

© Health Research and Educational Trust
DOI: 10.1111/j.1475-6773.2010.01166.x
RESEARCH ARTICLE (TH)

Chronic Care Improvement in Primary Care: Evaluation of an Integrated Pay-for-Performance and Practice-Based Care Coordination Program among Elderly Patients with Diabetes

Peter J. Fagan, Alyson B. Schuster, Cynthia Boyd, Jill A. Marsteller, Michael Griswold, Shannon M. E. Murphy, Linda Dunbar, and Christopher B. Forrest

Objective. To examine the effects of an intervention comprising (1) a practice-based care coordination program, (2) augmented by pay for performance (P4P) for meeting quality targets, and (3) complemented by a third-party disease management on quality of care and resource use for older adults with diabetes.

Data Sources/Study Setting. Claims files of a managed care organization (MCO) for 20,943 adults aged 65 and older with diabetes receiving care in Alabama, Tennessee, or Texas, from January 2004 to March 2007.

Study Design. A quasi-experimental, longitudinal study in which pre- and postdata from 1,587 patients in nine intervention primary care practices were evaluated against 19,356 patients in MCO comparison practices (>900). Five incentivized quality measures, two nonincentivized measures, and two resource-use measures were investigated. We examined trends and changes in trends from baseline to follow-up, contrasting intervention and comparison group member results.

Principal Findings. Quality of care generally improved for both groups during the study period. Only slight differences were seen between the intervention and comparison group trends and changes in trends over time.

Conclusions. This study did not generate evidence supporting a beneficial effect of an on-site care coordination intervention augmented by P4P and complemented by third-party disease management on diabetes quality or resource use.

Key Words. Pay for performance, care management, primary care, outcomes measurement, evaluation design, longitudinal study

There is little debate that realizing improvements in the quality of health care, especially for chronic conditions, will require changes in our physician

payment practices. There remains substantial controversy, however, in how to structure a payment system to align resource allocation with quality of care (IOM Committee on Quality of Health Care in America 2001). Much attention has been devoted to pay-for-performance (P4P) programs that link some portion of physician or practice compensation to the attainment of specific quality of care objectives (Rosenthal et al. 2004, 2005; Rosenthal and Frank 2006). P4P programs continue to expand in number and scope even though research on the effects of P4P on quality has produced mixed results (Rosenthal et al. 2005; Rosenthal and Frank 2006; Rosenthal 2007; Rosof, Sirio, and Kmetik 2008; Greene and Nash 2009). Health services leaders recommend P4P programs as part of the effort to address long-term health care spending growth (Antos et al. 2009).

In addition to restructuring financial incentives, improvement in the care of chronic conditions frequently involves disease management of affected patients (Fireman, Bartlett, and Selby 2004; Wagner 2004; Wagner and Reid 2007). Conventionally, disease management is provided by call center nurses who maintain telephonic contact with enrollees in order to assist them with the implementation of their chronic care regimens. While effort is made to have consistency between call center nurse and her patient panels, her contact with primary care providers is highly variable, and often not part of the disease management service mix. This lack of integration results in gaps in communication and coordination of care, both of which may have deleterious effects on the quality of care and patient outcomes.

In recent years, a “second generation” of disease management has sought to integrate the remote, telephonic services, such as those typically provided by nurse care managers, within the primary care medical home (Villagra 2004; Casalino 2005; American Academy of Family Physicians

Address correspondence to Peter J. Fagan, Ph.D., Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Johns Hopkins Healthcare LLC, 6704 Curtis Court, Glen Burnie, MD 21060; e-mail: pfagan@jhmi.edu. Alyson B. Schuster, M.P.H., M.B.A., Shannon M.E. Murphy, M.A., and Linda Dunbar, R.N., Ph.D., are with the Johns Hopkins Healthcare LLC, Glen Burnie, MD. Cynthia Boyd, M.D., M.P.H., is with the Division of Geriatric Medicine and Gerontology, Johns Hopkins University School of Medicine, Baltimore, MD. Jill A. Marsteller, Ph.D., M.P.P., is with the Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD. Michael Griswold, Ph.D., is with the University of Mississippi Medical Center, Jackson, MS. Christopher B. Forrest, M.D., Ph.D., is with the Children’s Hospital of Philadelphia, Philadelphia, PA.

2008). Co-locating the care management professional with the primary care clinicians may be a superior approach for improving care coordination, especially for patients with multimorbidity, which is the norm among elders (Boult et al. 2001; Boyd et al. 2005).

This article presents an evaluation of a chronic care improvement initiative that consisted of P4P practice-based care coordination augmented by third-party disease management with call center nurses dedicated to patients in intervention practices. The initiative (the “Intervention”) had three components: (1) a practice-based care coordination program using an on-site care coordinator (registered nurses [RNs], licensed practical nurses [LPNs], and medical assistants—a new position in the practices), (2) a P4P program in which practices were given a bonus payment for meeting specific quality indicator goals, and (3) a disease management program with call center nurses dedicated to patients in the intervention practices.

The aims of this paper were to examine the effects of the intervention on quality of care measures and resource use for patients with diabetes mellitus, the most common chronic condition that is care managed by third-party firms and for which there is an evidence base supporting its responsiveness to care management (Villagra and Ahmed 2004; Von Korff et al. 2005; Mangione et al. 2006; Weber and Neeser 2006). Moreover, diabetes is a condition that has behavioral change and care management potential in that it is relatively controllable by adherence to a medical regimen and lifestyle patterns that include healthy nutrition and activity (Hill-Briggs 2003).

This evaluation was guided by the following, general research question: “What is the intervention effect on incentivized quality measures, and what is the broader effect of the intervention on the overall quality of care for patients with diabetes?” Specifically, we examined the effects of the intervention on (1) quality of care for the incentivized care indicators for diabetes, (2) quality of care for nonincentivized care indicators for diabetes, and (3) utilization and medical costs incurred by the patients with diabetes in the intervention practices.

