.

Table 8. Coefficients and P-values for quality improvement regression on improvement category

	Contracts with Characteristic (N=416)	Coefficient	P-Value
Quality Improvement (QI) Activity			
Provider and/or patient portal systems	276	0.03	0.6830
Audit and feedback	386	-0.17	0.1140
Provider education	398	0.10	0.4090
Patient education	381	-0.02	0.9070
Changes in operation	390	-0.09	0.3720
Information systems	350	0.03	0.7220
Identification of care gaps	408	-0.03	0.8720
Facilitating access to services	327	-0.11	0.1620
Incentivizing members to get needed services	256	0.07	0.2490
Other methods of member engagement	367	0.07	0.3800
Benefit modifications	327	0.09	0.2820
Formulary modifications	347	0.10	0.3040
Provider incentive programs	**337**	**0.30**	**0.0010**
Compliance with NCQA accreditation standards	306	0.13	0.0680
Compliance with state and federal regulations	370	-0.18	0.1870
Physician credentialing and medical affairs supporting evidence based practices	321	0.02	0.7910
Overall utilization management	386	-0.12	0.3150
Service QI improvement activities	400	0.12	0.3190
Optimizing medical loss ratio	354	0.08	0.3820
Other QI activity	48	0.09	0.4360
Number of QI activities identified by Contract			
2	1	0.51	0.3090
3	7	0.16	0.6720
4	6	0.56	0.0820
5	11	0.19	0.4550
6	37	0.23	0.3540
7	22	0.39	0.0600
8	31	0.49	0.0110
9	75	0.18	0.3330
10	32	0.04	0.8140

	Contracts with Characteristic (N=416)	Coefficient	P-Value
Quality Improvement (QI) Activity			
11	186	0.01	0.9370
12	6	(omitted)	(omitted)
2013 Overall Star Rating			
Not available	2	1.16	0.0000
Plan too new to be measured	29	1.13	0.0000
Not enough data available	37	1.11	0.0000
2.5	31	0.94	0.0000
3	82	0.63	0.0000
3.5	106	0.47	0.0040
4	51	0.26	0.1210
4.5	70	0.11	0.4850
5	8	(omitted)	(omitted)
Other			
Aetna Parent Organization	23	-0.22	0.3460
Constant	N/A	-0.52	0.1450

Source: L&M/MPR MAO contract survey. MA data sources: Star Ratings from http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html.

Note: Of the 442 contracts answering the survey in 2013, 26 were no longer operating in 2015, leaving a total N of 416 for this analysis.

Provider incentive programs showed a strong correlation with improvement (significant at the 0.01 level). Of the surveyed contracts, the proportion of those who showed improvement was 0.3 higher (i.e., 30 percentage points higher) for contracts that engaged in provider incentive programs versus contracts that did not.

While these programs more commonly experienced Star Ratings improvements, there are limits on how broadly they could be replicated. Certain plan types are better suited to incentive programs than others. For example, PFFS plans, with less well-defined networks and a lack of network restrictions, have fewer opportunities to put such provider incentive programs in place. In addition, even HMOs with their more extensive physician contracting may find that these incentive programs are successful for only a subset of physicians. If an HMO's members make up a small proportion of an individual physician's patient panel, then that physician may not find it financially worthwhile, or even feasible, to make adjustments in response to a plan's incentive program.

Provider incentive programs were also discussed with the case study participants. These discussions also suggest limits on provider incentive programs. While most of the participating MA organizations offer financial incentives for meeting quality goals, they generally reported that arrangements are limited to only to the most engaged segments of their provider networks.

However, there appear to be opportunities to expand such programs. One case study organization reported that it does not yet offer a provider incentive program but would like to have one in the future. At this time, the organization does not track quality at the individual physician level but is exploring data collection approaches. The organization reported additional challenges that will need to be resolved, including some resistance to pay for performance (P4P) models within the network and the frequency with which members transition in and out of its products, which they feel complicates holding physicians accountable for the quality of care provided.

Finding 5. Enrollment changes are weakly related to changes in Star Ratings and not appreciably different between the QBP demonstration period and previous years

To assess whether enrollment changes are associated with changes in Star Ratings, we examine enrollment by overall Star Rating over time at the plan-level. Overall, our analyses show enrollment changes are weakly related to changes in Star Ratings, with a higher average enrollment change for plans that are part of 5-Star contracts and lower average enrollment change for those that are part of 2-Star contracts. There is, however, no statistically (or meaningfully) different relationship between increases in enrollment and changes in Star Ratings when comparing the demonstration period to previous years.

Unless otherwise noted, enrollment figures represent total beneficiary months[7] in a year for MA contracts, plans, or plan-county combinations. Because consolidations regularly occurred during the time period in question, longitudinal crosswalks from the HPMS were used to track plans and contracts over time. Changes in enrollment were then adjusted for plan consolidation. For instance, if two active plans in 2012 were consolidated into one plan in 2013, then the change in enrollment for the consolidated 2013 plan would be represented as the change between the sum of enrollment from the two 2012 plans and the total enrollment of the consolidated 2013 plan.

As contract Star Ratings were changing over time, we use these changes to identify a relationship between star ratings and changes in enrollment. The transition matrix in Table 9 captures this year-to-year contracts-level Star Ratings movement during the years 2009 to 2014. Year-to-year changes for any given level are meaningful; the most stable Star Ratings group is 5-Star contracts, of which 75 percent maintained a 5-Star Rating in the following year. Changes in enrollment may follow changes in star ratings for contracts. However, examining year-over-year changes in enrollment at the contract level is tenuous for accurately describing changes in enrollment. The plans comprising a given contract may change from year to year. As such, we look at plan-level enrollment changes.

[7] A beneficiary month is equal to one month that a beneficiary is enrolled. If a beneficiary is enrolled for nine months during the year then this beneficiary contributes nine beneficiary months. Twelve beneficiary months represent a beneficiary year. The enrollment analyses use total annual enrollment for the years 2009 through 2014. We did not use 2015 because enrollment data for the entire calendar year are not yet available. Where we use 2015 enrollment data elsewhere in this report, we utilize beneficiary months in March 2015 enrollment. For all other years, we use beneficiary years of enrollment in the calendar year, which are comparable to beneficiary months in one month of data.

Table 9. Year-to-year transitions of contracts between Star Ratings, 2009 to 2014

	Rating in the Following Year								
	≤2 Stars	**2.5 Stars**	**3 Stars**	**3.5 Stars**	**4 Stars**	**4.5 Stars**	**5 Stars**	**Not Rated**	**# Contracts**
≤2 Stars	0%	67%	33%	0%	0%	0%	0%	0%	6
2.5 Stars	2%	45%	47%	5%	0%	0%	0%	2%	259
3 Stars	0%	12%	62%	23%	2%	0%	0%	1%	633
3.5 Stars	0%	0%	13%	57%	24%	5%	0%	0%	445
4 Stars	0%	0%	0%	29%	46%	24%	0%	0%	237
4.5 Stars	0%	0%	1%	1%	20%	70%	9%	0%	152
5 Stars	0%	0%	0%	0%	5%	20%	75%	0%	20
Not rated	0%	7%	9%	6%	3%	1%	0%	73%	632
Contract count	9	240	640	522	273	200	31	469	2,384

Note: Table notes the proportion of contracts with a given Star Rating in one year, noted vertically, with each Star Rating listed horizontally in the following year. Contracts with the same Star Rating two years in a row are found on the diagonal.

To describe plan enrollment longitudinally, we identified year-to-year changes in plan enrollment using aggregated enrollment over consolidated plans, if consolidation occurred.[8] Table 10 displays the distribution of year-to-year plan-level enrollment changes for the 2009 to 2014 time period, using enrollment data from 2008 to 2014. On average, enrollment increased over this time period, with more than half of the plan-year combinations showing enrollment increases. Some differences in enrollment changes by Star Ratings warrant particular note:

- Plans in 2.5- to 4.5-Star rated contracts experienced mean increases in enrollment of roughly 5,000 beneficiary months, with the median increase less than 1,000 beneficiary months. Sixty-two to sixty-five percent of plan-years with these Ratings experience a year-over-year increase in enrollment.

- Plans in 5-Star contracts experienced larger average increases in enrollment than did other plans, with mean and median increases of roughly 10,000 and 2,000 beneficiary months, respectively. A larger proportion of 5-Star plan-years experienced enrollment increases, roughly 77 percent, than did plans that were a part of contracts with lower Star Ratings.

- Plans in contracts with Ratings of 2 Stars or below experienced a mean decrease in enrollment, though the majority (57 percent) of plans with Ratings of 2 or fewer Stars still experienced enrollment increases.

[8] For instance, if two plans in 2009 consolidated to become one plan in 2010, then we compared 2010 enrollment to 2009 enrollment *summed over the two original plans*.[8]

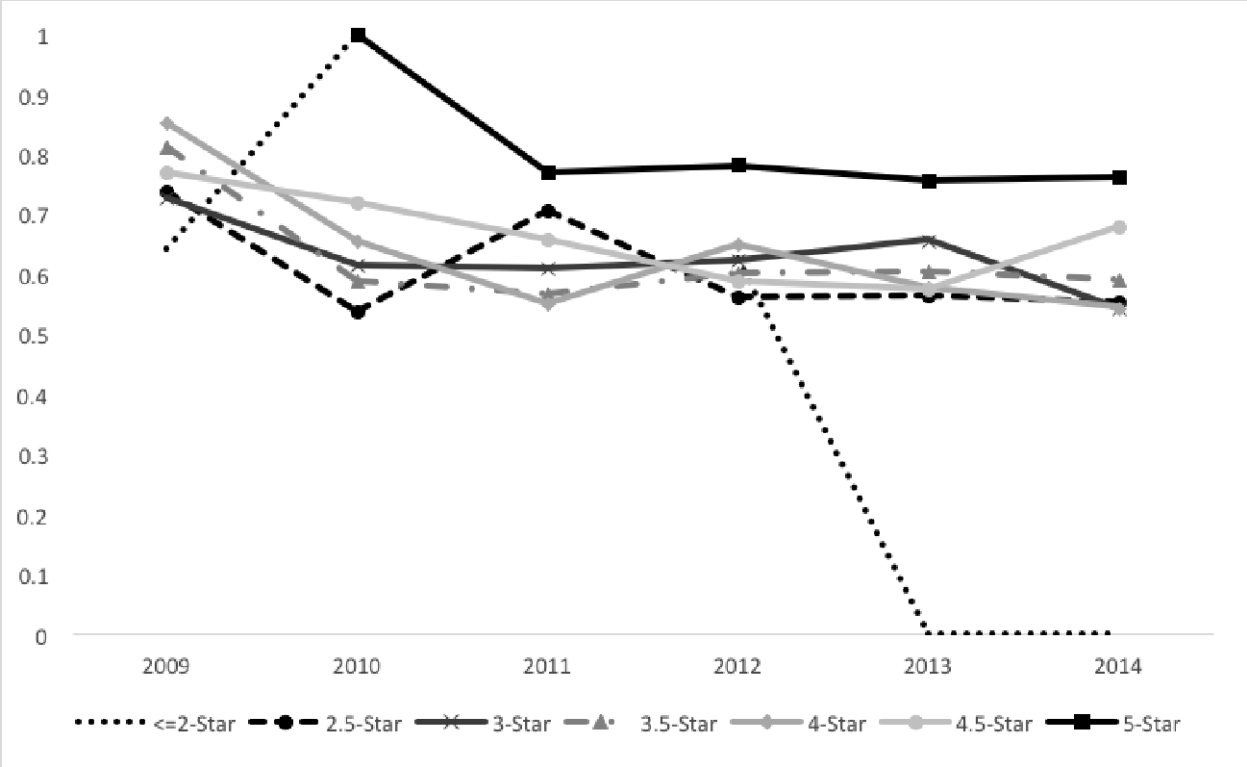
Table 10. Mean changes in plan enrollment by contract-level Star Rating, 2009 to 2014

	All plans	5 Star	4.5 Star	4 Star	3.5 Star	3 Star	2.5 Star	≤2 Star
Obs	15,407	306	1,689	1,949	3,774	3,863	1,783	67
Mean	4,966	10,135	4,636	4,487	6,627	4,505	5,384	-5,478
Std. Dev.	50,577	23,201	65,335	28,542	42,407	64,537	39,605	55,982
Percentile								
1st	-77,526	-18,198	-90,491	-70,721	-59,365	-105,398	-69,877	-387,311
5th	-17,010	-5,578	-12,104	-15,415	-18,921	-21,805	-17,894	-19,235
10th	-6,697	-2,525	-4,851	-7,978	-8,252	-8,358	-6,500	-8,422
25th	-622	30	-355	-956	-1,111	-989	-762	-1,122
Median	719	2,069	831	548	738	836	620	215
75th	6,175	9,843	5,903	5,688	6,871	7,592	6,238	4,566
90th	20,770	34,217	20,028	20,624	23,640	24,071	19,511	16,165
95th	41,415	53,444	35,242	42,042	47,719	45,744	38,752	42,636
99th	123,099	96,200	94,636	104,554	145,679	137,562	133,068	52,972
Proportion Increased Enrollment								
	0.652	0.768	0.654	0.625	0.623	0.640	0.629	0.567

Notes: Table shows the distribution of year-to-year changes in enrollment for MA plans over the 2009 to 2014 time period. Enrollment figures were aggregated over plans prior to consolidation, if plan consolidation occurred. Star Ratings reflect the Ratings achieved in the second year of a year-to-year change, such that conditioning on a Star Rating of x Stars shows enrollment changes from the previous year to the current year for a plan currently holding a Star Rating of x Stars.

Figure 12 displays the proportion of plans by contract-level Star Rating experiencing year-over-year enrollment increases in each year, 2009 to 2014. As noted, a higher proportion of 5-Star plans experienced an enrollment increase in each year than 2.5- to 4.5-Star plans any year. In addition, the proportion of plans with increasing enrollment does not consistently differ across 2.5- to 4.5-Star plans over the 2009 to 2014 time period. Finally, though they are few in number, the proportion of plans with 2 or fewer Stars that experienced increases in enrollment was similar to plans with higher Ratings in 2009, 2010, and 2012. However, in 2013 and 2014 zero 2-Star plans experienced increases in enrollment.

Figure 12. Proportion of plans with increased enrollment by Star Rating, 2009 to 2014



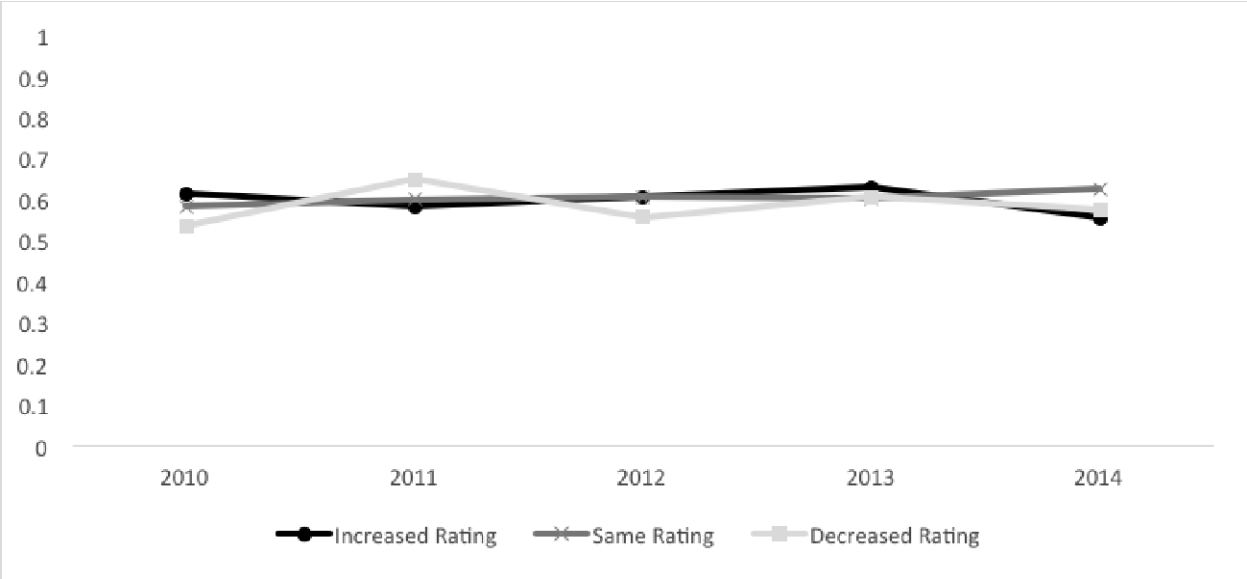
Source: Enrollment data were taken from the HPMS.

