Original Investigation

Effect of Pay-for-Performance Incentives on Quality of Care in Small Practices With Electronic Health Records
A Randomized Trial

Naomi S. Bardach, MD, MAS; Jason J. Wang, PhD; Samantha F. De Leon, PhD; Sarah C. Shih, MPH; W. John Boscardin, PhD; L. Elizabeth Goldman, MD, MCR; R. Adams Dudley, MD, MBA

IMPORTANCE Most evaluations of pay-for-performance (P4P) incentives have focused on large-group practices. Thus, the effect of P4P in small practices, where many US residents receive care, is largely unknown. Furthermore, whether electronic health records (EHRs) with chronic disease management capabilities support small-practice response to P4P has not been studied.

OBJECTIVE To assess the effect of P4P incentives on quality in EHR-enabled small practices in the context of an established quality improvement initiative.

DESIGN, SETTING, AND PARTICIPANTS A cluster-randomized trial of small (<10 clinicians) primary care clinics in New York City from April 2009 through March 2010. A city program provided all participating clinics with the same EHR software with decision support and patient registry functionalities and quality improvement specialists offering technical assistance.

INTERVENTIONS Incentivized clinics were paid for each patient whose care met the performance criteria, but they received higher payments for patients with comorbidities, who had Medicaid insurance, or who were uninsured (maximum payments: $200/patient; $100 000/clinic). Quality reports were given quarterly to both the intervention and control groups.

MAIN OUTCOMES AND MEASURES Comparison of differences in performance improvement, from the beginning to the end of the study, between control and intervention clinics for aspirin or antithrombotic prescription, blood pressure control, cholesterol control, and smoking cessation interventions. Mixed-effects logistic regression was used to account for clustering of patients within clinics, with a treatment by time interaction term assessing the statistical significance of the effect of the intervention.

RESULTS Participating clinics (n = 42 for each group) had similar baseline characteristics, with a mean of 4592 (median, 2500) patients at the intervention group clinics and 3042 (median, 2000) at the control group clinics. Intervention clinics had greater adjusted absolute improvement in rates of appropriate antithrombotic prescription (12.0% vs 6.1%, difference: 6.0% [95% CI, 2.2% to 9.7%], P = .001 for interaction term), blood pressure control (no comorbidities: 9.7% vs 4.3%, difference: 5.5% [95% CI, 1.6% to 9.3%], P = .01 for interaction term; with diabetes mellitus: 9.0% vs 1.2%, difference: 7.8% [95% CI, 3.2% to 12.4%], P = .007 for interaction term; with diabetes mellitus or ischemic vascular disease: 9.5% vs 1.7%, difference: 7.8% [95% CI, 3.0% to 12.6%], P = .01 for interaction term), and in smoking cessation interventions (12.4% vs 7.7%, difference: 4.7% [95% CI, −0.3% to 9.6%], P = .02 for interaction term). Intervention clinics performed better on all measures for Medicaid and uninsured patients except cholesterol control, but no differences were statistically significant.

CONCLUSIONS AND RELEVANCE Among small EHR-enabled clinics, a P4P incentive program compared with usual care resulted in modest improvements in cardiovascular care processes and outcomes. Because most proposed P4P programs are intended to remain in place more than a year, further research is needed to determine whether this effect increases or decreases over time.

TRIAL REGISTRATION clinicaltrials.gov Identifier: NCT00884013

Author Affiliations: Department of Pediatrics, University of California, San Francisco (Bardach); New York City Department of Health and Mental Hygiene, New York (Wang, De Leon, Shih); Department of Epidemiology and Biostatistics, University of California, San Francisco (Boscardin); Department of Internal Medicine, University of California, San Francisco (Goldman, Dudley); Philip R. Lee Institute for Health Policy Studies, University of California, San Francisco (Dudley).

Corresponding Author: Naomi S. Bardach, MD, MAS, 3333 California St, Ste 265, San Francisco, CA 94118 (bardachn@peds.ucsf.edu).

Innovations in technology and a greater focus on chronic disease management are changing the way health care is delivered.[1] The Affordable Care Act (ACA) includes payment reforms intended to facilitate substantive change and system redesign.[2] As health care evolves, it is important to understand how payment models influence performance in new care delivery environments.

In 2005, the New York City Department of Health and Mental Hygiene (DOHMH) established the Primary Care Information Project (PCIP) to improve preventive care for chronically ill patients in low–socioeconomic status neighborhoods. Funded through city, state, federal, and private foundation contributions of more than $60 million, PCIP codesigned and implemented in participating practices a prevention-oriented electronic health record (EHR) system with clinical decision support and disease registries and offered technical assistance and quality improvement visits.[3]

Most existing literature evaluates pay-for-performance (P4P) in large-group practices.[4-7] In contrast, the participating New York City practices were small (mostly 1-2 clinicians).[8] Small practices, where the majority of patients still receive care nationally,[9] historically have provided lower-quality care—especially solo practices[10]—and may have greater obstacles in improving care because they lack the scale and organizational structure to do so.[10,11] With widespread implementation of EHRs,[1] it is possible that EHR-enabled solo and small-group practices will be able to respond to P4P incentives and improve quality, but this has not been demonstrated.[12]

To address this gap in knowledge, we performed a cluster-randomized trial to assess the effect of P4P on preventive care processes and outcomes among practices participating in PCIP.