We hypothesized higher quality of medical care for enrollees with diabetes exposed to the intervention compared with others receiving usual primary care and conventional third-party disease management only. In examining the various nonincentivized measures, we sought to explore any positive halo effect of the intervention or any unintended consequences of the intervention such as decline in the nonincentivized quality of care measures (Casalino 1999; Forrest, Villagra, and Pope 2006).

METHODS

Intervention Design

Nine independent primary care practices responded to an initiative by a national MCO and a national disease management firm to collaborate on an intervention that consisted of P4P bonus payments for meeting quality of care measures, on-site care coordination, and dedicated call center disease management. The practices were selected by the MCO because they had leadership that was willing to champion locally the proposed quality improvement initiative. The practices, all physician owned, averaged 11.7 full-time employee (FTE) physicians (range 1–26). Four practices were multispecialty and five had electronic medical records (EMRs). Six practices were urban, and three were suburban. The intervention took place between August 2004 and February 2006.

The MCO and disease management organization agreed to place an on-site care coordinator in each intervention practice that was participating in the P4P care coordination program. A team of five call center nurses at the disease management company was assigned to serve only the intervention practices, responding to and initiating all calls to and from practices and their patients.

The on-site care coordinators were employed by the primary care practices, which were compensated for employee costs by the MCO. The role of the on-site coordinator was to alert physicians at the time of the patient's appointment to quality improvement opportunities, specifically for the P4P quality measures. Some of the alert information originated from the disease management call center nurses, and other information was collected by the on-site coordinator from the patient's medical record. If required, the on-site coordinator requested records from specialists or hospitals to confirm whether the patient had received certain services (i.e., eye exam, Hemoglobin A1c [HbA1c] test). An alert sheet for the visit provided the physician with patient-level information, demographics, lab tests due, new medications, treatment goals, etc.

At the conclusion of the visit, the on-site coordinator role was to review and convey appropriate information to the disease management call center nurses so that they could follow up with the member for relevant issues. For example, the on-site coordinator may have informed the call center nurse that the member was referred for an eye exam so that the call center nurse could contact the patient to ensure the eye exam had been completed. In essence, the on-site coordinators served as liaisons between the primary care clinicians and the call center nurses who promoted the disease management company treatment protocols directly with the members.

Pay for Performance

The maximum amount of the P4P bonus payments was 20 percent of the capitation fee for Medicare MCO patients in the physicians' panels. In none of the practices did the MCO enrollees exceed 10 percent of the practices' total number of patients. Thus, the P4P bonus represented a small percentage of the total annual practice revenue.

The MCO proposed 25 quality measures, which included indicators for preventive services, diabetes mellitus, congestive heart failure, coronary artery disease, and chronic obstructive pulmonary disease. These last four were the chronic conditions that the disease management company managed for the MCO population in both the intervention and the comparison practices. We limited our sample to patients with diabetes to ensure adequate disease-specific sample size, a sufficient number of metrics that could be constructed using claims data, and a robust evidence base for improving the care and outcomes of patients.

The MCO suggested and the participating practices agreed to a percentage goal for each measure. This allowed for variation in the practices' baselines and an opportunity for "buy-in" for the practices participating in the P4P care coordination program. Bonus payments were linked to the proportion of quality targets met. Practices were also offered an additional bonus if the medical loss ratio of the MCO enrollees decreased, sharing 50/50 with the MCO up to 33 percent of the savings.

Study Samples

The study design was quasi-experimental, involving an intervention group composed of all patients 65 and older with diabetes in the nine participating practices from Alabama, Tennessee, and Texas ($N = 1,587$), and a comparison group from the MCO that included other patients in those same states during the same time period who were not exposed to the intervention but retained a call center disease management program ($N = 19,356$). Diabetes status was determined by HEDIS 2005 criteria that includes diabetes diagnosis codes (ICD-9-CM and DRGs) and well as CPT and UB-92 revenue codes for outpatient/nonacute inpatient and acute inpatient/emergency services (National Committee for Quality Assurance [NCQA] 2004).

Data Sources

The data sources for the study were health plan administrative claims (medical and pharmacy) and enrollment files. Behavioral health claims were carved out

from the standard benefits. Although the study period ranged from January 2004 to March 2007, we used claims data from January 2003 to allow for an additional 12-month look-back period to assess the quality indicators during the 24-month preintervention period. Eight of the nine intervention sites began the P4P care coordination program between January and February 2006. For these practices, we had at least 24 months of preintervention (baseline) data and at least 12 months of postintervention data. The ninth site was a “pilot” P4P practice and started in August 2004. Repeated monthly observations were available for each participant while enrolled in the MCO during the study period.

Variable Construction

Independent Variable. The independent variable of interest was exposure to the intervention, a dichotomous indicator of whether the member received medical care in one of the nine P4P intervention practices or one of the comparison practices.

Time. Time was represented as the number of months since the start of the care coordination intervention. Time ranged from -25 to $+14$, with month 0 representing the starting intervention month in each site. For comparison group members, the 0 month was set at January 2006, representing the modal starting calendar month for the intervention sites. Two linear spline terms were included to allow changes in time trends across approximately yearly intervals (two baseline years and one follow-up year). The 12-month time frame was chosen as a substantively important period that was flexible enough to fit the data well. Regression discontinuity models and models with 6-month linear splines were also examined in sensitivity analyses.

Dependent Measures

Measurements of Care Quality. Using an intention-to-treat model, we examined the intervention effect on enrollees with diabetes using seven disease-specific quality measures and two measures of resource use: emergency department (ED) utilization and total health care costs. The quality of care measures included those that (1) coincided precisely with the P4P reward metrics for diabetes care and prevention, (2) were endorsed and promulgated by national health care organizations, and (3) could be operationalized with administrative and claims data.

