

UCSF

UC San Francisco Previously Published Works

Title

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

Permalink

<https://escholarship.org/uc/item/851993x3>

Journal

JAMA, 310(10)

ISSN

0098-7484

Authors

Bardach, Naomi S
Wang, Jason J
De Leon, Samantha F
[et al.](#)

Publication Date

2013-09-11

DOI

10.1001/jama.2013.277353

Peer reviewed



Published in final edited form as:

JAMA. 2013 September 11; 310(10): 1051–1059. doi:10.1001/jama.2013.277353.

Effect of pay-for-performance incentives on quality of care in small practices with electronic health records: A randomized trial

Naomi S. Bardach, MD, MAS¹, Jason J. Wang, PhD², Samantha F. De Leon, PhD², Sarah C. Shih, MPH², W. John Boscardin, PhD³, L. Elizabeth Goldman, MD, MCR⁴, and R. Adams Dudley, MD, MBA^{4,5}

¹Department of Pediatrics, University of California San Francisco

²New York City Department of Health and Mental Hygiene

³Department of Epidemiology and Biostatistics, University of California San Francisco

⁴Department of Internal Medicine, University of California San Francisco

⁵Philip R. Lee Institute for Health Policy Studies, University of California San Francisco

Abstract

Importance—Most evaluations of pay-for-performance (P4P) have focused on large-group practices. Thus, the effect of P4P in small practices, where many Americans 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 a P4P incentive on quality in EHR-enabled small practices in the context of an established quality improvement initiative.

Design—Cluster-randomized trial

Setting, Participants—Small (<10 clinicians) primary care practices in New York City from April 2009–March 2010. A city program had provided all participating practices with the same EHR with decision support and patient registry functionalities and quality improvement specialists offering technical assistance.

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

Main outcomes and measures—Differences in performance improvement, from beginning to end of the study, between control and intervention practices on aspirin or anti-thrombotic prescription, blood pressure control, cholesterol control, and smoking cessation. 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 practices (n=42 for each group) had similar baseline characteristics, with a mean (median) of 4592 (2500) patients at the incentive group practices and 3042 (2000) at the control group practices. Intervention practices had greater absolute adjusted improvement in rates of appropriate anti-thrombotic prescription (12.0% improvement vs. 6.0% improvement among controls, difference: 6.0% (2.2–9.7%), p=0.001 for intervention effect), blood pressure control improvement (no comorbidities: 9.7% vs. 4.2%, difference: 5.5% (1.6–9.3%), p=0.01 for intervention effect; with diabetes: 9.0% vs. 1.2%, difference: 7.8% (3.2–12.4%), p=0.007 for intervention effect; with diabetes and/or ischemic vascular disease: 9.5% vs. 1.7%, difference: 7.8% (3.0–12.6%), p=0.01 for intervention effect) and improvement in rates of smoking cessation interventions (12.4% vs. 7.7%, difference: 4.7% (–0.3–9.6%), p=0.02 for intervention effect). Intervention practices performed better on all measures for Medicaid and uninsured patients except cholesterol control, but no differences were statistically significant.

Conclusion and Relevance—Among small EHR-enabled practices, a P4P incentive program compared with usual care resulted in modest improvements in cardiovascular care processes and outcomes. Since 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.

ClinicalTrials.gov registration #—NCT00884013

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 (NYC) Department of Health and Mental Health (DOHMH) established the Primary Care Information Project (PCIP) to improve preventive care for chronically ill patients in low socio-economic status (SES) neighborhoods. Funded through city, state, federal and private foundation contributions of over \$60 million, PCIP co-designed and implemented in participating practices a prevention-oriented electronic health record (EHR) with clinical decision support and disease registries, and offered technical assistance and quality improvement (QI) visits.³

Most existing literature has evaluated pay-for-performance (P4P) in large-group practices,^{4–7} although the participating NYC 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 practices¹⁰—and may have greater obstacles to improving care because they have lacked 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

Eligible clinics were small practices (1–10 clinicians) participating in the PCIP. PCIP provided all clinics an EHR (eClinical Works) with clinical decision support (passive reminders on a side bar for each patient) for the measures in the study, and with patient registry and quality reporting capabilities.^{3,8,13} Clinic eligibility criteria included having at least 200 patients eligible for measurement, having at least 10% Medicaid or uninsured patients, and having used the EHR for at least three 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.

