REPORT TO CONGRESS

EVALUATION OF THE MEDICARE CARE MANAGEMENT PERFORMANCE DEMONSTRATION

Kathleen Sebelius
Secretary of Health and Human Services
2014

I. PURPOSE

Section 649 of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003 (P.L. 108-173) required the Secretary of Health and Human Services to establish a 3-year "pay-for-performance demonstration program with physicians to meet the needs of eligible [Medicare] beneficiaries through the adoption and use of health information technology and evidence-based outcome measures" (see the Appendix A). Section 649 authorized up to four demonstration sites to include two urban sites, one rural site, and one in a state "with a medical school with a Department of Geriatrics that manages rural outreach sites and is capable of managing patients with multiple chronic conditions, one of which is dementia." The statute also stipulated that "the Secretary shall ensure that the aggregate payments made by the Secretary do not exceed the amount which the Secretary estimates would have been paid if the demonstration program under this section was not implemented." This report summarizes the findings from the demonstration program.

II. DEMONSTRATION OBJECTIVES

The Medicare Care Management Performance (MCMP) demonstration was designed and conducted by the Centers for Medicare & Medicaid Services (CMS) as the first federally funded initiative to assess the impact of pay-for-performance on quality of care for small- and medium-sized primary care practices. Demonstration operations began July 1, 2007, and ended June 30, 2010. Physician practices were recruited from among primary care practices in Arkansas, California, Massachusetts, and Utah that participated in the Doctor's Office Quality-Information Technology (DOQ-IT) pilot project, a CMS-funded initiative to encourage use of electronic health records (EHRs). Practices of up to 10 physicians (with a few exceptions) were eligible for participation in the MCMP demonstration if they provided the majority of primary care services for at least 50 fee-for-service Medicare beneficiaries with congestive heart failure (CHF), coronary artery disease (CAD), diabetes, and other specified chronic conditions.[1]

Under the demonstration, physician practices could earn incentive payments through the following actions:

(1) reporting on the 26 clinical quality measures (see Appendix B) established by CMS for the baseline period (January 1-December 31, 2006),
(2) achieving specified performance standards on the 26 measures for each of the 3 years of the demonstration, and
(3) submitting data on the quality measures electronically through an EHR system certified by the Certification Commission for Healthcare Information Technology.[2]

Taking into account the various ways that practices could earn incentive payments, the maximum potential payment for reporting the baseline data and participating in all three demonstration

[1] In addition to the three primary target chronic conditions listed above, the other eligible conditions were Alzheimer's disease and other mental, psychiatric, and neurological disorders; any chronic cardiac/circulatory disease (e.g., arteriosclerosis, myocardial infarction, angina pectoris, stroke); any cancer; arthritis and osteoporosis; kidney disease; and lung disease. These conditions were identified through International Classification of Diseases, Ninth Revision, Clinical Modification diagnosis codes available in Medicare claims data (Wilkin et al., 2007).

[2] In contrast to the current Medicare and Medicaid EHR Incentive Programs that provide incentives for "meaningful use" of EHRs that are certified as meeting specific standards, the MCMP demonstration provided a bonus for electronic reporting, but did not explicitly require an EHR; electronic data submission in the MCMP demonstration included extracting and submitting data from any electronic database.

years was $38,500 per physician and $192,500 per practice.[3] Although the demonstration design did not provide extensive technical support to the practices on how to change their operations, some assistance was available.

The goals of the demonstration were to use financial incentives to improve the quality of care provided to eligible fee-for-service Medicare beneficiaries and encourage the implementation and use of health information technology (health IT) among primary care physicians. The specific objectives were to promote continuity of care, help stabilize medical conditions, prevent or minimize acute exacerbations of chronic conditions, and reduce adverse health outcomes among fee-for-service Medicare beneficiaries.

III. EVALUATION DESIGN AND METHODS

CMS contracted with Mathematica Policy Research to conduct an independent evaluation of the demonstration. The evaluation included two components: an implementation analysis and an outcomes analysis that compared outcomes for beneficiaries, physicians, and practices from the demonstration group (i.e., the "treatment" group) to those for the comparison group. The implementation analysis used quantitative data to describe demonstration participation and the characteristics of participating practices, and it assessed practices' responses to the incentives, including changes in their use of health IT and care management approaches throughout the demonstration. The analysis was based primarily on two rounds of site visits to a purposively selected sample of treatment group practices. In 2008, at the end of Year 1, 2-person teams visited 8 treatment group practices in each of the 4 states (a total of 32 practices); the teams visited 29 of these practices again in 2010 (excluding 3 that withdrew and were not available for interviewing). This analysis also used qualitative data from telephone interviews with comparison group practices,[4] practices that withdrew from the demonstration, and practices that had scored relatively high or low on their quality measures in Year 1. Additional implementation data included treatment group physicians' responses to a survey (described further below) about their perceptions of the demonstration.

The outcomes analysis relied on a quasi-experimental design. Each demonstration state was matched to non-demonstration states based on specific criteria that included demographics, extent of EHR use and pay-for-performance programs in the state, and other key characteristics (e.g., the ratio of specialists to general practice/family medicine physicians). Within these matched non-demonstration states, comparison group practices were chosen from participants in CMS's DOQ-IT initiative. Propensity score methods were used to match treatment group practices to comparison group practices in terms of key characteristics, such as size (both in terms of the number of Medicare beneficiaries and number of physicians at the practice), number of Medicare beneficiaries at the practice with the MCMP demonstration's target conditions (CAD, CHF, or diabetes), the amount of services used (including hospital and E&M visits) by beneficiaries in the practice, the practice's experience with health IT, and whether it was located in a medically underserved area.