Note: Lagged enrollment sums across all previous plans for plans experiencing consolidation.

Enrollment increases do not appear to be related to changes in Star Ratings

While Figure 12 documents enrollment changes in plans by Star Rating, the question remains whether *changes*, declines or improvements, in Star Ratings are related to increasing enrollment. A first descriptive look at this relationship (see Figure 13) arrays the proportion of plans in a given year with increased enrollment over the previous year against whether the plan's contract-level Rating increased, decreased, or stayed the same. The relative flatness of the lines in the figure shows that plans with increasing enrollment were similar over the entire study period. Furthermore, this relationship did not vary with whether plans experienced increased or decreased contract-level Star Ratings.

Figure 13. Proportion of plans with increasing, decreasing, or equal overall Star Ratings experiencing increased year-to-year enrollment, 2010 to 2014



Source: Enrollment data were taken from the HPMS.

Note: Lagged enrollment sums across all previous plans for plans experiencing consolidation.

To examine the relationship between enrollment and Star Ratings in more depth and control for additional plan characteristics potentially affecting enrollment, we estimated a series of linear probability models regressing an indicator variable (whether a plan observed an increase in enrollment) on year, plan, and contract-level Ratings covariates. Specification (1) of Table 11 below displays coefficients from a model that includes indicator variables for whether a plan's contract Rating increased or decreased from the previous year (coefficients shown), as well as year fixed effects (average differences for each year common to all plans) and indicator variables for each Star Rating level available in the previous year (lagged Ratings). As such, the coefficients for "Increased" and "Decreased" show the average difference in mean proportions for plans that experienced an increase or decrease in Star Rating, relative to plans with a consistent Star Rating from the previous year. The regression identifies these differences while controlling for yearly effects common to all plans and for effects common to all plans at each lagged Star Rating. The p-values for these coefficients suggest that the difference in proportions of plans with increased year-over-year enrollment do not differ statistically for plans with changes in contract-level Star Ratings relative to those that maintained the same Star Rating.

Specification (2) builds on (1) by adding controls for the number of plans in the previous year consolidated to the current year's plan (indicator variables for each quantity), indicator variables for whether the plan experienced a service area expansion or reduction (two indicators), and indicator variables for the number of years that a plan was in existence during this time period (indicator variables for 1 to 6 years). Specification (3) adds plan fixed effects to the model and specification (4) adds the additional interaction of the variables of interest, whether a plan's Star Rating increased or decreased from the previous year, with an indicator for whether the increase or decrease occurred during the QBP demonstration (2013 or 2014).

The directions of the coefficients in all specifications indicate that plans with declines in Star Ratings from the previous year are less likely to have increased enrollment, particularly in the years since the QBP demonstration was implemented. However, plans in contracts with increasing Star Ratings are no more or less likely than plans in contracts with stable Star Ratings to experience an increase in enrollment. The coefficient on the interaction term in specification (4) further suggests that plans in contracts with an increased Star Rating were less likely to experience enrollment increases, relative to plans in contracts with stable Star Ratings. Though it should be noted that none of the specifications indicate that these differences are statistically different from zero at the 5 percent level of significance.

Table 11. Coefficients from linear probability model regressing whether an enrollment increase occurred on changes in overall Star Ratings

	Model Specifications			
	(1)	**(2)**	**(3)**	**(4)**
Star Rating Relative to Previous Year				
Increased	0.005	0.006	-0.008	0.007
	p=0.787	p=0.759	p=0.534	p=0.667
Decreased	-0.036	-0.038	-0.033	-0.016
	p=0.125	p=0.088	p=0.065	p=0.458
Increased*QBP Demonstration Period				-0.033
				p=0.105
Decreased*QBP Demonstration Period				-0.045
				p=0.180
Control Variables				
Year Fixed Effect	X	X	X	X
Lagged Rating Indicators	X	X	X	X
Number of Plans Consolidated (Indicators)		X	X	X
Service Area Expansions and Reductions (Indicators)		X	X	X
Number of Years in Period (Indicators)		X		
Plan Fixed Effect			X	X

Note: Coefficients from linear probability models regressing an indicator variable for whether a plan experienced an increase in enrollment on whether the plan's contract experienced a Star Ratings increase or decrease from the previous year.

Despite the lack of findings when examining simple Ratings increases, differential effects may exist depending on the Star Rating value from the previous year. QBP bonus payments are determined based on last year's Rating (the lagged Rating), rather than the current year's Rating. Furthermore, the bonus amounts differ based upon the lagged Star Rating value. As such, we may see differential effects of Star Ratings changes depending upon the initial Star Rating values for plans.

Table 12 displays coefficients from two regressions where the dependent variable is, again, an indicator for whether enrollment increased. The two regressions build in additional controls, including: time fixed effects, indicators for the number of plans consolidated, indicators for service area expansions and reductions, and plan fixed effects. The specifications in Table 13 incorporate an interaction between whether the plan experienced a Rating increase/decrease in the following year and the lagged contract Star Rating for the plan (the value from the previous year). The coefficients in Table 13, therefore, show the difference in the proportion of plans increasing enrollment for plans in contracts that experienced increases or decreases in Star Ratings (relative to plans in contracts with stable Star Ratings), taking into account the Star Rating the contract previously held. The coefficients in specification (1) show that plans in contracts that moved downward from a 5-Star Rating were 17.9 percentage points less likely to experience enrollment increases than plans that were part of 5-Star rated contracts that remained at a 5-Star Rating. Plans in 4.5-Star contracts that increased to a 5-Star Rating were 10.2 percentage points more likely to experience an increase in enrollment than those that remained at a 4.5-Star Rating.

Unadjusted p-values are listed with each of the coefficients determining differences in proportions from plans remaining at the same value. Given that multiple comparisons are being made across the Star Rating groupings, we thus apply a Bonferroni adjustment for multiple comparisons, which suggests that critical p-values for statistical significance (equivalent to $\alpha = 0.05$) would be $0.05/12 = 0.004$ for specification (1) and $0.05/24 = 0.002$ for specification (2). After applying this adjustment, only one coefficient is deemed statistically significant; plans with a 2.5-Star Rating and a decrease in Star Rating after QBP was implemented were significantly less likely to experience an increase in enrollment than they were prior to the start of the QBP demonstration.

Table 12. Coefficients on interacted lagged Star Ratings values and increased or decreased Ratings in the following year, linear probability of increased enrollment

		Lagged Star Rating						
		5 Star	4.5 Star	4 Star	3.5 Star	3 Star	2.5 Star	2 Star
Specification	Dependent Variable = Indicator for Increased Enrollment							
(1)	Rating Increased	--	0.102	0.008	0.029	-0.041	-0.062	0.044
			p=0.014	p=0.825	p=0.182	p=0.066	p=0.054	p=0.693
	Rating Decreased	-0.179	0.056	-0.031	-0.065	-0.047	0.007	--
		p=0.077	p=0.158	p=0.359	p=0.087	p=0.168	p=0.959	
(2)	Rating Increased	--	0.056	0.036	0.094	-0.012	-0.084	0.191
			p=0.223	p=0.361	p=0.005	p=0.684	p=0.022	p=0.138
	Rating Decreased	-0.085	0.074	0.006	-0.067	-0.014	0.162	--
		p=0.190	p=0.127	p=0.880	p=0.114	p=0.704	p=0.257	
	Rating Increased*QBP Demonstration Period		0.168	-0.058	-0.095	-0.062	0.052	-0.169
			p=0.077	p=0.261	p=0.009	p=0.062	p=0.281	p=0.383
	Rating Decreased*QBP Demonstration Period	-0.124	-0.021	-0.087	0.031	-0.123	-0.577	--
		p=0.391	p=0.765	p=0.101	p=0.699	p=0.116	p=0.001	

Notes: Control variables for all specifications were year FE, lagged Rating indicators, number of plans consolidated indicators, service are expansions and reductions indicators, and plan FE. Coefficients from linear regression of the natural logarithm of enrollment on whether the plan's contract experienced a Star Ratings increase or decrease from the previous year, conditional on the plan's previous contract Star Rating. Bonferroni adjustment for multiple comparisons critical p-values for $\alpha = 0.05$ to be p=0.004 for specification (1) and p=0.002 for specification (2).

Because employer-based plans are restricted from enrollment by the general Medicare population, enrollment changes for these plans, according to Star Ratings values or changes, may differ systematically from non-employer based plans. To exclude potential differential effects of employer-based plans on the analyses, Table 13 presents coefficients of the same specifications presented in Table 12, but excludes employer plans from the analyses. The coefficients are similar between the two tables. However, the exclusion of employer plans from the analyses produces more statistically significant differences between plans with changing Star Ratings and those that remain at the same Star Rating after accounting for multiple comparisons. Specifically, Table 14 shows that, as before, 2.5-Star rated plans experiencing a decrease in Rating were less likely to have increased enrollment after the implementation of the QBP demonstration than they were prior to the demonstration. Additionally, non-employer 4.5-Star rated plans that increased in Rating to 5-Stars after the implementation of QBP were more likely to experience increased enrollment than they were prior to the QBP demonstration. Interestingly, non-employer 3.5-Star rated plans with an increase in Rating were less likely to have increased enrollment after the implementation of QBP demonstration than they were prior to it.

Without statistically significant differences, the coefficients of specification (2) in Table 13 provide some insight into whether differences in enrollment changes are related to quality changes observed prior to and during the QBP demonstration. Due to the number of coefficients listed in this table, coefficient-specific interpretations are provided in Appendix 6.

Table 13. Coefficients on interacted lagged Star Ratings values and increased or decreased Ratings in the following year, linear probability of increased enrollment, excluding employer plans

		Lagged Star Rating						
		5 Star	**4.5 Star**	**4 Star**	**3.5 Star**	**3 Star**	**2.5 Star**	**2 Star**
Specification	Dependent Variable = Indicator for Increased Enrollment							
(1)	Rating Increased	--	0.143	-0.021	0.046	-0.026	-0.009	0.055
			p=0.006	p=0.573	0.060	0.280	0.779	0.612
	Rating Decreased	-0.140	0.070	-0.030	-0.055	-0.061	0.021	--
		p=0.215	p=0.142	p=0.428	0.192	0.090	0.873	
(2)	Rating Increased	--	0.045	0.031	0.139	0.004	-0.026	0.177
			p=0.441	p=0.435	0.000	0.886	0.485	0.170
	Rating Decreased	-0.039	0.106	0.030	-0.038	-0.023	0.180	--
		p=0.649	p=0.077	p=0.531	p=0.427	0.553	0.215	
	Rating Increased*Post		0.383	-0.108	-0.138	-0.069	0.028	-0.141
			p=0.000	p=0.055	p=0.001	0.055	0.585	0.486
	Rating Decreased*Post	-0.119	-0.045	-0.131	-0.037	-0.130	-0.604	--
		p=0.424	p=0.575	p=0.024	p=0.676	0.100	0.000	

Notes: Control variables for all specifications were year FE, lagged Rating indicators, number of plans consolidated indicators, service are expansions and reductions indicators, and plan FE. Coefficients from linear regression of the natural logarithm of enrollment on whether the plan's contract experienced a Star Ratings increase or decrease from the previous year, conditional on the plan's previous contract Star Rating. Employer plans were excluded from the specifications. Bonferroni adjustment for multiple comparisons critical p-values for $\alpha = 0.05$ to be $p=0.004$ for specification (1) and $p=0.002$ for specification (2).

Finding 6. On average, plans receiving bonus payments during the QBP demonstration period had below-expected out-of-pocket costs relative to a linear trend from 2011 to 2015

Plans with higher star ratings receive a higher percentage of their rebate than plans with lower star ratings to lower beneficiaries' premiums or enhance benefits. The results show that on average, plans receiving bonus payments during the QBP demonstration period had OOP costs below what would be expected from a linear trend between 2011 and 2015. Plans in below 3-Star contracts, however, did not consistently have below-trend differences in OOP costs. As such, under the assumption that OOP costs would have grown linearly from 2011 to 2015 in the absence of the QBP demonstration, QBPs may have reduced Part C premiums, and expected non-premium and total out-of-pocket costs, from what they would have been in the absence of the QBP demonstration.

Cost sharing and expected out-of-pocket (OOP) costs for MA plans are available on the Medicare Plan Finder website. Using the Plan Finder data obtained from CGI Federal (the contractor managing the website) for payment years 2011 through 2015, we examined the relationship between contract-level Star Ratings/bonus payments and plan-level premiums and cost sharing. OOP costs were examined at the plan-segment level. This is the level at which plan bids are made, benchmarks are defined, and, importantly, the resulting bonus amounts are defined. Furthermore, OOP cost estimates are not available for employer-sponsored plans so the analyses examine non-employer plans only.

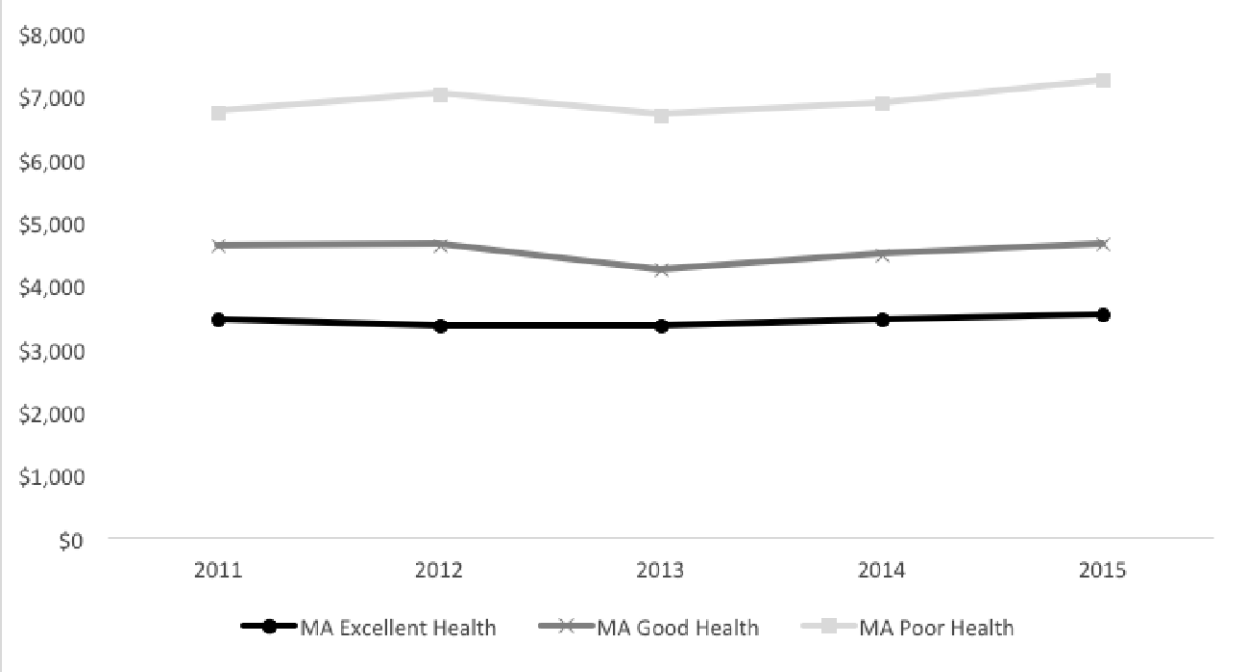
We merged the cost sharing and expected OOP from Medicare Plan Finder data with payment information received from CMS's Office of the Actuary, as well as plan consolidation and contract Star Ratings information from the HPMS in order to look at the relationships across OOP costs, Star Ratings, and payments. The Plan Finder cost data comprise information on non-employer based MA and PDP plans from calendar years 2011 to 2015 and include data on premiums required by plans; Part B, Part C, and Part D premiums, and expected annual costs of care (inpatient, outpatient prescription, dental, other services, and total annual out-of-pocket costs), conditional on three health states (excellent, good, and poor).