Five of the quality measures were incentivized by the P4P care coordination program: influenza vaccine, HbA1c testing, eye exam, low-density lipoprotein (LDL) screening, and nephropathy screening (NCQA 2004). In addition, there were two nonincentivized prescription process indicators: (1) avoiding (not prescribing) short-acting antihypertensive medications (Fick et al. 2003; Furmaga et al. 2005); and (2) prescribing an angiotensin converting enzyme (ACE)/angiotensin receptor blocker (ARB) medication for diabetics with renal insufficiency (American Diabetes Association 2009).

The quality outcome measures were dichotomous variables (1,0) signifying whether the quality indicator had been met for an eligible patient during the specific “look-back” period during which the quality measure should have been met. The look-back period for all the indicators was 12 months, as recommended by national clinical guidelines. For example, HEDIS technical specifications call for patients with diabetes to receive an LDL screening at least every 12 months (NCQA 2004).

In addition to the quality of care measures, we also examined the impact of the intervention on ED utilization and total paid costs by the MCO. ED utilization was chosen because both short- and long-term complications of diabetes have been used as preventable quality measures for both ED use and hospital admission (Grensenz, Lurie, and Ruder 2009). ED utilization that did not result in an admission was coded as a dichotomous indicator of whether the member had visited the ED in the previous 12 months. This operationalization was chosen due to the very low number of participants (1 percent) with any ED visits during the period of analysis.

Per member per month costs were calculated by summing all the paid medical (inpatient and outpatient) and pharmacy claims (allowed dollars) incurred by each member in a given month. Members’ medical costs are necessary information to determine the relative value of the health care services that the MCO members were receiving, including disease management, during the study period.

Statistical Analysis

Multivariate regression techniques using generalized estimating equations to account for clustering were used to assess the impact of the intervention on the outcome variables. Logistic models were used for analyzing quality indicator outcomes and ED utilization, while cost outcomes were analyzed using a gamma distribution and log-link to account for skewness. Effects of the

intervention were evaluated similarly to a difference in differences analysis, using relative differences between intervention and comparison groups in outcome trend changes. Intervention effect odds ratios (ORs) and relative cost ratios (RRs) were estimated using interactions between the time trend spline terms and the intervention exposure binary variable. In each analysis, we adjusted for age, gender, health risk (as measured by the ACG Dx Rx PM score) and state in the model to control against baseline group differences (Forrest et al. 2009). For a detailed description of the statistical analyses, see Appendix SA2. The Johns Hopkins University School of Medicine Institutional Review Board approved this research.

RESULTS

Study participants' average age was 74 years at the end of the study period or during the last month they were enrolled in the MCO and 58 percent were female (Table 1). The most prevalent comorbid conditions were coronary artery disease (75.6 percent), chronic obstructive pulmonary disease (36.6 percent), and congestive heart failure (27.6 percent). The morbidity burden

Table 1: Characteristics of Members in the Intervention and Comparison Samples

<i>Characteristics</i>	<i>Study Population</i>			<i>Baseline Differences p-Value</i>
	<i>Intervention</i>	<i>Comparison</i>	<i>Total</i>	
Members	1,587	19,356	20,943	
Mean months enrolled during study period (in months)	19.5	17.0	17.1	$p < .0001$
Mean age (in years)	74.6	73.9	74.0	$p < .0001$
Percent female	57.8	58.4	58.4	$p = .6103$
Morbidity (%)				
Diabetes only (%)	10.1	17.5	17.0	$p < .0001$
Chronic obstructive pulmonary disease	37.8	36.4	36.6	$p = .2751$
Congestive heart failure	31.3	27.3	27.6	$p < .0005$
Coronary artery disease	84.4	74.8	75.6	$p < .0001$
All four index conditions	17.8	15.1	15.3	$p = .0051$
ACG Dx Rx PM risk score	0.13	0.10	0.10	$p < .0001$
State of residence (%)				
Alabama	35.2	24.0	24.9	$p < .0001$
Tennessee	41.2	38.4	38.6	$p < .0001$
Texas	23.6	37.6	36.5	$p < .0001$

(Dx Rx PM) of the intervention group was higher ($p < .0001$) than the comparison group (56.2 percent).

Our primary hypothesis was that medical care, as measured by P4P-incentivized quality indicators, would be of higher quality for patients with diabetes exposed to the intervention compared with their nonexposed diabetic counterparts. Figures 1–4 illustrate the observed data and estimated changes in trends over time for each of the outcomes assessed. The 12-month linear spline models appear to fit the data adequately for all outcomes except the influenza vaccination quality measure, for which we report the 6-month spline model results. Table 2 provides a summary of the analysis results, with OR and RR representing the changes in trends over time for binary outcomes and costs, respectively.

As an illustrative example for interpretations on Figure 1 and Table 2, consider the LDL screening quality measure. Figure 1 shows that both the intervention and comparison groups were improving in the initial preintervention period (months -25 to -12). In the 12-month period just before implementation (months -12 to 0), the intervention group LDL screening

Figure 1: Monthly Trend for Incentivized Quality Indicators: Low-Density Lipoprotein Screening, Influenza Vaccine, and Hemoglobin A1c Testing