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

Randomization

Practices that agreed to participate were stratified by size (1–2 clinicians, 3–7 clinicians, or 8–10 clinicians), EHR “go-live” date, and NYC borough.

Intervention

We randomized participating practices to either financial incentives plus 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 reflects a conceptual model from Dudley et al.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 (e.g., spending more time with one 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 co-morbidities

or, as proxies for socioeconomic status, had Medicaid insurance or were uninsured (see Table 1).

Since 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 (~\$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 email and letters to all clinics regarding group assignment, including materials describing performance measures and their documentation in the EHR (all clinics, Appendix A) and the incentive structure (intervention group, Appendix B). Quality reports were sent to all practices quarterly (see intervention group—Appendix C, and control group—Appendix D), 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 cardiovascular risk (“the ABCS”: Aspirin or Anti-thrombotic 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 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. 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 ICD9 and CPT codes embedded in the EHR progress notes (see Appendix E). Patients with cholesterol tested in the five 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 practices, including patient panel size, we used data reported on the PCIP program agreements by practice 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 during the period. If there were multiple BP measurements recorded for a single patient before the study, the last pre-study BP was used to assess control at baseline. If there were then multiple BP measurements during the study period, the last BP in the study period was used to determine whether end-of-study control was achieved.

Statistical methods

Power calculations were based on Donner and Klar's 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 intra-cluster 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 on average 50 and the control group performance was 20%, using a two-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 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 1). 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 within-patient repeated measures for patients presenting for care during both the baseline and study measurement 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 two measurement occasions 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, as 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 12. To summarize the inference from this model, we present the odds ratios for the interaction term together with its 95% confidence interval and associated p-value.

In addition, we report performance in the groups at baseline and the end of study using adjusted probabilities, and the difference between the two groups in their change in adjusted probabilities from baseline to end-of-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 improvements¹⁴ and so performance changes across measures are potentially linked.^{22,23}

We performed two sensitivity analyses to address potential bias due to post-randomization drop out. First, using data from surveys completed by all participating clinic leads upon enrollment in the trial, we created propensity scores based on number of clinicians, percent Medicaid, percent Medicare, percent uninsured, time since implementation of EHR. We used the propensity scores to match the 7 control practices that dropped out with 7 control practices 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 incentive clinic (closed partway through the study), we duplicated, for each measure, the performance of the incentive 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 was 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 seven incentive clinics were used to represent the data from the seven missing control clinics, and data from one control clinic represented data from one missing incentive clinic. We then repeated the primary analyses.

All analyses were performed using STATA version 12.0 (Stata Corp, College Station, Texas). All statistical tests were two-sided with a 95% significance level.

The institutional review boards at the University of California San Francisco and the NYC DOHMH approved the study, with waivers of patient informed consent. Practice owners provided written informed consent for participation. The trial was registered at clinicaltrials.gov (NCT00884013).

Results

Patient and Clinic Characteristics

A total of 117 clinics were eligible; 84 clinics agreed to participate and were randomized. Incentive clinics reported a mean of 4592 patients/clinic (total=179,094 incentive clinic patients) and control clinics reported a mean of 3042 patients/clinic (total=118,626 control clinic patients, Figure 1, $p=0.45$ for comparison of means). Baseline clinic characteristics were similar in each group (Table 2). There was low to moderate performance at baseline in almost all ABCS, except for cholesterol control, which was >90% in both groups (Table 3). Baseline performance rates were higher in the intervention group for three of the seven measures (Table 3).

Information on baseline and final measure performance was available for 41 incentive and 35 control practices, with one incentive practice closing partway through the study, one

control practice withdrawing after randomization, and six control practices choosing not to allow study personnel to collect performance data (Figure 1).

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). The adjusted change in performance was statistically significantly higher in the intervention group than the control group for aspirin or antithrombotic prescription for patients with diabetes or ischemic vascular disease (12.0% for the intervention group vs. 6.1% for the control group, adjusted absolute difference in performance change between intervention and control: 6.0% [95% CI, 2.2% to 9.7%], $P=0.001$ for interaction term OR) and blood pressure control in patients with hypertension but without diabetes or ischemic vascular disease (9.7% for the intervention group vs. 4.3% for the control group, adjusted absolute difference: 5.5% [95% CI, 1.6% to 9.3%], $P=0.01$ for interaction term OR). There also was greater improvement in the intervention group on blood pressure control in patients with hypertension and diabetes (9.0% for the intervention group vs. 1.2% for the control group, adjusted absolute difference: 7.8% [95% CI, 3.2% to 12.4%], $P=0.007$ for interaction term OR), and hypertension and diabetes and/or ischemic vascular disease (9.5% for the intervention group vs. 1.7% for the control group, difference: 7.8% 95% CI, 3.0% to 12.6%), $P=0.01$ for interaction term OR) and smoking cessation interventions (12.4% for intervention group vs. 7.7% for control group, adjusted absolute difference: 4.7% [95% CI, -0.3% to 9.6%], $P=0.02$ 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%), $P=0.22$ for interaction term OR, Table 3).

