Medicare Advantage and Fee-for-Service Performance on Clinical Quality and Patient Experience Measures: Comparisons from Three Large States

Justin W. Timbie (iD)*, Andy Bogart, Cheryl L. Damberg, Marc N. Elliott* (iD)*, Ann Haas, Sarah J. Gaillot, Elizabeth H. Goldstein, and Susan M. Paddock*

Objective. To compare performance between Medicare Advantage (MA) and Fee-for-Service (FFS) Medicare during a time of policy changes affecting both programs.

Data Sources/Study Setting. Performance data for 16 clinical quality measures and 6 patient experience measures for 9.9 million beneficiaries living in California, New York, and Florida.

Study Design. We compared MA and FFS performance overall, by plan type, and within service areas associated with contracts between CMS and MA organizations. Case mix-adjusted analyses (for measures not typically adjusted) were used to explore the effect of case mix on MA/FFS differences.

Data Collection/Extraction Methods. Performance measures were submitted by MA organizations, obtained from the nationwide fielding of the Medicare Consumer Assessment of Healthcare Providers and Systems (MCAHPS) Survey, or derived from claims.

Principal Findings. Overall, MA outperformed FFS on all 16 clinical quality measures. Differences were large for HEDIS measures and small for Part D measures and remained after case mix adjustment. MA enrollees reported better experiences overall, but FFS beneficiaries reported better access to care. Relative to FFS, performance gaps were much wider for HMOs than PPOs. Excluding HEDIS measures, MA/FFS differences were much smaller in contract-level comparisons.

Conclusions. Medicare Advantage/Fee-for-Service differences are often large but vary in important ways across types of measures and contracts.

Key Words. Medicare Advantage, quality measurement, quality of care, patient experience, case mix adjustment

Thirty-three percent of Medicare beneficiaries were enrolled in a Medicare Advantage (MA) health plan in 2016, or nearly 17.6 million Medicare beneficiaries (Jacobson et al. 2017). The MA program, which has grown steadily over the past decade, attracts many beneficiaries because MA plans often offer supplemental benefits such as dental or vision care, and beneficiaries can choose from a variety of cost-sharing options in exchange for using a more restricted provider network. Beneficiaries report positive experiences with the program, and quality of care is high according to the 2016 MA Star Ratings, a five-star performance measurement system launched by the Centers for Medicare & Medicaid Services in 2008 (Centers for Medicare & Medicaid Services 2017a).

A limited number of comparisons of the performance of MA and FFS have been conducted (Landon et al. 2012; Gold and Casillas 2014; Biles, Casillas, and Guterman 2015); however, no comparisons of clinical quality or patient experience measures have been made using data more recent than 2009, despite numerous policy and payment initiatives affecting one or both programs. In particular, MA plans began receiving quality bonus payments (QBPs) based on their Star Ratings in 2012, which may have spurred MA plans to improve their performance (L&M Policy Research LLC 2016). On the other hand, the Affordable Care Act eliminated cost sharing for many preventive services beginning in 2010, which lowered barriers to beneficiaries' use of preventive services—especially among FFS beneficiaries who often face high cost-sharing requirements. Meanwhile, initiatives such as Meaningful Use and the Physician Quality Reporting System (PQRS) provided financial incentives to physicians participating in both MA and FFS to measure, report, and improve quality. It is unclear whether these policy changes have narrowed or widened performance between the two programs in recent years.

Prior studies documented higher performance of MA relative to FFS on clinical quality measures. For example, Brennan and Shepard (2010) found that MA outperformed FFS on 9 of 11 clinical process measures examined in 2006 and 2007, which included measures of medication management, mammography screening, cholesterol screening, and screening tests for

Address correspondence to Justin W. Timbie, Ph.D., RAND, 1200 S Hayes Street, Arlington, VA 22202; e-mail: jtimbie@rand.org. Andy Bogart, M.S., Cheryl L. Damberg, Ph.D., Marc N. Elliott, Ph.D., and Susan M. Paddock, Ph.D., are with RAND, Santa Monica, CA. Ann Haas, M.S., is with RAND, Pittsburgh, PA. Sarah J. Gaillot, Ph.D., and Elizabeth H. Goldstein, Ph.D., are with Centers for Medicare & Medicaid Services, Baltimore, MD.

beneficiaries with diabetes. Similarly, using data from 2009, Ayanian et al. (2013) showed that MA plans had higher performance than FFS on each of seven process measures they examined—five of which were used in Brennan and Shepard's study as well as two vaccination measures. While historically, FFS outperformed MA on measures of patient experience (Landon et al. 2004; Keenan et al. 2009; Mittler et al. 2010), Elliott et al. (2011) documented smaller differences between the two programs in 2007. Over the 8-year period beginning in 2007, both overall ratings of care and ratings of physicians were higher in MA, while ratings of specialists were comparable between the two systems (Elliott et al. 2016).

Although these studies and others provide key insights into the relative performance of each system, a 2014 review noted that most studies assessed a limited number of domains of quality and lacked subgroup analyses that could help identify potential sources of heterogeneity in performance (Landon et al. 2012; Gold and Casillas 2014; Biles, Casillas, and Guterman 2015). In particular, the magnitude of any differences in performance of MA health maintenance organizations (HMOs) and MA preferred provider organizations (PPOs) relative to FFS remains largely unknown despite the fact that nearly one-third of MA enrollees are now enrolled in MA-PPOs (Jacobson et al. 2017). Better understanding differences between MA-HMOs, MA-PPOs, and FFS in clinical quality and patient experience could help policy makers better understand the strengths and weaknesses of both systems and identify opportunities for quality improvement.

Using performance data from three large states in calendar year 2012, we compared the performance of MA and FFS both overall and separately for beneficiaries enrolled in MA-HMOs and MA-PPOs. We then examined contract-level differences in performance for the 114 MA contracts whose service areas are contained within the three states. Contracts are the units by which quality measures are reported and bonus payments are allocated within the MA Star Rating program. We compared individual contracts' scores with contract-specific comparison groups of FFS enrollees living within each contract's service area. We extend prior work by examining heterogeneity in performance by MA plan type and across individual contracts, and using a much larger set of performance indicators (including prescription drug performance measures that have not been used in comparisons previously). We also examine the role of case mix adjustment in explaining differences in performance between MA and FFS.