Methods

Setting and Participants

The institutional review boards at the University of California, San Francisco, and the New York City DOHMH approved the study, with waivers of patient informed consent. Clinic owners provided written informed consent for participation. Eligible clinics were small practices (1-10 clinicians) participating in the PCIP. The PCIP provided all clinics EHR software (eClinical Works) with clinical decision support (passive reminders on a sidebar for each patient) for the measures in the study, and with patient registry and quality reporting capabilities.[3,8,13] Clinic eligibility criteria were having at least 200 patients eligible for measurement, having at least 10% Medicaid or uninsured patients, and use of the EHR software for at least 3 months. Clinics were randomized in March 2009. Because the effect of P4P is contingent on clinicians knowing about the incentive, clinicians were not blinded to their group assignment.

The PCIP provided all practices with on-site quality improvement assistance, including coaching clinicians on EHR quality improvement features, supporting workflow redesign, and demonstrating proper EHR documentation of the study measures. The quality improvement coaches were blinded to clinic group assignment.

Randomization

Practices that agreed to participate were stratified by size (1-2 clinicians, 3-7 clinicians, or 8-10 clinicians), EHR implementation date, and New York City borough.

Intervention

We randomized participating clinics to either an intervention group receiving financial incentives and benchmarked quarterly reports of their performance or a control group receiving only quarterly reports. The financial incentive was paid to the practice at the end of the study. The clinicians in each practice decided whether to divide the incentive among themselves or to invest in the practice.

The incentive design reflected a conceptual model from a study by Frølich and coauthors.[14] We paid the incentive to the clinic, and we paid for a related set of measures to motivate clinicians to use practice-level mechanisms to enhance population-level disease management.[14] Clinicians may discount their estimates of expected revenue from the incentive if there is uncertainty about achieving the level of performance required.[14,15] Therefore, an incentive was paid for every instance of a patient meeting the quality goal, and clinicians were not penalized for patients who did not meet the quality goal. In addition, clinicians may better respond to incentives that recognize the opportunity cost of achieving the incentive relative to other work (eg, spending more time with a patient to achieve the metric rather than earning more money by seeing an additional patient).[14,15] To encourage physicians to improve care even for those patients for whom changing outcomes might require more resources (either because those patients were sicker or had lower socioeconomic status), we structured the incentive to give a higher payment when goals were met among patients with certain comorbidities or, as proxies for socioeconomic status, had Medicaid insurance, or were uninsured (see **Table 1**).

Because the differential amount of resources required to care for these populations is not known, we chose the baseline payment and the differential amounts based on informational interviews with clinicians and based on the Medicaid fee-for-service reimbursement at the time for a preventive visit for a healthy adult (approximately $18). The total amounts available to be awarded across a clinician's patient panel was expected to be approximately 5% of an average physician's annual salary.[16]

The study period was April 2009 to March 2010. In April 2009, study staff sent e-mails and letters to all clinics regarding group assignment, including materials describing performance measures and their documentation in the EHR (all clinics, eAppendix A in the Supplement) and the incentive structure (intervention group, eAppendix B in the Supplement). Quality reports were sent to all practices quarterly (see intervention group—eAppendix C, and control group—eAppendix D in the Supplement), with a final report delivered March 2010.

Objectives and Outcomes

The clinical areas targeted for P4P incentives were processes and intermediate outcomes that reduce long-term cardiovas-

Table 1. Incentive Payment Structure[a]

Clinical Preventive Service	Baseline Payment, $	Payment for High-Risk Patients, $			Total Possible Payment Per Patient, $
		Insurance: Uninsured/Medicaid	Comorbidity: IVD or DM	Combination of Insurance and Comorbidity: Uninsured/Medicaid and IVD or DM	
Aspirin[b]			20	20	20
Blood pressure control[c]	20	40	40	80	80
Cholesterol control[d]	20	40	40	80	80
Smoking cessation[e]	20	20	20	20	20

Abbreviations: DM, diabetes mellitus; IVD, ischemic vascular disease.

[a] Maximum reward possible per patient: $200.

[b] Aspirin measure: patients 18 years or older with IVD or 40 years or older with DM taking aspirin or another antithrombotic therapy (including cilostazol, clopidogrel bisulfate, warfarin sodium, dipyridamole).

[c] Blood pressure control: patients aged 18 to 75 years with hypertension, with blood pressure lower than 140/90 mm Hg (if without DM) or lower than 130/80 mm Hg (if with DM).

[d] Cholesterol control: male patients 35 years or older and female patients 45 years or older without IVD or DM who have a total cholesterol lower than 240 mg/dL or low-density lipoprotein lower than 160 mg/dL measured in the past 5 years (to convert cholesterol and low-density lipoprotein to millimoles per liter, multiply by 0.0259).

[e] Smoking cessation: patients 18 years or older identified as current smokers who received cessation counseling, referral for counseling, or prescription or increased dose of a cessation aid.

cular risk (the ABCS: aspirin or antithrombotic prescription, blood pressure control, cholesterol control, smoking cessation), summarized in Table 1. We included intermediate outcome measures (blood pressure and cholesterol control) because they are more proximate to better population health, whereas there is sometimes only a weak relationship between process measures and long-term outcomes.[17,18]

The primary outcome of interest was the differences between the incentive and control groups in the proportion of patients achieving the targeted measures. The secondary outcome was the differences between the incentive and control groups in the proportion of patients achieving the targeted measures among patients who were harder to treat because of comorbidities or insurance status. Health Maintenance Organization (HMO) Medicaid patients were not analyzed separately from other HMO patients because some clinics do not distinguish HMO Medicaid patients in the EHR.