For uninsured or Medicaid (non-HMO) patients, changes in measure performance were higher in intervention practices compared to controls (range: 7.91% points to 12.9% points) except in cholesterol control (-0.33% points), but the differences were not statistically significant (Table 4 (adjusted) and eTable 2 (unadjusted analyses)).

Each intervention practice received one end-of-study payment, with a total of \$692,000 paid across practices. The range of payments to practices was \$600–100,000 (median \$9,900; IQR: \$5,100–\$22,940), with a cap of \$100,000 per practice. Though payments were not made directly to clinicians, potential amounts per clinician across practices ranged from \$600–\$53,160/clinician (median: \$6323/clinician (IQR: \$3840–\$11470)).

Propensity score matching resulted in better balance of practice-level variables (eTable3 and eTable4). The propensity matched sensitivity analysis results were similar to the primary analyses, with larger incentive effects or effects within <1 percentage point of the effect sizes from the primary analyses. In the second sensitivity analysis, in which eight clinics perform the same as the opposite group, the three measures that remain statistically significant show that the intervention has an effect (eTable5).

Comment

In this cluster-randomized controlled study of P4P, 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 randomized controlled trial of P4P to focus specifically on independent small-group practices. The largest prior P4P studies that included small practices are observational studies of the Quality and Outcomes framework in the United Kingdom (UK). It is difficult to generalize those findings to the U.S. context, since UK small practices are nested in a national health system that employs the physicians, resulting in less fragmentation of payers, regulations, and incentives than in the U.S.

In terms of small practices in the US, there has been concern that such practices might not be able to respond to P4P incentives.^{10,11} This is important because 82% of U.S. physicians practice in groups of <10 clinicians.⁹ Under the CMS Meaningful Use program and the Affordable Care Act physician quality payment 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 been unusual in their IT 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 to the control group were similar to or better than results in RCTs in large medical group settings for process outcomes such as use of smoking cessation interventions (4.7% change in this study compared to 0.3% change²⁵ and 7.2% change²⁶), cholesterol testing,²⁷ and prescription of appropriate medications (no effect on appropriate asthma prescription²⁸ compared to the 6.0% increased anti-thrombotic prescriptions in this study under P4P). Further research designed to directly compare large practices to 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 Oprovides 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 over 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 this study are considered clinically meaningful, the greater improvements in the incentive group compared to controls on BP 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, this is the first trial of which we are aware 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 in patients who are often

considered harder to treat. As we did not have a group with non-tiered incentives, we cannot know whether the incentive design explains the outcomes achieved in difficult-to-treat patients.

While there were greater performance improvements in the incentive group for Medicaid non-HMO patients and uninsured patients (Table 4), these differences did not reach statistical significance. While 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 BP control is an especially important goal, incorporated into major public health programs such as Healthy People 2020.³¹ For instance, in the UKPDS trial, the number needed to treat (NNT) for controlling BP to prevent one diabetes-related death was 15 and the NNT to prevent one complication was 6.³² However, it has been difficult to achieve improvements in BP control.^{31,33} In our study, while the effect of the intervention was lower than the 10% improvement we estimated *a priori*, the absolute risk reduction for BP control among diabetics was 7.8% (NNT 13). This suggests that, for every 13 patients seeing incentivized clinicians, one more would achieve BP control. The 7.8% absolute change in BP control for patients with DM represents a 46% relative increase in BP control among intervention patients compared to the baseline of 16.8%. Further research is needed to determine whether this effect of the P4P intervention on BP control increases or decreases over time. However, this NNT to achieve BP control through incentives, taken together with the large relative increase in percent of patients with BP control and the potential effect of BP 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 post-randomization, with more control clinics leaving than incentive clinics. This may introduce a bias, if there are differential outcomes between missing and non-missing 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 three of the five statistically significant effects in the primary analysis remained significant. In a prior quality reporting program, health maintenance organizations 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 QI program. This may reflect a high level of intrinsic motivation to improve among practices in the study, as demonstrated by engagement with the QI specialists (Table 2). Even though it is possible that the QI visits contributed to overall improvement, the similar number of QI visits among incentive and control groups indicates that the incentive likely acted through an additional mechanism for improvement and that access to QI specialists does not explain the differential improvement seen in the incentive group.

