

Comparing Quality of Care in the Medicare Program

Niall Brennan, MPP; and Mark Shepard, BA

Despite a growing focus on measuring and reporting quality of care in Medicare to allow beneficiaries to make informed choices of providers and plans, little published information compares quality of care in traditional fee-for-service (FFS) Medicare and Medicare's private insurance option, Medicare Advantage (MA). By contrast, substantial resources exist for comparing quality among MA plans, which are presented prominently on Medicare's Web site in the same area used by beneficiaries to select a plan.¹ Due in part to this lack of data, debate among policy makers on the relative merits of MA and FFS has focused on payment rates rather than quality of care.²⁻⁴

Efforts to compare quality between MA and FFS have become a policy priority. After recommending MA-FFS comparisons for many years, the Medicare Payment Advisory Commission (MedPAC) issued detailed recommendations in March 2010 on methods for carrying out these comparisons, including an approach similar to the one we take.⁵ The Medicare Improvements for Patients and Providers Act of 2008 specifies that MA-FFS comparisons begin by March 2011, underlining the importance of pursuing currently feasible strategies.

We analyzed data on quality in FFS and MA programs during 2006-2007 using 11 measures of underuse of effective care from the Healthcare Effectiveness Data and Information Set (HEDIS). The HEDIS measures are reported annually for MA and commercial plans and form the basis of nationally recognized commercial plan rankings⁶ and quality ratings used to inform Medicare beneficiaries.¹ By contrast, HEDIS measures have not previously been available for the FFS population but were calculated for 2006-2007 for a special project by the Centers for Medicare & Medicaid Services (CMS). These data allow for one of the first national comparisons of MA and FFS on evidence-based clinical quality measures. Previous work compared MA and FFS on patient satisfaction and quality using the Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey, but measures were based on beneficiary recollection of receipt of recommended care like flu shots.^{7,8} Our administrative HEDIS measures complement the CAHPS comparison and allow for comparisons of rarer conditions like depression. Past work has also compared MA and FFS quality at the state and regional level, generally finding

higher quality care in MA managed care plans.⁹

Several issues complicated the comparison for certain measures, including variations in measure con-

Objective: To compare the clinical quality of care between Medicare fee-for-service (FFS) and Medicare Advantage (MA) programs.

Methods: We compared 11 Healthcare Effectiveness Data and Information Set (HEDIS) quality measures nationwide for MA managed care plans and the FFS program in 2006 and 2007. We adjusted FFS measures to match the geographic distribution of MA.

Results: Medicare Advantage plans scored substantially better (4-16 percentage points; median, 7.8 percentage points) on 8 measures, slightly better (1.5 percentage points) on 1 measure, and worse than FFS (2-5 percentage points; median, 4.1 percentage points) on 2 measures. The 8 measures on which MA scored substantially better were well established in the HEDIS measure set (introduced in the 1990s), whereas the other 3 were all newer (introduced in 2004-2005 data). Data and program differences complicated the comparison, but it is unlikely that they were large enough to explain the sizable MA-FFS gaps observed.

Conclusions: Quality measures showed large, though mixed, differences between MA and FFS. The dichotomy between older and newer measures in MA suggests a learning effect, with plans improving measurement and quality over time as measures become more familiar.

(*Am J Manag Care.* 2010;16(11):841-848)

In this issue
Take-Away Points / p842
www.ajmc.com
Full text and PDF

For author information and disclosures,
see end of text.

Take-Away Points

This study compared quality in traditional fee-for-service (FFS) Medicare and Medicare Advantage (MA) programs for 2006-2007.

- Relative performance on 11 clinical quality measures showed notable differences between FFS and MA, with neither program performing better on all measures.
- MA-FFS quality comparisons should be used to inform policy makers who set program provisions and beneficiaries choosing between traditional FFS Medicare and an MA plan.