Patients were identified for inclusion using *International Classification of Diseases, Ninth Revision* and *Current Procedural Terminology* codes embedded in the EHR progress notes (see eAppendix E in the Supplement). Patients with their cholesterol levels tested in the 5 years prior were included in the cholesterol measure. Patients were counted as achieving the measure goal based on blood pressure values, aspirin or other antithrombotic prescriptions, cholesterol values, and smoking cessation interventions documented in structured fields in the EHR, designed to be completed as part of clinicians' normal workflow, as previously described.[8]

Data Collection

To assess baseline differences in clinic characteristics between control and intervention groups, including patient panel size, we used data reported on the PCIP program agreements by participating clinicians. Both baseline and end-of-study performance data were collected electronically by PCIP staff at the end of the study. Clinics that exited the study did not contribute baseline or end-of-study data. Measure achievement was assessed using the final documentation in the EHR from the study period. If there were multiple blood pressure measure-

ments recorded for a single patient before the study, the last prestudy measurement was used to assess control at baseline. If there were then multiple blood pressure measurements during the study period, the last measurement in the study period was used to determine whether end-of-study control was achieved.
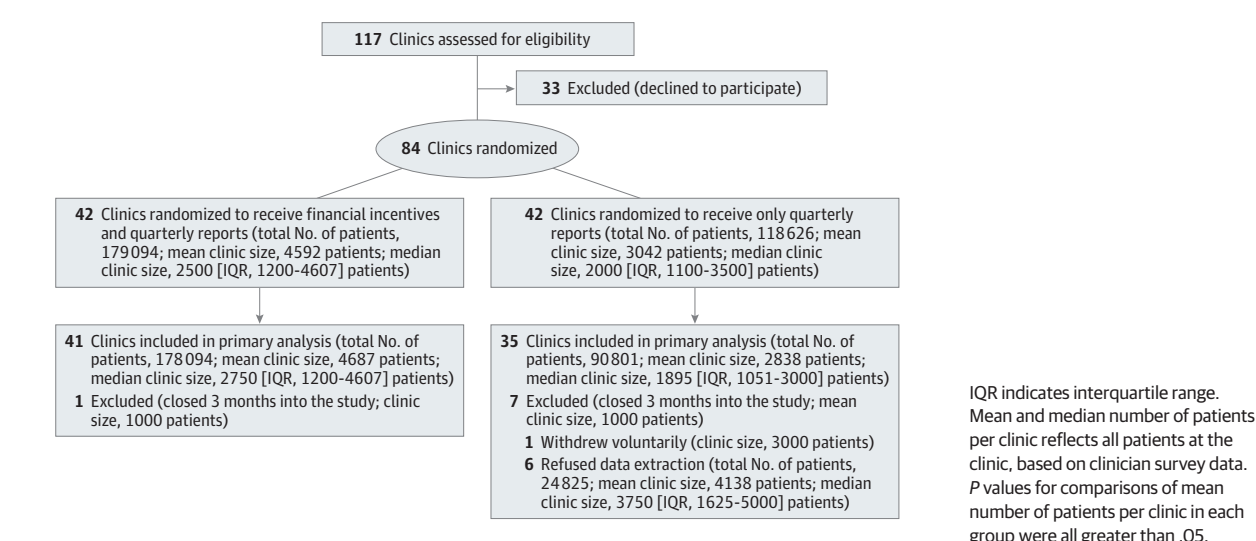
Statistical Methods

Power calculations were based on the Donner and Klar formula.[19] There was no peer-reviewed literature about the likely effects of an incentive of this size on our dependent variables, but we a priori estimated that the effect size would be approximately a 10% increase in the absolute level of performance. We used an intracluster correlation coefficient (ICC) of 0.1 as a conservative estimate based on prior published data on ICC for other process and outcome measures.[20] With 42 clinics per group, assuming that the number of patients per clinic per measure was an average of 50 patients and the control group performance was 20%, using a 2-sided test and 5% level of significance, we had 87% power to detect a 10% difference in performance across the measures (with 77% power if the control group performance was 50%). For the subgroup analysis of Medicaid non-HMO and uninsured patients, assuming that the number of patients per clinic per measure was 5 and the control group performance was 20%, we had 52% power to detect a 10% difference (with 41% power if the control group performance was 50%). We did not power the study to find a difference in the subgroup analysis.

For comparison of clinic and patient characteristics, the Wilcoxon rank sum test was used.

The unit of observation in this trial was the patient, but data were aggregated at the clinic level. Clinics that did not provide data were not included in the analysis (**Figure**). Patients were clustered within clinics, with variability in the number of patients per clinic. Because this can lead to larger clinics dominating results, we adjusted for clinic-level clustering. To accommodate the likely correlation of patient outcomes within clinic and to accommodate potential repeated measures for patients presenting for care during the baseline and study mea-

Figure. Flow of Clinics Through the Study

IQR indicates interquartile range. Mean and median number of patients per clinic reflects all patients at the clinic, based on clinician survey data. P values for comparisons of mean number of patients per clinic in each group were all greater than .05.

surement periods, we used multilevel mixed-effects logistic regression to model patient-level measure performance (achievement of the measure or failure) for each measure. The model included random intercepts for each clinic that were assumed to be constant across the baseline and the end of the study measurements and fixed-effects predictors of study group (intervention vs control group), time point (baseline vs follow-up), and the interaction of study group and time point. The primary interest lies in the interaction parameter, because it is a comparison between the study groups in the amount of change between the time points. This approach adjusts for the baseline differences in performance between groups. Computations were performed using the xtmelogit command in STATA, version 12 (Stata Corp). To summarize the inference from this model, we present the odds ratios (ORs) for the interaction term together with their 95% CIs and associated P values.