In addition, on 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 study⁸ 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 incentive 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 cardio-vascular 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

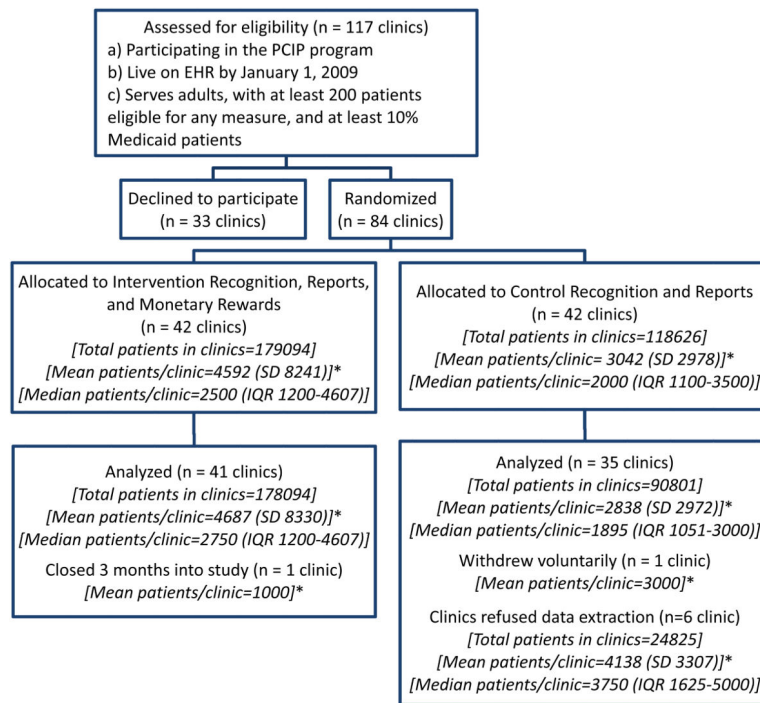
Acknowledgments

This work was supported by the Robin Hood Foundation, the Agency for Healthcare Research and Quality (R18 HS17059, R18 HS18275), the National Institute for Children's Health and Human Development (K23 HD065836) and NCRRC UCSF CTSI (KL2 RR024130-05). The funding sources were not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. None of the authors have conflicts of interest to disclose. Naomi S. Bardach, Jason Wang, and Samantha De Leon had full access to all the data in the study. Naomi S. Bardach takes responsibility for the integrity of the data and the accuracy of the data analysis. Presented in part at the AcademyHealth Meeting on June 13th, 2011 (Seattle, WA). The authors would like to thank the staff at eClinicalWorksTM for assisting the capture and retrieval of the data; PCIP staff for ensuring data were available for the study and the extensive outreach and communications to participating clinicians throughout the program. We would also like to acknowledge the following contributions, which were given without compensation: Thomas R. Frieden, MD, MPH (Centers of Disease Control and Prevention) and Farzad Mostashari, MD, MS (Office of the National Coordinator for Health Information Technology), for the inception and design of the Health eHearts program; Thomas A. Farley, MD, MPH (New York City Department of Health and Mental Hygiene), Amanda S. Parsons, MD, MBA (New York City Department of Health and Mental Hygiene) and Jesse Singer, DO, MPH (New York City Department of Health and Mental Hygiene) for their guidance and support of the Health eHearts program.