METHODS

Selection of MA Contracts

We used monthly contract service area files from the Centers for Medicare & Medicaid Services (CMS) to identify all MA contracts during calendar year 2012 ($n = 659$ contracts). From these, we identified the subset whose service areas were contained entirely within the states of New York ($n = 41$), Florida ($n = 36$), or California ($n = 37$), to provide a sample of contracts that were likely to enroll diverse beneficiary populations. The scope of our analyses was also limited to these three states because we had access to 100 percent FFS claims data only for beneficiaries living in these states. Enrollees living in these three states represented nearly a quarter of all Medicare beneficiaries in 2012. MA plan types (i.e., HMO or PPO) were identified using CMS administrative files.

Performance Measures

We selected 22 measures for the analysis that were reported by MA organizations to CMS in 2013 and used to generate the 2014 Star Ratings (see Table S1 for a description of each measure). Of this total, 16 were clinical quality measures, including 10 Healthcare Effectiveness Data and Information Set (HEDIS) measures (five measures for which administrative data must be used; and five measures for which medical record review could be used to supplement administrative data [i.e., "hybrid" method]); five measures relating to prescription drugs ("Part D measures"); and administration of annual flu vaccine, which is collected through the Medicare Consumer Assessment of Healthcare Providers and Systems (MCAHPS) survey. The HEDIS measures used in the analysis were selected to cover multiple domains of quality, including preventive care, chronic condition management, and hospital readmissions. We used six additional MCAHPS measures to assess beneficiaries' experiences with their providers and drug plan. CMS provided beneficiary-level data submitted by each contract reflecting performance during calendar year 2012.

Measuring FFS Performance

Fee-for-service residents living in the three states were used in all analyses. For contract-specific analyses, we created FFS comparison groups by selecting the

cohort of FFS beneficiaries who lived in each contract's service area and met a measure's eligibility criteria. We generated HEDIS measures for FFS beneficiaries using Medicare enrollment files and inpatient, outpatient, and carrier claims from 2010 through 2012. All HEDIS measures used HEDIS 2013 specifications. FFS enrollees included in the HEDIS analysis met the following criteria during the look-back period used to define eligibility for each measure: continuous residency in New York, Florida, or California (since the FFS claims extract used for the analysis was limited to these states only); continuous Part A and B eligibility; continuous enrollment in FFS; no hospice utilization; and no claims for which Medicare was a secondary payer. For measures in which a claim for a prescription drug qualified as a "success" for the measure, we required continuous enrollment in a prescription drug plan (PDP) for the entire measure look-back period. Part D performance measures for FFS beneficiaries enrolled in PDPs were derived from Prescription Drug Event claims. FFS patient experience was assessed using MCAHPS surveys fielded to Medicare FFS beneficiaries with or without Part D coverage. Four of these measures reflect FFS enrollee experiences with their FFS providers while two measures reflect their experiences with PDPs.

Estimating MA and FFS Performance

Medicare Advantage/Fee-for-Service differences on clinical quality and patient experience measures were estimated using logistic and linear regression, respectively. All analyses were conducted at the beneficiary level. Models used to estimate overall differences between MA and FFS included an indicator for MA and county fixed effects. Models used to compare MA-HMO, MA-PPO, and FFS included indicators for MA-HMO and MA-PPO and county fixed effects. Contract-specific analyses included fixed effects for each MA contract and county-specific FFS indicators (i.e., binary indicators of whether a beneficiary is both enrolled in FFS and resides in the county of interest). Parameterizing FFS as a set of county-specific indicators allowed FFS comparisons to be customized according to each contract's service area. After fitting each model, we converted estimates of MA/FFS differences into probabilities or point differences depending on the measure.

Patient experience measures and Plan All-Cause Readmissions were case mix adjusted for all analyses. For the six patient experience measures, we used the standard MCAHPS adjustment methodology, which includes adjustments for age (18–64, 65–69, 70–74, 75–79, 80–84, ≥85), education (eighth grade or less, some high school, high school, less than bachelor's degree,

bachelor's degree, postbachelor's degree), general health (excellent, very good, good, fair, poor), mental health (excellent, very good, good, fair, poor), dual eligibility, Part D low-income subsidy eligibility, and use of a proxy in responding to the survey. Plan All-Cause Readmissions were adjusted using HEDIS 2013 specifications.

We examined the extent to which case mix differences explained the observed MA/FFS differences for measures that were not already adjusted in our main analysis. We adjusted clinical quality measures for age (18–64, 65–69, 70–75, 76–79, 80–84, ≥85), gender, dual eligibility, Part D low-income subsidy eligibility, race/ethnicity, neighborhood socioeconomic status (SES) index, and hierarchical condition category (HCC) score. Each beneficiary's race/ethnicity was derived using the Medicare Bayesian Improved Surname and Geocoding method (Martino et al. 2013). To measure neighborhood SES, we used a six-item scale of neighborhood characteristics measured at the zip code level from the 2008 to 2012 American Community Survey (Bird et al. 2010). The six items included household income, poverty, receipt of public assistance, unemployment, household structure, and educational attainment. The HCC score is an index of predicted spending based on a beneficiary's sociodemographic characteristics and selected diagnoses measured from claims and represents a summary measure of comorbidity. Propensity-score weights, which were derived from a model that predicts enrollment in MA as a function of all case mix variables described above, were also included in all adjustment models to help improve balance between MA and FFS on the set of case mix variables. For both clinical quality and patient experience measures, we then applied county-specific weights (based on county-specific MA enrollment counts) to county-specific estimates of performance for each MA contract and its associated comparison group to account for differences in the geographic distribution of MA and FFS enrollees within each contract's service area.