[3] In contrast to these maximum payments, the current Medicare and Medicaid EHR Incentive Programs do not set a maximum level per practice; the maximum per eligible professional is $44,000, with additional incentives available for Medicare eligible professionals that practice in a Health Professional Shortage Area.

[4] Poor participation of comparison practices in the telephone interviews limited their ability to contribute materially to the evaluation's findings.

At the start of the demonstration, 700 eligible practices were enrolled in the treatment group; however, 23 were excluded from the analysis because they did not complete the 2007 (baseline) Office Systems Survey (OSS)—a survey of practices' use of health IT—and were missing key variables needed for the matching process. The remaining 677 practices were included in the evaluation and matched to 548 comparison group practices in 9 non-demonstration states.[5] Some comparison group practices were matched to more than one treatment group practice.

For assessing changes in quality of care, service use, and Medicare cost outcomes, a "difference-in-differences" analytic approach was used. Changes in beneficiaries' outcomes for each demonstration year (Year 1, Year 2, and Year 3) were compared to a baseline period (calendar year 2006) for both treatment and matched comparison group practices. Since data for multiple years were needed in this analysis, practices with beneficiaries assigned to them for only 1 year were not included, resulting in 665 treatment practices (representing 190,827 beneficiaries at baseline) and 525 comparison practices (representing 134,817 beneficiaries at baseline) for the claims-based analysis. Annualized measures of expenditures and service use were constructed for beneficiaries assigned to treatment or comparison group practices for the baseline, Year 1, Year 2, and Year 3 periods. Data were collected from Medicare claims for 7 of the demonstration's 26 incentivized quality-of-care measures, and proxies for 3 additional measures were derived from Medicare Part D data.[6,7] In addition, two quality measures were constructed related to inappropriate drug use. For the claims-based analysis, beneficiary-level characteristics were controlled for, including demographic characteristics (age, sex, and race), chronic condition diagnoses, and dual-eligibility status.

To evaluate 6 of the other quality-of-care measures as well as the beneficiary's perspective on other aspects of quality of care, including care coordination and satisfaction with care, a stratified random sample of the beneficiaries assigned to treatment and comparison group practices was surveyed in 2009. For both the treatment and comparison groups, response rates were high (77 and 78 percent, respectively): overall, 4,387 beneficiaries responded. To analyze beneficiary survey outcomes, logit models were used that controlled for the beneficiary's demographic characteristics and preferred language, the reason for Medicare entitlement, chronic conditions, Medicare service use before the demonstration, and the characteristics of the practice to which the beneficiary was assigned.

To measure quality of care from the physician's perspective, a stratified random sample of the physicians from treatment and comparison group practices was surveyed between 2009 and 2010. At least one physician per practice was included in this survey. Response rates varied between the treatment and comparison groups, with approximately 62% responding in the treatment group, and 51% in the comparison group; overall, 1,338 physicians responded. The survey asked questions about physicians' experiences with using health IT, patient-provider communication, care coordination, and satisfaction with care. To analyze survey outcomes, logit models were used that controlled for physician demographic and socioeconomic characteristics as well as practice characteristics.

---

[5] The matched non-demonstration states were Nebraska and Texas (for Arkansas), Arizona (for Southern California), Oregon and Washington (for Northern California), New York and Connecticut (for Massachusetts), and Idaho and Colorado (for Utah).

[6] For purposes of scoring and payment, treatment group practices were allowed to use chart review to exclude patients from the denominator for a particular measure for "medical reasons" (e.g., drug contraindications), "patient reasons" (e.g., economic or social circumstances), or "systems reasons" (e.g., the lack of availability of a vaccine). However, these exclusions were not recorded in the claims data and thus are not reflected in claims-based analyses.

[7] The three proxy measures were constructed for those beneficiaries who were enrolled in Part D, approximately half the sample, for the baseline and first 2 years of the demonstration.

Measures of practices' health IT use, including their use of EHRs and 33 specific EHR functions, were drawn from the OSS, a survey of practices' use of health IT. The OSS was administered to treatment and comparison group practices in 2007 and 2010; the final analysis sample included 311 treatment group practices and 173 comparison group practices. Analysis of the OSS included a description of the change in treatment group practices' health IT use over the demonstration period among practices that completed both rounds of the OSS. Because only 32 percent of comparison group practices responded to the OSS, a formal treatment-comparison analysis was not conducted.

A synthesis analysis examined the factors (with a particular focus on health IT) that were related to treatment group practices' overall performance in the demonstration. It used both cross-tabulation and multivariate analysis to explore the relationship between practices' EHR use and change in their quality scores over the demonstration. Quality measures in this analysis included both practice-level quality composite scores (based on the 26 incentivized quality measures) and claims-based measures (including 20 related to quality of care such as preventable hospitalizations and 2 related to expenditures; see Appendix B).

IV. EVALUATION FINDINGS[8]

Beneficiary Characteristics

Data are based on beneficiaries who were assigned to practices at baseline:

- Approximately 60% of beneficiaries in the treatment and comparison groups were 65-79 years old.
- Approximately 60% of beneficiaries in the treatment and comparison groups were women.
- Approximately a quarter of beneficiaries had coronary artery disease, about 12% had congestive heart failure, more than a third had diabetes, and about three-quarters had other chronic conditions.
- Most of the beneficiaries were white (83-97%, depending on the site).