OOP cost estimates, though received from CGI Federal, were created by Fu Associates using methods developed by CMS and data from the Medicare Current Beneficiary Survey (MCBS). MCBS populations used in each year's calculation differed, depending on availability of MCBS data. Each OOP estimate was based on two years of MCBS beneficiaries. Commonly, but not always, these two cohorts overlapped one year in two consecutive years of OOP cost calculations.[9]

[9] Exceptions to the overlapping cohorts exist in the time period studied. For estimating CY 2012 costs, the 2006 and 2007 MCBS were used, while CY 2013 costs were estimated using the 2008 and 2009 MCBS. Also, for both CY 2014 and CY 2015 cost estimation, the 2009 and 2010 MCBS were used. Other variations in the OOP calculations also occurred during the 2011 to 2015 period. Starting in 2012, the estimates for Original Medicare (OM), or Medicare FFS, were estimated in parallel to the MA cost estimates. Prior to 2012, OOP estimates for Medicare FFS were calculated separately using slightly different assumptions. Also, minor differences in categorization of expenses—mapping utilization to benefit areas—exist from year-to-year. The year-to-year differences in the cost estimation do not include major methodological overhauls not mentioned and any differences are assumed to be small on total OOP costs. Documentation for each year's expected OOP costs are found at https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/OOPCResources.html

For non-employer plans eligible to receive a QBP bonus, Figure 14 displays the average expected total out-of-pocket costs by year, conditional on the health status of beneficiaries. Total expected out-of-pocket costs include premiums and cost sharing. Across all health status levels, the same pattern emerges over the time period; flat, or slightly declining average expected total out-of-pocket costs persist from 2011 to 2013, with slightly increasing costs from 2013 to 2015.[10] For beneficiaries in "good" or "poor" health, the expected out-of-pocket costs were lower in 2013 than in 2012 or 2014; for beneficiaries in "excellent" health, the average expected out-of-pocket costs in 2013 was statistically indistinguishable from the average in 2012.

Figure 14. Mean total expected MA plan out-of-pocket costs by health status, 2011 to 2015



Source: Expected out-of-pocket costs were taken from CGI Federal data.

Note: Figure displays the mean expected out-of-pocket costs for non-employer MA plans eligible to receive a QBP bonus during the demonstration period.

Table 14 displays mean premiums for plans in each year, from 2011 to 2015. While cost sharing makes up the vast majority of total expected out-of-pocket costs, premiums may be internalized by beneficiaries as having greater importance—they are the first costs beneficiaries face, and are more readily understood as fixed costs that do not vary according to health. The means show that Part C premiums dropped slightly over this period, primarily due to an increase in the proportion of plans with $0 premiums. Also, the proportion of plans offering Part D coverage increased during the period, from 73 percent in 2011 and 2012 to 77 percent in 2015. The mean premium charged for the Part D coverage, however, was also increasing slightly over the time period. Finally, the proportion of plans including a reduction in the Medicare Part B premium was low throughout the

[10] For additional information, including standard deviations and indicators for pair-wise statistical differences, see Appendix 7. Mean Out-of-Pocket Expenditures.

period, peaking in 2012 at 6 percent. When offered, the average reduction during the 2011 to 2015 period ranged from $46 to $59 from the Part B premium.

Table 14. Mean MA plan premiums, 2011 to 2015

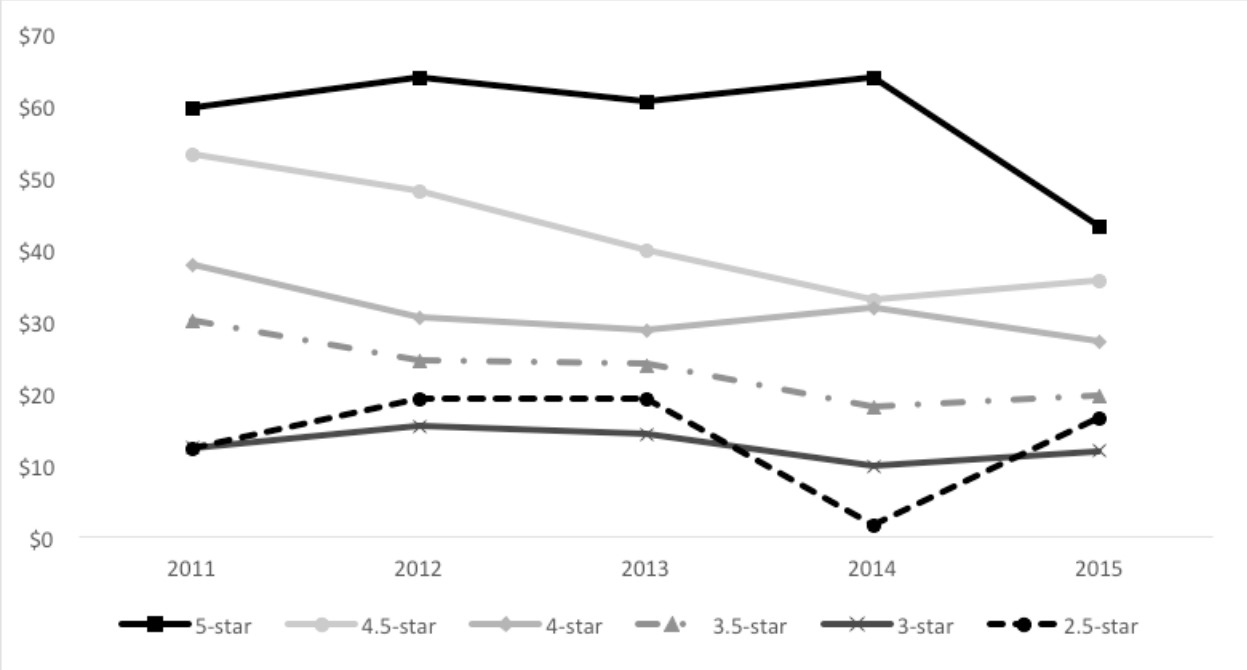
	2011	2012	2013	2014	2015
	(N=1,825)	(N=2,180)	(N=2,297)	(N=2,192)	(N=2,212)
Mean Part C premium	$26	$25	$24	$23	$24
(Std Dev)	*($42)*	*($41)*	*($42)*	*($40)*	*($42)*
Proportion with $0 Part C premium	53%	55%	57%	57%	57%
Mean Part C premium if greater than $0	$56	$56	$57	$53	$56
(Std Dev)	*($47)*	*($45)*	*($48)*	*($46)*	*($49)*
Proportion with Part D plan	73%	73%	74%	75%	77%
Mean Part D premium if applicable	$18	$18	$18	$21	$21
(Std Dev)	*($19)*	*($20)*	*($21)*	*($24)*	*($25)*
Mean Part B premium reduction if applicable	$52	$53	$59	$53	$46
(Std Dev)	*($31)*	*($35)*	*($34)*	*($31)*	*($33)*
Proportion with Part B premium reduction	4%	6%	5%	4%	4%

Source: Plan premium cost information was taken from CGI Federal annual estimates of out-of-pocket costs.

Note: Means were calculated over non-employer plans eligible to receive a bonus under the QBP demonstration if Star Ratings requirements were met.

Premiums charged by plans varied with the contract Star Ratings. Figure 15 shows the mean Part C premiums for plans with each overall Star Rating in 2011 through 2015. The average premiums are higher for plans with higher current overall Star Ratings. This effect is due to differences in the proportion of plans with $0 premiums (higher Star Ratings have a lower proportion of $0 premium plans) *and* premiums for those greater than $0 (higher-rated plans have higher premiums if premiums are non-zero).

Figure 15. MA plans' mean Part C premiums by overall Star Rating, 2011 to 2015

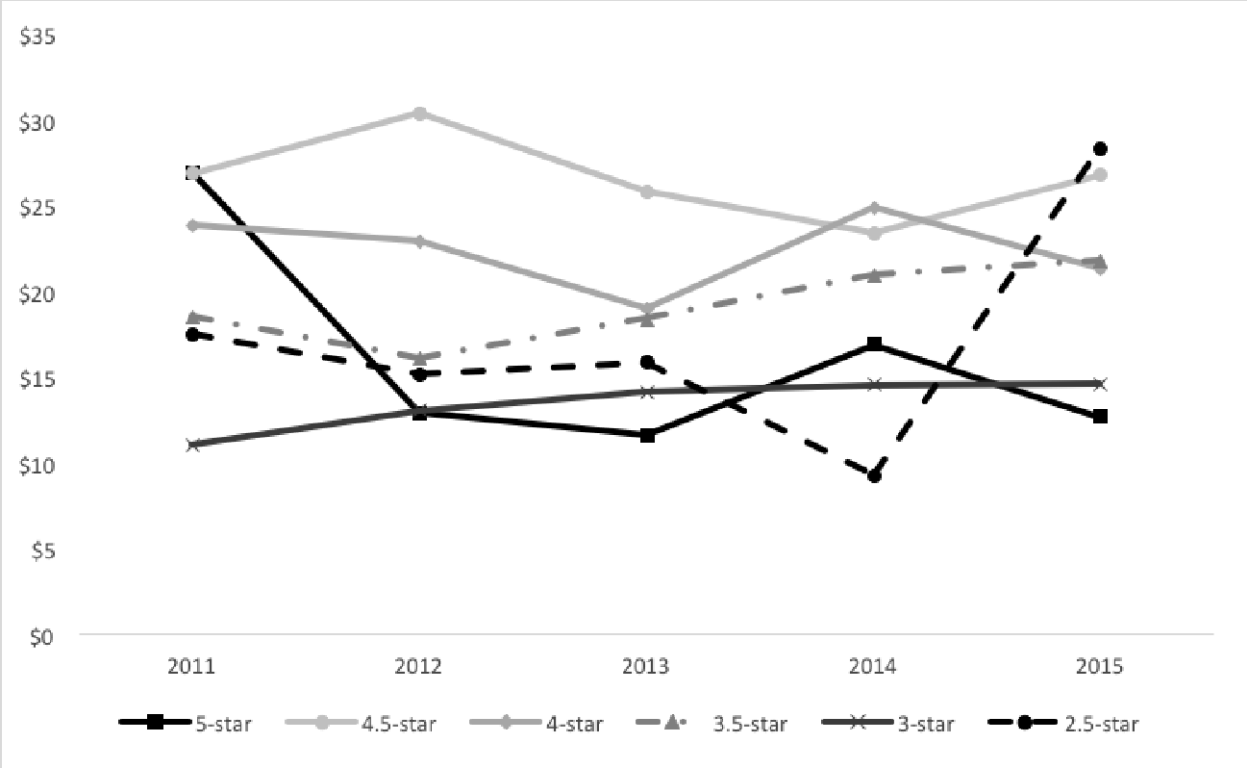


Source: Premium data were taken from CGI Federal data.

Note: Figure displays the mean Part C premiums for non-employer MA plans eligible to receive a QBP bonus during the demonstration period conditional on the current contract Star Rating.

Figure 16 displays the mean Part D premiums for plans that offer Part D coverage, by overall Star Rating. As with Part C premiums, there is a tendency for higher-rated plans to have higher average premiums. Plans under 5-Star rated contracts, however, do not follow this tendency and include Part D premium averages lower than 3.5- to 4.5-Star Ratings after 2011. There are not clear trends in the proportion of MA plans including Part D coverage by Star Rating (proportions not shown). However, plans with overall Ratings of 3.5-Stars or higher are slightly more likely to include Part D coverage than plans with Ratings of 3 or fewer Stars. The difference becomes greater in later years.

Figure 16. MA plans' mean Part D premiums by overall Star Rating, 2011 to 2015

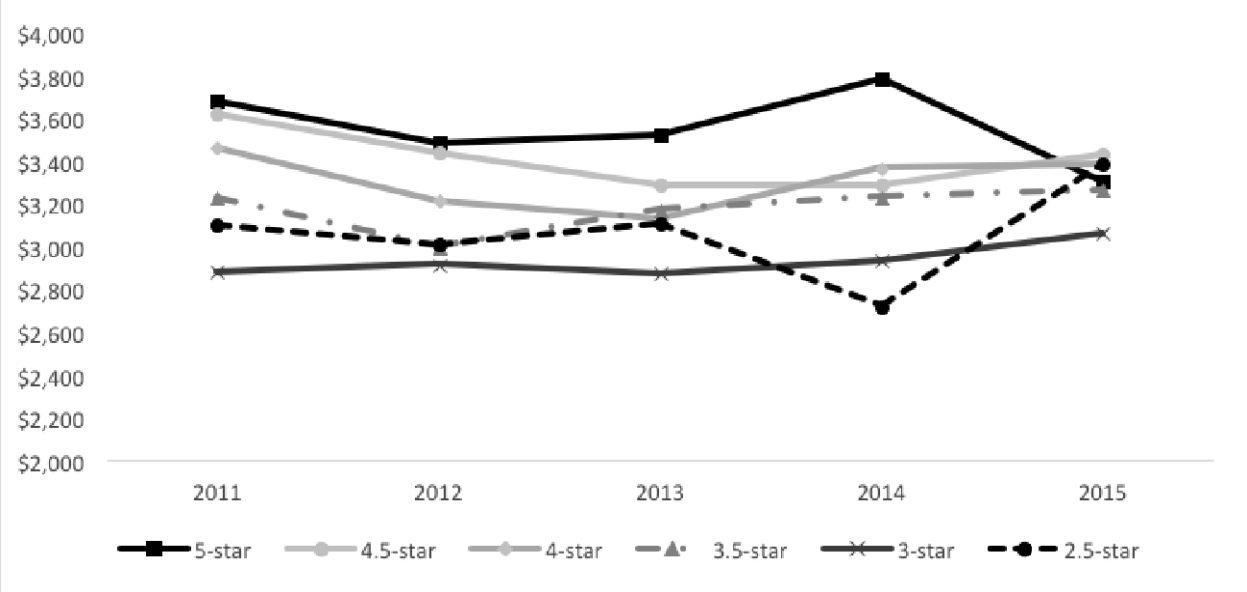


Source: Premium data were taken from CGI Federal data.

Note: Figure displays the mean Part D premiums for non-employer MA plans eligible to receive a QBP bonus during the demonstration period and included Part D benefits conditional on the contract Rating.

Figure 17, Figure 18, and Figure 19 display the mean, non-premium, expected OOP costs for plans by contract Star Rating and year for beneficiaries in "excellent," "good," and "poor" health. As with premium costs, there is a tendency for higher-rated plans to have higher mean OOP costs. The trend is clearest for beneficiaries in "excellent" health, becomes less well defined for beneficiaries in "good" health, and does not hold for beneficiaries in "poor" health.

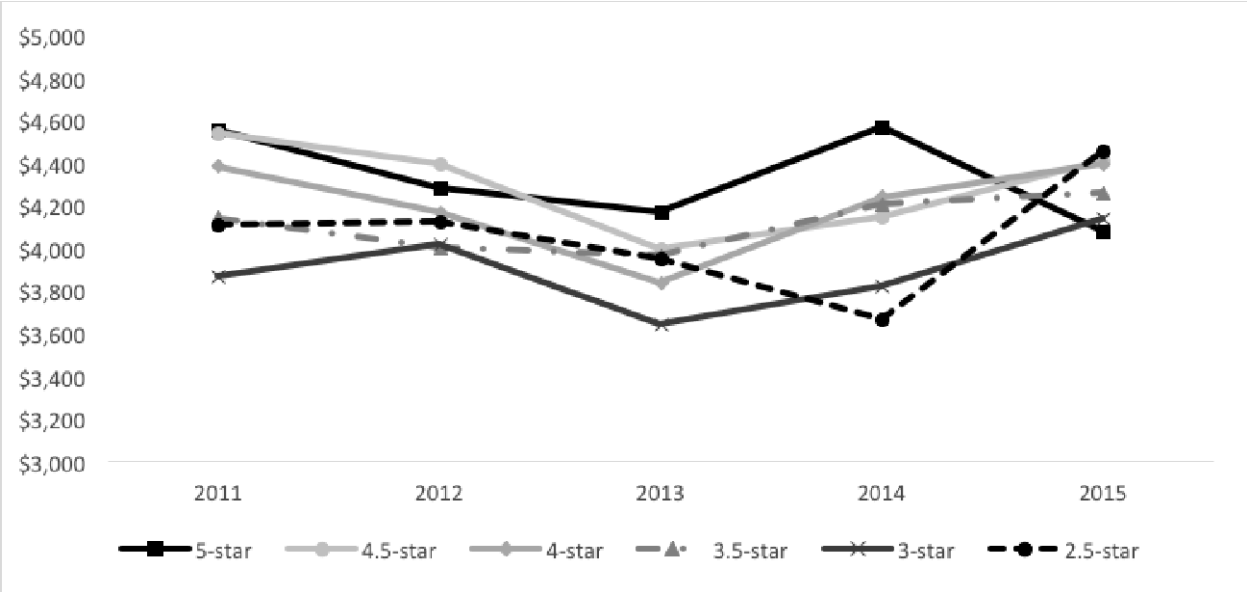
Figure 17. Mean non-premium out-of-pocket costs for beneficiaries in "excellent" health by overall Star Rating, 2011 to 2015



Source: Expected costs data were taken from CGI Federal data.