We determined the statistical significance of our contract-level MA/FFS differences by testing linear contrasts of regression parameters that were customized to each contract's service area counties. Each contrast tested whether each MA/FFS difference was statistically different from 0 using a two-sided t-test. To illustrate the magnitude of these differences, we also converted MA/FFS difference estimates into contract-level effect sizes. An effect size threshold of 0.2 standard deviations in absolute value, which represents a widely used benchmark for at least a "small" effect size (Cohen 1988), was used to identify contract-level differences that were at least "small" in magnitude.

RESULTS

Nearly 10 million Medicare beneficiaries living in California, Florida, or New York, who met eligibility criteria for at least one performance measure, were included in the analysis. MA enrollees were more likely to be under age 80, black, and Hispanic, and were far less likely than FFS beneficiaries to be dually eligible for Medicare and Medicaid (Table 1). Overall, MA and FFS beneficiaries lived in communities that had similar socioeconomic profiles according to six zip-code-level measures (e.g., median household income, receipt of public assistance) and a composite measure of socioeconomic status. Although our analysis was limited to three states, the sample includes performance data from 79 unique MA organizations that entered into 114 contracts with CMS to provide services to MA enrollees (Table 2).

Medicare Advantage outperformed FFS on each of the 16 clinical quality measures we examined, although the magnitude of the difference varied by type of measure. Among the 10 HEDIS measures, MA outperformed FFS by as little as 2.3 percentage points (Plan All-Cause Readmissions) to as much as 41.9 percentage points (Colorectal Cancer Screening). Differences were large for HEDIS measures that were reported using administrative data as well as measures that were eligible for reporting using the hybrid method. While MA also outperformed FFS on all five Part D measures, the differences were much smaller and none exceeded 3.3 percentage points. Performance on patient experience measures was somewhat mixed with MA outperforming FFS on four measures (Getting Appointments and Care Quickly, Rating of Health Care Quality, Rating of Drug Plan, and Getting Needed Prescription Drugs) and FFS outperforming MA on one measure (Getting Needed Care). We found no statistically significant differences on the Care Coordination measure. Case mix adjusting clinical quality measures did not substantively change the observed patterns (Table 3, Panel 2). Case mix adjustment helped to improve the performance of FFS by several percentage points on most HEDIS measures, but it had a negligible effect in closing the gap between MA and FFS on Part D measures.

We found significant differences in performance by plan type—with MA-HMO plans outperforming MA-PPO plans on nearly every measure (Table 4). While MA-PPOs tended to perform better than FFS, this was not universally true. For example, FFS had higher performance on Glaucoma Screening and two of the three medication adherence measures. Although we had much less power to detect differences between MA-PPOs and FFS on

Table 1:   Characteristics of Beneficiaries Included in MA/FFS Comparisons

	MA Beneficiaries (N = 3,571,743)	FFS Beneficiaries (N = 6,352,239)
Beneficiary characteristics		
Age: 64 or under, n (%)	409,209 (11.5)	738,801 (11.6)
65–69	951,576 (26.6)	1,409,475 (22.2)
70–74	760,544 (21.3)	1,315,099 (20.7)
75–79	598,284 (16.8)	1,065,546 (16.8)
80–84	454,212 (12.7)	886,896 (14.0)
85 and older	397,918 (11.1)	936,422 (14.7)
Female, n (%)	2,124,275 (59.5)	3,751,061 (59.1)
Race/ethnicity: Asian/Pacific Islander, mean % (SD %)*	7.3 (24.2)	6.4 (22.1)
Black	9.5 (27.9)	7.1 (23.2)
Hispanic	17.4 (33.9)	11.1 (26.7)
Multiracial/American Indian/Alaska Native	0.2 (1.5)	0.2 (1.9)
White	65.5 (43.8)	75.3 (38.2)
Medicaid dual eligibility, n (%)	675,056 (18.9)	1,973,707 (31.1)
Part D low-income subsidy eligibility, n (%)	109,212 (3.1)	118,875 (1.9)
Hierarchical Condition Category Score, mean (SD)	1.08 (0.83)	1.11 (0.89)
State of residence: California, n (%)	1,959,820 (54.9)	2,683,793 (42.2)
Florida	913,749 (25.6)	1,959,147 (30.8)
New York	698,174 (19.5)	1,709,299 (26.9)
Medicare Advantage Contract Type: HMO, n (%)	3,153,211 (88.3)	–
PPO	365,395 (10.2)	–
Other	53,137 (1.5)	–
Zip-code-level characteristics, mean (SD)		
Socioeconomic Status index	−0.11 (0.98)	−0.02 (0.99)
Percent of individuals aged 25 + with less than HS diploma	16.8 (11.1)	15.3 (10.9)
Percent of males aged 16 + unemployed	11.2 (4.2)	10.9 (4.5)
Percent of individuals with annual income below poverty	14.5 (8.3)	14.3 (8.5)
Percent of households with public assistance income	3.2 (2.5)	2.9 (2.6)
Percent female headed households with children	10.9 (6)	10.2 (5.9)
Median annual household income (in $1,000)	60.1 (22.6)	60.7 (24.7)

Notes: MA enrollees are included in this table if they were enrolled in an MA contract that operated exclusively within the states of California, Florida, or New York in 2012. FFS beneficiaries were included if they lived in any of the three states during 2012. All beneficiaries were eligible for the denominator of at least one of the 16 clinical quality measures or six patient experience measures included in the analysis.

*Beneficiaries' race/ethnicities are derived using an indirect estimation methodology as described in Martino et al. (2013).

HMO, health maintenance organization; HS, high school; PPO, preferred provider organization; SD, standard deviation.