Demonstration Participation

- Approximately three-fourths of the 677 enrolled practices actively participated by submitting quality data for Year 3. The overall attrition rate was low, likely because the requirements for participation were minimal (beyond submitting data).
    - 12% of enrolled practices formally withdrew during the demonstration. Practices that provided reasons for withdrawing cited limited resources available to participate (18 of 29 practices) and the burdensome data submission effort (12 of 29). The decision may have been influenced in some cases by staff turnover and the inability of some practices' EHRs to generate the necessary quality data.
    - 10% became inactive because they did not submit Year 3 quality data and thus were ineligible for an incentive payment.
- During the first round of site visits in 2008, 21 of 32 practices reported that the data submission process was arduous; however, during the second round of site visits in 2010,

[8] Statistical significance was assessed using a two-tailed test at the 10-percent level of significance for the null hypothesis that a particular estimate was equal to zero.

22 of 29 practices reported that it had become less burdensome because they had mastered the process.
    - Reasons cited for why the data submission process was burdensome included the time required to enter and review data as well as a lack of technical support.
- During telephone interviews, staff at seven of the ten high-performing practices reported that their higher quality scores were due to their combined commitment to achieving and documenting high-quality performance. Involving all types of staff in the effort was viewed as crucial.
- During telephone interviews, four of the seven lower-performing practices reported that they took action in response to their quality scores; however, their Year 3 scores were lower than their Year 1 scores for the composite scores for each condition (diabetes, CAD, and CHF) as well as preventive care. Of the three remaining practices, two appeared uninterested in the demonstration, and one had closed.

Incentive Payments

- The average practice incentive payment increased from about $14,300 in Year 1 to more than $18,000 in Years 2 and 3.
    - Higher total incentive payments per practice were generally received by practices of larger size, in terms both of number of physicians and of higher beneficiary caseload, and by those that used health IT prior to the demonstration.
    - Larger practices did not have higher average payments per beneficiary.
- The percentage of practices that received the electronic reporting incentive increased from 23% in Year 1 to 29% in Year 3.

Treatment Group Physician Response to the Demonstration

Data are based on 696 treatment group physicians who responded to the physician survey:

- >90% strongly agreed or agreed that the MCMP demonstration targeted important medical conditions, used well-accepted and appropriate measures of care quality, and encouraged adoption and use of EMRs.
- >80% strongly agreed or agreed that the MCMP demonstration required a reasonable level of effort to report annual quality data.
- 72% strongly agreed or agreed that the MCMP demonstration had rules for rewarding quality care that were easy to understand.
- 56% strongly agreed or agreed that the MCMP demonstration provided sufficient financial rewards.
- 27% reported that the demonstration improved their clinical decision-making.
- 38% said that the demonstration increased the amount of time they spent educating patients with chronic conditions.
- 48% reported that they had improved their adherence to recommended clinical practice guidelines or evidence-based medicine.
- 69% would recommend the MCMP demonstration to their colleagues, and 20% reported that the MCMP demonstration was better than other pay-for-performance programs that they had experienced.

Coordination and Management of Care

Descriptive results based on a pre-post analysis of the treatment group practices:

- 25 of 29 visited practices made modest changes in care management between Year 1 and Year 3, and these changes tended to focus on care for patients with diabetes.
- 5 of the 29 visited practices reported improving their documentation of care and paying closer attention to the guidelines for care of beneficiaries with chronic conditions.
- The number of visited practices reporting use of features such as EHR-generated alerts and reminders increased from 13 in 2008 to 16 in 2010.
    - Although the number of visited practices reporting use of features such as EHR-generated alerts and reminders increased, processes for following up with patients after these messages were not always in place, due to reasons such as limited visit time and lack of additional payment for care management activities.

Results based on a difference-in-differences analysis of the treatment group practices relative to a set of matched comparison group practices:

- Relative to the comparison group, treatment group physicians rated care coordination more favorably in terms of satisfaction with care coordination (74% vs. 66%), and they were less likely to report inadequate access to patient information (50% vs. 60%) and that patients received the wrong drug or dosage or had a drug-drug interaction (21% vs. 28%) (N=1,312).
- Beneficiaries did not commonly perceive problems related to care coordination.
    - Relative to the comparison group, treatment group patients reported being less satisfied with communication between their providers (69% vs. 73%) (N=3,937); this negative finding was driven by results from California, which had an 11-percentage-point difference favoring the comparison group.
- Beneficiaries in both the treatment and the comparison groups generally reported positively on their interactions with physicians, and there were no statistically significant treatment-comparison group differences.
- Generally, beneficiaries in both the treatment and the comparison groups reported similar levels of health-related knowledge and behavior.
    - Relative to the comparison group, treatment group patients with diabetes were more likely to report they examined their feet daily in the last 12 months (48% vs. 42%) (N=1,232).

Quality of Care

Descriptive results based on a pre-post analysis of the treatment group practices:

- For average practice-level scores, based on the 449 practices that submitted data in both Year 1 and Year 3, 18 of the 26 quality-of-care measures improved between Year 1 and Year 3.
    - Many of the improvements were small (i.e., <2 percentage points).
    - 11 of these improvements were statistically significant:
        - More patients with CAD were prescribed a lipid-lowering therapy (88% vs. 85%).