Note: Figure displays the mean non-premium related out-of-pocket costs for beneficiaries of non-employer MA plans in "excellent" health.

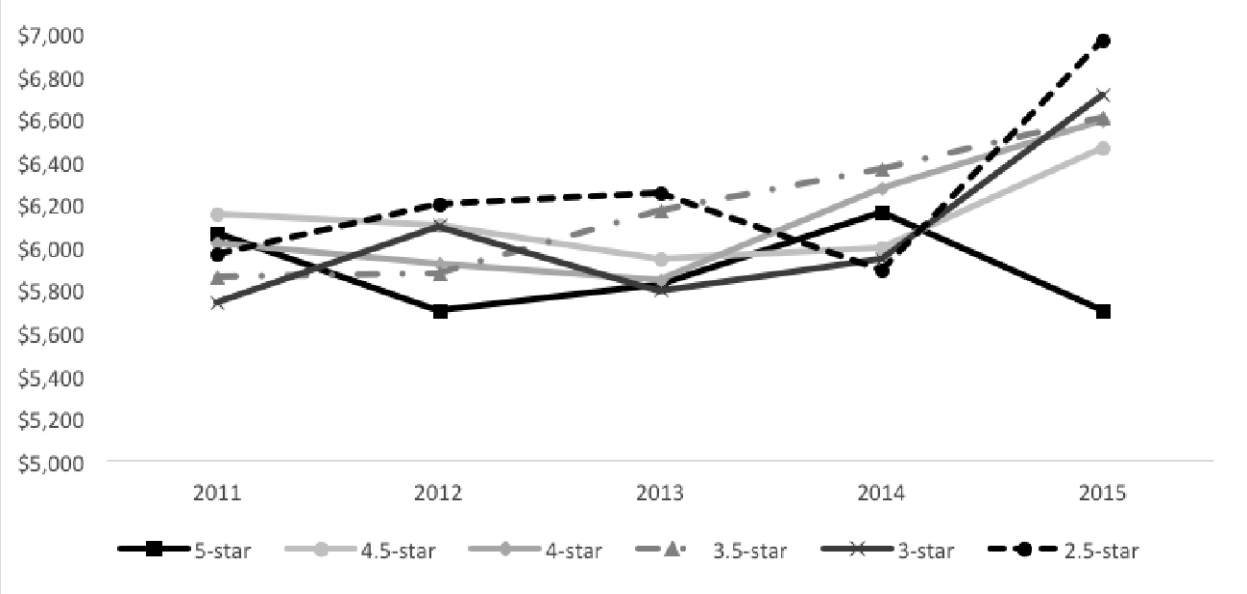
Figure 18. Mean non-premium out-of-pocket costs for beneficiaries in "good" health by overall Star Rating, 2011 to 2015



Source: Expected costs data were taken from CGI Federal data.

Note: Figure displays the mean non-premium related out-of-pocket costs for beneficiaries of non-employer MA plans in "good" health.

Figure 19. Mean non-premium out-of-pocket costs for beneficiaries in "poor" health by overall Star Rating, 2011 to 2015



Source: Expected costs data were taken from CGI Federal data.

Note: Figure displays the mean non-premium related out-of-pocket costs for beneficiaries of non-employer MA plans in "poor" health.

QBPs possibly reduced Part C premiums and expected out-of-pocket costs, though difficult to confirm

While the descriptive correlation between contract-level Star Ratings and OOP costs are noted above, the question remains whether the bonus payments translated into changes in OOP costs. We are not able to ascertain a causal estimate of the effect, though two components of the system help to illuminate differences occurring during the QBP years. First, the QBP demonstration was limited to the 2012-2014 period. Second, not all contracts received bonus payments during the period and higher rated contracts received higher bonuses, all else equal.

If QBP demonstration bonus payments were related to premiums and cost sharing, we would expect to see an effect during the demonstration payment years, 2012 through 2014, relative to non-demonstration years. Time trends common to all health plans, MA and non-MA, may have occurred during the QBP period driving changes in premiums or other OOP costs. If this occurred, then we may inaccurately ascribe these changes to the QBP program. If an appropriate comparison existed, then we may measure changes common to both MA plans and comparisons net of the commonly experienced changes. However, appropriate comparisons to MA premiums and costs are not available. In the absence of appropriate comparisons we descriptively examine differences in OOP costs during the QBP period, 2012 to 2014, relative to costs at two years outside of the demonstration, 2011 and 2015. Furthermore, we examine how any observed changes vary with bonus amounts, which differed across MA plans, during the time period.

First, to describe mean costs during this time, we estimated a linear regression using data from 2011 to 2015. The outcomes of interest (premium amounts, non-premium OOP costs, and total OOP costs) were regressed on a linear time trend, indicator variables for each of the QBP demonstration years (2012, 2013, 2014), and plan-segment fixed effects. The coefficients of interest are the indicators for years 2012 to 2014, which show how the mean outcomes varied from a linear trend from 2011 and 2015. Negative (positive) coefficients indicate that the mean outcomes were below (above) the linear trend established between 2011 and 2015 values and that costs were less than (greater than) expected during 2012 through 2014.

Table 15 displays the fitted regression coefficients from the linear time trend models for Part C and Part D premiums, as well as non-premium and total annual expected costs, conditional on health status. The results indicate that, indeed, the mean costs typically fell below the linear trend established between 2011 and 2015 in each of the QBP demonstration years. For instance, the linear trend between Part C premiums in 2011 and 2015 suggests that Part C premiums would have increased by $1.13 per year. However, Part C premiums were, on average, $1.24, $0.94, and $1.12 below this expectation in 2012, 2013, and 2014, respectively.

Across premium and cost sharing outcomes, the mean costs fell significantly below the linear trend in all but two occasions, Part D premiums in 2014 and total OOP costs in 2012 for beneficiaries in "poor" health.

As a reference for the practicality of linear growth in OOP costs over the time period in question, we present expected OOP costs for Medicare FFS beneficiaries. Medicare FFS beneficiaries did not directly experience changes in OOP costs due to the QBP demonstration; they are, however, similar in age and have expected OOP costs, conditional on health, in the CGI Federal data. As shown in Table 15, expected OOP costs in 2012 to 2014 fell both above and below a linear trend between the expected OOP costs of 2011 and 2015. Medicare FFS beneficiaries in all health statuses faced higher expected OOP costs than beneficiaries of the same health status in of MA plans. However, FFS beneficiaries in "Excellent" and "Good" health faced lower growth trends between 2011 and 2015 than did MA plan beneficiaries, on average.

Table 15. Difference in MA plans' mean premium, non-premium OOP, and total OOP costs in QBP demo years from linear trend

Outcome	Part C Premium	Part D Premium[1]	Non-premium OOP Costs			Total OOP Costs		
Health	All	All	"Excellent"	"Good"	"Poor"	"Excellent"	"Good"	"Poor"
2012	-1.24***	-1.02***	-145.26***	-71.97***	-34.83***	-131.62***	-19.70**	86.27***
2013	-0.94***	-1.38***	-130.11***	-343.14***	-217.63***	-136.45***	-386.77***	-287.88***
2014	-1.12***	0.57*	-25.34***	-99.61***	-130.19***	-44.72***	-131.44***	-180.97***
Linear trend	1.13***	2.23***	58.72***	59.79***	151.86***	59.77***	60.23***	208.08***
Constant	22.72***	14.95***	3121.29***	4098.95***	5923.48***	3373.62***	4506.97***	6575.60***

*Notes: Table displays regression coefficients from a regression of cost outcomes on a linear trend over the 2011-2015 period. The sample includes non-employer MA plans eligible to receive a QBP bonus during the demonstration period. Coefficients on individual years, 2012-2014, show the difference in means for these years from the linear trend. [1]When fitting the regression on Part D premium amounts, the sample was restricted to plans with Part D coverage. ***Indicates statistical significance at the 1% level. **Indicates statistical significance at the 5% level. *Indicates statistical significance at the 10% level.*

Table 16. FFS expected OOP costs relative to a linear trend

Outcome	FFS Total OOP Costs		
Health	"Excellent"	"Good"	"Poor"
2012	-5.10	273.60	731.70
2013	-119.40	-369.60	79.80
2014	-137.70	-244.80	-356.10
Linear trend	6.30	19.20	591.90
Constant	4209.60	6522.00	10795.20

Notes: Table displays the expected total OOP costs for Medicare FFS beneficiaries in "Excellent," "Good," and "Poor" health from 2011 to 2015. The expected 2011 OOP costs are represented as "Constant." The expected costs in 2012, 2013, and 2014 are calculated as the constant plus one, two, or three times the linear trend value, respectively, plus the value displayed in the year's respective row. The expected costs in 2015 are calculated as the constant plus four times the linear trend value.

The second characteristic of the QBP program leveraged for analysis is that higher rated plans receive higher bonuses, all else equal. That is, if two plans are operating in the same area with equivalent beneficiaries and submit the same bid then the higher rated plan receives a higher payment under the QBP program. However, if the service area *or* beneficiary populations differ then the payments received by a lower rated plan may be higher than those of a higher rated plan for reasons unrelated to the QBP program. As such, we rely on the Star Ratings used to generate the bonus payment amount to identify the variation of interest in our analysis. Specifically, we examined differences from the linear growth trends of out-of-pocket costs by lagged Star Ratings, the Ratings on which bonuses were paid.

By examining difference from trends by plans of similar Star Ratings, we are able to see how expected OOP costs differ from trend during the QBP period for similar plans; comparing across Star Ratings rather than within Star Ratings will increase the likelihood of comparing plans that differ in meaningful, but unobserved, ways—high rated plans differ from low rated plans in unobserved ways—and, without a clear identification strategy we may inaccurately attribute observed differences to differences in the payment rather than beneficiary populations or another unobserved attribute.

Table 17 through Table 20 display results of regressions, as above, where the dependent variable is a cost component (premium amount, non-premium out-of-pocket costs, or total annual out-of-pocket costs) and the independent variables include a constant, a linear time trend, indicator variables for years during the QBP demonstration, and plan fixed effects. The fitted coefficients on the indicator variables are of interest and represent the difference in mean costs from the linear trend, set by 2011 and 2015, for the QBP year. Regressions were run separately for plans with lagged Star Ratings in three groups: 4 to 5 Stars, 3 to 3.5 Stars, and fewer than 3 Stars.[11]

If bonus payments alter the premiums and cost sharing constructs of plans and assuming that costs would have changed linearly in the absence of the QBP program, then we should see greater differences between the linear trends for plans under higher-rated contracts than for plans under lower-rated contracts. Since the regressions were run separately, the constant (2011 mean of the outcome) and the linear trend to the 2015 mean are allowed to differ across the lagged Star Ratings groups and health status. As such, the differences from trends are specific to the Star Ratings groups and beneficiary health status. These may differ for an important reason to the analysis, plans under contracts with Star Ratings of 4 Stars or higher received bonuses in 2015 under the ACA bonus arrangement, once the QBP demo had concluded and payment rules reverted to what was originally legislated under the ACA. If the bonuses in 2015 altered the mean 2015 outcomes for these plans, then the trend line will differ from that associated with plans under lower-rated contracts for this reason. Additionally, the linear trends are fitted on only two years, 2011 and 2015, and results would be sensitive to unrelated variance in either year. Finally, again, differences in trends assume that the outcomes would be changing linearly during the time period, in the absence of the QBP demonstration.

[11] Star Ratings were actual overall Ratings. Plans receiving an overall Star Rating value because it was too new to be rated or included too few beneficiaries were not included in the analyses. As noted earlier in the report and discussed in the methods section, an overall Star Rating for 2010 was calculated by L&M Policy Research for the purpose of this project using the same rules applied to create the 2011 overall Star Ratings.

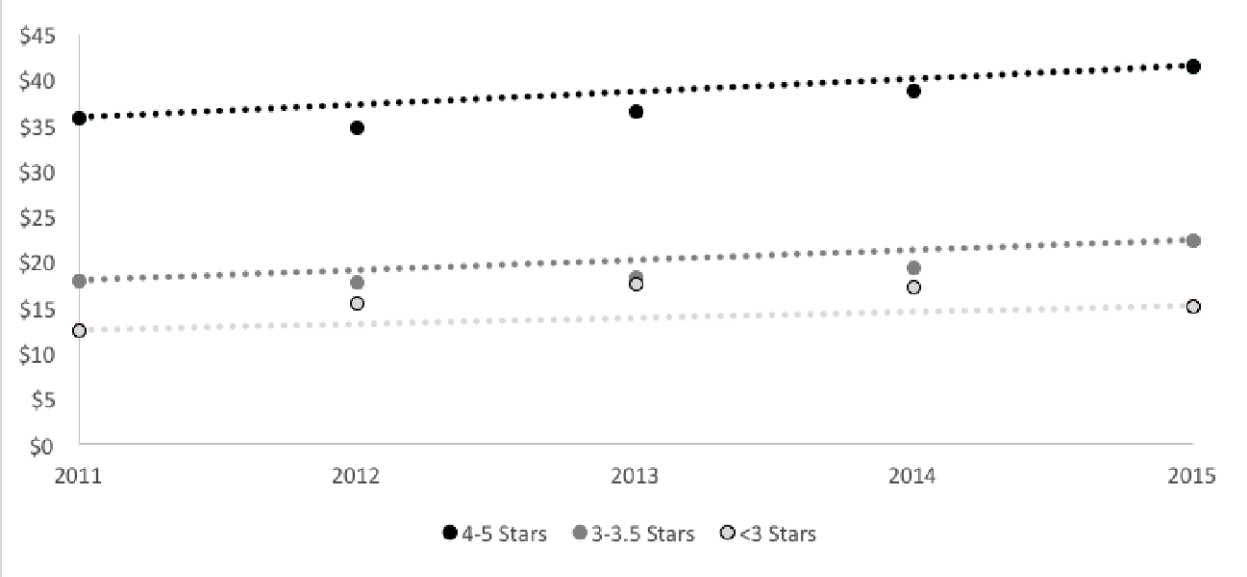
Table 17 displays the fitted regression coefficients for models where the Part C premium is the dependent variable. The means across lagged Star Ratings groups indicate that the mean Part C premiums were lower than the linear trend for plans under contracts with lagged overall Ratings of 3- to 3.5-Stars or 4- to 5-Stars. Conversely, mean Part C premiums for plans with lagged overall Ratings of fewer than 3-Stars were above the linear trend in 2012 and 2013. These differences suggest that plans receiving a QBP bonus payment offered lower-than-expected premiums during the QBP demonstration and plans that did not receive such a bonus did not. Figure 20 displays these results.

Table 17. Difference in MA plans' mean Part C premiums in QBP demo years from linear trends by overall Star Rating

	4-5 Stars	3-3.5 Stars	<3 Stars
2012	-2.45***	-1.27***	2.42**
2013	-2.06***	-1.91***	3.75**
2014	-1.27**	-1.89***	2.82
Linear trend	1.45	1.12	0.7
Constant	35.66	17.9	12.33

*Notes: Table displays regression coefficients from a linear regression of Part C premiums on a linear trend over the 2011-2015 period. The sample includes non-employer MA plans eligible to receive a QBP bonus during the demonstration period. Coefficients on individual years, 2012-2015, show the difference in means for these years from the linear trend. ***Indicates statistical significance at the 1% level. **Indicates statistical significance at the 5% level. The constants and linear trends were all statistically different from zero at the 1% level of significance; asterisks are suppressed in the table.*

Figure 20. Expected Part C premium, fixed effect regression results with linear trend comparison



Source: Expected costs data were taken from Medicare Plan Finder website (through its contractor CGI Federal).
Note: Figure displays the expected Part C premiums in each year from regressions of Part C premiums on a linear time trend, indicator variables for 2012, 2013, and 2014, and plan-segment fixed effects.

Table 19 displays fitted regression coefficients for models where the Part D premium was the dependent variable. The sample was restricted to plans with Part D coverage. The means for plans with a lagged Rating of 3- to 3.5-Stars were below the linear trend during the QBP demonstration years. Plans with lagged Star Ratings higher, 4- to 5-Stars, or lower, less than 3-Stars, did not differ significantly from trend in 2012 or 2013 and were significantly higher than the linear trend in 2014.