Table 2:   MA and FFS Sample Sizes by State

	California	*Florida*	*New York*
Number of MA organizations	29	21	29
Number of MA contracts	37	36	41
Number of MA-HMOs	29	25	26
Number of MA-PPOs	1	6	4
Number of MA Other	7	5	11
Number of MA beneficiaries	1,959,820	913,749	698,174
Number of FFS beneficiaries	2,683,793	1,959,147	1,709,299

Notes: An MA organization is a managed care organization that enters into a prepaid contract with the Centers for Medicare & Medicaid Services (CMS) to provide services for Medicare beneficiaries. Quality measures for the MA Star Ratings program are reported at the contract level.

patient experience measures, the same general pattern emerged in which MA-PPO performance more closely resembled that of FFS than that of MA-HMOs.

When examining contract-level results, we observed a few notable differences. While MA contracts were far more likely to outperform their FFS comparison groups on the 10 HEDIS measures than vice versa, on Part D measures, MA contracts outperformed FFS on only two of six measures (High-Risk Medication and Diabetes Treatment) (Table 5, Panel 1). On all three medication adherence measures, MA plans were equally likely to outperform FFS as vice versa. Adjusting the clinical quality measures or applying county-level weights to better match MA and FFS beneficiary samples had little to no effect on these results.

Beneficiaries enrolled in MA and FFS reported similar patient experience in contract-level comparisons; however, where the largest differences did exist, they were more likely to favor FFS than vice versa (Table 5, Panel 2). For example, 44.6 percent of MA contracts had lower scores than FFS on Getting Needed Care, whereas only 1.2 percent of contracts outperformed FFS. Similarly, while nearly three-quarters of MA contracts either had no statistically significant differences or differences that did not meet our 0.2 standard deviation threshold on the Care Coordination measure, nearly all other contracts had lower performance than FFS.

In sensitivity analyses, we used a higher threshold to identify contracts with "large" differences in performance relative to FFS (at least 0.8 SD; Cohen 1988). The MA advantage on HEDIS administrative-only measures narrowed considerably, but MA was still far more likely to outperform FFS than vice versa (Table S2). For Part D measures, we found far fewer MA contracts

Table 3: Overall Differences in Performance between MA and FFS on Clinical Quality and Patient Experience Measures

Measure	N MA	N FFS	Unadjusted Analysis*			Adjusted Analysis†		
			MA Mean	FFS Mean	Difference (MA-FFS)	MA Mean	FFS Mean	Difference (MA-FFS)
Clinical quality								
HEDIS administrative data-only measures								
Breast cancer screening	358,598	687,038	80.2	58.8	21.3	77.9	59.8	18.1
Glaucoma screening	1,872,719	3,997,698	75.4	66.2	9.2	74.2	65.8	8.4
Osteoporosis management for fractures	33,163	44,879	39.8	12.5	27.3	37.1	13.2	23.9
Rheumatoid arthritis management	32,887	47,115	76.9	65.3	11.6	75.7	67.0	8.7
Plan all-cause readmissions	327,895	966,808	–	–	–	10.1	12.4	−2.3
HEDIS hybrid measures								
Colorectal cancer screening	259,301	2,492,106	86.1	44.2	41.9	81.7	46.3	35.4
Cholesterol management LDL-C screening	63,579	318,824	96.0	86.9	9.2	95.0	87.0	8.0
Diabetic eye exam performed	94,932	783,364	76.4	54.1	22.3	73.0	55.6	17.4
Diabetic cholesterol screening	28,658	783,364	90.6	82.4	8.2	90.1	82.7	7.4
Diabetes care: Nephropathy care	99,441	516,620	94.3	81.0	13.3	93.1	82.2	10.9
MCAHPS measure								
Annual flu vaccine	32,137	28,100	71.4	67.9	3.5	–	–	–
Part D measures								
High-risk medication	3,053,440	3,164,305	4.2	6.4	−2.2	–	–	–
Diabetes treatment	645,341	740,612	86.0	82.7	3.3	85.7	82.3	3.4
Medication adherence for diabetes medications	449,830	498,695	77.2	75.1	2.1	77.3	74.6	2.6
Medication adherence for hypertension	1,360,312	1,568,631	79.0	75.9	3.0	78.8	76.0	2.8
Medication adherence for cholesterol	1,415,438	1,593,436	73.8	71.5	2.3	73.9	71.2	2.6

Continued

Table 3. *Continued*

Measure	N MA	N FFS	Unadjusted Analysis*			Adjusted Analysis†		
			MA Mean	FFS Mean	Difference (MA-FFS)	MA Mean	FFS Mean	Difference (MA-FFS)
Patient experience								
Getting needed care	24,000	15,768	—	—	—	83.9	85.7	−1.8
Getting appointments and care quickly	29,503	18,517	—	—	—	71.4	69.1	2.2
Care coordination	32,324	19,676	—	—	—	84.0	84.5	−0.5
Rating of health care quality	28,210	17,320	—	—	—	85.5	84.3	1.3
Rating of drug plan	29,991	8,606	—	—	—	86.7	81.8	4.9
Getting needed prescription drugs	29,212	8,640	—	—	—	91.4	87.3	4.2

Notes: All MA/FFS differences are statistically significant ($p < .001$) with the exception of the Care Coordination measure ($p = .11$). Lower scores are better for Plan All-Cause Readmissions and High-risk Medication measures. The hybrid method uses both administrative data and medical record review to measure performance.

*The Plan All-Cause Readmissions measure and all patient experience measures require case mix adjustment.

†All clinical quality measures are adjusted for age (18–64, 65–69, 70–75, 76–79, 80–84, ≥85), gender, dual eligibility, Part D low-income subsidy eligibility, race/ethnicity, neighborhood socioeconomic status (SES) index, and hierarchical condition category (HCC) score. The HCC score is an index of predicted spending based on a beneficiary's sociodemographic characteristics and selected diagnoses measured from claims and represents a summary measure of comorbidity. To measure neighborhood SES, we used a six-item scale of neighborhood characteristics measured at the zip-code-level from the 2008 to 2012 American Community Survey. The six items included household income, poverty, receipt of public assistance, unemployment, household structure, and educational attainment. All case mix adjustment models included propensity scores. Propensity scores are estimates of the probability of enrollment in MA as a function of the case mix adjustment variables. Case mix adjustment for the High-risk Medication measure is considered inappropriate because performance is dependent entirely on a physician's prescribing behavior. Plan All-Cause Readmissions were adjusted according to HEDIS specifications. All patient experience measures are case mix adjusted using the standard CAHPS case mix adjustment methodology, which includes adjustment for age (18–64, 65–69, 70–74, 75–79, 80–84, ≥85), education (eighth grade or less, some high school, high school, less than bachelor's degree, bachelor's degree, postbachelor's degree), general health (excellent, very good, good, fair, poor), mental health (excellent, very good, good, fair, poor), dual eligibility, Part D low-income subsidy eligibility, and use of a proxy in responding to the survey.