        - More patients with CAD had their most recent low-density lipoprotein (LDL) cholesterol <130 mg/dL (76% vs. 74%).
        - More patients with diabetes had one or more blood tests for hemoglobin A1c (94% vs. 93%).
        - Fewer patients with diabetes had their most recent A1c level >9 percent (16% vs. 17%).
        - More patients with diabetes had a dilated retinal exam (61% vs. 58%).
        - More patients with diabetes had their last blood pressure measurement below 140/90 mmHg (among those who received a test) (54% vs. 50%).
        - More patients with diabetes had at least one LDL cholesterol test (86% vs. 85%).
        - More patients with CHF were prescribed beta-blocker therapy, among those who also had left ventricular systolic dysfunction (88% vs. 85%).
        - More patients with specified chronic diseases had breast cancer screenings during the current or previous year, among those under age 69 (73% vs. 69%).
        - More patients with specified chronic diseases had colorectal cancer screenings during recommended period (60% vs. 51%).
        - More patients with specified chronic diseases had pneumonia vaccinations, among those with a chronic condition over age 65 (65% vs. 56%).
    - Of the measures with percentage-point declines, two were statistically significant, although both were small (i.e., <2 percentage points): fewer patients with diabetes had at least one test for microalbumin (or had medical attention for existing nephropathy or microalbuminuria or albuminuria) (80% vs. 81%), and fewer patients with specified chronic diseases had an influenza vaccination between September and February of the year prior to the measurement year, among those over age 50 (71.5% vs. 73.4%).
    - The trend in composite measures showed incremental positive improvement (of about 1 to 2 percentage points) between Year 1 and Year 3, with practices following the guidelines more than 80% of the time and following them more often in Year 3 than in Year 1 (83% vs. 81%).

Results based on a difference-in-differences analysis of the treatment group practices relative to a set of matched comparison group practices:

- In the claims-based analysis, the treatment and comparison groups were similar in terms of the 7 quality-of-care measures at baseline, and there were few statistically significant difference-in-differences estimates for these measures.
    - Statistically significant estimates were negative for left ventricular ejection fraction tests in Year 2 for patients with CAD (-4.2 percentage points; N=10,037), lipid tests in Year 1 and eye examinations in Year 3 for patients with diabetes (-2.0 percentage points for each measure; N=211,010), and breast cancer screenings in Year 3 for patients with specified chronic diseases (-2.0 percentage points; N=126,852), indicating that—for these measures—the comparison group improved more than the treatment group.

- There were very few statistically significant difference-in-differences estimates for the preventable hospitalization measures.
    - Treatment group patients experienced a small yet statistically significant increase in three preventable hospitalization measures relative to the comparison group in Year 3 only—cardiac hospitalizations among beneficiaries with CAD (16 additional hospitalizations per 1,000 beneficiaries; N=326,128), and cardiac hospitalizations as well as hospitalizations for microvascular complications among beneficiaries with diabetes (7 and 1 additional hospitalizations per 1,000 beneficiaries, respectively; N=485,328).
- The analysis with prescription drug event data did not show any statistically significant improvement in the treatment group patients having claims for any of the three classes of drugs recommended for beneficiaries with CAD compared with the comparison group.
- Of the seven quality measures assessed by self-report through the beneficiary survey which was administered in 2009 (N=4,207), four indicated that treatment group patients were statistically significantly less likely to report receiving the recommended interventions.
    - Fewer treatment than comparison group patients with CHF reported having their weight measured at their last visit (89% vs. 96%).
    - Fewer treatment than comparison group patients reported being asked whether they had ever received a pneumonia vaccination (61% vs. 65%).
    - Fewer treatment than comparison group patients reported having received a flu vaccination within the past 2 years (83% vs. 88%).
    - Fewer treatment than comparison group patients reported receiving appropriate colon cancer screening within the past 5 years (71% vs. 73%).
- Although both beneficiary groups expressed widespread satisfaction with their experience with care, the treatment group was statistically significantly less likely to be satisfied for several measures (N=4,207).
    - Relative to the comparison group, treatment group patients were less likely to be satisfied with the attention they received from the provider during office visits (81% to 84%) and with how well the provider explained what to expect with their health or chronic conditions in the future (68% to 71%).
    - Relative to the comparison group, treatment group patients were also less likely to be satisfied with the provider's accessibility (69% to 73%) and their ability to get an appointment as soon as they wanted (65% to 69%).
    - Negative findings were generally driven by results from California; for example, self-reported satisfaction with their providers' explanations of what to expect with their health or chronic conditions in the future was 7 percentage points lower for treatment group patients relative to the comparison group.
- Satisfaction with the quality of care provided to Medicare beneficiaries with chronic conditions was widespread, with approximately 96% of both treatment and comparison group physicians reporting being very satisfied or somewhat satisfied with the quality of care their patients with chronic conditions received (N=1,312).
    - Relative to the comparison group, treatment group physicians were statistically significantly more likely to report being very satisfied or somewhat satisfied with patients' receipt of recommended preventive services (86% vs. 80%) and with their level of Medicare reimbursement (24.5% vs. 17.4%).

Use of Health IT

Descriptive results based on a pre-post analysis of the treatment group practices:

- Of 311 treatment group practices that completed both rounds of the OSS, the proportion using an EHR increased from 68% in 2007 to 83% in 2010.
    - By 2010, 95% used some form of health IT, such as EHRs, stand-alone products for electronic prescribing, and stand-alone registries.
    - Statistically significant gains were made in the proportion of practices that used their EHR to review and act on care reminders (from 53% in 2007 to 73% in 2010), prescribe electronically (from 69% to 80%), and use registry functions to track patients with chronic conditions (from 46% to 60%).
    - Practices that adopted EHRs during the demonstration had lower scores on the use of EHR functions in comparison to practices with preexisting EHRs (e.g., overall score: 38 vs. 66, on a 100-point scale).
    - Practices made progress in the completeness of EHR information, the communication of care outside the practice, and use of clinical decision support.
- Of 29 visited practices, 22 had EHRs at the start of the demonstration, and one of the seven without an EHR in 2007 adopted an EHR during the demonstration.
    - 8 of the 22 visited practices that used EHRs reported that they were using the EHR's features more or better than in the past.
    - Six of the visited practices had implemented EHR-generated alerts and reminders, two had implemented diagnostic test ordering, one had implemented diagnostic test viewing, and two had implemented a process for generating lists of patients needing follow-up care.
    - Seven of the visited practices established or increased use of a patient web portal outside their EHR.
    - Two of the visited practices adopted stand-alone chronic condition registries to track performance on various measures.
- Visited practices reported that several factors encouraged the use of health IT, including the availability of financial and technical support as well as having a health IT champion and experience with health IT.
    - Multiple factors outside the MCMP demonstration, including other federal and private-sector initiatives, influenced the increase in use of health IT among visited demonstration practices. Fifteen of 29 practices reported that they were involved in private-sector pay-for-performance programs, and 3 of 29 practices reported basing their health IT decisions on potential meaningful use incentives from the American Recovery and Reinvestment Act Medicare and Medicaid EHR Incentive Programs (a majority of practices had some level of awareness of the meaningful use incentive programs). Additionally, like other physicians and physician practices, practices that submitted data for MCMP incentive payments were also eligible for payment under the Physician Quality Reporting System—formerly the Physician Quality Reporting Initiative—in which they could receive an additional 1.5-2.0% of estimated allowed charges for covered Part B Physician Fee Schedule services each year.

  - Additionally, practices affiliated with a larger medical group, health system, or IPA were influenced in their use of health IT by the strategic plans of these larger organizations, which focused on creating efficient systems. Practices associated with larger organizations also benefited from dedicated resources.
  - Barriers to use of health IT included the perceived time burden, lack of technical support, lack of a cohesive practice culture, financial burden, reluctance on the part of some older physicians, lack of comfort with electronic systems, hearsay about EHR users' negative experiences, and concerns about interoperability between systems and data security.

Results based on a difference-in-differences analysis of the treatment group practices relative to a set of matched comparison group practices (N=1,312):

- In the physician survey, 97% of treatment and comparison group physicians reported that they used or planned to use EHRs to record and manage patient care, and the majority stated that they used or planned to use basic EHR functions in the next 12 months.
- Based on the physician survey, relative to the comparison group, treatment group physicians were statistically significantly less likely to use or plan to use EHRs to enter laboratory, radiology, or diagnostic test orders (67% vs. 83%) and issue reminders to patients (29% vs. 41%).
- Relative to the comparison group, treatment group physicians were statistically significantly more likely to report having more email exchanges with beneficiaries over the past year than in previous years (19% vs. 14%).
- Treatment group physicians reported statistically significantly fewer barriers to health IT use than comparison group physicians, including:
  - start-up costs (57% vs. 68%),
  - maintenance costs (61% vs. 68%),
  - skepticism about the effectiveness/usefulness of EHRs (45% vs. 52%),
  - lack of training/technical support (58% vs. 65%),
  - amount of time necessary to use the system (68% vs. 79%),
  - time/ability to incorporate old records into the new system (70% vs. 78%), and
  - poor return on investment (47% vs. 56%).

Results based on a synthesis analysis of the factors related to the treatment group practices' overall performance in the demonstration:

- Analyses of six practice-level composite measures, which were based on the practices' 26 reported measures, showed no consistent evidence that EHR adoption and use were linked to an improvement in quality. Although practices' level of EHR use was statistically significantly related to an improvement in the diabetes composite score, it was also statistically significantly related to a decline in the CAD intermediate composite score.
- Treatment group practices that adopted EHRs by 2007 (N=178), which was the beginning of the demonstration, had lower quality scores during Year 3 than those that did not (N=79) for three of the six composite measures: the CAD process, CHF process, and preventive services composite measures. However, practices that adopted EHRs by 2007 (N=170) increased their quality score over time for the proportion of CAD patients with LDL under control (<130 mg/dL) by 4 percentage points whereas those practices that had

not adopted an EHR by the end of the demonstration (N=36) had a 3 percentage-point decline.

- Analyses of the claims-based quality measures provided some support for the hypothesis that health IT adoption and use improves care quality.
    - 2 of 20 claims-based measures showed a statistically significant relationship between EHR adoption and quality. Compared to beneficiaries with diabetes served by practices that did not use EHRs, beneficiaries with diabetes served by practices that did use EHRs were 5 percentage points more likely to receive a urine test and 3 percentage points more likely to receive a dilated eye exam (N=34,029).
    - 6 of 20 claims-based measures showed a statistically significant relationship between greater EHR use and quality. Greater EHR use was positively associated with decreased preventable hospitalizations for diabetes (N=76,836), reduced hospitalizations for patients with chronic conditions (N=198,518), more frequent breast cancer screening (N=3,369), and more frequent urine tests and dilated eye exams for diabetes patients (N=34,029).
        - However, a higher level of EHR use was also associated with increased preventable hospitalizations for microvascular complications for patients with diabetes (N=76,836).
    - Additionally, two measures showed a statistically significant relationship between greater EHR use and reduced inpatient and total Medicare expenditures (N=198,518).
- Analysis of Year 1 demonstration payments and practices' quality scores did not show strong, consistent evidence to suggest that practices invested Year 1 payments or bonuses in quality improvements (N=491).
    - Practices that participated in the Physician Quality Reporting System improved their quality scores by 3 percentage points, whereas those that did not participate improved by only 1 percentage point.
    - Practices that received the Year 1 electronic reporting bonus improved their quality scores by 4 percentage points, compared to the 2-percentage-point improvement among practices that did not receive the bonus.
    - However, practices receiving a higher-than-average first-year payment per physician improved their overall quality score by 2 percentage points, while those that received a lower-than-average payment improved their score by 3 percentage points.
- For the synthesis analysis, the majority of the relationships tested were not statistically significant; therefore, the results of these analyses should be interpreted collectively and cautiously.