Medicare Advantage Sample

We compared these FFS data with concurrent HEDIS measures publicly reported by MA plans and audited by the National Committee for Quality Assurance (NCQA).¹⁴ Because private FFS plans were exempt from quality reporting requirements at that time, we excluded them and

limited our analysis to managed care plans including HMOs, point-of-service plans, and preferred provider organizations (PPOs). In addition, we excluded MA plans centered outside of the 50 states plus the District of Columbia. The final data included plans with total enrollment of approximately 6.0 million in 2006 and 6.5 million in 2007.

struction within the HEDIS framework, data limitations, and underlying program differences between MA and FFS. However, we argue that these data represent a valuable first step that shows how Medicare can better use existing resources to monitor FFS quality and inform beneficiaries who are choosing between MA and FFS. We also suggest ways in which future efforts could improve upon this comparison.

limited our analysis to managed care plans including HMOs, point-of-service plans, and preferred provider organizations (PPOs). In addition, we excluded MA plans centered outside of the 50 states plus the District of Columbia. The final data included plans with total enrollment of approximately 6.0 million in 2006 and 6.5 million in 2007.

Quality Measures

All of the data are process measure rates defined according to HEDIS specifications. These were constructed by using claims data to identify the subset of enrollees (called the denominator or eligible population) for whom a treatment or screening was clinically recommended. The measure rate was the fraction of this denominator population who received the recommended care in accordance with the measure definition.¹⁵ We studied 11 of the 12 HEDIS measures analyzed by the GEM project (Table 1), excluding only colon cancer screening because of an insufficient look-back period. HEDIS specifications allow a colonoscopy to have been performed in the past 9 years and a flexible sigmoidoscopy or double contrast barium enema to have been performed in the past 4 years. But the GEM study only analyzed Medicare claims over a 3-year period from 2005-2007.

STUDY DATA AND METHODS

Fee-for-Service Sample

We analyzed 11 quality measures for Medicare FFS in 2006-2007 as calculated and published by CMS for the Generating Medicare Physician Quality Performance Measurement Results (GEM) project.¹⁰ This project was primarily intended to measure quality at the medical group practice level, but CMS also produced population-level measures. Our data were aggregated measures at the national, state, and zip code levels, covering all beneficiaries continuously enrolled in Medicare Parts A and B, and for some measures Part D, during the measurement years. Measures were constructed using CMS's Parts A, B, and D claims databases.

The measures were constructed by CMS to conform to HEDIS specifications that require only administrative claims data to calculate. Data limitations necessitated a few minor modifications. One was a shorter look-back period for denominator exclusions because CMS analyzed data only for 2005-2007. Another was that for beneficiaries not enrolled in Part D, diabetes could be identified only from diagnoses in encounter data, not from use of diabetes medication.

HEDIS requires pharmacy claims data for 5 of the measures (Table 1), which are available only for the approximately 50% of FFS beneficiaries enrolled in stand-alone Part D plans. For these measures, the FFS data apply only to the population enrolled in Parts A, B, and D. Although this population differs from the population enrolled in Parts A and B, which is used for the other measures,¹¹ the MA-FFS comparison is still of interest.

Beyond the MA-FFS comparison, these data present a snapshot of the national quality of care in FFS, updating results for other quality measures earlier in the decade.^{12,13}

To analyze nationwide quality in each program, we summed numerators and denominators across plans (MA) or states (FFS), producing a national rate for each measure, following HEDIS 2008 Technical Specifications.¹⁵ (We used this formula to determine the measure denominator.) This method differs from NCQA's practice of taking raw averages across plan scores (irrespective of plan size), but it produces a more accurate national picture of quality for the average beneficiary.

Administrative and Hybrid Measures

There is an important variation in the construction of 6 of the 11 measures (see Table 1) arising from the different ways FFS Medicare and MA plans operate. For these 6 measures, HEDIS allows (but does not require) plans to calculate measure rates on a random sample of the denominator population, using medical chart review to determine whether this sample received appropriate care—a proce-