References

1. Medicare and Medicaid Programs; Electronic Health Record Incentive Program; Final Rule 42 CFR Parts 412, 413, 422 et al. 2010; 75:44314–44588. <http://www.gpo.gov/fdsys/pkg/FR-2010-07-28/pdf/2010-17207.pdf>.
2. Strokoff, SL. Office of the Legislative Council tC. Public Law 111–148. 2010. Patient Protection and Affordable Care Act of 2010; p. 288-291.
3. Mostashari F, Tripathi M, Kendall M. A tale of two large community electronic health record extension projects. *Health Aff (Millwood)*. Mar-Apr;2009 28(2):345–356. [PubMed: 19275989]
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. [PubMed: 20731816]
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. [PubMed: 21901722]
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*. Apr; 2010 45(2):553–564. [PubMed: 20059568]
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*. Feb; 2010 16(2):e35–42. [PubMed: 20148608]
8. Parsons A, McCullough C, Wang J, Shih S. Validity of electronic health record-derived quality measurement for performance monitoring. *Journal of the American Medical Informatics Association*. Jul 1; 2012 19(4):604–609. [PubMed: 22249967]
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*. May 1; 2011 18(3):271–275. [PubMed: 21486885]
10. Tollen LA. Physician Organization in Relation to Quality and Efficiency of Care: A Synthesis of Recent Literature. *The Commonwealth Fund*. 2008; 89 <http://www.commonwealthfund.org/Publications/Fund-Reports/2008/Apr/Physician-Organization-in-Relation-to-Quality-and-Efficiency-of-Care--A-Synthesis-of-Recent-Literatu.aspx>.
11. Crosson FJ. The delivery system matters. *Health Aff (Millwood)*. Nov-Dec;2005 24(6):1543–1548. [PubMed: 16284026]
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. *Annals of Internal Medicine*. 2012; 157(12):889–899. [PubMed: 23247940]
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. [PubMed: 21962009]
14. Frolich A, Talavera JA, Broadhead P, Dudley RA. A behavioral model of clinician responses to incentives to improve quality. *Health Policy*. Jan; 2007 80(1):179–193. [PubMed: 16624440]
15. Dudley, RA.; Frolich, A.; Robinowitz, DL.; Talavera, JA.; Broadhead, P.; Luft, HS. *Strategies To Support Quality-based Purchasing: A Review of the Evidence*. Rockville (MD): 2004.
16. Colleges AoAM. [Accessed July 19, 2012] Specialty Information: Family Medicine. 2010. https://www.aamc.org/students/medstudents/cim/specialties/63820/cim_pub_fp.html
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*. Jul 5; 2006 296(1): 72–78. [PubMed: 16820549]
18. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA*. Dec 13; 2006 296(22):2694–2702. [PubMed: 17164455]
19. Donner A, Klar N. Cluster randomization trials. *Statistical Methods in Medical Research*. Apr; 2000 9(2):79–80.

20. Beck CA, Richard H, Tu JV, Pilote L. Administrative Data Feedback for Effective Cardiac Treatment: AFFECT, a cluster randomized trial. *JAMA*. Jul 20; 2005 294(3):309–317. [PubMed: 16030275]
21. Ridker PM, Cannon CP, Morrow D, et al. C-reactive protein levels and outcomes after statin therapy. *N Engl J Med*. Jan 6; 2005 352(1):20–28. [PubMed: 15635109]
22. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ*. Apr 18; 1998 316(7139):1236–1238. [PubMed: 9553006]
23. Bacchetti P. Peer review of statistics in medical research: the other problem. *BMJ*. May 25; 2002 324(7348):1271–1273. [PubMed: 12028986]
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. [PubMed: 12243635]
25. Roski J, Jeddloh 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. *Preventive medicine*. Mar; 2003 36(3):291–299. [PubMed: 12634020]
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*. Oct 13; 2008 168(18):1993–1999. [PubMed: 18852400]
27. Young GJ, Meterko M, Beckman H, et al. Effects of paying physicians based on their relative performance for quality. *J Gen Intern Med*. Jun; 2007 22(6):872–876. [PubMed: 17443360]
28. Mullen KJ, Frank RG, Rosenthal MB. Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *The Rand journal of economics*. Spring;2010 41(1):64–91. [PubMed: 21667575]
29. Eijkenaar F, Emmert M, Scheppach M, Schoffski O. Effects of pay for performance in health care: A systematic review of systematic reviews. *Health Policy*. Feb 4.2013
30. Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *JAMA*. Feb 21; 2007 297(7):740–744. [PubMed: 17312294]
31. [Accessed April 10, 2013] Healthy People 2020: Heart Disease and Stroke Objectives. 2013. <http://www.healthypeople.gov/2020/topicsobjectives2020/objectiveslist.aspx?topicId=21>
32. UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. *BMJ*. Sep 12; 1998 317(7160):703–713. [PubMed: 9732337]
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*. Mar; 2006 47(3):345–351. [PubMed: 16432045]
34. Werner RM, Goldman LE, Dudley RA. Comparison of change in quality of care between safety-net and non-safety-net hospitals. *JAMA*. May 14; 2008 299(18):2180–2187. [PubMed: 18477785]
35. Woolhandler S, Ariely D, Himmelstein DU. Why pay for performance may be incompatible with quality improvement. *BMJ*. 2012; 345:e5015. [PubMed: 22893567]
36. Casalino LP. The unintended consequences of measuring quality on the quality of medical care. *N Engl J Med*. Oct 7; 1999 341(15):1147–1150. [PubMed: 10511617]