Table 18. Difference in MA plans' mean Part D premiums in QBP demo years from linear trends by overall Star Rating

	4-5 Stars	3-3.5 Stars	<3 Stars
2012	-0.47	-2.26***	-0.52
2013	0.34	-2.83***	0.54
2014	1.79***	-0.75*	5.82**
Linear trend	2.98	1.9	-0.19
Constant	16	14.33	14.21

*Notes: Table displays regression coefficients from a linear regression of Part D premiums on a linear trend over the 2011-2015 period. The sample includes non-employer MA plans offering Part D coverage and eligible to receive a QBP bonus during the demonstration period. Coefficients on individual years, 2012-2015, show the difference in means for these years from the linear trend. ***Indicates statistical significance at the 1% level. **Indicates statistical significance at the 5% level. *Indicates statistical significance at the 10% level. The constants and linear trends were all statistically different from zero at the 1% level of significance; asterisks are suppressed in the table.*

The regressions showing differences in mean costs from trend for non-premium out-of-pocket costs, Table 20, and total expected out-of-pocket costs, Table 21, suggest similar findings. For beneficiaries in "Excellent" health, plans with lagged Star Ratings lower than 3-Stars had mean costs below trend in 2011 and 2012. Costs for beneficiaries in "Excellent" health of 4- to 5-Star plans or 3- to 3.5-Star plans were below trend during the entire QBP demonstration period and to a greater extent than plans with fewer than 3 Stars. These findings suggest that plans receiving bonus payments during the QBP period may have structured benefits for beneficiaries in "Excellent" health to have lower-than-trend costs during the demonstration, with reductions exceeding non-bonus plans.

Table 19. Differences in MA plans' mean expected non-premium out-of-pocket costs in QBP demonstration years from linear trends by overall Star Rating, conditional on beneficiary health

	4-5 Stars	3-3.5 Stars	<3 Stars	4-5 Stars	3-3.5 Stars	<3 Stars	4-5 Stars	3-3.5 Stars	<3 Stars
	"Excellent"			"Good"			"Poor"		
2012	-164.42***	-165.81***	-73.41***	-127.81***	-93.45***	25.04	-110.49***	-75.44***	18.72
2013	-140.93***	-179.47***	-59.45	-360.36***	-406.08***	-303.22***	-232.10***	-307.55***	-399.15***
2014	-13.6	-75.17***	62.52	-89.78***	-158.33***	-26.63	-125.41***	-216.93***	-385.84***
Linear trend	57.22	63.49	51.96	46.23	69.72	62.61	93.34	184.93	290.5
Constant	3305.66	3034.09	2987	4269.88	4016.15	3997.84	6041.91	5856.36	5878.43

*Notes: Table displays regression coefficients from a linear regression of expected non-premium cost sharing on a linear trend over the 2011-2015 period. The sample includes non-employer MA plans eligible to receive a QBP bonus during the demonstration period. Coefficients on individual years, 2012-2015, show the difference in means for these years from the linear trend. ***Indicates statistical significance at the 1% level. **Indicates statistical significance at the 5% level. The constants and linear trends were all statistically different from zero at the 1% level of significance; asterisks are suppressed in the table.*

Table 20. Differences in MA plans' mean expected total out-of-pocket costs in QBP demonstration years from linear trends by overall Star Rating, conditional on beneficiary health

	4-5 Stars	3-3.5 Stars	<3 Stars	4-5 Stars	3-3.5 Stars	<3 Stars	4-5 Stars	3-3.5 Stars	<3 Stars
	"Excellent"			"Good"			"Poor"		
2012	-131.09***	-160.95***	-67.32***	-26.67	-60.49***	63.27*	117.65***	4.8	140.23**
2013	-141.50***	-185.15***	-75.71**	-413.64***	-441.42***	-352.69***	-308.88***	-376.33***	-426.84***
2014	-40.17***	-87.64***	41.9	-133.65***	-181.05***	-64.84	-180.13***	-270.44***	-388.42***
Linear trend	62.5	64.40073	58.243	51.78828	70.6579	70.87491	188.2919	235.983	332.8245
Constant	3581.85	3263.315	3217.635	4734.223	4382.462	4387.833	6736.026	6447.873	6535.779

*Notes: Table displays regression coefficients from a linear regression of expected total beneficiary out-of-pocket costs on a linear trend over the 2011-2015 period. The sample includes non-employer MA plans eligible to receive a QBP bonus during the demonstration period. Coefficients on individual years, 2012-2015, show the difference in means for these years from the linear trend. ***Indicates statistical significance at the 1% level. **Indicates statistical significance at the 5% level. The constants and linear trends were all statistically different from zero at the 1% level of significance; asterisks are suppressed in the table.*

The regression coefficients are similar for expected cost sharing among beneficiaries in "good" or "poor" health. Cost sharing and total costs for plans under contracts with lagged Ratings of 4- to 5-Stars or 3- to 3.5-Stars are below the linear trends, with the exception of 2012, when mean costs were not significantly below trend for beneficiaries in "good" health (4- to 5-Star Ratings) and "poor" health (3- to 3.5-Star Ratings). Interestingly, expected costs for "good" or "poor" health beneficiaries in plans that did not receive bonus payments (those with overall Ratings of less than 3-Stars) are also below trend in 2013 and 2014 by magnitudes similar to plans that received bonuses.

The regressions revealing differences in mean premium and cost-sharing values during the QBP demonstration from a linear trend show that plans with Star Rating values of 3-Stars or higher were more likely than plans which were ineligible for QBP demonstration bonus to have mean costs below trend during the QBP demonstration period. As such, under the assumption that out-of-pocket costs would have grown linearly from 2011 to 2015 in the absence of the demonstration, bonus payments may have reduced Part C premiums and expected out-of-pocket costs from what they would have been in the absence of the QBP demonstration. While non-premium OOP costs are estimated using data from the MCBS, they display similar changes in growth during the QBP period for plans with higher, rather than lower, Star Ratings and thus higher bonus payments as do plan premiums, which are known. However, we are not able to identify causal effects of the bonus payments during the demonstration period on either premiums or non-premium OOP costs.

CONCLUSION

As the Affordable Care Act was being drafted, there were spirited discussions across key stakeholders about the impact of planned MA payment reductions on beneficiaries, notably, that MA contracts would leave the market as they did in the wake of BBA '97 payment cuts, and MA enrollments would decline. In fact, across the QBP demonstration period, average Star Ratings improved, more beneficiaries enrolled in higher-rated plans, and more beneficiaries had access to higher-rated plans. While there is no definitive way to attribute these changes (in whole or in part) to the QBP demonstration itself, evaluation analyses do show that the demonstration at least did not diffuse or reverse trends: Star Rating and plan enrollment increases that began prior to the demonstration continued throughout the demonstration period.

Evidence from measure-level analyses, which included comparisons to population groups not subject to the QBP demonstration (commercial, Medicaid, Medicare FFS, and stand-alone PDPs), suggests that patterns observed in MA during the QBP demonstration period were generally similar to those in other payer sectors. On most measures examined, MA contracts saw an increase in average scores between the pre-demonstration period and the end of the demonstration. While no notable reductions in quality measure scores or enrollment are observed during this period, relative to general trends, it remains unclear whether the demonstration may have dampened or propelled the magnitude of these trends. It is possible that as value-based payments become more common among many payers, industry-wide emphasis on quality spills over and affects all patients, regardless of payer.

Data from surveys and interviews with MA organizations, point to the impact of the QBP demonstration on operations and how it may have amplified incentives of the MA payment changes in the ACA. Our survey of MA organizations found that the budget for quality improvement activities increased between 2010 (when the demonstration was announced) and 2013 (when the survey was conducted) for 88 percent of survey respondents. In interviews with MA organizations, several noted that the QBP demonstration was beneficial in strengthening the business case for organizations to focus on and invest in quality improvement initiatives, given payments were now tied to Star Ratings. As we noted in our case study report with six MA organizations, while most of them reported that their engagement in QI and focus on Star Ratings preceded the announcement of the MA QBP Demonstration, several indicated that the demonstration provided important incentives for improving Star Ratings. Nevertheless, case study interviews also revealed that organizations could not always tease out the effect of the demonstration from the effect the of the QBP payments generally.

METHODS

The evaluation includes multiple data sources and analytic activities, which are described in more detail below.

Data sources

Medicare Advantage Contracts

To establish the universe of contracts of interest we identified and retained all plans that had information on Part C premiums in the Health Plan Management System (HPMS). Over time the number of these contracts fluctuates from a high of 634 contracts in 2009 to a low of 525 in 2015, as shown in Table 22 In addition to most contract characteristics used in the analysis, all plan characteristics are from HPMS.

Table 21. Number of distinct contracts in each Star Ratings year, 2009 to 2015

Star Rating year	Contracts of interest	Contracts with an overall Rating	Contracts that were too new to be measured	Contracts with not enough data	Missing information
2009	634	339	N/A	N/A	295
2010	583	400	N/A	N/A	183
2011	531	385	90	55	1
2012	546	431	75	39	1
2013	553	440	62	51	0
2014	566	424	57	85	0
2015	525	389	59	77	0

Note: L&M calculated the overall Ratings in 2009 and 2010 using from measure scores according to 2011 technical specifications and did not specify which contracts were 'Too new to be measured' or 'Did not have enough data available.' Contract E5088 was present in the contracts of interest for 2011 and 2012 but did not have an Overall Rating.

Star Ratings

The overall Star Ratings derive from publicly available data sets retrieved from CMS website.[12] For each year the master excel datasets cleaned and merged to the contract list of interest identified from HPMS. Data for certain contracts were not available for the Star Rating analysis for a given year if they did not meet the CMS thresholds for minimum enrollment or were too new to be rated. In these cases, their Star Ratings data could be included for other years once those criteria were met and a Rating calculated.

[12] As of May 2015 the CMS website is: http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html

In order to further document pre-demonstration trend the team estimated synthetic overall Ratings for 2009 and 2010. The methodology used followed those specified on pages two and three of the 2011 Part C Technical Notes document, with one distinction.[13] We tested out approach on the 2011 data and correctly estimated 95 percent of overall Ratings.[14] For these two years the team did not designate which contracts were "Too New to be Measured" or where there was "Not Enough Data Available" and instead chose to retain those contracts separately as missing, as shown in the last column of Table 21.

After calculating these synthetic Ratings the team appended them to the overall Star Ratings file, merging this dataset to the MA contracts of interest. Nearly all of the MA contracts of interest were present in the Overall Star Ratings dataset; the only exception being Employer/Union Only Direct Contract PFFS contract Desert Healthcare Employee Benefits Trust (E5088), which is present in the dataset for 2011 and 2012. For those two years this contract was assigned a missing Overall Star Rating.

Wherever possible the team used the spring overall Ratings as these were updated to reconcile any ongoing appeals contracts may have made pertaining to their measures or Ratings. If the spring overall Ratings were not available the team defaulted to the Fall Ratings. Table 23 details the Star Ratings used in each year.

Table 22. Available Ratings data release dates for Star Rating years 2009 to 2015

Star Rating Year	Available Rating		
	Part C	Part D	Overall
2009*	Fall	Spring	Estimated
2010*	Fall	Fall	Estimated
2011	Spring	Spring	Spring
2012	Spring	Spring	Spring
2013	Spring	Spring	Spring
2014	Spring	Spring	Spring
2015	Fall	Fall	Fall

Enrollment

Monthly enrollment files for each year, 2009 through 2014 were pulled from HPMS and used to generate an average monthly enrollment estimate at the plan-county, county, and contract level—

[13] The only deviation from the 2011 technical notes was adjusting the number of measures required for an overall Rating for 2009 and 2010. Rather than using the same number of measures specified in the 2011 technical specifications for 2009 and 2010, we changed the minimum number to half the total count of measures that the organization type is required to report in each of those years. This approach was taken because the total number of measures is different in 2009, 2010, and 2011 and so the minimum number of measures needed to calculate an overall Rating should be adjusted accordingly.

[14] More information on the team's calculation of synthetic Ratings can found in the memo to 'Calculating synthetic Overall Ratings for Star Ratings Year 2009, 2010' to CMS dated March 13, 2015.

for non-out-of-area enrollment only. Enrollment statistics for 2015 were taken from the March 2015 enrollment file. These files are used for enrollment weighting and for the enrollment analysis. The team either used the contract enrollment estimate from the same calendar year or lagged the enrollment by a constant, depending on the analysis.

MA Measure-Level Data

The Medicare Advantage measure-level data derive from the same publicly available data sets as the overall Ratings.[15] To conduct the measure level analyses the team weighted these measures using enrollment in the year the data were collected. (e.g., 2012 Star Ratings year HEDIS measures weighted by the enrollment in the year data were collected (Star Ratings year minus two)).

Comparison groups (FFS CAHPS, NCQA, PDP) measure level data

The team used a series of comparison groups to benchmark the MA measures. These comparison groups include data from the National Committee for Quality Assurance (NCQA), the Fee-for-Service Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys, and Medicare standalone Prescription Drug Plan (PDP) contracts. Specifically, these data and their enrollment weights were arrived at using the following procedures:

- Medicaid and private insurance data for select measures were purchased from NCQA for HEDIS years 2008 – 2013. These data retained state level enrollment by organization id collected at the same time as the measures themselves
- Fee-for-Service CAHPS survey data were obtained from CMS for calendar years 2008 – 2013. The team used publicly available county level MA penetration estimates to calculate an enrollment weight[16]
- MA standalone PDP data for select measures were retrieved from the same publicly available datasets as the Overall Ratings and MA measures, for Star Ratings years 2009 – 2015. PDP enrollment was retrieved from HPMS

QI Activities

As the data analyses rely primarily on administrative data, the evaluation team conducted a survey of MA contracts to capture high-level information on how MAOs perceive the demonstration and planned for or implemented changes in quality initiatives. Based on the results of the key informant interviews, the team developed a draft questionnaire to capture information on (1) how Medicare Advantage organizations (MAOs) perceive the demonstration and its effect on quality, (2) how they are planning for or implementing changes in quality initiatives, (3) how they identify factors that help or hinder the capacity to achieve quality improvements, and (4) what influences the decisions to make changes in quality initiatives.

[15] As of May 2015 the CMS website is: http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html

[16] MA penetration datasets are available here: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData/MA-State-County-Penetration.html

After the questionnaire was drafted, researchers conducted a pre-test in June–July 2012 with individuals from eight health plans that had been identified to participate in expert/stakeholder interviews in fall 2011 and winter 2012. The self-administered pre-test questionnaire contained 48 questions. Respondents were asked to complete the self-administered questionnaire and note how long it took. Respondents were also asked to participate in a debriefing interview to discuss any comments or concerns. The questionnaire was finalized in consultation with CMS.

As part of the Office of Management and Budget clearance process, the project description and estimated burden was published in the Federal Register and open to public comment for 60 days. Additional changes were made to the questionnaire based on public comments prior to OMB approval, including shortening the questionnaire and making an electronic version available to MAOs.

Development of the sampling frame

The universe of MA contracts came from CMS's contract and plan contact database maintained in the Health Plan Management System (HPMS). Using 2013 contract information, researchers included the following "contract types[17]:"

- Local CCP
- Regional CCP
- PFFS
- MSA

This generated 558 contracts. Researchers removed contracts with a status of "withdrawn" or "terminated," leaving 522 "active" and 11 "consolidation" contracts, for a total sample of 533. The 533 contracts were held by 165 different parent organizations.

Data collection procedures

Advance letter

At the start of data collection, we sent an advance letter printed on CMS letterhead and signed by Cynthia Tudor, Ph.D., CMS's Director of the Medicare Drug Benefit and C&D Data Group, introducing the study and providing contact information for both CMS and Mathematica staff. Letters were sent via FedEx (or by Priority Mail when the address was a P.O. Box) to the CEO of each parent health plan organization. An email was sent a few days later to the CEO and "general contact" for the organization with the advance letter included as an attachment.

[17] Excluded contracts include: 1876 Cost, Chronic Care, Demo, Employer/Union Only Direct Contract PDP, HCPP – 1833 Cost, National PACE, PDP, POS Contractor

Initial calls

We began initial calls to the CEOs approximately one week after the advance letters and emails were sent. The purpose of the initial calls was to (1) confirm the number of plans under each contract as indicated by CMS, (2) ask if quality improvement activities differed in any meaningful way across plans for any one contract (warranting sending more than one questionnaire for those contracts), and (3) determine the appropriate respondent for the questionnaire and how it should be sent (by mail or by email with a fillable PDF version attached).