In addition, we report performance in the groups at baseline and the end of the study using adjusted probabilities and report the difference between the 2 groups in their change in adjusted probabilities from baseline to the end of the study (difference in differences) to summarize the effect in a manner more easily interpretable to readers. As done in other trials with multiple tests for related outcomes with consistent results across tests, we did not adjust for multiple comparisons.[21-23] The conceptual model underlying P4P supports this, positing that system-level interventions are required to achieve improvements,[14] and so performance changes across measures are potentially linked.[22,23]

We performed 2 sensitivity analyses to address potential bias due to postrandomization drop out. First, using data from surveys collected from each clinic upon enrollment in the trial, we created propensity scores based on number of clinicians, percentage of Medicaid patients, percentage of Medicare patients, percentage of uninsured patients, and time since implementation of EHRs. We used the propensity scores to match the 7 control clinics that dropped out with 7 control clinics that

participated. We made a conservative assumption that the control clinics that dropped out had the same performance as their propensity score-matched control clinics.[24] For the missing intervention clinic (closed partway through the study), we duplicated, for each measure, the performance of the intervention clinic that had the lowest performance improvement. We chose the lowest performances to generate the most conservative estimate of the incentive effect. We then repeated the primary analyses.

In the second sensitivity analysis, we referred to the randomization strata from the original study design and assumed that each clinic whose data were missing would have performed exactly the same as the paired clinic in its randomization stratum. This puts a conservative bound on the effects of the intervention because data from 7 intervention clinics were used to represent the data from the 7 missing control clinics and data from 1 control clinic represented data from 1 missing intervention clinic. We then repeated the primary analyses.

All analyses were performed using STATA, version 12. All statistical tests were 2-sided with a 95% significance level.

Results

Patient and Clinic Characteristics

Of the 117 eligible clinics, 84 clinics agreed to participate and were randomized. Intervention clinics reported a mean of 4592 (median, 2500) patients per clinic for a total of 179 094 patients and the control clinics reported a mean of 3042 (median, 2000) patients per clinic for a total of 118 626 patients (P = .45 for comparison of means; Figure). Baseline clinic characteristics were similar in each group (Table 2). There was low to moderate performance at baseline in almost all measures, except for cholesterol control, which was more than 90% in both groups (Table 3). Baseline performance rates were higher in the intervention group for 3 of the 7 measures (Table 3).

Table 2. Baseline Characteristics of Intervention and Control Clinics and Patients

Characteristics, mean (SD)[a]	Intervention Group	Control Group	P Value[b]
Patient characteristics			
Age, y	45.8 (6.7)	46.6 (4.8)	.62
Male, %	42.0 (8.6)	39.8 (10.5)	.48
Clinic characteristics			
Clinicians, median (IQR)	1 (1-2)	1 (1-2)	.77
Patients	4592 (8241)	3042 (2978)	.45
Median (IQR)	2500 (1200-4607)	2000 (1100-3500)	
Time since EHR implementation, mo	9.93 (4.47)	9.57 (4.44)	.81
Quality improvement specialist visits	5.17 (3.43)	4.24 (2.73)	.25
Insurance, %			
Commercial	33.8 (23.9)	32.1 (21.6)	.89
Medicare	25.6 (22.0)	26.8 (17.6)	.32
Medicaid	35.3 (28.3)	35.7 (24.8)	.88
Uninsured	4.3 (4.8)	4.7 (4.9)	.60

Abbreviation: IQR, interquartile range.

[a] Data are reported for the clinics randomized (42 intervention clinics, 42 control clinics).

[b] Comparisons made with Wilcoxon rank sum testing.

Table 3. Change in Performance in Incentive and Control Groups for All Insurance Types[a]

	Baseline Performance, % (95% CI)		End-of-Study Performance, % (95% CI)		Between-Group Differences, % (95% CI)		
	Control[b]	Intervention[b]	Control[b]	Intervention[b]	Absolute Change	Odds Ratio[c]	P Value[c]
Aspirin therapy, with IVD or DM	54.4 (47.6 to 61.2)	52.6 (46.0 to 59.1)	60.5 (54.0 to 67.0)	64.6 (58.7 to 70.5)	6.0 (2.2 to 9.7)	1.28 (1.10 to 1.50)	.001
Blood pressure control							
No IVD or DM	31.8 (20.6 to 43.1)	52.1 (40.2 to 64.0)	36.1 (24.1 to 48.0)	61.8 (50.5 to 73.0)	5.5 (1.6 to 9.3)	1.23 (1.05 to 1.44)	.01
IVD	46.0 (31.1 to 60.7)	68.4 (55.7 to 81.2)	57.3 (43.1 to 71.5)	70.8 (59.5 to 82.2)	−9.1 (−22.1 to 3.9)	0.71 (0.40 to 1.24)	.23
DM	10.4 (5.8 to 15.0)	16.8 (10.4 to 23.3)	11.6 (6.6 to 16.6)	25.8 (17.4 to 34.4)	7.8 (3.2 to 12.4)	1.52 (1.12 to 2.07)	.007
IVD or DM	16.9 (10.2 to 23.6)	28.9 (19.8 to 38.0)	18.6 (11.5 to 25.7)	38.4 (28.2 to 48.6)	7.8 (3.0 to 12.6)	1.37 (1.07 to 1.75)	.01
Cholesterol control	90.5 (88.6 to 92.5)	91.4 (89.4 to 93.3)	92.0 (90.6 to 93.4)	91.6 (90.1 to 93.1)	−1.2 (−3.2 to 0.7)	0.86 (0.67 to 1.09)	.22
Smoking cessation intervention[d]	19.1 (12.5 to 25.7)	17.1 (11.1 to 23.2)	26.8 (18.6 to 35.0)	29.5 (21.0 to 38.1)	4.7 (−0.3 to 9.6)	1.30 (1.04 to 1.63)	.02