Service Use and Expenditures

Results based on a difference-in-differences analysis of the treatment group practices relative to a set of matched comparison group practices (N=1,190):

- There was no clear pattern of changes to Medicare utilization rates (only statistically significant differences are highlighted below).

  - Hospitalizations increased, but only during Year 3 for treatment group patients relative to the comparison group.
  - Emergency room use increased during Years 1 and 3, but not in Year 2, for treatment group patients relative to the comparison group, which was driven largely by increases in Utah.
  - Relative to the comparison group, treatment group patients' physician office visits increased during Years 1 and 2, but not in Year 3.
  - Relative to the comparison group, the treatment group patients' outpatient visits decreased during Years 1 and 2, but not in Year 3.
- Total Medicare expenditures statistically significantly increased for the treatment group patients relative to the comparison group in Year 2 ($315 per beneficiary per year) and Year 3 ($311 per beneficiary per year).
- Expenditures for all Part A services (excluding home health) statistically significantly increased for the treatment group patients relative to the comparison group in Year 2 (by $232) and Year 3 (by $275).
- Among Part B services, the demonstration was associated with a statistically significant increase ($70) in spending for physician services in Year 2.
- The two groups were not statistically significantly different in their expenditures for home health services.
- Incentive payments were more than $26 million; since the treatment group practices had higher costs than the comparison group (see above), the demonstration does not appear to have been budget neutral.
  - After accounting for incentive payments, the demonstration was associated with excess spending of approximately $163 million.

V. CONCLUSIONS

Overall, the MCMP demonstration helps to identify opportunities and challenges associated with pay-for-performance programs that may inform current and future value-based purchasing efforts. Cooperation in the MCMP was high among practices participating in the demonstration (treatment groups): three-quarters of treatment group practices participated by submitting quality measures during all three years and received incentive payments, and practices reported that the data submission process became less burdensome over time. Visited treatment group practices also reported improving aspects of care management through increased documentation and attention to the care guidelines, although many found electronic reporting of the quality measures to be a challenge.

Over the course of the demonstration, practices increased their use of health IT; among the subset of practices that completed both rounds of the OSS, the percentage of treatment group practices using an EHR rose from 68 percent in 2007 (the start of the demonstration) to 83 percent in 2010 (the end of the demonstration). The percentage of practices using EHRs to manage care also increased, with EHRs more frequently being used to review and act on reminders of care as well as to track patients with chronic conditions, though it is unclear whether these changes were in response to the demonstration. The changes that treatment group practices made to care management and to health IT use may have led to some improvements in the quality of care they provided, as the overall percentage of the time that practices followed

evidence-based guidelines rose from 81% to 83%. The quality scores for 5 of the 26 quality-of-care measures statistically significantly rose by more than 3 percentage points, and none decreased by that amount. Nearly half the treatment group physicians said the demonstration improved their adherence to recommended clinical practice guidelines, and the vast majority believed the demonstration targeted important medical conditions, used well-accepted and appropriate measures of care quality, and required a reasonable level of effort.

Relative to physicians in comparison group practices, those in treatment group practices reported fewer barriers to adopting and using EHRs, but they did not report being more likely to use their EHRs, or to use specific features of their EHRs. Treatment group physicians rated care coordination higher along three of seven measures examined and were more satisfied with beneficiaries' receipt of recommended services relative to the comparison group physicians. However, treatment group patients were no more likely to report receiving the clinical interventions related to the demonstration's quality process measures. Satisfaction with quality and coordination of care was widespread among both beneficiary groups; however, the treatment group patients were somewhat less likely to be satisfied than the comparison group patients. For measures where the treatment group was less likely to report being satisfied with care, the results were driven mostly by responses from California and its comparison states (i.e., Arizona, Oregon, and Washington). There was no baseline survey, so it is unclear whether beneficiary satisfaction changed due to the demonstration or some other aspect of the medical practices or health care policy in the state.

In claims-based analyses of the change in quality measures over the demonstration, there were few treatment-comparison differences; for one measure in Year 1 and Year 2 and two measures in Year 3, the comparison group's performance improved more than the treatment group's. Similarly, for preventable hospitalization measures, the few statistically significant changes favored the comparison group. Relative to the comparison group, Medicare costs increased during Years 2 and 3 for the treatment group. Because treatment group practices had higher (rather than lower) costs than comparison group practices during Years 2 and 3, the demonstration does not appear to have been budget neutral. Rather, the difference-in-differences analysis suggests that the demonstration was associated with excess spending of $163 million, after accounting for incentive payments.

Because Medicare costs fluctuate greatly over time and there is no reason to expect a pay-for-performance demonstration to lead to a decline in quality of care, especially for services or processes that are directly incentivized, the few statistically significant results for Medicare cost and service use may have been random. That is, the observed treatment-control differences may have been driven by chance fluctuations in service use and costs, and not due to the MCMP demonstration itself. It is also possible that factors in the treatment and comparison groups' environments—such as private-sector pay-for-performance programs, medical home pilots, secular trends in health IT adoption and use, and a culture of change among groups or systems in many of the hospitals in which physicians were affiliated—influenced the evaluation results, which makes it difficult to draw definitive conclusions. Data on these factors were not available and could not be accounted for in the analysis.