Table 4: Differences in Performance between MA and FFS by MA Plan Type

Measure	MA-HMO Mean	MA-PPO Mean	FFS Mean	MA-HMO – FFS Difference Estimate	p Value	MA-PPO – FFS Difference Estimate	p Value	MA-HMO – MA-PPO Difference Estimate	p Value
Clinical quality									
HEDIS administrative data-only measures									
Breast cancer screening	82.4	65.2	58.6	23.8	<.001	6.5	<.001	17.3	<.001
Glaucoma screening	76.8	62.7	66.1	10.7	<.001	–3.3	<.001	14.0	<.001
Osteoporosis management for fractures	42.1	19.1	12.3	29.8	<.001	6.9	<.001	23.0	<.001
Rheumatoid arthritis management	77.6	73.7	65.3	12.3	<.001	8.4	<.001	3.9	<.001
Plan all-cause readmissions	9.9	11.0	12.4	–2.5	<.001	–1.4	<.001	–1.1	<.001
Hedis hybrid measures									
Colorectal cancer screening	86.5	60.4	44.1	42.4	<.001	16.2	<.001	26.2	<.001
Cholesterol management LDL–C screening	96.3	89.1	86.8	9.5	<.001	2.3	<.001	7.2	<.001
Diabetic eye exam performed	77.2	61.8	54.1	23.1	<.001	7.7	<.001	15.4	<.001
Diabetic cholesterol screening	91.0	87.9	82.4	8.6	<.001	5.6	<.001	3.0	<.001
Diabetes care: Nephropathy care	94.6	87.3	80.9	13.7	<.001	6.4	<.001	7.3	<.001
MCAHPS measure									
Annual flu vaccine	72.3	64.4	67.9	4.4	<.001	–3.5	.09	7.8	<.001
Part D measures									
High-risk medication	4.0	5.1	6.4	–2.4	<.001	–1.3	<.001	–1.1	<.001
Diabetes treatment	86.2	84.8	82.7	3.5	<.001	2.1	<.001	1.4	<.001
Medication adherence for diabetes medications	77.6	73.5	75.0	2.5	<.001	–1.5	<.001	4.1	<.001
Medication adherence for hypertension	79.3	76.2	75.9	3.4	<.001	0.3	.03	3.1	<.001
Medication adherence for cholesterol	74.3	70.2	71.5	2.8	<.001	–1.3	<.001	4.1	<.001

Table 4. *Continued*

Measure	MA-HMO Mean	MA-PPO Mean	FFS Mean	MA-HMO – FFS Difference		MA-PPO – FFS Difference		MA-HMO – MA-PPO Difference	
				Estimate	p Value	Estimate	p Value	Estimate	p Value
Patient experience									
Getting needed care	83.8	84.4	85.7	−1.8	<.001	−1.3	.19	−0.6	.59
Getting appointments and care quickly	71.5	70.9	69.1	2.3	<.001	1.8	.07	0.6	.57
Care coordination	84.0	84.3	84.5	−0.6	.09	−0.2	.83	−0.4	.71
Rating of health care quality	85.6	84.8	84.2	1.3	<.001	0.6	.44	0.8	.31
Rating of drug plan	87.4	80.9	81.6	5.8	<.001	−0.7	.48	6.5	<.001
Getting needed prescription drugs	91.8	88.3	87.2	4.6	<.001	1.1	.35	3.5	.004

Notes: No clinical quality measure is case mix adjusted with the exception of Plan All-Cause Readmissions (according to HEDIS specifications). All patient experience measures are case mix adjusted using the standard CAHPS case mix adjustment methodology (see Table 3 notes).

Table 5:   Contract-Level Differences in Performance between MA and FFS

	Without Case Mix Adjustment or Geographic Weighting			With Case Mix Adjustment			With Geographic Weighting[†]		
	Better Than FFS	No Difference	Worse Than FFS	Better Than FFS	No Difference	Worse Than FFS	Better Than FFS	No Difference	Worse Than FFS
Measure	% of Contracts			% of Contracts			% of Contracts		
Clinical quality*									
HEDIS administrative data-only measures									
Breast cancer screening	89.9	7.2	2.9	91.3	5.8	2.9	89.9	7.2	2.9
Glaucoma screening	62.5	13.9	23.6	68.1	13.9	18.1	63.9	13.9	22.2
Osteoporosis management for fractures	77.8	20.4	1.9	66.7	33.3	0.0	77.8	20.4	1.9
Rheumatoid arthritis management	49.2	42.4	8.5	40.7	49.2	10.2	49.2	42.4	8.5
Plan all-cause readmissions	–	–	–	53.5	45.1	1.4	53.5	45.1	1.4
HEDIS hybrid measures									
Colorectal cancer screening	90.3	5.6	4.2	90.3	6.9	2.8	90.3	5.6	4.2