Abbreviations: DM, diabetes mellitus; IVD, ischemic vascular disease.

[a] All values are adjusted for clustering within clinics using mixed-effects logistic regression modeling. Unadjusted values and numerators and denominators are in eTable 1 in the Supplement. The statistical model creates odds ratios, and we changed this to more easily interpretable adjusted percentages using the predicted estimates from the model. The P values indicate whether the effects are statistically significant.

[b] Baseline rates differed between control vs intervention clinics (P < .05) for blood pressure control, no comorbidities, blood pressure control in patients with IVD, and blood pressure control for IVD or DM.

[c] Odds ratios and P values for the interaction term of study group and study year, with odds ratios greater than 1 indicating that the intervention group had greater improvement compared with baseline than the control group and odds ratios less than 1 indicating that the control group had greater improvement compared with baseline than the intervention group. P values of less than .05 were considered statistically significant.

[d] Smoking cessation interventions measure: patients 18 years or older identified as current smokers, receipt of cessation counseling, referral for counseling, or prescription or increased dose of a cessation aid, documented in the electronic health record.

Information on baseline and final measure performance was available for 41 intervention and 35 control clinics, with 1 intervention clinic closing partway through the study, 1 control clinic withdrawing after randomization, and 6 control clinics choosing not to allow study personnel to collect performance data (Figure).

Effectiveness of Incentive

Performance improved in both groups during the study, with positive changes from baseline for all measures (Table 3), with larger changes in the unadjusted analysis (eTable 1 in the Supplement). The adjusted change in performance was sta-

tistically significantly higher in the intervention group than the control group for aspirin or antithrombotic prescription for patients with diabetes or ischemic vascular disease (adjusted absolute change in performance, 12.0% for the intervention group vs 6.1% for the control group; absolute difference in performance change between intervention and control, 6.0% [95% CI, 2.2% to 9.7%]; P = .001 for interaction term OR) and blood pressure control in patients with hypertension but without diabetes or ischemic vascular disease (adjusted absolute change in performance, 9.7% for the intervention group vs 4.3% for the control group; absolute difference, 5.5% [95% CI, 1.6% to 9.3%]; P = .01 for interaction term OR). There also was greater

Table 4. Change in Performance in Incentive and Control Groups for Medicaid (non-HMO) and Uninsured Patients[a]

	Baseline Performance, % (95% CI)		End-of-Study Performance, % (95% CI)		Between-Group Differences, % (95% CI)[b]		
	Control[c]	Intervention[c]	Control[c]	Intervention[c]	Absolute Change	Odds Ratio[d]	P Value[d]
Aspirin therapy, with IVD or DM	42.0 (31.8 to 52.1)	39.4 (30.7 to 48.2)	49.6 (40.2 to 59.0)	56.4 (48.0 to 64.8)	9.4 (−2.0 to 20.8)	1.46 (0.92 to 2.33)	.11
Blood pressure control							
No IVD or DM	32.4 (20.4 to 44.5)	45.1 (32.7 to 57.6)	38.5 (26.2 to 50.8)	58.1 (46.2 to 70.1)	7.0 (−2.6 to 16.6)	1.30 (0.86 to 1.96)	.22
IVD	61.8 (27.3 to 96.3)	65.6 (35.1 to 96.2)	56.3 (24.6 to 87.9)	65.2 (39.3 to 91.2)	5.2 (−37.3 to 47.6)	1.23 (0.20 to 7.52)	.81
DM	15.1 (6.9 to 23.2)	21.4 (12.7 to 30.1)	13.5 (6.7 to 20.4)	30.5 (21.1 to 40.0)	10.8 (−0.6 to 22.1)	1.84 (0.82 to 4.14)	.14
IVD or DM	21.2 (11.3 to 31.2)	26.4 (16.3 to 36.5)	18.8 (10.3 to 27.4)	37.0 (26.2 to 47.8)	12.9 (1.0 to 24.9)	1.90 (0.96 to 3.75)	.07
Cholesterol control	91.2 (85.6 to 96.8)	90.5 (85.7 to 95.3)	92.2 (89.2 to 95.2)	91.2 (88.2 to 94.2)	−0.3 (−7.5 to 6.8)	0.95 (0.40 to 2.28)	.91
Smoking cessation intervention[e]	8.42 (2.1 to 14.8)	11.3 (3.6 to 19.0)	17.1 (6.4 to 27.8)	27.9 (13.5 to 42.2)	7.9 (−2.7 to 18.5)	1.35 (0.74 to 2.47)	.32

Abbreviations: DM, diabetes mellitus; HMO, health maintenance organization; IVD, ischemic vascular disease.