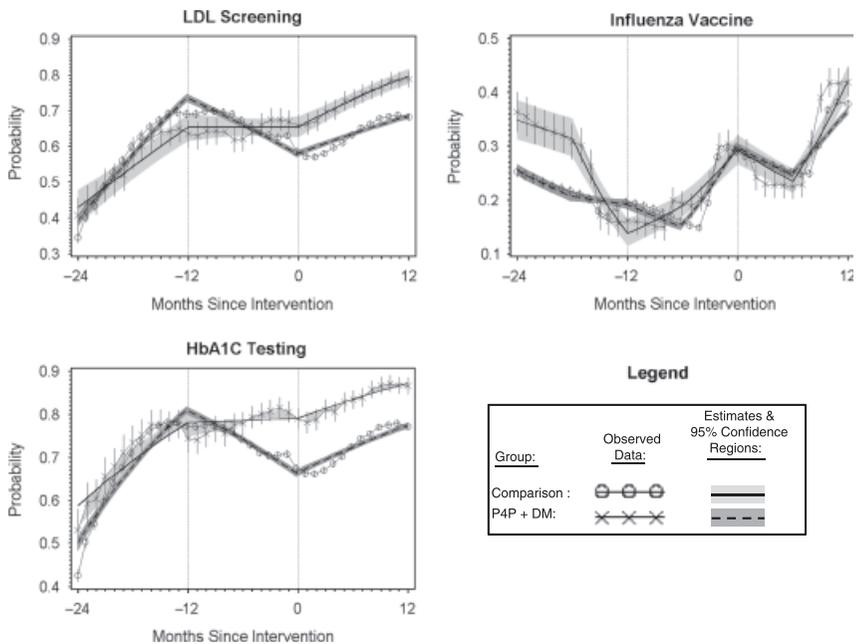
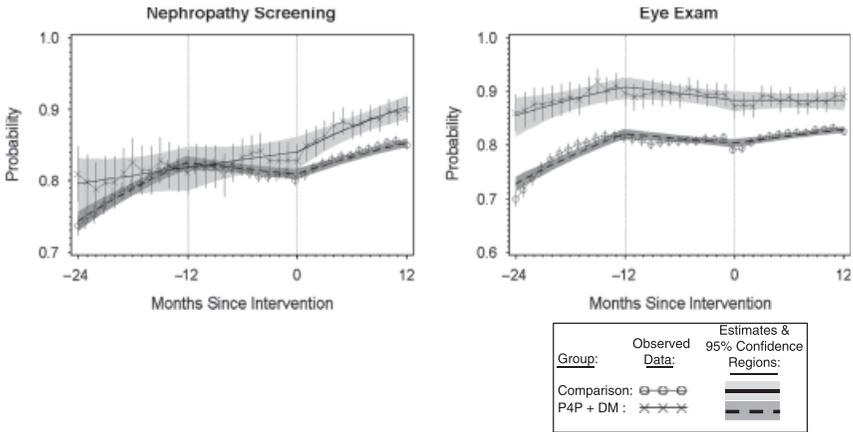


Figure 2: Monthly Trend for Incentivized Quality Indicators: Nephropathy Screening and Eye Exam



rates plateaued and rates in the comparison group actually decreased. During the intervention period (months 0–14) both groups again saw improvements in LDL screening rates. The improvement over the observed – 12 to 0 month plateau for the intervention group was a doubling in the odds of LDL screening (OR = 1.99, 95 percent CI [1.44, 2.75]) while the intervention period

Figure 3: Monthly Trend for Nonincentivized Quality Indicators: ACE Script for Those with Renal Insufficiency and No Short-Acting Medications

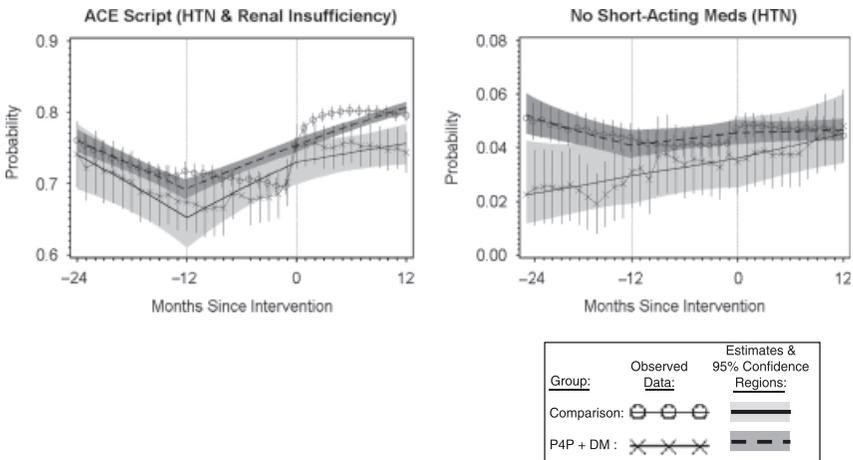
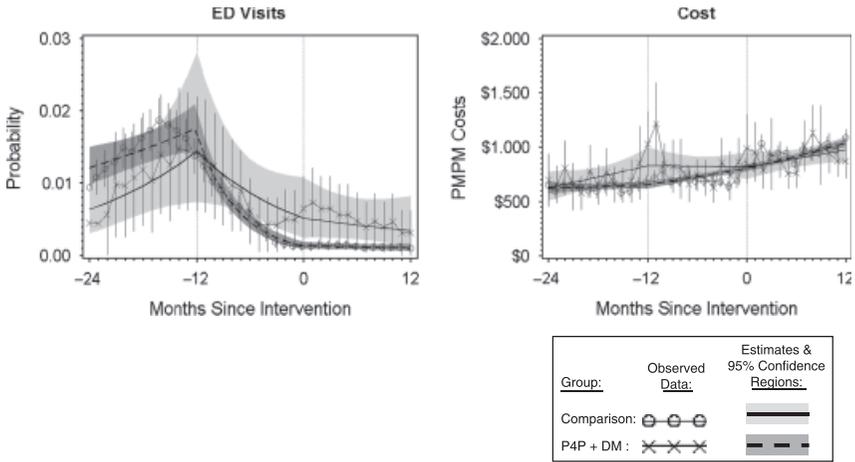


Figure 4: Monthly Trend for Emergency Department Visits and Per Member Per Month Costs



improvement over the observed – 12 to 0 month decrease for the comparison group was a threefold increase in the odds of screening (OR = 3.23 [2.94, 3.55]). Thus, the improvement in LDL screening for the intervention group was 40 percent smaller ($p = .005$) than the improvement for the comparison group (1.99/3.23); treatment effect OR = 0.62 (0.44, 0.86).