The initial calls were made over four weeks beginning in early July. None of the health plans we contacted reported any significant differences in quality improvement initiatives across plans under the same contract, and therefore, multiple surveys did not have to be completed for any one contract.

Questionnaire mailing

Organizations were offered the option of receiving the questionnaire in hardcopy form via FedEx or by email in the form of a fillable PDF. The vast majority of organization contacts (95 percent) requested that the questionnaire be sent via email.

Organizations with multiple contracts were instructed to complete a questionnaire for each contract. Having an electronic version of the questionnaire reduced the burden for multiple-contract organizations, as they could save their responses to a questionnaire completed for one contract and make slight changes for a different contract when there was only minimal variation. However, if there were no differences in quality improvement activities or initiatives among contracts, respondents were instructed to list all of the relevant contracts to which the questionnaire refers. As questionnaires were received, we recorded the date of receipt and completion status in a tracking database. For organizations that were not reached during the initial call phase, the questionnaire was sent via email to the general contact and quality contact.

Reminder emails/Reminder calls

Approximately two to three weeks after the questionnaire was sent, reminder emails were sent to non-responders (those who had not yet responded and had not refused). The reminder email retained the instructions from the first email along with the PDF as an attachment and asked that respondents return the questionnaire(s) as soon as possible.

Reminder calls were made approximately one week after sending the reminder email. If these questionnaires were not received within a week (or two weeks for contacts with four or more contracts) of the reminder call, another call was made to the non-responders.

Response Rate

We received completed questionnaires for 442 of the 533 contracts for an overall, unweighted response rate of 83 percent. All 533 contracts were considered eligible. That is, none of the contracts in the sample were found to be inactive, nor did any organization indicate that they did not have MA contracts. Several organizations indicated that their contracts were new and/or very

small, making them ineligible for Star Ratings, but these contracts were deemed eligible for the survey because they were still part of the demonstration.

Of the 533, nine percent of contracts (48) refused to respond. The 48 refusals came from 13 different parent organizations, with one organization accounting for 21 of the refusal cases. Eight percent of the contracts (43) were eligible non-completes. These were cases that we were unable to reach or that were successfully contacted and did not refuse participation but did not submit a questionnaire by the end of the field period. Of the 43 contracts that did not complete questionnaires, one parent organization had 25 contracts that were part of the sample and accounted for 14 of the non-completes.

Data processing

Receipt and quality control

The team used a tracking system to follow the progress of each contact attempt throughout the study: (1) initial call, (2) survey mailing, (3) reminder emails/reminder calls. Results from the initial calls were tracked at the organization level. As respondents were identified and questionnaires sent, we recorded the date and mode of delivery. Reminder email dates were also tracked in the database.

As questionnaires were returned, Quality Control staff reviewed each completed survey and checked for any skip errors, range problems, and/or any other inconsistencies. These were either resolved by the Quality Control staff based on specifications that were provided by the project director, or by contacting the project director with questions. In a few cases, follow-up contacts were needed.

Data extraction

After all the questionnaires were reviewed, the data from the PDF files were extracted directly into a spreadsheet to be converted to SAS. All but one questionnaire was submitted by email. For that response, data were entered directly into the fillable PDF instrument for ease of extraction.

Payment Data

To calculate the payment amounts paid under the QBP demonstration, the CMS Office of the Actuary (OACT) provided the evaluation team with (1) data on Medicare program payments made to each plan and the number of beneficiary months by county for each year, 2012 through 2014 and (2) an estimate of what payments would have been under the statutory ACA payment formula assuming that the bid made under the QBP would have been the same under the ACA, regardless of the differences in benchmarks. OACT also assumed that risk-adjustment and enrollment did not differ under these two scenarios. To calculate the amount paid as a result of the QBP demonstration, we take the difference payments made under the QBP and estimated payments under the ACA, that reflect OACT's assumptions. Payment data were merged with contract-level overall Star Ratings data retrieved from CMS website to categorize payments by Star Ratings for the relevant Star Ratings year.

Out-of-pocket cost information

As described the L&M team used many existing plan and contract characteristics available in the HPMS, however, the team was also provided data from CMS contractors: CGI and Fu Associates. Together these two contractors provided L&M with estimated out-of-pocket costs for 2008-2014, which is not available on HPMS. The data from the two sources are provided conditional on health, but are not appropriate for comparisons across the two sources. Fu Associates OOP data, available from 2008 through 2010, conditions expected costs on age (six groupings) and health (five groupings). Data from CGI Federal, available from 2010 through 2015, conditions expected costs on health (three groupings) and are the same as data available on the Medicare Plan Finder Web tool. As such, we used data from CGI Federal as it provides out-of-pocket cost expectations during the QBP period. The data provided by CGI and Fu Associates included only non-employer plans. As a result, only non-employer plans were included in the analyses of out-of-pocket expenditures.

Data analysis

Descriptive analyses

The Star Ratings data analysis captured changes in the Ratings over time, both for the entire QBP contract pool and stratified by selected MAO characteristics of interest. Specific analyses included:

- Year-over-year and descriptive analyses of changes in overall Star Ratings score distributions from Star Rating years, 2009-2015.
- Description of changes in overall Star Rating distribution for the MA population as a whole, and for population segments based on the following contract variables:
    - Type of contract
    - Tax status
    - Percent of contract enrollees enrolled in SNPs within the contract
    - Areas of the country where beneficiaries have access to a four Star or more rated contract:
        - A longitudinal tracking of MA contract Ratings changes over time. Contracts were tracked in time period *t* and *t+1,* the probability of their improving or declining in the following was calculated and plotted on line graphs
        - Year-over-year descriptive analyses of allocation of quality bonus payments by Star Ratings.

Tracking MA measures and their comparison groups over time

The MA measures and their respective comparison groups were enrollment weighted and plotted over time to create a visual comparison of trends. Enrollment from the data collection year was used to weight the Medicare data. Enrollment from the year the data were released (the year following the data collection) was used to weight the NCQA data.

Model building for QI activities

The team estimated a simple linear regression in order to isolate any potential effects of the quality improvement (QI) activities contracts noted as important during the survey on their Star Ratings. To do this the team categorized the plans according to whether they improved or declined/stayed the same between overall Star Ratings year 2013 and 2015. The team used the following independent variables for the model:

- The number of activities the plan noted as important (which ranged from zero to twelve)
- The contract's overall Star Rating in 2013
- An indicator variable for parent organization Aetna, which was present for 23 of the 416 contracts that completed the survey and had a Star Rating in 2013 and 2015

In estimating this model the team found that the QI activity 'Provider incentive programs' was the only statistically significant QI activity and associated with contracts that improved their Overall Rating between 2013 and 2015.

Model building for enrollment analysis

Analyses of enrollment changes required data from several sources. First, crosswalk data from the HPMS were used to identify the contract and plan ID that each plan had in the following year. Second, enrollment data were merged with the crosswalk information and using year-to-year changes in contract and plan associations, we created one year lagged values of enrollment, summing enrollment over all plans in the previous year operating under the same plan ID in the current year. Third, the enrollment files were merged with Star Ratings information, including synthetic Star Ratings for 2009 and 20110. Fourth, the plans in each year, and the lagged plans, were merged with plan information from the HPMS. The resulting files, for each year 2009 to 2014, included observations with current year information for a plan and lagged enrollment and plan information from one year prior. Finally, an analytic dataset was created by appending each of the files that contain current year information and lagged information. Only plans with non-zero-enrollment in one year of study and those eligible for QBP bonus during the demonstration period were kept in the analytic file.

Analyses of enrollment included providing descriptive statistics at both the contract and plan level. At the contract and plan level:

- We identified mean enrollment for contracts and plans conditional on the contract Star Rating.
- We provide information on year-to-year changes in Star Ratings for contracts in the analyses over the entire 2009 to 2014 period.
- As plan consolidation was common during the 2009 to 2014 period, we provide counts of the number of consolidations occurring for each year, conditional on the number of plans consolidating into one plan.

Enrollment changes are described by showing not only mean changes for each Star-Rating, but also several points in the distribution of changes (1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th

percentiles. Additionally, we show for each Star Rating the number of plans experiencing year-to-year enrollment growth, by Star Rating and year.

Whether plans experienced enrollment growth is a focus of the analyses. Linear regressions on a limited dependent variable (1=year-over-year increase in enrollment) are presented with Star-Ratings changes, serving as indicators of quality changes, being the covariates of interest. The regressions are descriptive, demonstrating how the proportion of plans with increasing enrollment changes conditional on whether the plans experienced changes in Star Ratings. Several specifications are presented to test the robustness of the findings. The preferred specification includes year fixed effects, lagged Star Ratings indicators, indicators for the number of plans consolidated, indicators for a service area expansion or reduction in the year, and plan-level fixed effects.

The final regressions presented are, again, descriptive. The regressions interact the variables of interest, whether the plan experienced a Rating increase or decrease, with the lagged Star Rating value, the level from which the changes occurred.

Model building for Premium and Out-of-Pocket Cost Analyses

Using data obtained from CGI Federal on cost sharing and expected out-of-pocket (OOP) costs for MA plans for payment years 2011 through 2015, we examine the relationship between contract-level Star Ratings and bonus payments and plan-level premiums, and cost sharing. The cost data from CGI Federal, which excluded employer plans, were merged with contract Ratings from 2011 to 2015 for plans that would otherwise be eligible for QBP bonus payments during the QBP demonstration. From 2012 to 2014, these included all non-employer plans that would receive a bonus if rated three Stars or higher. The analyses include providing descriptive statistics for premium and cost sharing, conditional on health status. Descriptive statistics are presented for each year and each year conditioned on contract Star Ratings.

To describe how mean premiums and cost sharing varied during the QBP period, we estimated a linear regression using data from 2011 to 2015. The outcomes of interest (premium amounts, non-premium OOP costs, and total OOP costs) were regressed on a linear time trend, indicator variables for each of the QBP demonstration years (2012, 2013, 2014), and plan-level fixed effects. The coefficients of interest are the indicators for years 2012 to 2014, which show how the mean outcomes varied from a linear trend between 2011 and 2015. Negative (positive) coefficients indicate that the mean outcomes were below (above) the linear trend established between 2011 and 2015 values and that costs were less than (greater than) expected during years where the bonus payments were made. Interpreting the coefficients in a meaningful manner relies on the assumption that the outcome costs would have changed linearly in the absence of the QBP demonstration.

APPENDIX 1. SAMPLE PAYMENT CALCULATION

The differences between the QBP and the ACA payment are a function of the differences in the level of bonus payment amounts for each Star Rating, but they are also a result of how the quality bonuses are applied in formula that determined payments to MA plans. The ACA made several changes to the way MA plan payments are calculated, including tying county benchmarks, which are the maximum amounts that the Medicare program will pay to provide Part A and B coverage to Medicare beneficiaries in each county, to projected county-level Medicare fee-for-service spending. During the phase in of the new benchmark calculation, benchmarks will be a blend of the pre-ACA and ACA benchmarks. The phase-in period began in 2012 and lasts two, four, or six years; the larger the difference between a county's pre-ACA and ACA benchmarks, the longer the phase-in period.

The following are simplified examples of how payments are calculated under the ACA formula and under the QBP demonstration under two scenarios: (1) when a plan's bid amount is equal to the county benchmark amount and (2) when a plan's bid amount is below the QBP benchmark; the differences between the ACA and QBP amounts under the two scenarios is the amount paid under the QBP demonstration.

In these simplified examples, we will calculate the amount of the bonus payment for a 3.5-Star contract with a single county service area in 2012. All the relevant information for this example is in Table 23. (The actual values for each county are published annually in the MA ratebooks.)

Table 23. Values to calculate hypothetical payments under the ACA and QBP payment formulas

Relevant values for example	2012
3.5 Star ACA bonus percentage	0%
3.5 Star QBP bonus percentage	3.5%
County transition-blend percentage	50%
Pre-ACA county rate	$734
Pre-ACA transition blend %	50%
FFS rate excluding phase-out IME	$726
Quartile percent	1.075
FFS transition blend %	50%

Example Payment Calculation 1

The county benchmark (rate) is a blend of the pre-ACA amount and the ACA amount. Under the QBB demonstration, the quality bonus percentage is applied to both the pre-ACA amount and the ACA amount, as shown underlined in the equation below:

$$2012\ Rate_{QBP} =$$

$$[(2012\ Pre\text{-}ACA\ Rate) \times (1 + \underline{bonus\ \%}) \times (Pre\text{-}ACA\ transition\ blend\ \%)] +$$

[(2012 FFS rate excluding phase-out IME) x (quartile % + bonus %) x (FFS transition blend %)]

In our example using a 3.5 Star plan, the bonus percent under the QBP demonstration is 3.5 percent:

$$2012\ Rate_{QBP} = [(\$734) \times (1+.035) \times (.50)] + [(\$726) \times (1.075 + .035) \times (.50)]$$

$$2012\ Rate_{QBP} = \$783$$

Under the ACA, the quality bonus percentage is applied *only* to the ACA amount.

$Rate_{ACA}$= [(2012 Pre-ACA Rate) x (Pre-ACA transition blend %)] +

[(2012 FFS rate excluding phase-out IME) x (quartile % + bonus %) x (FFS transition blend %)]

And in our example using a 3.5 Star contract, the bonus percent under the ACA would be zero, resulting in a rate of $757 per beneficiary per month:

$$Rate_{ACA} = [(\$734) \times (.50)] + [(\$726) \times (1.075) \times (.50)]$$

$$Rate_{ACA} = \$757$$

Under a scenario where the plan bid the benchmark in this county under these rates formulas (and for the sake of simplicity assume no risk adjustment), the effect of the quality bonus payment demonstration would be the QBP rate calculated above ($783) minus the ACA rate calculated above ($757) or $26 per beneficiary per month in 2012. This calculation shows that the bonus percentage and the portion of the rate formula to which the bonus percentage applies affect the QBP bonus amount.

Example Payment Calculation 2

Now, consider a second scenario where the plan bid was below the QBP benchmark--$760 per beneficiary per month. In that case, the plan would receive as payment its bid plus the rebate amount. When a plan's actual big is below the case-mix adjusted benchmark, it received a portion of that difference in the form of a rebate. A plan is required return the rebate to its enrollees as supplemental benefits or lower premiums. Premiums savings may be applied to the Part B premium (in which case the government retains the amount for that use), the Part D premium, or the premium for the total package that may include supplemental benefits.[18] The ACA also changed the rebate calculation. Prior to the ACA, plans received 75 percent of the difference regardless of Star Rating. The ACA tied rebates to performance Starting in 2012 and phased the change in the rebate calculation in in 2012 and 2013. The rebate percentages under the ACA and the phase in values are shown in Table 24.

[18] Medicare Payment Advisory Commission, "Medicare Advantage Program Payment System: Payment Basics," October 2014. Available at http://www.medpac.gov/documents/payment-basics/medicare-advantage-program-payment-system-14.pdf?sfvrsn=0.

Table 24. MA rebate and phase-in percentages under the ACA

	2012	2013	2014
ACA rebate percentage			
<3.5 Star Rating	50%	50%	50%
3.5 to <4.5 Star Rating	65%	65%	65%
4.5+Star Rating	70%	70%	70%
New plan	Like 3 Star (50%)	Like 3 Star (50%)	Like 3.5 Star (65%)
Low enrollment	Like 4.5 Star (70%)	Like 3 Star (50%)	Like 3 Star (50%)
Rebate phase-in %			
Pre-ACA weight	67%	33%	0%
ACA weight	33%	67%	100%

In our example of a 3.5 Star plan in 2012, the payment under the QBP in 2012 would be $760 plus a rebate amount of $16.49 that the plan could use to provide additional benefits, such as reduced cost sharing or coverage of services that are not part of the Medicare benefit package. The rebate amount is calculated in 2012, using a blend of the ACA and pre-ACA formulas:

$$2012\ Rate_{QBP} = \$760 + [(\$783-\$760) \times .75 \times .67] + [(\$783-\$760) \times .65 \times .33]$$

$$2012\ Rate_{QBP}+rebate = \$760+\$16.49$$

$$2012\ Rate_{QBP}+rebate = \$776.49$$

Assuming that the plan bid the same dollar amount ($760) under the ACA, the payment under the ACA in 2012 would still be $757, because the bid was $3 above the benchmark. Beneficiaries would have to make up the difference by paying extra premiums in this scenario. In this scenario, the amount of the QBP bonus payment would equal $776.49 minus $757, or $19.49 per beneficiary per month in 2012.