■ **Table 1.** HEDIS Quality Measures^a

Measure Name	Description of Recommended Care	Measure Notes		
		Admin/Hybrid	Requires Part D Coverage	First Year in HEDIS
Annual monitoring for persistent medications	Annual receipt of appropriate blood tests for patients on persistent medications (eg, diuretics)	Admin	Yes	2005
Antidepressant medication management (acute phase)	Taking antidepressants continuously during 12-week acute phase of new depression episode	Admin	Yes	1998
Antirheumatic drug therapy	Receipt of a disease-modifying antirheumatic drug for patients with rheumatoid arthritis	Admin	Yes	2005
Breast cancer screening (age 42-69 years)	Mammogram during measurement year or year prior for women age 42-69 years	Admin	No	1996
Persistence of beta-blockers	Persistence on a beta-blocker drug for 180 days postdischarge for AMI	Admin	Yes	2004
Beta-blockers after heart attack	Receipt of an ambulatory beta-blocker prescription at discharge for AMI	Hybrid	No	1996
Cardiovascular disease: LDL testing	Annual LDL test for patients with ischemic vascular disease or an acute cardiovascular event	Hybrid	No	1998
Diabetics: eye exams	Eye exam during measurement year or prior year exam negative for retinopathy for diabetics age 18-75 years	Hybrid	No	1996
Diabetics: A1C testing	At least 1 A1C screening during measurement year for diabetics age 18-75 years	Hybrid	No	1998
Diabetics: LDL testing	At least 1 LDL screening during measurement year for diabetics age 18-75 years	Hybrid	No	1998
Diabetics: nephropathy	Screening or ongoing treatment for nephropathy during measurement year for diabetics age 18-75 years	Hybrid	Yes	1998

A1C indicates glycosylated hemoglobin; Admin, administrative; AMI, acute myocardial infarction; HEDIS, Healthcare Effectiveness Data and Information Set; LDL, low-density lipoprotein.

^a“Measurement year” refers to the year for which the quality measure is applicable (2006 or 2007 in our data), and “prior year” is the single year before the measurement year. Measures that required prescription drug plan coverage were calculated for fee-for-service on the subset of fee-for-service beneficiaries continuously enrolled in Parts A, B, and D.

ture called the hybrid method. Because their claims data often are incomplete, HMOs and point-of-service plans typically use the hybrid method, which significantly boosts their quality scores above the administrative-only calculation.¹⁶ By contrast, PPOs (as well as FFS in the GEM study) typically lack the requisite medical chart data, so NCQA requires them to follow the administrative-only specification. (This requirement has been removed starting with HEDIS 2010.¹⁷)

Although this methodologic difference makes sense in the context of plans’ data and reimbursement practices, it could bias our FFS rates downward if FFS-reimbursed physicians fail to submit claims for all procedures or omit important diagnosis codes. (Upward bias also is possible if FFS-reimbursed physicians submit claims for procedures not actually performed.) To address this issue, we observed whether the 6 hybrid measures showed different trends than the 5 administrative-only measures, which are constructed identically in MA and FFS.

■ **Table 2.** Sample Characteristics^a

Category	2006		2007	
	FFS	MA	FFS	MA
Enrollment demographics, %				
Male	44.5	42.1	44.7	42.0
Disability insurance (under age 65 y)	17.4	9.2	17.9	9.7
Dual eligible	17.3	12.5	17.4	13.4
Metropolitan area resident	76.1	95.1	76.4	94.6
Enrollment by census division, %				
New England	5.6	3.5	5.6	4.0
Middle Atlantic	13.7	22.0	13.6	21.7
East North Central	17.3	7.2	17.0	7.3
West North Central	7.4	4.9	7.4	4.9
South Atlantic	20.8	14.0	20.8	14.1
East South Central	7.3	3.7	7.3	3.8
West South Central	10.8	6.6	10.8	7.2
Mountain	5.5	8.7	5.6	9.1
Pacific	11.6	29.4	11.9	27.8
Enrollment by plan type (MA), %				
HMO/POS	—	93.8	—	91.2
PPO	—	6.2	—	8.8
Number of plan contracts (MA)				
	—	264	—	309
Total enrollment (in measured plans)	35,176,538	5,978,584	34,842,196	6,454,358
Enrollment in PDPs (FFS)	16,213,246	—	16,850,205	—

FFS indicates fee-for-service; MA, Medicare Advantage; PDP, prescription drug plan; POS, point of service; PPO, preferred provider organization.
^aData are for FFS (all Part A and B enrollees) and MA (plans included in our sample) for July of 2006 and 2007. Our MA sample includes all plans reporting quality measures, except private FFS plans.

We also compared rates for FFS and MA PPOs, neither of which uses the hybrid method.

Geographic Adjustment

Differences between national MA and FFS quality measures are partly due to MA-FFS differences within the same areas and partly due to their different distributions of beneficiaries across areas. Assuming geographic enrollment differences are primarily driven by factors unrelated to quality, it is important to control for geographic variation to isolate the “within-area” quality difference. We did this in 2 ways.