[a] All values are adjusted for clustering within clinics using mixed-effects logistic regression modeling. Unadjusted values and numerators and denominators are in eTable 2 in the Supplement. The statistical model creates odds ratios, and we changed this to more easily interpreted adjusted percentages using the predicted estimates from the model.

[b] The absolute between-group differences in performance change between the intervention and control groups indicate whether the incentive overall changed the performance on each measure.

[c] Baseline rates did not differ between control vs intervention clinics (P > .05 for all).

[d] Odds ratios and P values are for the interaction term of study group and study year, with positive values indicating that the intervention group had greater improvement compared with baseline than the control group and negative values indicating that the control group had greater improvement compared with baseline than the intervention group. The P values in the last column indicate whether the effects are statistically significant.

[e] Smoking cessation interventions: patients 18 years or older identified as current smokers, receipt of cessation counseling, referral for counseling, or prescription or increased dose of a cessation aid, documented in the electronic health record.

improvement in the intervention group on blood pressure control in patients with hypertension and diabetes (adjusted absolute change in performance, 9.0% for the intervention group vs 1.2% for the control group; absolute difference, 7.8% [95% CI, 3.2% to 12.4%]; *P* = .007 for interaction term OR), hypertension and diabetes or ischemic vascular disease (adjusted absolute change in performance, 9.5% for the intervention group vs 1.7% for the control group; absolute difference, 7.8% [95% CI, 3.0% to 12.6%]; *P* = .01 for interaction term OR), and smoking cessation interventions (adjusted absolute change in performance, 12.4% for the intervention group vs 7.7% for the control group; absolute difference, 4.7% [95% CI, −0.3% to 9.6%]; *P* = .02 for interaction term OR). There was no statistically significant difference between groups on cholesterol control in the general population (adjusted absolute difference, −1.2% [95% CI, −3.2% to 0.7%]; *P* = .22 for interaction term OR) (Table 3).

For uninsured or Medicaid (non-HMO) patients, changes in measured performance were higher in the intervention clinics than the control clinics (range of adjusted absolute differences, 7.9% to 12.9%), except in cholesterol control (absolute adjusted difference, −0.33%), but the differences were not statistically significant (**Table 4** [adjusted] and eTable 2 in the Supplement [unadjusted analyses]).

Each intervention clinic received 1 end-of-study payment, with a total of $692 000 paid across all intervention clinics. The range of payments to clinics was $600 to $100 000 (median, $9900; interquartile range [IQR], $5100-$22 940), with a cap of $100 000 per clinic. Although payments were not made directly to clinicians, potential amounts per clinician across

practices ranged from $600 to $53 160 per clinician (median, $6323 per clinician; IQR, $3840-$11 470).

Propensity score matching resulted in a better balance of practice-level variables (eTable 3 and eTable 4 in the Supplement). The propensity-matched sensitivity analysis results were similar to the primary analyses, with larger incentive effects or effects within less than 1 percentage point of the effect sizes from the primary analyses. In the second sensitivity analysis, in which 8 clinics perform the same as the opposite group, the 3 measures that remain statistically significant show that the intervention has an effect (eTable 5 in the Supplement).

Discussion

In this cluster-randomized study of P4P incentives, we found that EHR-enabled small practices were able to respond to incentives to improve cardiovascular care processes and intermediate outcomes.

To our knowledge, this is the first clinical trial of P4P incentives to focus specifically on independent small-group practices. The largest prior P4P studies that included small practices were observational studies of the Quality and Outcomes Framework in the United Kingdom. It is difficult to generalize those findings to the US context, because small practices in the United Kingdom are nested in a national health system that employs the physicians, resulting in less fragmentation of payers, regulations, and incentives than in the United States.

In terms of small practices in the United States, there has been concern that such clinics might not be able to respond to

P4P incentives.[10,11] This is important because 82% of US physicians practice in groups of fewer than 10 clinicians.[9] Under the Centers for Medicaid & Medicare Services (CMS) Meaningful Use program and the ACA value-based payment modifier programs,[1,2] small practices are facing financial and regulatory pressure to abandon paper-based records and to improve chronic disease management. Thus, although the small practices in our study may have had greater information technology capacity relative to their peers, they likely are more representative of what small practices will look like in the future.

Our study does not address the issue of whether small and large practices achieve different results. However, the improvements in the intervention group compared with the control group were similar to, or better than, results from clinical trials in large medical group settings for process outcomes, such as use of smoking cessation interventions (4.7% change in our study compared with 0.3% change[25] and 7.2% change[26] in previous studies), cholesterol testing,[27] and prescription of appropriate medications (no effect on appropriate asthma prescription[28] compared with the 6.0% increased antithrombotic prescriptions in our study in the intervention group). Further research designed to directly compare large practices with EHR-enabled small practices will be needed to determine whether modern small practices can achieve results similar to larger practices.

The P4P literature varies in how incentives are paid and what influence they have.[29] Our study provides new evidence on approaches that have not previously been tried.[30] These include paying for performance on each patient, rather than paying based on percentage performance within the practice panel. This means that patients in whom meeting the target may be difficult do not threaten the panel-wide reimbursement. Depending on whether the effect sizes found in our study are considered clinically meaningful, the greater improvements in the intervention group compared with the control group on blood pressure control in all patients and smoking cessation in all patients provide supporting evidence that this incentive structure can be effective in the context of EHR-enabled small practices.