APPENDIX 2. STAR RATINGS CUT POINT ANALYSIS

As noted in the introduction, evaluating changes in overall Star Ratings from year to year is confounded by changes to the way these composite are measured. These year-to-year measurement changes include:

- Changes to the set of measures that are included in the overall Ratings. Each year new measures are included and measures are retired or become display measures. Relatedly, contracts' Star Ratings may reflect a different set of measures from year to year, depending on which measures they report.
- Changes to some measure specifications. For example, the specific population to which a measure applies may change.
- Changes to the individual measure Star Rating cut points. The cut points are the thresholds for translating scores on individual measures to a Star Rating for each individual measure. A contract could have earned the same raw score for a given measure, but this could have translated to a different Star Rating score across years.

To assess whether changing cut points for translating individual measure scores to Star Ratings values made it easier or more difficult to obtain a particular Star Rating from one year to the next, we examined 36 Part C measures that were fairly consistently collected for the 2011 to 2015 Ratings. As shown in Table 25, there is movement in both directions, with some Star Ratings being easier to obtain in the next year and some being more difficult. (The list of measures analyzed follows the table.) For these 36 measures, it becomes more difficult to obtain the same Star Rating on more measures moving from one year to the next. For example, in 2012, it was easier to obtain a 2-Star Rating on 8 of 20 measures than it was in 2011 and more difficult to obtain a 2-Star Rating for the remaining 12 measures. The least movement is in the 4-Star category, where CMS has published the pre-determined 4-Star threshold prior to the Ratings. Starting with the 2016 Star Ratings, CMS will not use predetermined 4-Star thresholds.[19]

Table 25. Analysis of changes to individual measure Star Rating cut points for 36 measures, 2012 to 2015

To obtain:	Star Ratings year	Count measures	Easier than previous year	Same as previous year	More difficult than previous year
2-Stars	2012	20	8	0	12
	2013	34	12	3	19
	2014	36	9	8	19
	2015	36	10	6	20
3-Stars	2012	21	6	6	9

[19] Centers for Medicare & Medicaid Services, Trends in Part C & D Star Rating Measure Cut Points, November 18, 2014. Available at https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/Downloads/2015-Trends-in-Part-C-and-D-Rating-Measure-Cut-Points-.pdf.

To obtain:	Star Ratings year	Count measures	Easier than previous year	Same as previous year	More difficult than previous year
	2013	34	9	8	17
	2014	35	12	11	12
	2015	34	11	7	16
4-Stars	2012	21	0	19	2
	2013	34	2	27	5
	2014	36	4	28	4
	2015	36	5	23	8
5-Stars	2012	20	9	3	8
	2013	34	12	7	15
	2014	36	11	10	15
	2015	36	12	18	16

1. Breast Cancer Screening
2. Colorectal Cancer Screening
3. Cardiovascular Care - Cholesterol Screening
4. Diabetes Care - Cholesterol Screening
5. Glaucoma Testing
6. Annual Flu Vaccine
7. Improving or Maintaining Physical Health
8. Improving or Maintaining Mental Health
9. Monitoring Physical Activity
10. Adult BMI Assessment
11. Care for Older Adults – Medication Review
12. Care for Older Adults – Functional Status Assessment
13. Care for Older Adults – Pain Screening
14. Osteoporosis Management in Women who had a Fracture
15. Diabetes Care – Eye Exam
16. Diabetes Care – Kidney Disease Monitoring
17. Diabetes Care – Blood Sugar Controlled
18. Diabetes Care – Cholesterol Controlled
19. Controlling Blood Pressure
20. Rheumatoid Arthritis Management
21. Improving Bladder Control
22. Reducing the Risk of Falling
23. Plan All-Cause Readmissions
24. Getting Needed Care
25. Getting Appointments and Care Quickly
26. Customer Service
27. Overall Rating of Health Care Quality

28. Overall Rating of Plan
29. Care Coordination
30. Complaints Tracking Module
31. Beneficiary Access and Performance Problems
32. Members Choosing to Leave the Plan
33. Health Plan Quality Improvement
34. Plan Makes Timely Decisions about Appeals
35. Reviewing Appeals Decisions
36. Call Center - Foreign Language interpreter and TTY/TDD availability

APPENDIX 3. ADDITIONAL OVERALL STAR RATINGS DATA

For all tables in this appendix, we calculated the overall Ratings in 2009 and 2010 using from measure scores according to 2011 technical specifications. We used publicly available Star Ratings data accessed from http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html. Other characteristics were identified using data from HPMS. Enrollment was obtained from monthly enrollment files, with the exception of 2015, for which we used the March 2015 enrollment data. Note that the totals in the tables reflect the sum of the rows in the tables; for those tables excluding contract too new and too few observations, those values are also excluded from the totals.

Table 26. Distribution of MA contracts by overall Star Ratings, 2009 to 2015

	Total contracts	Fewer than 3 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars	Contract too new	Too few Obs.
2009	339	70	128	79	44	18	0	0	0
2010	400	80	156	81	56	26	1	0	0
2011	530	43	165	100	41	33	3	55	90
2012	545	71	143	118	51	39	9	39	75
2013	553	63	126	130	62	51	8	51	62
2014	566	18	108	141	87	60	10	85	57
2015	525	28	73	135	85	59	9	77	59

Table 27. Distribution of MA enrollment by overall Star Ratings, 2009 to 2015

	Total enrollees	Fewer than 3 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars
2009	9,985,284	1,568,257	3,551,133	2,753,585	934,699	1,177,610	--
2010	10,401,884	1,664,665	3,997,357	1,891,194	1,390,474	1,452,624	5,570
2011	10,383,085	916,291	3,603,900	2,984,373	1,072,850	1,688,428	117,243
2012	12,489,924	1,136,570	3,462,768	4,417,339	1,273,707	1,037,667	1,161,873
2013	14,005,973	690,481	2,951,700	5,352,269	1,809,702	2,009,213	1,192,608
2014	15,324,499	185,694	2,443,456	4,854,480	3,292,231	3,158,212	1,390,426
2015	16,073,357	339,428	1,601,617	3,700,711	5,717,501	3,228,123	1,485,977

Table 28. Distribution of MA contracts by overall Star Ratings and tax status, 2009 to 2015

	Count contracts	Fewer than 3 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars
2009 (n=339)							
For Profit	255	66	112	52	24	1	0
Non-Profit	84	4	16	27	20	17	0
2010 (n=400)							
For Profit	301	79	137	53	26	6	0
Non-Profit	99	1	19	28	30	20	1
2011 n=(385)							
For Profit	294	39	154	68	23	10	0
Non-Profit	91	4	11	32	18	23	3
2012 (n=431)							
For Profit	333	66	123	101	26	16	1
Non-Profit	98	5	20	17	25	23	8
2013 (n=440)							
For Profit	343	57	108	108	46	22	2
Non-Profit	97	6	18	22	16	29	6
2014 (n=424)							
For Profit	322	15	93	118	63	31	2
Non-Profit	102	3	15	23	24	29	8
2015 (n=388)							
For Profit	285	22	65	105	64	28	1
Non-Profit	103	5	8	30	21	31	8

Table 29. Distribution of MA contracts by overall Star Ratings and organization type, 2009 to 2015

	Count contracts	Fewer than 3 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars	Contract too new	Too few obs
2009 (n=339)									
Local CCP	300	54	115	72	42	17	0	0	0
PFFS	26	12	5	6	2	1	0	0	0
Regional CCP	13	4	8	1	0	0	0	0	0

	Count contracts	Fewer than 3 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars	Contract too new	Too few obs
2010 (n=400)									
Local CCP	356	67	137	74	51	26	1	0	0
PFFS	33	9	13	6	5	0	0	0	0
Regional CCP	11	4	6	1	0	0	0	0	0
2011 (n=530)									
Local CCP	495	39	154	94	40	33	3	42	90
PFFS	22	1	4	5	1	0	0	11	0
Regional CCP	13	3	7	1	0	0	0	2	0
2012 (n=545)									
Local CCP	510	67	126	115	50	39	9	29	75
PFFS	22	2	8	2	1	0	0	9	0
Regional CCP	13	2	9	1	0	0	0	1	0
2013 (n=553)									
Local CCP	524	58	113	124	62	50	8	49	60
PFFS	18	5	5	4	0	0	0	2	2
Regional CCP	11	0	8	2	0	1	0	0	0
2014 (n=566)									
Local CCP	543	16	98	134	85	59	10	85	56
PFFS	12	1	4	4	2	0	0	0	1
Regional CCP	11	1	6	3	0	1	0	0	0
2015 (n=525)									
Local CCP	507	27	68	127	82	59	9	77	58
PFFS	8	0	2	3	2	0	0	0	1
Regional CCP	10	1	3	5	1	0	0	0	0

Table 30. Distribution of MA contracts by overall Star Ratings and plan types, 2009 to 2015

	Count contracts	Fewer than 3 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars	Contract too new	Too few obs
2009 (n=339)									
HMO/HMOPOS	228	47	78	53	34	16	0	0	0
Local PPO	71	7	36	19	8	1	0	0	0
PSO (State License)	1	0	1	0	0	0	0	0	0
PFFS	26	12	5	6	2	1	0	0	0
Regional CCP	13	4	8	1	0	0	0	0	0
2010 (n=400)									
HMO/HMOPOS	272	53	101	56	38	23	1	0	0
Local PPO	83	14	35	18	13	3	0	0	0
PSO (State License)	1	0	1	0	0	0	0	0	0
PFFS	33	9	13	6	5	0	0	0	0
Regional CCP	11	4	6	1	0	0	0	0	0
2011 (n=530)									
HMO/HMOPOS	351	32	108	69	30	24	3	21	64
Local PPO	143	7	45	25	10	9	0	21	26
PSO (State License)	1	0	1	0	0	0	0	0	0
PFFS	22	1	4	5	1	0	0	11	0
Regional CCP	13	3	7	1	0	0	0	2	0
2012 (n=545)									
HMO/HMOPOS	359	55	86	71	39	31	9	15	53
Local PPO	150	12	40	43	11	8	0	14	22
PSO (State License)	1	0	0	1	0	0	0	0	0
PFFS	22	2	8	2	1	0	0	9	0
Regional CCP	13	2	9	1	0	0	0	1	0
2013 (n=553)									
HMO/HMOPOS	370	50	78	79	40	36	8	31	48
Local PPO	154	8	35	45	22	14	0	18	12

	Count contracts	Fewer than 3 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars	Contract too new	Too few obs
PFFS	18	5	5	4	0	0	0	2	2
Regional CCP	11	0	8	2	0	1	0	0	0
2014 (n=566)									
HMO/HMOPOS	393	15	71	83	58	40	10	70	46
Local PPO	150	1	27	51	27	19	0	15	10
PFFS	12	1	4	4	2	0	0	0	1
Regional CCP	11	1	6	3	0	1	0	0	0
2015 (n=525)									
HMO/HMOPOS	383	19	58	83	59	42	9	70	43
Local PPO	124	8	10	44	23	17	0	7	15
PSO (State License)	0	0	0	0	0	0	0	0	0
PFFS	8	0	2	3	2	0	0	0	1
Regional CCP	10	1	3	5	1	0	0	0	0

Table 31. Overall Star Ratings by share of SNP plans under contract, 2009 to 2015

	Count contracts	Fewer than 3 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars
2009 (n=339)							
Zero Percent SNP	260	47	93	67	36	17	0
Between 0-100% SNP	46	14	22	6	4	0	0
100% SNP	33	9	13	6	4	1	0
2010 (n=400)							
Zero Percent SNP	303	55	114	64	48	21	1
Between 0-100% SNP	54	17	21	9	3	4	0
100% SNP	43	8	21	8	5	1	0
2011 n=(385)							
Zero Percent SNP	306	28	124	85	34	32	3
Between 0-100% SNP	35	10	20	5	0	0	0
100% SNP	44	5	21	10	7	1	0
2012 (n=431)							

	Count contracts	Fewer than 3 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars
Zero Percent SNP	325	41	101	98	40	36	9
Between 0-100% SNP	60	22	26	9	2	1	0
100% SNP	46	8	16	11	9	2	0
2013 (n=440)							
Zero Percent SNP	341	36	89	109	53	46	8
Between 0-100% SNP	48	12	21	9	3	3	0
100% SNP	51	15	16	12	6	2	0
2014 (n=424)							
Zero Percent SNP	314	7	67	107	71	53	9
Between 0-100% SNP	56	5	20	17	8	5	1
100% SNP	54	6	21	17	8	2	0
2015 (n=388)							
Zero Percent SNP	279	13	38	99	72	48	9
Between 0-100% SNP	60	8	16	20	8	8	0
100% SNP	49	6	19	16	5	3	0

Table 32. Share of counties with a 4-Star (or higher) rated contract, 2009 to 2015

County has at least one 4-Star rated contract?	2009	2010	2011	2012	2013	2014	2015
No	68%	47%	37%	44%	8%	5%	2%
Yes	32%	53%	63%	56%	92%	95%	98%

APPENDIX 4. TRANSITION MATRICES

The quality of care provided in 2011 may not equal that contract's Rating because of the lag between the data collection year (t-2 and t-1, depending on the measure) and the Star Ratings year (t). One-year transition matrices show contracts' Star Rating in year *t+1* given its Star Rating in year *t*.[20] As a result, there may potentially be a "survivor effect" if the contracts that remained in the data were significantly different than contracts that entered or exited from the market. However, the number of contracts that remained in the data was much larger than the contracts that entered or exited the market, which would make a "survivor effect" relatively small. Note that for all tables in this appendix, contracts that were too new to be measured or that did not have enough data to calculate a Rating are excluded from this analysis.

Table 33. One-year transition matrix, 2009 to 2010

	2010							
2009	<=2 Stars	2.5 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars	Missing
<=2 Stars	0%	25%	25%	0%	0%	0%	0%	50%
2.5 Stars	0%	52%	39%	2%	0%	0%	0%	8%
3 Stars	0%	13%	69%	11%	2%	0%	0%	5%
3.5 Stars	0%	0%	20%	53%	22%	1%	0%	4%
4 Stars	0%	0%	0%	30%	45%	25%	0%	0%
4.5 Stars	0%	0%	0%	0%	28%	67%	0%	6%
5 Stars	--	--	--	--	--	--	--	--

Table 34. One-year transition matrix, 2010 to 2011

	2011							
2010	<=2 Stars	2.5 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars	Missing
<=2 Stars	0%	0%	0%	0%	0%	0%	0%	100%
2.5 Stars	0%	36%	47%	5%	0%	0%	0%	12%
3 Stars	0%	7%	74%	16%	0%	0%	0%	3%
3.5 Stars	0%	0%	13%	58%	16%	6%	0%	6%
4 Stars	0%	0%	0%	34%	44%	20%	0%	2%
4.5 Stars	0%	0%	0%	8%	16%	60%	12%	4%
5 Stars	0%	0%	0%	0%	0%	100%	0%	0%

[20] The data included in the transition matrices represent contracts that remained from one year to the next and includes contracts that changed contract numbers, as indicated on the publicly available plan crosswalks. Available at http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData/Plan-Crosswalks.html.