First, we weighted the state-level FFS rates to match the distribution of the MA measure’s denominator population across states. (The MA data are at the plan level, but almost all MA managed care plans are heavily concentrated [$>95\%$] in 1 state, making it possible to allocate each plan’s denominator to a single state. For plans with enrollment in more than 1 state, we allocated each measure’s denominator across states using the enrollment distribution, which is available

at the plan-state level.) The adjusted MA-FFS difference is equal to a weighted average of the 51 within-state quality differences. This approach controls for state-level differences but misses intrastate variation (eg, between urban and rural areas).

Second, we preliminarily controlled for substate geography by weighting the FFS measure denominator populations to match the county-level distribution of MA enrollees. Although adjusting at a smaller geographic level is preferable, this adjustment has 2 limitations. First, the distribution of MA enrollment may differ from the distribution in each measure’s denominator population (although the 2 distributions should be correlated), but the latter was not available at the county level. Second, county-level adjustment was not feasible for 4 measures, for which most zip code-level FFS rates have been suppressed because they were based on fewer than 11 beneficiaries. Because of these limitations, we report both the state-level and county-level geographic adjustments.

Quality of Care in Medicare

Table 3. HEDIS Quality Measures in Medicare Advantage (MA) and Medicare Fee-for-Service (FFS) Plans, 2006-2007^a

Measure Name	2006, %						2007, %					
	MA Overall	Un-adjusted FFS	FFS (Geographically Adjusted)		MA PPOs Only	FFS (Geographically Adjusted)	MA Overall	Un-adjusted FFS	FFS (Geographically Adjusted)		MA PPOs Only	FFS (Geographically Adjusted)
			State	County					State	County		
Annual monitoring of persistent medications	84.5	89.6	89.4 [4.9] ^b	88.3 [3.7] ^b	85.4	89.7 [4.2] ^b	86.7	90.2	90.0 [3.4] ^b	88.9 [2.2] ^b	87.1	90.4 [3.3] ^b
Antidepressant medication management	69.4	57.9	56.3 [-13.1] ^b	— ^c	— ^c	— ^c	69.0	56.5	54.6 [-14.5] ^b	— ^c	— ^c	— ^c
Antirheumatic drug therapy	70.0	70.0	68.5 [-1.5] ^b	— ^c	70.3	68.0 [-2.3] ^b	72.1	72.3	70.7 [-1.4] ^b	— ^c	74.3	70.7 [-3.6] ^b
Breast cancer screening	75.0	59.4	58.7 [-16.3] ^b	55.5 [-19.5] ^b	71.8	59.4 [-12.4] ^b	73.3	59.3	58.8 [-14.5] ^b	55.7 [-17.6] ^b	70.0	59.4 [-10.6] ^b
Persistence of beta-blockers	73.7	76.4	75.5 [1.8] ^b	— ^c	— ^c	— ^c	75.2	80.7	80.1 [5.0] ^b	— ^c	— ^c	— ^c
Beta-blockers after heart attack	94.2	87.8	87.4 [-6.8] ^b	— ^c	— ^c	— ^c	— ^d	89.9	— ^d	— ^d	— ^d	— ^d
Cardiovascular disease: LDL testing	89.5	81.3	81.5 [-8.0] ^b	80.6 [-8.9] ^b	87.5	82.4 [-5.1] ^b	89.3	81.6	82.0 [-7.3] ^b	81.5 [-7.9] ^b	86.0	83.0 [-3.0] ^b
Diabetics: eye exams	61.2	52.4	53.1 [-8.1] ^b	52.9 [-8.4] ^b	54.0	53.3 [-0.7] ^b	57.5	51.8	52.3 [-5.2] ^b	52.4 [-5.1] ^b	50.3	52.4 [2.1] ^b
Diabetics: A1C testing	87.4	80.4	80.6 [-6.9] ^b	79.5 [-7.9] ^b	85.0	81.3 [-3.7] ^b	85.4	80.8	81.0 [-4.4] ^b	80.3 [-5.1] ^b	80.1	81.4 [1.3] ^b
Diabetics: LDL testing	84.6	76.9	77.3 [-7.3] ^b	77.2 [-7.4] ^b	82.1	78.3 [-3.8] ^b	84.6	77.6	78.1 [-6.5] ^b	78.3 [-6.3] ^b	81.9	78.8 [-3.1] ^b
Diabetics: nephropathy	88.7	78.4	78.5 [-10.2] ^b	79.0 [-9.7] ^b	84.2	78.1 [-6.2] ^b	85.1	76.9	77.3 [-7.8] ^b	78.2 [-6.9] ^b	81.4	77.0 [-4.4] ^b