**Figure 1.**

Flow Diagram of Study Clinics

Abbreviations: PCIP: Primary Care Information Project, through the Department of Mental Health and Hygiene in New York City. EHR: electronic health record.

*Mean and median number of patients/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 >0.05.

Table 1

Incentive Payment Structure

Clinical Preventive Service	Base Payment		Payment for High-Risk Patients		Total Possible Payment per Patient
	Insurance: Commercial Comorbidity: No IVD or DM	Insurance: Uninsured/Medicaid	Comorbidity: IVD or DM	Combination of Insurance and Comorbidity: Uninsured/Medicaid and IVD or DM	
Aspirin	-	-	\$20	\$20	\$20
BP control	\$20	\$40	\$40	\$80	\$80
Cholesterol Control	\$20	\$40	\$40	\$80	\$80
Smoking Cessation	\$20	\$20	\$20	\$20	\$20

Maximum reward possible per patient: \$200

Aspirin measure: Patients ages 18 years or older with IVD or ages 40 years or older with DM on aspirin or another anti-thrombotic therapy (including cilostazol, clopidogrel bisulfate, warfarin sodium, dipyridamole).

Blood pressure (BP) control: Patients 18–75 years of age with Hypertension, with BP <140/90 (if without DM) or <130/80 (if with DM).

Cholesterol control: Male patients ≥ 35 years of age and female patients ≥ 45 years of age without IVD or DM who have a total cholesterol < 240 or LDL < 160 measured in the past 5 years.

Smoking Cessation: Patients ages 18 years or older identified as current smokers who received cessation counseling, referral for counseling, or prescription or increase dose of a cessation aid.

Abbreviations: IVD: Ischemic vascular disease; DM: Diabetes Mellitus.

Table 2

Baseline Characteristics of Incentive and Control Clinics and Patients

Characteristics	Incentive Group	Control Group	P value
Patient Characteristics			
Age, mean y (SD)	45.8 (6.7)	46.6 (4.8)	0.62
Male, mean % (SD)	42.0 (8.6)	39.8 (10.5)	0.48
Clinic characteristics			
Clinicians, median No. (IQR)	1 (1–2)	1 (1–2)	0.77
Patients, mean No. (SD)	4592 (8241) 2500 (1200–4607)	3042 (2978) 2000 (1100–3500)	0.45
Time since EHR implementation, mean (SD), months	9.93 (4.47)	9.57 (4.44)	0.81
Quality improvement specialist visits, mean (SD)	5.17 (3.43)	4.24 (2.73)	0.25
Insurance, mean % (SD)			
Commercial	33.8% (23.9)	32.1% (21.6)	0.89
Medicare	25.6% (22.0)	26.8% (17.6)	0.32
Medicaid	35.3% (28.3)	35.7% (24.8)	0.88
Uninsured	4.3% (4.8)	4.7% (4.9)	0.60

Data are reported for the clinics randomized (42 intervention clinics, 42 control clinics). Comparisons made with Wilcoxon rank sum testing.