If one compares trends in influenza vaccination from the 6-month period just before the intervention period (months – 6 to 0) with trends in the latter half of the intervention period (months 7–14), the estimated improvement in the intervention group is approximately 80 percent greater; OR = 1.79 (1.37, 2.35), as shown in Table 2. However, as Figure 1 illustrates, both intervention and comparison groups appear to follow similar influenza vaccination trajectories across the entire study period and the potential intervention effect appears to be attributed to a slight slowing of the improvement in the comparison group, rather than a substantive increase in the improvement of the intervention group. This 6-month comparison for influenza vaccination was the only component of the five incentivized quality measures where any positive intervention effect was observed.

For the remaining four incentivized measures (HbA1c testing, nephropathy screening, eye examination, and LDL screening), both groups had absolute increases in the probability of meeting the quality measures in the follow-up period (see Figures 1 and 2, and Table 2). When comparing the differences between the intervention and the comparison groups in the

Table 2: Results Comparing Baseline to Follow-Up for Each Study Group and the Relative Change between Study Groups across Study Periods

<i>Measures</i>	<i>Intervention Group</i>	<i>Comparison Group</i>	<i>Intervention Relative to Comparison</i>
Pay-for-performance (P4P)-incentivized quality indicators			
Influenza vaccine in last 12 months among diabetics*	↑ Improved OR = 1.30 95% CI (1.01, 1.68)	↓ Worsened OR = 0.73 95% CI (0.67, 0.79)	↑ Positive effect OR = 1.79 95% CI (1.37, 2.35)
Hemoglobin A1c testing in last 12 months among diabetics	↑ Improved OR = 1.66 95% CI (1.14, 2.43)	↑ Improved OR = 3.76 95% CI (3.42, 4.13)	↓ Negative effect OR = 0.44 95% CI (0.30, 0.65)
Eye exam in last 12 months among diabetics	↔ No evidence of change OR = 1.33 95% CI (0.84, 2.11)	↑ Improved OR = 1.36 95% CI (1.22, 1.51)	↔ No evidence of effect OR = 0.98 95% CI (0.61, 1.58)
Low-density lipoprotein screen in last 12 months among diabetics	↑ Improved OR = 1.99 95% CI (1.44, 2.75)	↑ Improved OR = 3.23 95% CI (2.94, 3.55)	↓ Negative effect OR = 0.62 95% CI (0.44, 0.86)
Nephropathy screening in last 12 months among diabetics	↔ No evidence of change OR = 1.45 95% CI (0.95, 2.19)	↑ Improved OR = 1.52 95% CI (1.36, 1.70)	↔ No evidence of effect OR = 0.95 95% CI (0.62, 1.46)
Quality indicators not P4P-incentivized			
Script for ACE among hypertensive members with renal insufficiency and diabetes	↔ No evidence of change OR = 0.76 95% CI (0.55, 1.04)	↔ No evidence of change OR = 1.00 95% CI (0.89, 1.12)	↔ No evidence of effect OR = 0.76 95% CI (0.54, 1.06)
No short-acting meds among hypertensive members with diabetes [†]	↔ No evidence of change OR = 1.02 95% CI (0.55, 1.91)	↔ No evidence of change OR = 0.92 95% CI (0.77, 1.10)	↔ No evidence of effect OR = 1.11 95% CI (0.58, 2.13)
Utilization and cost measures			
Emergency department visits among diabetics [†]	↔ No evidence of change OR = 1.79 95% CI (0.29, 10.92)	↓ Worsened OR = 10.46 95% CI (4.62, 23.66)	↔ No evidence of effect OR = 0.17 95% CI (0.02, 1.5)
Total medical costs	↔ No evidence of change RR = 1.20 95% CI (0.87, 1.66)	↔ No evidence of change RR = 1.05 95% CI (0.96, 1.14)	↔ No evidence of effect RR = 1.15 95% CI (0.82, 1.61)

Notes. ↑, significant increase in odds across study periods; ↓, significant decrease in odds; ↔, no significant change: $\alpha = .05$; OR, odds ratio.

*Results are displayed for the comparison between the last 6 months of baseline period and the last 6 months of the follow-up period.

[†]A decrease is an improvement in quality (i.e., positive intervention effect).

change from baseline to follow-up, there was either no significant difference between the intervention and comparison groups (nephropathy screening and eye exam), or the improvement was significantly lower for the intervention sites (HbA1c testing, $p < .0001$; LDL screening, $p < .01$).

The second research question asked, “Does the intervention provide higher quality of care for *nonincentivized* quality indicators for diabetic patients than that received in comparison practices with standard call center-based DM?” Using filled prescriptions as a marker for written prescriptions, two prescription measures were used to examine the question (see Figure 3). There was no evidence of an intervention main effect between the patients of the intervention and comparison practices in either avoiding short-acting antihypertensive medication (OR = 1.11 [0.58, 2.13]) or prescribing an ACE for those with renal insufficiency (OR = 0.76 [0.54, 1.06]).

The third and final research question asked whether the intervention affected resource use as measured by (1) ED utilization and (2) total medical costs (see Figure 4). For ED utilization, both groups’ trends demonstrated a decreasing likelihood of a visit during each of the study periods. The change in odds from baseline to follow-up was not significantly different between the groups (OR = 0.17 [0.02, 1.25]), although the comparison group’s 105 percent increase (from OR = 0.08 at baseline to 0.84 at follow-up) was statistically significant (OR = 10.46 [4.62, 23.66]).