A1C indicates glycosylated hemoglobin; HEDIS, Healthcare Effectiveness Data and Information Set; LDL, low-density lipoprotein; PPO, preferred provider organization.
^a Bracketed terms are MA-FFS differences. "Geographically Adjusted" refers to adjustment of the FFS measures to match the geographic distribution of the corresponding measure for MA—all MA plans for the first set of comparisons and MA PPO plans for the second set of comparisons (see text for full description).
^b $P < .01$.
^c Insufficient data were available to calculate these measures.
^d Beta-blocker therapy after a heart attack was retired as a HEDIS measure in 2007.

Sociodemographic Differences

Traditionally—for instance, in NCQA publications¹⁸ and in MA quality ratings presented to Medicare beneficiaries¹—HEDIS process measures have not been case mix adjusted because they apply to a clinically similar denominator population. However, the different characteristics of MA and FFS enrollees may raise concerns. Because we did not have quality measures stratified by demographics, it was impossible to adjust for case mix. Instead, we used enrollment-level differences as a proxy to assess the potential magnitude of demographic differences.

RESULTS

Sample Characteristics

Table 2 shows the demographics of enrollees in FFS and in MA plans included in our sample. Fee-for-service enrolls slightly more males and significantly more under age 65 years disabled and people dually eligible for Medicare and Medicaid. (We defined dual eligibility broadly to include any beneficiary whose Part B premium is paid by a state Medicaid program.) Medicare enrollees are more concentrated in metropolitan areas and in the Pacific and Middle Atlantic census divisions.

National Quality in Medicare Advantage Versus Fee-for-Service

Table 3 shows quality measures for MA and for FFS adjusted to match the geographic composition of MA, as described in Geographic Adjustment, above. We also report unadjusted FFS quality rates, because these are of independent interest and illustrate that the geographic adjustments did not make a large difference. All of the Table 3 differences are statistically different from zero ($P < .01$), because the national rates are based on extremely large sample sizes. Because we had essentially complete program data rather than a random sample, significance should be interpreted as rejecting the hypothesis that likelihood of obtaining appropriate care is uncorrelated with program enrollment.

The comparison presents a mixed picture. Focusing on the differences geographically adjusted across states, 8 quality measures were 4 to 16 percentage points (median, 7.8 percentage points) higher in MA in 2006 and 2007. Breast cancer screening was dramatically higher (approximately 15 percentage points) in MA. Quality of diabetes care also was higher in MA, with rates 4 to 10 percentage points higher on the 4 measures studied. Notably, all 8 measures on which MA scored substantially higher are long-established measures included in HEDIS since the 1990s (see Table 1).

By contrast, FFS showed better results on the 3 measures introduced into HEDIS in 2004-2005. Monitoring for patients on persistent medications and persistence of beta-blocker therapy were 2 to 5 percentage points (median, 4.1 percentage points) higher in 2006 and 2007, and antirheumatic drug therapy was only slightly (1.5 percentage points) higher in MA.

Adjusting FFS to match MA's county-level enrollment distribution did not materially change these results. Although some differences moderated toward zero, other differences were strengthened and in no case did the sign or statistical significance of the difference change.

MA-FFS quality differences were all in the same direction in 2007 as in 2006, although most differences narrowed. All 3 of the more recently introduced measures showed rapid 1-year improvements of 1.4 to 2.1 percentage points in MA. However, all of the well-established measures showed declines in MA between 2006 and 2007, particularly 3 of the diabetes measures, which fell by more than 2 percentage points.

Sociodemographic Differences

Our data did not allow adjustment for case mix, but we note that the sociodemographic differences reported in Table 1 were likely too small to explain most of the large MA-FFS differences reported above. For instance, differences in dual-eligible enrollment were only 4 to 5 percentage points, and

the county-level geographic adjustments reported above already controlled for metropolitan residency, with little effect for most measures.

Another concern is that the breast cancer screening comparison may be biased because the fraction of under age 65 years disabled beneficiaries in FFS is 8.2 percentage points higher than that in MA, and younger women are less likely to obtain biennial mammograms.⁵ For instance, in 2006 MA data (not reported in Table 3), women age 42 to 51 years were 19.3 percentage points less likely to have a mammogram than women age 52 to 69 years. But even if the mammogram gap between those age 65+ years and those under 65 years were this large, it would explain just 1.6 percentage points, or about one tenth, of the MA-FFS gap.