In addition, to our knowledge, this is the first trial in which there is greater payment for meeting a target when patient factors make meeting the target more difficult. We found that improvement for patients with diabetes or with multiple comorbidities was similar to that of the population without comorbidities (Table 3). This implies that this incentive structure may have been effective, in that clinicians were successful with patients who are often considered harder to treat. Because we did not have a group with nontiered incentives, we cannot know whether the incentive design explains the outcomes achieved in patients who are difficult to treat.

Although there were greater performance improvements in the intervention group for Medicaid non-HMO patients and uninsured patients (Table 4), these differences did not reach statistical significance. Although the interpretation must be that this study identified no significant associations with performance improvement in this subgroup, it is possible that the study was underpowered to identify a difference. A larger trial might have been able to detect a significant difference.

An important aspect of this study was providing incentives to improve intermediate outcomes, rather than just processes, and doing so specifically in patients with more risk factors. Achieving better blood pressure control is an especially important goal, incorporated into major public health programs such as Healthy People 2020.[31] For instance, in the UK Prospective Diabetes Study (UKPDS) trial, the number needed to treat (NNT) for controlling blood pressure to prevent 1 diabetes-related death was 15, and the NNT to prevent 1 complication was 6.[32] However, it has been difficult to achieve improvements in blood pressure control.[31,33] In our study, although the effect of the intervention was lower than the 10% improvement that we estimated a priori, the absolute risk reduction for blood pressure control among patients with diabetes was 7.8% (NNT, 13). This suggests that, for every 13 patients seeing incentivized clinicians, 1 more patient would achieve blood pressure control. The 7.8% absolute change in blood pressure control for patients with diabetes mellitus represents a 46% relative increase in blood pressure control among intervention patients compared with the baseline of 16.8%. Further research is needed to determine whether this effect of the P4P intervention on blood pressure control increases or decreases over time. However, this NNT to achieve blood pressure control through incentives, taken together with the large relative increase in percentage of patients with blood pressure control and the potential effect of blood pressure control on risk of ischemic vascular events, suggests a reasonable opportunity to reduce morbidity and mortality through P4P as structured in this study.

Several limitations of this study warrant mention. Some clinics exited the program after randomization, with more control clinics leaving than intervention clinics. This may introduce a bias, if there are differential outcomes between missing and nonmissing clinics. The estimates of the effects of the intervention were robust to sensitivity analyses. The sensitivity analysis assuming that the control clinics with missing data performed similarly to propensity-matched control clinics did not change the number of statistically significant findings or their direction. In the sensitivity analysis based on the more extreme assumption that clinics with missing data performed exactly the same as clinics in the opposite study group, we found that 3 of the 5 statistically significant effects in the primary analysis remained significant. In a prior quality reporting program, HMOs that dropped out had lower performance,[24] so it is possible that the missing clinics had lower performance than the analyzed control clinics. If a similar reporting bias was present in our study, this would underestimate the incentive effect. However, if the missing clinics did not perceive the need to stay in the program because they were high performers, their performance may have been higher than what we assumed in our sensitivity analyses, which would have led to an overestimate of the incentive effect.

Additionally, this intervention occurred in the setting of a voluntary quality improvement program. This may reflect a high level of intrinsic motivation to improve among practices in the study, as demonstrated by engagement with the quality improvement specialists (Table 2). Even though it is possible that the quality improvement visits contributed to over-

all improvement, the similar number of visits among intervention and control groups indicates that the incentive likely acted through an additional mechanism for improvement and that access to quality improvement specialists does not explain the differential improvement seen in the intervention group.

Another study within the PCIP program found that clinician documentation for some of the measures did not identify all eligible patients and all patients who achieved the goals.[8] However, most measures were well documented in the prior study,[8] and the improvement in both groups on the measures over time implies that there may have been improved documentation in both groups, rather than only in the intervention group.

There have been reports that incentives can have unintended consequences.[34] Examples include causing clinicians to focus on what is measured and incentivized at the expense of other important clinical activities and undermining of intrinsic motivation through an emphasis on the financial rationale for performing well.[35,36] In this study, we have no data about whether the incentives used caused these effects. Further research is needed to determine the balance between the positive effects we could measure and any potential unintended consequences.

Conclusion

We found that a P4P program in EHR-enabled small practices led to modest improvements in cardiovascular processes and outcomes. This provides evidence that, in the context of increasing uptake of EHRs with robust clinical management tools, small practices may be able to improve their quality performance in response to an incentive.

REFERENCES

1. Centers for Medicare & Medicaid Services. Federal register: part 2. http://www.gpo.gov/fdsys/pkg/FR-2010-07-28/pdf/2010-17207.pdf. Accessed August 8, 2012.

2. Strokoff SL. Patient Protection and Affordable Care Act of 2010. http://housedocs.house.gov/energycommerce/ppacacon.pdf. Accessed August 13, 2013.

3. Mostashari F, Tripathi M, Kendall M. A tale of 2 large community electronic health record extension projects. *Health Aff (Millwood)*. 2009;28(2):345-356.

4. Van Herck P, De Smedt D, Annemans L, Remmen R, Rosenthal MB, Sermeus W. Systematic review: effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv Res*. 2010;10:247.

5. Scott A, Sivey P, Ait Ouakrim D, et al. The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database Syst Rev*. 2011;(9):CD008451.

6. Chung S, Palaniappan L, Wong E, Rubin H, Luft H. Does the frequency of pay-for-performance payment matter? experience from a randomized trial. *Health Serv Res*. 2010;45(2):553-564.