The analysis of the average total medical cost trends per member per month indicated that there was no significant difference ($p = .42$) between intervention and comparison members with diabetes on the change in cost trend from baseline to follow-up. The expected cost trend for members in the intervention group demonstrated a slightly (15 percent) higher yearly increase from baseline to follow-up than found for the comparison group, although this difference was not statistically significant (RR = 1.15 [0.82, 1.61]).

DISCUSSION

This study evaluated the impact of a novel approach for improving the care of patients with diabetes consisting of a practice-based care coordination program augmented with P4P and complemented with a dedicated (to the practice) third-party disease management program. Our evaluation compared the quality of care and resource use between members with diabetes exposed to the intervention with nonexposed counterparts. We found that patients with diabetes in both the intervention and comparison practices experienced

improvements in the quality of diabetes care across a range of indicators. This upward quality trend could be due to nonspecific secular effects on all these practices and/or an effect of the third-party disease management on diabetes care that was available in both groups (Agency for Healthcare Research and Quality 2005).

In addition, we were unable to detect a consistent effect of the intervention on diabetes care or resource use. Similar to other studies on P4P, our results yielded weak and mixed effects on a set of quality indicators, with only one out of five incentivized measures demonstrating significant improvement. With MCO patients constituting less than 10 percent of the intervention practices' patient panels, this study lends further support to the conclusion that P4P bonus payments that are small relative to total physician income may not be salient enough to shape physician and office practice behavior (Dudley 2005).

Lastly, the fact that there was no difference between the two study groups in the nonincentivized measures of quality of care, utilization, and cost measures supports the conclusion that there was no evidence of a positive halo effect of the P4P incentives observed in nonincentivized measures, nor negative unintended consequences associated with the intervention. In all these conclusions, we recognize that measuring quality indicators is not equivalent to measuring the improvement of health status of a population.

The major limitation of the study is the lack of data available about the comparison practices ($N > 900$), preventing an examination of the representativeness of the comparison group as a valid counterfactual group for the intervention practices' patients. There was some evidence of baseline differences between the groups in terms of morbidity, age, length of health plan enrollment. We suspect that the higher morbidity burden in the intervention group might be due to heightened surveillance and coding before the intervention. We proceeded with the comparison, however, relying on the fact that the comparison group was composed of nonintervention MCO patients ($N = 19,356$) in the three states and we made adjustments for age, gender, health risk, and state in the analyses to control for baseline differences.

One of the strengths of this analysis is the longitudinal evaluation of the intervention. A longitudinal statistical model with a comparison group protects inferences against historical, maturation, and regression to the mean threats to validity, which are inherent problems of observational studies with simple pre-post-measurement with no comparison group (Wilson and MacDowell 2003; Linden 2006). For example, if we examined a pre-post-point estimate analysis of LDL screening for diabetic patients in the

intervention group only, we would have concluded a significant treatment effect. Longitudinal data analysis with a comparison group prevents this interpretation from going unchallenged.

The data available for the study were limited to administrative claims and membership files. Researchers have used claims data with apparent success (Parente et al. 2004; Wennberg et al. 2004; Solberg et al. 2006; Gilmore et al. 2007). Others have cited the risk of claims and electronic records data underestimating quality measures (Baker et al. 2007; Pawlson, Scholle, and Powers 2007) and have suggested criteria such as aggregating to the practice level for the use of claims data (Fuhlbrigge et al. 2008).

The optimal quality standards for older adults with chronic diseases, and particularly those with multiple conditions, are still under development. While claims-based measurements have limitations and do not address what was offered to patients and the extent of shared decision making in a clinical encounter, they do provide worthwhile information about these processes. We judged that in this study the data would be comparable for both intervention and comparison practices because the data came from the same MCO administrative claims data warehouse. In addition, our use of multiple repeated measures to some extent avoids the underestimations of single-point frequencies in pre–post-analyses that use claims data only.

We agree with those who argue that the context of an intervention or quality improvement initiative should be described and, where warranted, associated with the effects (Pawson and Tilley 1997; Christianson 2007). Accounts of practice “champions” who were effective leaders in integrating both the P4P and the quality improvement initiative would improve our understanding of the probable mechanism of causality. In the course of conducting this study, we heard anecdotes that described such individuals in the intervention practices. Similarly, practice factors such as number of physicians and patient–physician ratio, support staff, EMR use, and within-state geographic variability that impedes or enables access to services are factors that would have contributed to our understanding of the findings (Beich et al. 2006). However, the number of practices in the tri-state universe of those that treated MCO enrollees made these data impossible to obtain with the resources available.

Our interviews with the on-site coordinators and the call center nurses indicated that there was much variation from the original protocol for the on-site coordinators and for the working relationship between the practices and the call center (Marsteller et al. 2008). There was considerable variation in structure among intervention practice sites regarding how the coordinators functioned within the practice and with the call center nurses. In our inter-

views with the on-site coordinators, some reported a high level of job satisfaction while others said that the lack of clarity about their role posed adjustment challenges. The lack of a consistent functioning of the on-site coordinator across practices may have contributed to this lack of intervention effect.

In summary, our study did not lend support to the use of on-site care coordination augmented by P4P financial incentives and complemented with third-party disease management as a means to improve quality of care for older patients with diabetes. Further research involving interventions with clearer coordinator roles and/or larger financial incentives for physicians may demonstrate an improved intervention effect. The benefit of P4P payment systems may also differ for care measures related to other chronic conditions, although our preliminary findings for measures related to care for chronic obstructive pulmonary disease, coronary artery disease, and chronic heart failure were similar to those for diabetes. The positive lessons reinforced by this study related to design and analysis are (1) using comparison groups in the evaluation of quality improvement initiatives, especially those that involve efforts to improve the value (quality/cost) of health care services, is essential; and (2) using longitudinal data analyses with their helpful graphics rather than point estimate comparisons contributes to the valid interpretation of the data.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Paula Norman, B.S., for the extensive data cleaning, data definitions, and management—enabling MCO claims data to be used in analytical databases. Melissa Sherry, B.S., B.A., for her administrative assistance in the latter stages of the research project.