Table 35. One-year transition matrix, 2011 to 2012

	2012							
2011	<=2 Stars	2.5 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars	Missing
<=2 Stars	--	--	--	--	--	--	--	--
2.5 Stars	10%	61%	27%	2%	0%	0%	0%	0%
3 Stars	1%	19%	54%	24%	1%	0%	0%	0%
3.5 Stars	0%	0%	21%	53%	19%	7%	0%	0%
4 Stars	0%	0%	2%	24%	44%	29%	0%	0%
4.5 Stars	0%	0%	3%	0%	21%	58%	18%	0%
5 Stars	0%	0%	0%	0%	33%	0%	67%	0%

Table 36. One-year transition matrix, 2012 to 2013

	2013							
2012	<=2 Stars	2.5 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars	Missing
<=2 Stars	0%	80%	20%	0%	0%	0%	0%	0%
2.5 Stars	2%	53%	42%	2%	0%	0%	0%	3%
3 Stars	0%	11%	59%	28%	1%	0%	0%	0%
3.5 Stars	0%	0%	8%	64%	24%	4%	0%	0%
4 Stars	0%	0%	0%	30%	44%	26%	0%	0%
4.5 Stars	0%	0%	0%	0%	21%	76%	0%	3%
5 Stars	0%	0%	0%	0%	0%	22%	78%	0%

Table 37. One-year transition matrix, 2013 to 2014

	2014							
2013	<=2 Stars	2.5 Stars	3 Stars	3.5 Stars	4 Stars	4.5 Stars	5 Stars	Missing
<=2 Stars	--	--	--	--	--	--	--	--
2.5 Stars	2%	17%	69%	12%	0%	0%	0%	0%
3 Stars	0%	4%	51%	41%	3%	0%	0%	0%
3.5 Stars	0%	1%	7%	53%	32%	6%	0%	1%
4 Stars	0%	0%	0%	26%	52%	23%	0%	0%
4.5 Stars	0%	0%	0%	2%	20%	71%	8%	0%
5 Stars	0%	0%	0%	0%	0%	25%	75%	0%

Table 38. One-year transition matrix, 2014 to 2015

	2015							
2014	**<=2 Stars**	**2.5 Stars**	**3 Stars**	**3.5 Stars**	**4 Stars**	**4.5 Stars**	**5 Stars**	**Missing**
<=2 Stars	--	--	--	--	--	--	--	--
2.5 Stars	0%	38%	62%	0%	0%	0%	0%	0%
3 Stars	0%	10%	52%	32%	1%	0%	0%	5%
3.5 Stars	0%	3%	12%	65%	14%	3%	0%	2%
4 Stars	0%	2%	1%	29%	54%	13%	0%	0%
4.5 Stars	0%	2%	0%	0%	28%	65%	3%	2%
5 Stars	0%	0%	0%	0%	0%	30%	70%	0%

APPENDIX 5. MEASURE-LEVEL ANALYSIS

Table 39 shows the national average enrollment-weighted score for measures that were measured consistently over time, though some have small specification changes, and that were included in the overall Star Ratings measure during some part of the time period, 2009 through 2015. As discussed in the body of the report, we did not find a pattern of greater improvement in the demonstration period for MA contracts compared to a comparison group of commercial plans, Medicaid plans, FFS Medicare, or stand-alone PDPs on a select set of comparable measures. Table 40 shows the change in the scores over time for a broader set of measures than shown in the body of the report. It shows that average enrollment-weighted scores for a set of measures generally increased between the pre-QBP and QBP demo periods. However, these changes cannot be attributed to the QBP demonstration. We note that the Adult BMI assessment measure had a particularly large increase (measured in percentage point differences) over the demonstration period than other measures and compared to the pre-demonstration period. In addition, three SNP measures: (1) Care for older adults-pain screening, (2) Care for older adults-functional status assessment, and (3) Care for older adults-medication review, showed the large increases in the demonstration period, but data for these measures are not available prior to Star Ratings year 2012, so we cannot examine the trend prior to the QBP demo.

Table 39. Average enrollment-weighted scores for select measures, 2009 to 2015

Measure	2009	2010	2011	2012	2013	2014	2015	Percentage point change 2009-2012**	Percentage point change 2012-2015**
Part C Measures									
Colorectal cancer screening	53.2	58.9	56.9	58.0	63.8	66.5	69.1	4.9	11.1
Cholesterol screening for members with heart disease	88.5	--	89.6	90.2	90.5	90.4	90.9	1.7	0.7
Cholesterol screening for members with diabetes	86.5	--	89.0	89.1	90.2	90.2	90.8	2.6	1.7
Flu vaccine[a]	69.9	67.4	65.6	69.6	70.0	72.2	74.0	-0.3	4.4
Improving/maintaining physical health	59.4	66.3	66.6	66.6	65.3	66.7	68.3	7.2	1.8
Improving/maintaining mental health	80.9	77.8	77.2	77.6	76.8	77.5	79.7	-3.2	2.1
Monitoring physical activity	46.5	47.2	47.3	48.0	49.8	50.8	51.2	1.5	3.1
Adult BMI assessment	--	--	42.1	53.4	69.4	82.4	91.2	11.3	37.8
Care for older adults-Medication review	--	--	--	68.4	71.6	82.2	87.7	n/a	19.3

Measure	2009	2010	2011	2012	2013	2014	2015	Percentage point change 2009-2012**	Percentage point change 2012-2015**
Care for older adults- Functional status assessment	--	--	--	51.7	57.3	71.5	78.2	n/a	26.5
Care for older adults- Pain screening	--	--	--	45.2	53.2	70.9	86.7	n/a	41.5
Osteoporosis management	21.3	22.8	23.4	24.1	25.9	28.4	32.4	2.8	8.3
Diabetes care - Eye exam	61.0	--	66.0	66.2	67.5	68.5	70.7	5.2	4.5
Diabetes care - Kidney disease screening	87.4	--	88.6	89.6	90.1	90.8	92.0	2.1	2.5
Diabetes care - Blood sugar controlled	76.1	--	73.6	75.3	76.4	77.6	79.5	-0.8	4.2
Diabetes care - Cholesterol controlled	51.2	--	52.8	53.9	55.9	56.2	58.3	2.7	4.4
Controlling blood pressure	61.4	62.6	61.9	63.0	65.2	65.4	69.0	1.6	6.0
Rheumatoid arthritis management	70.5	71.6	72.9	74.4	75.3	77.5	78.7	4.0	4.3
Improving bladder control	36.3	35.9	36.3	36.2	35.5	35.4	36.5	-0.1	0.3
Reducing the risk of falling	55.1	55.0	56.0	58.1	58.0	60.2	59.0	3.0	0.9
All-cause readmission*	--	--	--	12.5	12.6	11.8	10.4	n/a	-2.2
Ease of getting needed care and seeing specialists	84.0	84.3	85.1	85.5	85.6	86.1	85.0	1.5	-0.5
Getting appointments and care quickly	73.8	74.4	74.3	76.1	76.0	76.2	76.7	2.2	0.6
Customer service	87.1	87.8	88.4	87.7	88.5	88.4	88.2	0.6	0.5
Overall rating of health care quality	84.0	84.4	84.7	86.2	86.4	86.4	86.6	2.2	0.4
Members' overall rating of health plan	84.4	84.5	83.9	85.7	86.4	86.5	86.4	1.3	0.6
Complaints about health plan*	--	2.0	0.5	0.4	0.3	0.2	0.2	-1.6	-0.2
Voluntary disenrollment	--	9.7	11.8	9.8	8.2	8.2	8.8	0.1	-1.0

Measure	2009	2010	2011	2012	2013	2014	2015	Percentage point change 2009-2012**	Percentage point change 2012-2015**
Timely decisions about appeals	88.5	86.1	91.6	89.9	88.7	89.8	92.4	1.4	2.4
Appeal decisions upheld	82.1	78.7	76.3	78.5	83.6	86.4	88.9	-3.6	10.4
Pneumonia vaccine[b]	67.3	66.9	67.7	69.8	70.7	70.2	71.0	2.5	1.1
Breast cancer screening[c]	69.3	70.4	71.5	71.4	71.8	73.6	75.4	2.1	4.0
Glaucoma testing[d]	60.7	62.3	65.0	65.7	68.2	70.3	--	5.0	4.6
Part D Measures									
Members' overall rating of drug plan	83.6	84.1	83.3	85.0	85.5	85.5	85.7	1.4	0.6
Getting needed prescriptions	90.2	90.1	90.3	91.5	91.6	91.5	91.3	1.3	-0.2
Diabetes treatment	--	84.1	84.6	84.4	84.9	85.4	86.3	0.3	1.9
Medication adherence - Oral diabetes meds	--	--	--	74.0	75.2	76.0	77.9	n/a	3.8
Medication adherence - ACEIs/ARBs	--	--	--	73.8	75.5	77.6	79.8	n/a	6.0
Medication adherence – Statins	--	--	--	69.2	70.1	71.9	74.8	n/a	5.6
Appeals decisions upheld	--	82.8	80.3	87.8	94.1	92.2	93.6	5.0	5.8
Timely decisions about appeals	--	88.2	91.3	87.9	94.5	93.5	97.4	-0.3	9.5
Getting information from the drug plan	81.0	81.3	81.1	83.1	84.3	84.0	82.6	2.1	-0.4

Sources: Publicly available measure-level data from CMS website and enrollment data from CMS monthly enrollment files, except 2015, for which we used March 2015 enrollment data.

*Notes: A higher number represents improvement, except for those measures marked with an *. We excluded measures from this analysis that were included in Star Ratings only after the QBP demo was announced and Part D measures that do not count towards the overall Star Ratings measure (because they are duplicative of Part C measures). **Where a value was not available for all years, we took the difference between the first and last year available for the pre- and demo-period. [a]The measure changed in 2015 to "had a flu shot since July;" formerly it had been "since September." [b]The measure was not included in Star Ratings since 2012, but display data are available for 2013-2015. [c]The measure was not included in Star Ratings in 2015, but was included as a display measure in 2015. In addition, the eligible population changed in 2011 from female enrollees age 50 to 69 to those age 40 to 69. [d]The eligible population changed in 2011 from enrollees age 50 to 80 to those age 50 to 75.*

APPENDIX 6. ADDITIONAL ENROLLMENT ANALYSIS DATA

Table 40 displays the frequency with which a given number of plans, listed in the header, consolidated into one plan in the given year. For example, the column labeled "2 plans" shows that two plans consolidated into one plan 181 times from 2008 into 2009—i.e., 362 plans consolidated into 181 plans.

Table 40. Frequency of plans consolidating to one plan in the following year

	Number of plans in previous year consolidating into one plan						
	No Consolidation	2 plans	3 plans	4 plans	5 plans	6 plans	7+ plans
2009	3,815	181	25	36	7	7	6
2010	3,130	283	30	23	11	12	5
2011	2,265	194	43	27	1	8	14
2012	2,836	96	7	6	0	1	2
2013	3,060	96	8	1	0	0	0
2014	2,903	90	10	5	2	1	2

Note: Table shows the frequency with which a given number of plans, listed in the header, consolidated into one plan in the given year. Consolidation information was taken from annual crosswalk files available in the HPMS.

Table 41. Interpretations of changes in proportions of plans with increasing enrollment prior to and during the QBP demonstration

Previous Star Rating	Change in Star Rating	Prior to QBP	During QBP
5 Stars	Decrease	The proportion of 5-Star rated plans that experienced an enrollment increase with a Ratings decrease was 3.9 percentage points lower than it was for 5-Star plans that remained 5-Star plans.	5-Star plans with Ratings decreases were an additional 11.9 percentage points less likely to experience an accompanying increase in enrollment.
4.5 Stars	Increase	The proportion of 4.5-Star rated plans that experienced an enrollment increase with a Ratings increase was 4.5 percentage points higher than it was for 4.5-Star plans that remained 4.5-Star plans.	4.5 Star plans with Ratings increases were an additional 38.3 percentage points more likely to experience an accompanying increase in enrollment during the QBP period. **The difference is statistically significant.
	Decrease	The proportion of 4.5-Star rated plans that experienced an enrollment increase after a Ratings decrease was 10.6 percentage points *higher* than	The difference in proportion of plans with increased enrollment between 4.5 Star plans with a Ratings decrease and those that remained at 4.5 Stars moved closer to zero during the QBP demonstration, at a

Previous Star Rating	Change in Star Rating	Prior to QBP	During QBP
		it was for 4.5-Star plans that remained 4.5-Star plans.	10.6 – 4.5 = 6.1 percentage point gap.
4 Stars	Increase or Decrease	The proportions of 4-Star rated plans that experienced an enrollment increase along with a Star Ratings increase or decrease were similar (~3 %-pts difference) to the proportion for 4-Star plans with a consistent year-to-year Rating.	The proportion of 4-Star rated plans that experienced an enrollment increase accompanying a Star Ratings increase or decrease decreased during the QBP period.
3.5 Stars	Increase	The proportion of 3.5-Star rated plans that experienced an enrollment increase with a Ratings increase was 13.9 percentage points higher than it was for 3.5-Star plans without a Ratings change. *The difference is statistically significant.	The difference in proportion of plans with increased enrollment between 3.5 Star plans with a Ratings increase and those that remained at 3.5 Stars was reduced to zero during the QBP period. *The reduction is statistically significant.
	Decrease	The proportion of 3.5-Star rated plans that experienced an enrollment increase with a Ratings decrease was 3.8 percentage points lower than it was for 3.5-Star plans that remained 3.5-Star plans.	5-Star plans with Ratings decreases were an additional 3.7 percentage points less likely to experience an accompanying increase in enrollment.
3 Stars	Increase	The proportions of 3-Star rated plans that experienced an enrollment increase were nearly the same, a 0.4 percentage point difference, for 3-Star plans with a Ratings increase and those without a Ratings change.	The proportion of 3-Star plans that experienced an enrollment increase was *lower* for 3-Star plans with an increase in Rating during the QBP period.
	Decrease	The proportion of 3-Star rated plans that experienced an enrollment increase with a Ratings decrease was 2.3 percentage points lower than it was for 3-Star plans that remained 3-Star plans.	3-Star plans with Ratings decreases were an additional 13 percentage points less likely to experience an accompanying increase in enrollment.
2.5 Stars	Increase	The proportion of 2.5-Star rated plans that experienced an enrollment increase along with a Star Ratings increase was 2.6 lower that the proportion with increased enrollment for 2.5-Star plans	The difference in proportion of plans with increased enrollment between 2.5 Star plans with a Ratings increase and those that remained at 2.5 Stars was reduced to zero during the QBP period, after an offsetting 2.8 percentage point increase in the difference.

Previous Star Rating	Change in Star Rating	Prior to QBP	During QBP
		with a consistent year-to-year Rating.	
	Decrease	The proportion of 2.5-Star rated plans that experienced an enrollment increase after a Ratings decrease was 18 percentage points *higher* than it was for 2.5-Star plans that remained 2.5-Star plans.	2.5-Star rated plans with a decrease in Star Rating became much less likely to experience an increase in enrollment during the QBP period (60.4 percentage points). *The difference is statistically significant.
2 Stars	Increase	The proportion of 2-Star rated plans that experienced an enrollment increase with a Ratings increase was 17.7 percentage points higher than it was for 2-Star plans without a Ratings change.	The difference in proportion of plans with increased enrollment between 2-Star plans with a Ratings increase and those that remained at 2 Stars was reduced to closer to zero during the QBP period. Note that a small number of plans carried 2-Star Ratings during this period.

Note: Descriptions in this table are applicable to the coefficients for specification (2) listed in Table 13.

APPENDIX 7. MEAN OUT-OF-POCKET EXPENDITURES

Table 42. Mean out-of-pocket expenditures by type of expenditure, year, and health

	2011	2012	2013	2014	2015
	(N=1,825)	(N=2,180)	(N=2,297)	(N=2,192)	(N=2,212)
Mean total out-of-pocket expenditures *(Std Dev)*					
Excellent Health	$3,466[A]	$3,364[B]	$3,358[B]	$3,452[A]	$3,530
	($792)	*($849)*	*($847)*	*($843)*	*($859)*
Good Health	$4,620[A]	$4,634[A]	$4,238	$4,483	$4,645[A]
	($1036)	*($1187)*	*($1036)*	*($1065)*	*($1076)*
Poor Health	$6,748[A]	$7,028	$6,695[A]	$6,895	$7,242
	($1680)	*($2089)*	*($1789)*	*($1867)*	*($1981)*
Mean non-premium out-of-pocket expenditures *(Std Dev)*					
Excellent Health	$3,207[A]	$3,081[B]	$3,107[B]	$3,222[A]	$3,297
	($739)	($766)	($781)	($790)	($795)
Good Health	$4,180[A]	$4,133[A,B]	$3,867	$4,121[B]	$4,296
	($817)	($864)	($868)	($904)	($900)
Poor Health	$5,993[A]	$6,092	$5,985[A]	$6,169	$6,532
	($1040)	($1105)	($1162)	($1154)	($1201)
Mean Part B, Part C, Part D premium *(Std Dev)*					
All Beneficiaries	$159	$138	$143[A]	$144[A]	$146[A]
	($57)	*($58)*	*($59)*	*($59)*	*($60)*
Mean Part C premium *(Std Dev)*					
All Beneficiaries	$26[A]	$25[A,B]	$24[A,B]	$23[B]	$24[A,B]
	($42)	*($41)*	*($42)*	*($40)*	*($42)*

A, B = Numbers in the same row with the same superscript (e.g. A or B) are not statistically significantly different from one another.

L&M POLICY RESEARCH