Administrative and Hybrid Measures

Among the 5 administrative-only measures, 2 (antidepressant management and breast cancer screening) were substantially higher in MA, and 2 were higher in FFS (annual monitoring for persistent medications and persistence of beta-blocker therapy); antirheumatic drug therapy was similar in the 2 programs. All 6 hybrid measures were higher in MA, although not by as much as antidepressant management or breast cancer screening.

To assess the impact of hybrid data collection on the results, we compared quality measures between MA PPOs and FFS (adjusted to match PPOs' state-level distribution), neither of which use the hybrid method. Most of the PPO-FFS differences were in the same direction as in the broader MA-FFS comparison, including 8 of the 10 comparisons of hybrid measures. However, the gaps on the hybrid measures were usually smaller, and FFS did better on 2 diabetes measures in 2007. This comparison suggests the hybrid method is a significant factor, but most MA-FFS differences persisted in an administrative-only comparison.

DISCUSSION

To our knowledge, this is the first study to conduct a national comparison of MA and FFS Medicare based on HEDIS clinical quality measures. The results are relevant both for the policy process and for individual beneficiaries selecting between FFS and MA plans. We analyzed how data limitations and program differences raise issues of comparability, many of which can be improved with more complete data.

Our study builds on the HEDIS measurement of Medicare managed care plans that has been required since 1997 and has been used for quality ratings to inform beneficiaries since 1999.¹⁹ Using the data we studied, CMS could construct similar ratings for FFS in a beneficiary's state or metropolitan area

(as MedPAC recommends) to supplement existing comparisons among MA plans.

FFS historically was not designed to measure quality or influence beneficiary choices through ratings. In recent years, Medicare has introduced quality measurement for providers, including hospitals, physicians, skilled nursing facilities, and home health agencies, but measures for comparing MA to FFS are less well developed.⁵ With the mandate in the Medicare Improvements for Patients and Providers Act and the health reform legislation to expand value-based purchasing, comparing MA and FFS has gained greater policy importance. MedPAC's March 2010 report on ways to implement such a comparison by March 2011 underscores the timeliness of our study. Nonetheless, our comparison is limited to ambulatory care process measures. Comparing MA and FFS on risk-adjusted outcome measures, as in recent work at a state and local level,^{20,21} also will be valuable.

Our data showed significant and typically large differences between MA and FFS. The differences were stable from 2006 to 2007 and robust to weighting the FFS data to match MA's geographic distribution at the state or county level. Of the 11 measures, MA performed substantially better on 8, slightly better on 1, and worse on 2. These results present a more mixed picture than results from the CAHPS survey earlier in the decade, which found MA beneficiaries more likely to receive all 3 preventive services studied.^{8,9} Similarly, a study using 1990's Medicare Current Beneficiary Survey data found MA beneficiaries more likely to receive most but not all preventive services examined.¹⁹

One pattern in our results was that MA performed better on all 8 measures introduced to HEDIS in the late 1990s, whereas FFS performed better or only slightly worse on the 3 measures introduced in 2004-2005. Medicare Advantage also showed rapid improvement in the 3 newer measures from 2006 to 2007 (and into 2008 based on NCQA reports¹⁸), but showed declines for all 8 older measures. If this dichotomy is not coincidental, it suggests a learning effect in MA, or less favorably a "teaching to the test" effect. Newly introduced measures may have lower scores in MA initially, but these scores quickly increase as plans learn to ensure effective care delivery and complete measurement of existing care. Although the 2 mechanisms for the learning effect have different implications, they cannot be disentangled in our data. Other work on this topic¹⁹ has used a stable source of measurement data and did not find more rapid increases in quality measures in MA compared with FFS when HEDIS was introduced in the late 1990s, suggesting that improvements in measurement may be more important than improvements in care delivery.

Our analysis has several comparability limitations. The most significant, but also easiest to address in future efforts,

are measurement differences. For instance, CMS could in future FFS calculations extend look-back periods indefinitely, limit the breast cancer screening measure to women over age 52 years, and limit all measures to enrollees in Parts A, B, and D to address several concerns.