Disclosures: The Health Industry Forum (HIF), Heller School for Social Policy, and Management, Brandeis University provided funding for this study.

The MCO provided funding for the stipends (U.S.\$50) for 48 physicians who completed an online survey.

Linda Dunbar serves on an Advisory Board of the disease management company.

While Johns Hopkins Medicine International, LLC, has a multiyear contract with the disease management company used in this article, this study was not a product of that contract and the disease management company played no role in the design, analysis, or report of the study.

Disclaimers: None.

REFERENCES

- Agency for Healthcare Research and Quality. 2005. "National Healthcare Quality Report, 2005" [accessed August 19, 2010]. Available at <http://www.ahrq.gov/qual/nhqr05/nhqr05.htm>
- American Academy of Family Physicians. 2008. "Joint Principles of the Patient-Centered Medical Home." *Delaware Medical Journal* 1: 21–2.
- American Diabetes Association. 2009. "Standards of Medical Care in Diabetes." *Diabetes Care* 32 (S1): S13–61.
- Antos, J., J. Bertko, M. Chernew, D. Cutler, D. Goldman, M. McClellan, E. McGlynn, M. Pauly, L. Schaeffer, and S. Shortell. 2009. *Bending the Curve: Effective Steps to Address Long-Term Health Care Spending Growth*. Washington, DC: Engelberg Center for Health Care Reform.
- Baker, D. W., S. D. Persell, J. A. Thompson, N. S. Soman, K. M. Burgner, D. Liss, and K. S. Kmetik. 2007. "Automated Review of Electronic Health Records to Assess Quality of Care for Outpatients with Heart Failure." *Annals of Internal Medicine* 146 (4): 270–7.
- Beich, J., D. P. Scanlon, J. Ulbrecht, E. W. Ford, and I. A. Ibrahim. 2006. "The Role of Disease Management in Pay-for-Performance Programs for Improving the Care of Chronically Ill Patients." *Medical Care Research and Review* 63 (1): 96S–116S.
- Boult, C., L. Boult, L. Morishita, B. Dowd, R. L. Krane, and C. F. Urdangarin. 2001. "A Randomized Clinical Trial of Outpatient Geriatric Evaluation and Management." *Journal of the American Geriatric Society* 49: 351–9.
- Boyd, C. M., J. Darer, C. Boult, L. P. Fried, L. Boult, and A. W. Wu. 2005. "Clinical Practice Guidelines and Quality of Care for Older Patients with Multiple Comorbid Diseases: Implications for Pay for Performance." *The Journal of the American Medical Association* 294 (6): 716–24.
- Casalino, L. P. 1999. "The Unintended Consequences of Measuring Quality on the Quality of Medical Care." *The New England Journal of Medicine* 341 (15): 1147–50.
- . 2005. "Disease Management and the Organization of Physician Practice." *The Journal of the American Medical Association* 293 (4): 485–8.
- Christianson, J. B. 2007. "Evaluating Pay-for-Performance in Medicaid through Real-World Observation." *Health Affairs (Project Hope)* 26 (4): w528–31.
- Diggle, P. J., P. Heagerty, K. Liang, and S. L. Zeger. 2002. *Analysis of Longitudinal Data*. 2nd Edition. New York: Oxford University Press.
- Dudley, R. A. 2005. "Pay-for-Performance Research: How to Learn What Clinicians and Policy Makers Need to Know." *The Journal of the American Medical Association* 294 (14): 1821–3.
- Fick, D. M., J. W. Cooper, W. E. Wade, J. L. Waller, J. R. Maclean, and M. H. Beers. 2003. "Updating the Beers Criteria for Potentially Inappropriate Medication Use in Older Adults: Results of a US Consensus Panel of Experts." *Archives of Internal Medicine* 163 (22): 2716–24.
- Fireman, B., J. Bartlett, and J. Selby. 2004. "Can Disease Management Reduce Health Care Costs by Improving Quality?" *Health Affairs* 23 (6): 63–75.