Table 5. Continued

Measure	Without Case Mix Adjustment or Geographic Weighting			With Case Mix Adjustment			With Geographic Weighting[†]		
	Better Than FFS	No Difference	Worse Than FFS	Better Than FFS	No Difference	Worse Than FFS	Better Than FFS	No Difference	Worse Than FFS
	% of Contracts			% of Contracts			% of Contracts		
Cholesterol management LDL-C screening	57.6	40.9	1.5	53.0	47.0	0.0	57.6	40.9	1.5
Diabetic eye exam performed	79.5	19.2	1.4	76.7	20.5	2.7	79.5	19.2	1.4
Diabetic cholesterol screening	80.8	17.8	1.4	79.5	20.5	0.0	80.8	17.8	1.4
Diabetes care: Nephropathy care	91.8	8.2	0.0	86.3	13.7	0.0	91.8	8.2	0.0
MCAHPS measure									
Annual influenza vaccination	19.3	62.7	18.1	–	–	–	19.3	62.7	18.1
Part D measures									
High-risk medication	52.7	36.4	10.9	–	–	–	54.5	33.6	11.8
Diabetes treatment	64.2	29.2	6.6	64.2	33.0	2.8	65.1	28.3	6.6
Medication adherence for diabetes medications	22.3	54.4	23.3	31.1	58.3	10.7	23.3	54.4	22.3
Medication adherence for hypertension	34.9	35.8	29.4	40.4	45.0	14.7	36.7	33.9	29.4

Table 5. *Continued*

Measure	Without Case Mix Adjustment or Geographic Weighting			With Case Mix Adjustment			With Geographic Weighting[†]		
	Better Than FFS	No Difference	Worse Than FFS	Better Than FFS	No Difference	Worse Than FFS	Better Than FFS	No Difference	Worse Than FFS
	% of Contracts			*% of Contracts*			*% of Contracts*		
Medication adherence for cholesterol	31.5	37.0	31.5	34.3	50.0	15.7	33.3	35.2	31.5
Patient experience									
Getting needed care	—	—	—	1.2	54.2	44.6	1.2	54.2	44.6
Getting appointments and care quickly	—	—	—	18.1	69.9	12.0	18.1	69.9	12.0
Care coordination	—	—	—	2.4	72.3	25.3	2.4	72.3	25.3
Rating of health care quality	—	—	—	15.7	75.9	8.4	15.7	75.9	8.4
Rating of drug plan	—	—	—	16.3	81.3	2.5	20.0	80.0	0.0
Getting needed prescription drugs	—	—	—	18.8	77.5	3.8	18.8	77.5	3.8

Notes: All differences reported in the table met two criteria: (1) $p < .05$ for a significance test of the null hypothesis that MA = FFS, and (2) the MA/FFS difference (measured in effect sizes) ≥ 0.2 or ≤ -0.2.

*See Table 3 notes for case mix adjustment methods.

[†]County-level geographic weights were applied so that the FFS comparison population matched the geographic distribution of the MA population enrolled in each contract. No measures in this panel are case mix adjusted with the exception of Plan All-Cause Readmissions and the six patient experience measures.

with "large" differences; however, for patient experience measures, we found roughly the same proportion of contracts either outperforming or underperforming FFS as when using the lower threshold.

DISCUSSION

Medicare Advantage contracts operating within three large, diverse states provided substantially higher quality of care than FFS for all 16 clinical quality measures we examined. The differences were largest on HEDIS measures, and much smaller on measures of prescription drug prescribing and adherence. Differences in performance were mixed on patient experience measures where MA performance was higher on four measures, FFS outperformed MA on one measure (Getting Needed Care), and no differences were found on the Care Coordination measure. While case mix adjustment helped improve scores for FFS to some extent, it did not substantially narrow the gap in performance on clinical quality measures. Differences in performance were consistently larger for MA-HMOs compared with MA-PPOs, and for a small number of measures FFS outperformed MA-PPOs. Finally, contract-level comparisons indicate that, within individual service areas, MA and FFS performances are far more balanced on medication adherence measures and patient experience measures, but differences remained large on all HEDIS measures.

The magnitude of the performance differences on HEDIS measures is consistent with findings from a prior analysis using 2009 performance data (Ayanian et al. 2013). Among measures common to both analyses, we found that MA outperformed FFS on breast cancer screening rates (21.3 percentage point difference vs. 13.5 percentage points in the prior analysis), diabetic eye examinations (22.3 percentage points vs. 17.1 percentage points), diabetic cholesterol tests (8.2 percentage points vs. 9 percentage points), and cholesterol screening for patients with cardiac conditions (9.2 percentage points vs. 7 percentage points). If MA/FFS performance differences in these three states are generalizable nationally, then these results suggest that quality differences could be widening over time as a result of recent incentive programs to stimulate quality improvement within Medicare that, as of 2012, were more likely to affect MA than FFS. However, smaller MA/FFS differences in our contract-level analysis indicate that these overall results may be driven by a small number of high-performing contracts.

The performance differences estimated in the current analysis are not entirely comparable with those of prior studies because of differences in data

sources and methods. First, while previous studies used 20 percent samples of FFS beneficiaries or 100 percent samples of beneficiaries attributed to group practices, we used a three-state sample, which allowed us to generate FFS comparison groups that comprised 100 percent of FFS beneficiaries living in each contract's service area. Unlike some prior analyses, our analyses used patient-level data, which allowed us to examine the role of case mix differences in explaining observed differences. Moreover, our analysis is the first to incorporate Medicare Part D claims data, which is needed to accurately define measure numerators and denominators for a growing number of quality measures, including 8 of the 16 measures used in the current analysis. Finally, we included 100 percent of beneficiaries enrolled in MA or FFS—including those under age 65—who were excluded in some prior analyses. Although we only used a three-state sample, each of these methodological enhancements allows a more rigorous examination of differences in performance between the two programs.