7. Chung S, Palaniappan LP, Trujillo LM, Rubin HR, Luft HS. Effect of physician-specific pay-for-performance incentives in a large group practice. *Am J Manag Care*. 2010;16(2):35-42.

8. Parsons A, McCullough C, Wang J, Shih S. Validity of electronic health record-derived quality measurement for performance monitoring. *J Am Med Inform Assoc.* 2012;19(4):604-609.

9. Rao SR, Desroches CM, Donelan K, Campbell EG, Miralles PD, Jha AK. Electronic health records in small physician practices: availability, use, and perceived benefits. *J Am Med Inform Assoc*. 2011;18(3):271-275.

10. Tollen LA. Physician organization in relation to quality and efficiency of care: a synthesis of recent literature. http://www.commonwealthfund.org/usr_doc/Tollen_physician_org_quality_efficiency_1121.pdf. Accessed August 13, 2013.

11. Crosson FJ. The delivery system matters. *Health Aff (Millwood)*. 2005;24(6):1543-1548.

12. Houle SKD, McAlister FA, Jackevicius CA, Chuck AW, Tsuyuki RT. Does performance-based remuneration for individual health care practitioners affect patient care? a systematic review. *Ann Intern Med*. 2012;157(12):889-899.

13. Amirfar S, Taverna J, Anane S, Singer J. Developing public health clinical decision support systems (CDSS) for the outpatient community in New York City: our experience. *BMC Public Health*. 2011;11:753.

14. Frølich A, Talavera JA, Broadhead P, Dudley RA. A behavioral model of clinician responses to incentives to improve quality. *Health Policy*. 2007;80(1):179-193.

15. Dudley RA, Frølich A, Robinowitz DL, Talavera JA, Broadhead P, Luft HS. *Strategies To Support Quality-based Purchasing: A Review of the Evidence*. http://www.ncbi.nlm.nih.gov/books/NBK43997/. Accessed August 13, 2013.

16. Modern Healthcare. *Modern Healthcare*'s 17th Annual Physician Compensation Survey. http://www.modernhealthcare.com/article/20100719/DATA/100719985. Published July 19, 2010. Accessed August 20, 2013.

17. Bradley EH, Herrin J, Elbel B, et al. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. *JAMA*. 2006;296(1):72-78.

18. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA*. 2006;296(22):2694-2702.

19. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. New York, NY: Oxford University Press; 2000.

20. Beck CA, Richard H, Tu JV, Pilote L. Administrative Data Feedback for Effective Cardiac Treatment: AFFECT, a cluster randomized trial. *JAMA*. 2005;294(3):309-317.

21. Ridker PM, Cannon CP, Morrow D, et al; Pravastatin or Atorvastatin Evaluation and Infection Therapy-Thrombolysis in Myocardial Infarction 22 (PROVE IT-TIMI 22) Investigators. C-reactive protein levels and outcomes after statin therapy. *N Engl J Med*. 2005;352(1):20-28.

22. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ*. 1998;316(7139):1236-1238.

23. Bacchetti P. Peer review of statistics in medical research: the other problem. *BMJ*. 2002;324(7348):1271-1273.

24. McCormick D, Himmelstein DU, Woolhandler S, Wolfe SM, Bor DH. Relationship between low quality-of-care scores and HMOs' subsequent public disclosure of quality-of-care scores. *JAMA*. 2002;288(12):1484-1490.

25. Roski J, Jeddeloh R, An L, et al. The impact of financial incentives and a patient registry on preventive care quality: increasing provider adherence to evidence-based smoking cessation practice guidelines. *Prev Med*. 2003;36(3):291-299.

26. An LC, Bluhm JH, Foldes SS, et al. A randomized trial of a pay-for-performance program targeting clinician referral to a state tobacco quitline. *Arch Intern Med*. 2008;168(18):1993-1999.

27. Young GJ, Meterko M, Beckman H, et al. Effects of paying physicians based on their relative performance for quality. *J Gen Intern Med*. 2007;22(6):872-876.

28. Mullen KJ, Frank RG, Rosenthal MB. Can you get what you pay for? pay-for-performance and the quality of healthcare providers. *Rand J Econ*. 2010;41(1):64-91.

29. Eijkenaar F, Emmert M, Scheppach M, Schöffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy*. 2013;110(2-3):115-130.

30. Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *JAMA*. 2007;297(7):740-744.

31. Healthy People 2020. Heart disease and stroke objectives. http://www.healthypeople.gov/2020 /topicsobjectives2020/objectiveslist.aspx ?topicId=21. Accessed April 10, 2013.

32. UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. *BMJ*. 1998;317(7160):703-713.

33. Okonofua EC, Simpson KN, Jesri A, Rehman SU, Durkalski VL, Egan BM. Therapeutic inertia is an impediment to achieving the Healthy People 2010 blood pressure control goals. *Hypertension*. 2006;47(3):345-351.

34. Werner RM, Goldman LE, Dudley RA. Comparison of change in quality of care between safety-net and non–safety-net hospitals. *JAMA*. 2008;299(18):2180-2187.

35. Woolhandler S, Ariely D, Himmelstein DU. Why pay for performance may be incompatible with quality improvement. *BMJ*. 2012;345:5015.

36. Casalino LP. The unintended consequences of measuring quality on the quality of medical care. *N Engl J Med*. 1999;341(15):1147-1150.