On the hybrid issue, MedPAC recommended that administrative-only measure comparisons begin immediately but that hybrid measures be adjusted (perhaps by drawing upon electronic health records in FFS and encounter data in MA that will be collected starting in 2012) before comparing them.⁵ Although we acknowledge the potential bias in comparing hybrid measures, our comparison of FFS and MA PPOs suggests significant differences remain with an administrative-only comparison. Given the potential for benchmarking initial differences and observing trends in relative quality, we think comparing hybrid measures can be valuable immediately.

Comparability issues related to enrollment differences go to the heart of what is meant by "relative quality" in MA and FFS. Traditionally, researchers have in mind a treatment effect of moving a fixed population from one program to the other. Only randomization can reliably measure such treatment effects, although even randomization suffers from quality spillovers through doctors treating both MA and FFS patients. Limiting measures to clinically comparable subsets of the population has been the feasible alternative to randomization used by HEDIS, and we followed this approach. In addition, because Medicare beneficiaries choose among local alternatives, we adjusted FFS measures to match the distribution of MA across states or, where feasible, counties.

Some researchers have suggested the importance of sociodemographic case mix adjustment, arguing some groups face greater barriers to effective care independent of plan quality. Although this is no doubt true, case mix adjustment may be inappropriate if sociodemographics are not independent of plan selection.²² Previous research on commercial plans found case mix adjustment had little effect on HEDIS measures for most plans, although it had important effects for a few plans.^{23,24} Future work should assess its importance for the MA-FFS comparison.

These qualifications aside, this study provides a significant first step in comparing quality between MA and FFS. As policy makers grapple with improving quality in Medicare, this and future analyses can inform their decisions.

Acknowledgments

We thank the editor and 2 anonymous referees for extremely helpful comments. We also thank Mark McClellan, PhD, Joachim Roski, PhD, Larry Kocot, JD, and Kristina Lowell, PhD, for their critical review and editing of earlier versions of this manuscript.

Author Affiliations: From The Brookings Institution (NB), Washington, DC; Department of Economics (MS), Harvard University, Cambridge, MA.

Mr Brennan is currently employed by the Centers for Medicare & Medicaid Services, Washington, DC.

Funding Source: This study was funded by the National Science Foundation.

Author Disclosures: The authors (NB, MS) report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (NB, MS); acquisition of data (NB, MS); analysis and interpretation of data (NB, MS); drafting of the manuscript (NB, MS); critical revision of the manuscript for important intellectual content (NB, MS); statistical analysis (NB, MS); provision of study materials or patients (NB); obtaining funding (MS); administrative, technical, or logistic support (NB); and supervision (NB).

Address correspondence to: Niall Brennan, MPP, Office of Policy, Centers for Medicare & Medicaid Services, 200 Independence Ave SW, Washington, DC 20024. E-mail: niall.brennan@cms.hhs.gov.