Findings from the site visits suggest that incentive payments might not have been large enough, or disbursed promptly enough, to induce practices to improve their quality scores. Due to the annual nature of the quality measures and lags in Medicare claims data, the 1-year interval between when practices submitted quality scores and when they received incentive payments may have prevented practices from acting on their quality scores in ways that could be reflected

in subsequent years' quality scores. It is possible that a stronger relationship between health IT use and care quality will emerge over time, after practices have become accustomed to using the new technology, or as improved, more user-friendly EHR systems become available to small- and medium-sized primary care practices.

The evaluation has several notable limitations. First, it used a quasi-experimental, rather than experimental, design. Thus, it could not control fully for practice characteristics, the environment and other programs in each of the states that changed over time, or natural fluctuations in Medicare service use and costs. Second, it would not be appropriate to generalize from this evaluation to a national level, as the demonstration was conducted in only four states. Importantly, the results presented above are pooled across the states, but a fair amount of variation existed between the states. Additionally, both treatment and comparison group practices already had experience with health IT due to their participation in the DOQ-IT initiative. Finally, only a small proportion of treatment group practices were included in site visits, which limits the ability to make robust statements about participating practices' experiences in the demonstration.

Future policies and programs might be able to generate greater improvements in quality of care than were generated by the MCMP demonstration by providing practices with stronger incentives, timely feedback to their performance, and more technical assistance to effectively use EHRs. In general, national policy has been moving toward reducing the key barriers that were found in the implementation analysis by (1) providing higher incentive payments than the MCMP demonstration that may help offset the high costs of adopting and upgrading EHR systems (American Recovery and Reinvestment Act of 2009 [ARRA]-funded Medicare and Medicaid EHR Incentive Programs); (2) improving the availability of technical support (ARRA-funded Regional Extension Centers); and (3) taking steps to require that available EHR products meet high standards for capability, functionality, and security through a rigorous certification process (ARRA-mandated EHR Certification regulations and ARRA-funded testing tools). The MCMP demonstration and evaluation provides confirmation that sustained implementation of these more recent strategies may be needed to promote effective use of health IT in small- and medium-sized primary care practices.

REFERENCES

Wilkin, John C., C. William Wrightson, David Knutson, Erika G. Yoshino, Anahita S. Taylor, and Kerry E. Moroz. 2007. *Medicare Care Management Performance Demonstration Design Report*. Columbia, MD: Actuarial Research Corporation.

APPENDIX A

SEC. 649. MEDICARE CARE MANAGEMENT PERFORMANCE DEMONSTRATION

(a) ESTABLISHMENT.

(1) IN GENERAL.—The Secretary shall establish a pay-for-performance demonstration program with physicians to meet the needs of eligible beneficiaries through the adoption and use of health information technology and evidence-based outcomes measures for

(A) promoting continuity of care;

(B) helping stabilize medical conditions;

(C) preventing or minimizing acute exacerbations of chronic conditions; and

(D) reducing adverse health outcomes, such as adverse drug interactions related to polypharmacy.

(2) SITES.—The Secretary shall designate no more than 4 sites at which to conduct the demonstration program under this section, of which

(A) 2 shall be in an urban area;

(B) 1 shall be in a rural area; and

(C) 1 shall be in a State with a medical school with a Department of Geriatrics that manages rural outreach sites and is capable of managing patients with multiple chronic conditions, one of which is dementia.

(3) DURATION.—The Secretary shall conduct the demonstration program under this section for a 3-year period.

(4) CONSULTATION.—In carrying out the demonstration program under this section, the Secretary shall consult with private sector and non-profit groups that are under taking similar efforts to improve quality and reduce avoidable hospitalizations for chronically ill patients.

(b) PARTICIPATION.

(1) IN GENERAL.—A physician who provides care for minimum number of eligible beneficiaries (as specified by the Secretary) may participate in the demonstration program under this section if such physician agrees, to phase-in over the course of the 3-year demonstration period and with the assistance provided under subsection (d)(2)

(A) the use of health information technology to manage the clinical care of eligible beneficiaries consistent with paragraph (3); and

(B) the electronic reporting of clinical quality and outcomes measures in accordance with requirements established by the Secretary under the demonstration program.

(2) SPECIAL RULE.—In the case of the sites referred to in subparagraphs (B) and (C) of subsection (a)(2), a physician who provides care for a minimum number of beneficiaries with two or more chronic conditions, including dementia (as specified by the Secretary), may

participate in the program under this section if such physician agrees to the requirements in subparagraphs (A) and (B) of paragraph (1).

(3) PRACTICE STANDARDS.—Each physician participating in the demonstration program under this section must demonstrate the ability

(A) to assess each eligible beneficiary for conditions other than chronic conditions, such as impaired cognitive ability and co-morbidities, for the purposes of developing care management requirements;

(B) to serve as the primary contact of eligible beneficiaries in accessing items and services for which payment may be made under the Medicare program;

(C) to establish and maintain health care information system for such beneficiaries;

(D) to promote continuity of care across providers and settings;

(E) to use evidence-based guidelines and meet such clinical quality and outcome measures as the Secretary shall require;

(F) to promote self-care through the provision of patient education and support for patients or, where appropriate, family caregivers;

(G) when appropriate, to refer such beneficiaries to community service organizations; and

(H) to meet such other complex care management requirements as the Secretary may specify.

The guidelines and measures required under subparagraph (E) shall be designed to take into account beneficiaries with multiple chronic conditions.

(c) PAYMENT METHODOLOGY.—Under the demonstration program under this section the Secretary shall pay a per beneficiary amount to each participating physician who meets or exceeds specific performance standards established by the Secretary with respect to the clinical quality and outcome measures reported under subsection (b)(1)(B). Such amount may vary based on different levels of performance or improvement.