- Forrest, C. B., K. W. Lemke, D. P. Bodycombe, and J. P. Weiner. 2009. "Medication, Diagnostic, and Cost Information as Predictors of High-Risk Patients in Need of Care Management." *The American Journal of Managed Care* 15 (1): 41–8.
- Forrest, C. B., V. G. Villagra, and J. E. Pope. 2006. "Managing the Metric vs Managing the Patient: The Physician's View of Pay for Performance." *American Journal of Managed Care* 12 (3): 83–5.
- Fuhlbrigge, A., V. Carey, J. Finkelstein, P. Lozano, T. Inui, and K. Weiss. 2008. "Are Performance Measures Based on Automated Medical Records Valid for Physician/Practice Profiling of Asthma Care?" *Medical Care* 46 (6): 620–6.
- Furmaga, E. M., P. A. Glassman, F. E. Cunningham, and C. B. Good. 2005. "Reducing the Use of Short-Acting Nifedipine by Hypertensives Using a Pharmaceutical Database." In *Advances in Patient Safety: From Research to Implementation. (AHRQ Publication No. 05-0021-3 ed.)*, edited by K. Henriksen, J. B. Battles, E. S. Marks, and D. I. Lewin, pp. 277–89. Rockville, MD: Agency for Healthcare Research and Quality.
- Gilmore, A. S., Y. Zhao, N. Kang, K. L. Ryskina, A. P. Legorreta, D. A. Taira, and R. S. Chung. 2007. "Patient Outcomes and Evidence-Based Medicine in a Preferred Provider Organization Setting: A Six-Year Evaluation of a Physician Pay-for-Performance Program." *Health Services Research* 42 (6): 2140–59.
- Greene, S. E., and D. B. Nash. 2009. "Pay for Performance: An Overview of the Literature." *American Journal of Medical Quality: The Official Journal of the American College of Medical Quality* 24 (2): 140–63.
- Grensenz, C., N. Lurie, and T. Ruder. 2009. *Ambulatory Care Sensitive Hospitalizations and Emergency Department Visits in Baltimore City*. Santa Monica, CA: Rand Corporation.
- Hill-Briggs, F. 2003. "Problem Solving in Diabetes Self-Management: A Model of Chronic Illness Self-Management Behavior." *Annals of Behavioral Medicine* 25 (3): 182–93.
- IOM Committee on Quality of Health Care in America. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: Institute of Medicine, National Academy Press.
- Linden, A. 2006. "What Will It Take for Disease Management to Demonstrate a Return on Investment? New Perspective on an Old Theme." *American Journal of Managed Care* 12 (4): 217–22.
- Mangione, C. M., R. B. Gerzoff, D. F. Williamson, W. N. Steers, E. A. Kerr, A. F. Brown, et al. 2006. "The Association between Quality of Care and the Intensity of Diabetes Disease Management Programs." *Annals of Internal Medicine* 145 (2): 107–16.
- Marsteller, J. A., P. Fagan, A. Schuster, Y. Hsu, L. Dunbar, L. Rosenthal, et al. 2008. "Integration of Care Management into Primary Care Practice: Care Manager Perspectives." *Poster Session Presented at Academy Health Annual Research Meeting*, Orlando, FL.
- National Committee for Quality Assurance (NCQA). 2004. *HEDIS^(R) 2005 Technical Specifications*. Washington, DC: National Committee for Quality Assurance.

- Parente, S. T., S. S. Kim, M. D. Finch, L. A. Schloff, T. S. Rector, R. Seifeldin, and J. D. Haddox. 2004. "Identifying Controlled Substance Patterns of Utilization Requiring Evaluation Using Administrative Claims Data." *American Journal of Managed Care* 10 (11): 783–90.
- Pawlson, L. G., S. H. Scholle, and A. Powers. 2007. "Comparison of Administrative-Only versus Administrative Plus Chart Review Data for Reporting HEDIS Hybrid Measures." *American Journal of Managed Care* 13 (10): 553–8.
- Pawson, R., and N. Tilley. 1997. *Realistic Evaluation*. London: Sage Publications.
- Rosenthal, M. B. 2007. "P4P: Rumors of its Demise may be Exaggerated." *American Journal of Managed Care* 13 (5): 238–9.
- Rosenthal, M. B., R. Fernandopulle, H. R. Song, and B. Landon. 2004. "Paying for Quality: Providers' Incentives for Quality Improvement." *Health Affairs* 23 (2): 127–41.
- Rosenthal, M. B., and R. G. Frank. 2006. "What Is the Empirical Basis for Paying for Quality in Health Care?" *Medical Care Research and Review* 63 (2): 135–57.
- Rosenthal, M. B., R. G. Frank, Z. Li, and A. M. Epstein. 2005. "Early Experience with Pay-for-Performance: From Concept to Practice." *The Journal of the American Medical Association* 294 (14): 1788–93.
- Rosof, B. M., C. A. Sirio, and K. S. Kmetik. 2008. "Pay-for-Performance System for English Physicians." *The New England Journal of Medicine* 359 (20): 2176–7.
- Solberg, L. I., K. I. Engebretson, J. M. Sperl-Hillen, P. J. O'Connor, M. C. Hroschowski, and A. L. Crain. 2006. "Ambulatory Care Quality Measures for the 6 Aims from Administrative Data." *American Journal of Medical Quality* 21 (5): 310–6.
- Villagra, V. G. 2004. "Perspective: Integrating Disease Management into the Out-patient Delivery System during and after Managed Care." *Health Affairs*, Web exclusive, 281–3. doi:10.1377/hlthaff.W4.281.
- Villagra, V. G., and T. Ahmed. 2004. "Effectiveness of a Disease Management Program for Patients with Diabetes." *Health Affairs* 23 (4): 255–66.
- Von Korff, M., W. Katon, E. H. B. Lin, G. Simon, E. Ludman, M. Oliver, et al. 2005. "Potentially Modifiable Factors Associated with Disability among People with Diabetes." *Psychosomatic Medicine* 67 (2): 233–40.
- Wagner, E. H. 2004. "Chronic Disease Care." *British Medical Journal* 328 (7433): 177–8.
- Wagner, E. H., and R. J. Reid. 2007. "Are Continuity of Care and Teamwork Incompatible?" (Editorial). *Medical Care* 45 (1): 6–7.
- Weber, C., and K. Neeser. 2006. "Using Individualized Predictive Disease Modeling to Identify Patients with the Potential to Benefit from a Disease Management Program for Diabetes Mellitus." *Disease Management: DM* 9 (4): 242–56.
- Wennberg, J. E., E. S. Fisher, T. A. Stukel, and S. M. Sharp. 2004. "Use of Medicare Claims Data to Monitor Provider-Specific Performance among Patients with Severe Chronic Illness" [accessed August 19, 2010]. *Health Affairs*, Web exclusive, 5–17. Available at <http://content.healthaffairs.org/cgi/content/abstract/hlthaff.var.5v1>
- Wilson, T., and M. MacDowell. 2003. "Framework for Assessing Causality in Disease Management Programs: Principles." *Disease Management: DM* 6 (3): 143–58.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: Statistical Methodology.

Table S3: Analysis Results for Trends, Changes in Trends, and Differences in Trend Changes by Outcome Measure.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.