REFERENCES

1. **Centers for Medicare & Medicaid Services.** Medicare plan finder. <http://www.medicare.gov/find-a-plan/questions/home.aspx>. Accessed July 1, 2010.
2. **Gold M.** Medicare's private plans: a report card on Medicare Advantage. *Health Aff (Millwood)*. 2009;28(1):w41-w54.
3. **Zarabozo C, Harrison S.** Payment policy and the growth of Medicare Advantage. *Health Aff (Millwood)*. 2009;28(1):w55-w67.
4. **General Accounting Office.** Increased Spending Relative to Medicare Fee-for-Service May Not Always Reduce Beneficiary Out-of-Pocket Costs. February 2008. GAO-08-359. <http://www.gao.gov/new.items/d08359.pdf>. Accessed March 30, 2009.
5. **Medicare Payment Advisory Commission.** Report to the Congress, Chapter 6: Report on Comparing Quality Among Medicare Advantage Plans and Between Medicare Advantage and Fee-for-Service Medicare. March 2010. http://www.medpac.gov/chapters/Mar10_Ch06.pdf. Accessed July 1, 2010.
6. **US News & World Report.** *America's best health plans*. November 2009. <http://health.usnews.com/health-plans>. Accessed July 1, 2010.
7. **Landon BE, Zaslavsky AM, Bernard SL, Cioffi MJ, Cleary PD.** Comparison of performance of traditional Medicare vs Medicare managed care. *JAMA*. 2004;291(14):1744-1752.
8. **Keenan PS, Elliott MN, Cleary PD, Zaslavsky AM, Landon BE.** Quality assessments by sick and healthy beneficiaries in traditional Medicare and Medicare managed care. *Med Care*. 2009;47(8):882-888.
9. **Barton MB, Dayhoff DA, Soumerai SB, Rosenbach ML, Fletcher RO.** Measuring access to effective care among elderly Medicare enrollees in managed and fee-for-service care: a retrospective cohort study. *BMC Health Serv Res*. 2001;1(1):11.
10. **Centers for Medicare & Medicaid Services.** Generating Medicare Physician Quality Performance Measurement Results (GEM) project. <http://www.cms.gov/GEM/>. Accessed January 15, 2009.
11. **Kaiser Family Foundation.** *Medicare Fact Sheet: The Medicare Prescription Drug Benefit*. November 2009. <http://www.kff.org/medicare/upload/7044-10.pdf>. Accessed July 2, 2010.
12. **Fisher ES, Goodman DC, Chandra A.** *Regional and Racial Variation in Health Care Among Medicare Beneficiaries: A Brief Report of the Dartmouth Atlas Project*. Robert Wood Johnson Foundation. December 2008. http://www.dartmouthatlas.org/downloads/reports/AF4Q_disparities_Dec2008.pdf. Accessed February 23, 2009.
13. **Jencks SF, Huff ED, Cuerdon T.** Change in the quality of care delivered to Medicare beneficiaries 1998-1999 to 2000-2001 [published correction appears in *JAMA*. 2003;289(20):2649]. *JAMA*. 2003;289(3):305-312.
14. **Centers for Medicare & Medicaid Services.** HEDIS Public Use Files. <https://www.cms.gov/MCRAdvPartDenrolData/HEDIS/list.asp>. Accessed March 4, 2009.
15. **National Committee for Quality Assurance.** Technical Specifications. Washington, DC: National Committee for Quality Assurance; October 2007. *HEDIS 2008; vol 2*.
16. **Pawlson LG, Scholle SH, Powers A.** Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures. *Am J Manag Care*. 2007;13(10):553-558.
17. **Centers for Medicare & Medicaid Services.** 2010 HEDIS, HOS and CAHPS Measures for Reporting by Medicare Advantage Organizations. Memorandum to Medicare Advantage Quality Contacts and Medicare Compliance Officers. December 2, 2009. https://www.cms.gov/PrescriptionDrugCovContra/Downloads/MemoHEDIS2010Reporting-Measures_12.02.09.pdf. Accessed July 1, 2010.
18. **National Committee for Quality Assurance.** The State of Health Care Quality 2009. http://www.ncqa.org/Portals/0/Newsroom/SOHC/SOHC_2009.pdf. Accessed July 2, 2010.
19. **Bundorf MK, Choudhry K, Baker L.** Health plan performance measurement: does it affect quality of care for Medicare managed care enrollees? *Inquiry*. 2008;45(2):168-183.
20. **Lemieux J, Chovan T, Chen C, Carpenter L, Buck K, Heath K.** Working Paper: Using State Hospital Data to Compare Readmission Rates in Medicare Advantage and Medicare's Traditional Fee-for-Service Program. America's Health Insurance Plans, Center for Policy and Research. May 2010. <http://www.ahipresearch.org/pdfs/9State-Readmits.pdf>. Accessed July 2, 2010.
21. **Lemieux J, Chovan T, Chen C.** Working Paper: Comparisons of Utilization in Two Large Multi-State Medicare Advantage HMOs and Medicare Fee-for-Service in the Same Service Areas. America's Health Insurance Plans, Center for Policy and Research. December 2009. <http://www.ahipresearch.org/pdfs/MAvsFFS-CO9and10.pdf>. Accessed July 2, 2010.
22. **Romano PS.** Should health plan quality measures be adjusted for case mix? *Med Care*. 2000;38(10):977-980.
23. **Zaslavsky AM, Hochheimer JN, Schneider EC, et al.** Impact of sociodemographic case mix on the HEDIS measures of health plan quality. *Med Care*. 2000;38(10):981-992.
24. **Zaslavsky AM, Epstein AM.** How patients' sociodemographic characteristics affect comparisons of competing health plans in California on HEDIS quality measures. *Int J Qual Health Care*. 2005;17(1):67-74. ■