(d) ADMINISTRATION.

(1) USE OF QUALITY IMPROVEMENT ORGANIZATIONS.—The Secretary shall contract with quality improvement organizations or such other entities as the Secretary deems appropriate to enroll physicians and evaluate their performance under the demonstration program under this section.

(2) TECHNICAL ASSISTANCE.—The Secretary shall require in such contracts that the contractor be responsible for technical assistance and education as needed to physicians enrolled in the demonstration program under this section for the purpose of aiding their adoption of health information technology, meeting practice standards, and implementing required clinical and outcomes measures.

(e) FUNDING.

(1) IN GENERAL.—The Secretary shall provide for the transfer from the Federal Supplementary Medical Insurance Trust Fund established under section 1841 of the Social

Security Act (42 U.S.C. 1395t) of such funds as are necessary for the costs of carrying out the demonstration program under this section.

(2) BUDGET NEUTRALITY.—In conducting the demonstration program under this section, the Secretary shall ensure that the aggregate payments made by the Secretary do not exceed the amount which the Secretary estimates would have been paid if the demonstration program under this section was not implemented.

(f) WAIVER AUTHORITY.—The Secretary may waive such requirements of titles XI and XVIII of the Social Security Act (42 U.S.C. 1301 et seq.; 1395 et seq.) as may be necessary for the purpose of carrying out the demonstration program under this section.

(g) REPORT.—Not later than 12 months after the date of completion of the demonstration program under this section, the Secretary shall submit to Congress a report on such program, together with recommendations for such legislation and administrative action as the Secretary determines to be appropriate.

(h) DEFINITIONS.—In this section:

(1) ELIGIBLE BENEFICIARY.—The term ‘‘eligible beneficiary’’ means any individual who—

(A) is entitled to benefits under part A and enrolled for benefits under part B of title XVIII of the Social Security Act and is not enrolled in a plan under part C of such title; and

(B) has one or more chronic medical conditions specified by the Secretary (one of which may be cognitive impairment).

(2) HEALTH INFORMATION TECHNOLOGY.—The term ‘‘health information technology’’ means email communication, clinical alerts and reminders, and other information technology that meets such functionality, interoperability, and other standards as prescribed by the Secretary.

APPENDIX B

TABLE 1. Quality Measures Incentivized by the MCMP Demonstration

Whether Patients with CAD: Were prescribed antiplatelet therapy Were prescribed a lipid-lowering therapy Were prescribed beta-blocker therapy, among those with prior myocardial infarction Received a lipid profile Had most recent LDL cholesterol <130 mg/dl Were prescribed ACE inhibitor therapy, among those who also have diabetes and/or LVSD
Whether Patients with CHF: Had left ventricular function results recorded Had left ventricular ejection fraction tested (among those hospitalized with heart failure) Had weight measurement recorded Had patient education class on disease management and health behavior change during one or more visits within a six-month period Were prescribed beta-blocker therapy, among those who also have LVSD Were prescribed ACE inhibitor therapy, among those who also have LVSD Were prescribed warfarin therapy, among those with paroxysmal or chronic atrial fibrillation
Whether Patients with Diabetes: Had blood test for HbA1c Had most recent A1c level >9 percent Had blood pressure below 140/99 mm Hg Had LDL cholesterol test Had most recent LDL cholesterol <130 mg/dl Had test for microalbumin Had dilated retinal exam Had foot exam
Whether Patients with Specified Chronic Conditions Received Preventive Care Measures, Including: Blood pressure measurement during last office visit Breast cancer screening during current or previous year, among those younger than 69 Colorectal cancer screening during recommended period Influenza vaccination during September through February of year prior to measurement year, among those older than 50 Pneumonia vaccination, among those with a chronic condition older than 65

ACE = angiotensin converting enzyme; CAD = coronary artery disease; CHF = congestive heart failure; LVEF=left ventricular ejection fraction; LVSD = left ventricular systolic dysfunction; MCMP=Medicare Care Management Performance Demonstration; UTI = urinary tract infection.

TABLE 2. Claims-Based Measures

Quality Measures
Among Beneficiaries with CAD, Received: A lipid profile
Among Beneficiaries with Diabetes, Had: Blood test for hemoglobin A1c LDL cholesterol test Urine test for protein (microalbuminuria) Dilated retinal exam
Among Beneficiaries with CHF, Had: Test for left ventricular ejection fraction
Among Female Beneficiaries Aged 40 to 69, Had: Breast cancer screening
Claims-Based Drug Compliance Measures
Among Patients with CAD, Had a Claim for: Antiplatelet drugs Lipid-lowering drugs ACE inhibitors (among those with diabetes and/or LVSD)
Hospitalizations
Hospitalizations Any hospitalization Number of hospitalizations
Among Beneficiaries with CAD, Number of Hospitalizations Related to: Cardiac problems
Among Beneficiaries with Diabetes, Number of Hospitalizations Related to: Cardiac problems Diabetes Peripheral vascular or extremity complications Microvascular complications
Among Beneficiaries with CHF, Number of Hospitalizations Related to: Fluid/electrolyte problems CHF
Among Beneficiaries with Any Chronic Disease, Number of Hospitalizations for: Pneumonia or UTI
Expenditure Measures
Total expenditures
Inpatient expenditures

ACE = angiotensin converting enzyme; CAD = coronary artery disease; CHF = congestive heart failure; LVEF=left ventricular ejection fraction; LVSD = left ventricular systolic dysfunction; MCMP=Medicare Care Management Performance Demonstration; UTI = urinary tract infection.