The differences in performance we observed may be due to a number of factors. First, the two programs faced widely different financial incentives in 2012—the year when the earliest wave of Medicare Shared Savings Program Accountable Care Organizations were launched in the FFS program. Under the MA QBP program, contracts that achieve high levels of performance according to the Star Rating system are eligible for incentive payments—an additional $26 per beneficiary per month on average in payment year 2012 (L&M Policy Research LLC 2016). In addition, MA's capitation-based payment system also provides strong incentives for contracts to avoid costly specialty and acute care. As a result, MA contracts may contract more selectively with providers to ensure high-quality networks and may offer additional financial incentives to their network physicians to encourage high levels of performance. MA contracts may also have dedicated staff to provide outreach and encouragement to enrollees to seek preventive care. While incentive programs were available to FFS providers in 2012, such as the Physician Quality Reporting System and Meaningful Use program, participation rates vary across states and the financial incentives tend to be small relative to the MA QBPs (Centers for Medicare & Medicaid Services, 2016; L&M Policy Research LLC 2016). Differences in MA/FFS performance on Part D measures may result from FFS physicians having less direct access to beneficiaries' pharmacy claims information relative to MA contracts, which may make it more difficult to achieve higher levels of coordination and performance on Part D measures.

Differences in data sources or measurement error may also explain some of our findings—particularly for HEDIS measures reported using the hybrid method, where FFS performance is likely to be underestimated because medical record reviews were not feasible. Despite this bias, these estimates can still help to assess whether performance gaps are growing or narrowing over time. In addition, unlike administrative data systems used by many commercial plans, Medicare FFS claims do not include laboratory values, which may lower performance on measures such as Medical Attention for Nephropathy (for which a positive urine macroalbumin test is one way to achieve a "success" on the measure). Other explanations may include differences in the interpretation of measure specifications (such as differences in the restrictiveness of provider specialty criteria when identifying valid numerator events for some HEDIS measures) or inaccurate definition of the denominator population. Both issues could be assessed through audits of data submitted by MA contracts.

Although we examined the impact of case mix adjustment on our results, systematic differences in the characteristics of beneficiaries enrolled in MA and FFS may yet explain some of the observed differences in performance. In particular, we lack patient-level information on the wide range of clinical, behavioral, environmental, and other factors that may affect performance and may differ between the two programs. In addition, Medicare enrollment files do not contain information on beneficiaries' supplemental insurance coverage (e.g., Medigap or employer-sponsored coverage) that covers beneficiary cost sharing. Beneficiaries who lack supplemental coverage— nearly 14 percent of all Medicare beneficiaries in 2010 (Cubanski et al. 2015) —may have a lower propensity to seek care, which could result in lower estimates of performance for FFS (Capps and Dranove 2011). Although we adjusted for a beneficiary's socioeconomic status using three measures (i.e., dual eligibility status, Part D low-income subsidy eligibility, and neighborhood SES), they may not be an adequate proxy for a beneficiary's cost-sharing burden. Nevertheless, unmeasured case mix differences would have to be substantial to account for the large gaps in performance between MA and FFS, suggesting that they are unlikely to be the only factor contributing to the observed differences.

The differences in performance we observed between MA-HMOs and MA-PPOs may be due to differences in care management or care coordination practices between the two types of plans. To the extent that MA-HMO provider networks are narrower in scope, include more tightly integrated groups of providers, and have shared electronic health records or

greater health information exchange, MA-HMOs may be able to provide more effective population management. In addition, MA-HMOs may be more likely than MA-PPOs to contract with physicians using risk-based contracts and thus provide stronger incentives for achieving high levels of quality. The differences between MA-HMOs and MA-PPOs were large and deserve further analysis—especially as MA-PPO enrollment grows over time.

Our study had a number of limitations. First, although our analyses covered nearly 25 percent of the MA enrollee population in 2012, our analysis was limited to contracts operating within three large states because our FFS data files were limited to these states. Our findings may not be representative of the performance of the MA program overall if contracts operating in these states provide higher or lower quality of care than the MA average. Second, we conducted comparisons on only two dimensions of performance—clinical quality and patient experience—and did not assess differences in spending or health outcome, which may be needed to provide a comprehensive summary of differences in performance between the two programs. Among these, only health outcomes are currently included in MA Star Ratings (using the Medicare Health Outcomes Survey) but are not systematically collected for the FFS population. Third, FFS performance may be underestimated on all HEDIS measures eligible for the "hybrid" reporting method; however, reviewing medical records for the FFS beneficiary cohort was not feasible. Fourth, CMS administrative data do not currently include information on social risk factors, such as income, employment, or other factors that might differ between MA and FFS. While we adjusted for race/ethnicity, dual eligibility, Part D low-income subsidy eligibility, and neighborhood measures of socioeconomic status, these measures are unlikely to fully capture the set of social risk factors that might contribute to observed differences in performance. Finally, we used zip-code-level measures of neighborhood SES to adjust performance estimates, which do not account for differences in SES within zip codes.

Given the large differences in performance, we observed on many measures, systematic and ongoing monitoring of these differences at a national level should remain a high priority. As FFS physicians expand their participation in advanced alternative payment models spurred by incentives included in CMS's Quality Payment Program (Centers for Medicare & Medicaid Services 2017b), current differences in performance between MA and FFS could attenuate, while improving the quality of care to all Medicare beneficiaries.

CONCLUSION

Using recent data from three large states, we found that MA outperformed FFS on nearly all clinical quality and most patient experience measures. Performance differences were much larger between MA-HMOs and FFS relative to the differences between MA-PPOs and FFS across most measures and were generally smaller in contract-level comparisons with the exception of HEDIS measures. These findings are generally consistent with prior studies that found higher performance in MA for most HEDIS measures. Identifying factors explaining lower performance on measures of clinical quality in the FFS system—including the potential role of bias from unmeasured case mix differences—should be a high priority. Repeating these analyses on an ongoing basis with the full set of MA contracts could improve the generalizability of these analyses while helping to better monitor program-wide trends as value-based purchasing expands within the Medicare program and beyond.

ACKNOWLEDGMENTS

REFERENCES

Ayanian, J. Z., B. E. Landon, A. M. Zaslavsky, R. C. Saunders, L. G. Pawlson, and J. P. Newhouse. 2013. "Medicare Beneficiaries More Likely to Receive Appropriate Ambulatory Services in HMOs Than in Traditional Medicare." *Health Affairs (Millwood)* 32 (7): 1228–35.

Biles, B., G. Casillas, and S. Guterman. 2015. "Variations in County-Level Costs Between Traditional Medicare and Medicare Advantage Have Implications for Premium Support." *Health Affairs (Millwood)* 34 (1): 56–63.

Bird, C. E., T. Seeman, J. J. Escarce, R. Basurto-Davila, B. K. Finch, T. Dubowitz, M. Heron, L. Hale, S. S. Merkin, M. Weden, and N. Lurie. 2010. "Neighbourhood Socioeconomic Status and Biological 'Wear and Tear' in a Nationally Representative Sample of US Adults." *Journal of Epidemiology and Community Health* 64 (10): 860–5.

Brennan, N., and M. Shepard. 2010. "Comparing Quality of Care in the Medicare Program." *American Journal of Managed Care* 16 (11): 841–8.

Capps, C., and D. Dranove. 2011. "Intended and Unintended Consequences of a Prohibition on Medigap First-Dollar Benefits" [accessed on February 6, 2017]. Available at http://www.protectmedigap.org/pdf/Capps and Dranove - Medigap cost sharing.pdf

Centers for Medicare & Medicaid Services. 2016. "2014 Reporting Experience Including Trends (2007-2015): Physician Quality Reporting System" [accessed on May 6, 2017]. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/2014_PQRS_Experience_Rpt.pdf

Centers for Medicare & Medicaid Services. 2017a. "Part C and D Performance Data" [accessed on May 8, 2017]. Available at https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html

Centers for Medicare & Medicaid Services. 2017b. "Quality Payment Program" [accessed on May 8, 2017]. Available at https://qpp.cms.gov/

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edition. Hillsdale, NJ: Lawrence Erlbaum.

Cubanski, J., C. Swoope, C. Boccuti, G. Jacobson, G. Casillas, S. Griffin, and T. Neuman. 2015. "A Primer on Medicare: Key Facts about the Medicare Program and the People It Covers" [accessed on February 6, 2017]. Available at http://kff.org/report-section/a-primer-on-medicare-what-types-of-supplemental-insurance-do-beneficiaries-have/

Elliott, M. N., A. M. Haviland, N. Orr, K. Hambarsoomian, and P. D. Cleary. 2011. "How do the Experiences of Medicare Beneficiary Subgroups Differ between Managed Care and Original Medicare?" *Health Services Research* 46 (4): 1039–58.

Elliott, M. N., B. E. Landon, A. M. Zaslavsky, C. Edwards, N. Orr, M. K. Beckett, J. Mallett, and P. D. Cleary. 2016. "Medicare Prescription Drug Plan Enrollees Report Less Positive Experiences Than Their Medicare Advantage Counterparts." *Health Affairs (Millwood)* 35 (3): 456–63.

Gold, M., and G. Casillas. 2014. "What Do We Know about Health Care Access and Quality in Medicare Advantage Versus the Traditional Medicare Program?" [accessed on January 16, 2017]. Available at http://kff.org/medicare/report/what-do-we-know-about-health-care-access-and-quality-in-medicare-advantage-versus-the-traditional-medicare-program/

Jacobson, G., A. Damico, T. Neuman, and M. Gold. 2017. "Medicare Advantage 2017 Spotlight: Enrollment Market Update" [accessed on August 20, 2017]. Available at http://www.kff.org/medicare/issue-brief/medicare-advantage-2017-spotlight-enrollment-market-update/

Keenan, P. S., M. N. Elliott, P. D. Cleary, A. M. Zaslavsky, and B. E. Landon. 2009. "Quality Assessments by Sick and Healthy Beneficiaries in Traditional Medicare and Medicare Managed Care." *Medical Care* 47 (8): 882–8.

L&M Policy Research LLC. 2016. "Evaluation of the Medicare Quality Bonus Payment Demonstration" [accessed on January 16, 2017]. Available at https://innovation.cms.gov/Files/reports/maqbpdemonstration-finalevalrpt.pdf

Landon, B. E., A. M. Zaslavsky, S. L. Bernard, M. J. Cioffi, and P. D. Cleary. 2004. "Comparison of Performance of Traditional Medicare vs. Medicare Managed Care." *Journal of the American Medical Association* 291 (14): 1744–52.

Landon, B. E., A. M. Zaslavsky, R. C. Saunders, L. G. Pawlson, J. P. Newhouse, and J. Z. Ayanian. 2012. "Analysis of Medicare Advantage HMOs Compared with Traditional Medicare Shows Lower Use of Many Services during 2003-09." *Health Affairs (Millwood)* 31 (12): 2609–17.

Martino, S. C., R. M. Weinick, D. E. Kanouse, J. A. Brown, A. M. Haviland, E. Goldstein, J. L. Adams, K. Hambarsoomian, D. J. Klein, and M. N. Elliott. 2013. "Reporting CAHPS and HEDIS Data by Race/Ethnicity for Medicare Beneficiaries." *Health Services Research* 48 (2 Pt 1): 417–34.

Mittler, J. N., B. E. Landon, E. S. Fisher, P. D. Cleary, and A. M. Zaslavsky. 2010. "Market Variations in Intensity of Medicare Service Use and Beneficiary Experiences With Care." *Health Services Research* 45 (3): 647–69.

SUPPORTING INFORMATION

Additional supporting information may be found online in the supporting information tab for this article:

Appendix SA1: Author Matrix.

Table S1: Description of Performance Measures Used in the Analysis.

Table S2: Frequency of "Large" Differences in Performance between MA and FFS in Unadjusted Analyses.

Table S3a: Regression Coefficients and Standard Errors for Case-Mix Adjusted Analyses: HEDIS Administrative Data-Only Measures.

Table S3b: Regression Coefficients and Standard Errors for Case-Mix Adjusted Analyses: HEDIS Hybrid Measures.

Table S3c: Regression Coefficients and Their Standard Errors for Case-Mix Adjusted Analyses: Part D Measures.

Table S3d: Regression Coefficients and Their Standard Errors for Case-Mix Adjusted Analyses: MCAHPS.