Table 3

Change in Performance in Incentive and Control Groups for All Insurance Types

	Baseline Performance Adjusted ^a % (95%CI)		End of Study Performance Adjusted ^a % (95%CI)		Between Group Differences in Performance Change		
	Control ^b	Incentive ^b	Control ^b	Incentive ^b	Absolute Adjusted ^a Change in Incentive – Change in Control (95%CI)	Adjusted ^a Odds Ratio, Interaction Term for Study Group and Study Year (95%CI) ^c	P value ^c
Aspirin therapy, with IVD or DM	54.4% (47.6–61.2)	52.6% (46.0–59.1)	60.5% (54.0–67.0)	64.6% (58.7–70.5)	6.0% (2.2–9.7%)	1.28 (1.10–1.50)	0.001
BP control, no IVD or DM	31.8% (20.6–43.1)	52.1% (40.2–64.0)	36.1% (24.1–48.0)	61.8% (50.5–73.0)	5.5% (1.6–9.3%)	1.23 (1.05–1.44)	0.01
BP control, with IVD	46.0% (31.1–60.7)	68.4% (55.7–81.2)	57.3% (43.1–71.5)	70.8% (59.5–82.2)	-9.1% (-22.1–3.9%)	0.71 (0.40–1.24)	0.23
BP control, with DM	10.4% (5.8–15.0)	16.8% (10.4–23.3)	11.6% (6.6–16.6)	25.8% (17.4–34.4)	7.8% (3.2–12.4%)	1.52 (1.12–2.07)	0.007
BP control, with IVD or DM	16.9% (10.2–23.6)	28.9% (19.8–38.0)	18.6% (11.5–25.7)	38.4% (28.2–48.6)	7.8% (3.0–12.6%)	1.37 (1.07–1.75)	0.01
Cholesterol control	90.5% (88.6–92.5)	91.4% (89.4–93.3)	92.0% (90.6–93.4)	91.6% (90.1–93.1)	-1.2% (-3.2–0.7%)	0.86 (0.67–1.09)	0.22
Smoking cessation intervention ^d	19.1% (12.5–25.7)	17.1% (11.1–23.2)	26.8% (18.6–35.0)	29.5% (21.0–38.1)	4.7% (-0.3–9.6%)	1.30 (1.04–1.63)	0.02

BP: Blood Pressure; IVD: ischemic vascular disease; HTN: hypertension; DM: diabetes mellitus.

^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. The statistical model creates odds ratios and we change this to more easily interpretable adjusted percentages using the predicted estimates from the model. The p-values in the last column indicate whether the effects are statistically significant.

^b Baseline rates differed between control vs. intervention practices ($p < 0.05$) 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 > 1 indicating that the incentive group had greater improvement compared to baseline than the control group and odds ratios < 1 indicating that the control group had greater improvement compared to baseline than the incentive group. P-values of < 0.05 considered statistically significant.

^d Smoking cessation interventions measure: For patients ages 18 years or older identified as current smokers, receipt of cessation counseling, referral for counseling, or prescription or increase dose of a cessation aid, documented in the EHR.

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

	Baseline Performance		End of Study Performance		Between Group Differences in Performance Change		Adjusted ^a Odds Ratio, Interaction Term for Study Group and Study Year (95%CI) ^c	P value ^e
	Control ^b	Incentive ^b	Control ^b	Incentive ^b	Absolute Adjusted ^a Change in Incentive – Change in Control	Year		
Aspirin therapy, with IVD or DM	42.0% (31.8–52.1)	39.4% (30.7–48.2)	49.6% (40.2–59.0)	56.4% (48.0–64.8)	9.4% (–2.0–20.8%)	1.46 (0.92–2.33)	0.11	
Blood pressure control, no IVD or DM	32.4% (20.4–44.5)	45.1% (32.7–57.6)	38.5% (26.2–50.8)	58.1% (46.2–70.1)	7.0% (–2.6–16.6%)	1.30 (0.86–1.96)	0.22	
Blood pressure control, with IVD	61.8% (27.3–96.3)	65.6% (35.1–96.2)	56.3% (24.6–87.9)	65.2% (39.3–91.2)	5.2% (–37.3–47.6%)	1.23 (0.20–7.52)	0.81	
Blood pressure control, with DM	15.1% (6.9–23.2)	21.4% (12.7–30.1)	13.5% (6.7–20.4)	30.5% (21.1–40.0)	10.8% (–0.6–22.1%)	1.84 (0.82–4.14)	0.14	
Blood pressure control, with IVD or DM	21.2% (11.3–31.2)	26.4% (16.3–36.5)	18.8% (10.3–27.4)	37.0% (26.2–47.8)	12.9% (1.0–24.9%)	1.90 (0.96–3.75)	0.07	
Cholesterol control	91.2% (85.6–96.8)	90.5% (85.7–95.3)	92.2% (89.2–95.2)	91.2% (88.2–94.2)	–0.3% (–7.5–6.8%)	0.95 (0.40–2.28)	0.91	
Smoking cessation intervention ^d	8.42% (2.1–14.8)	11.3% (3.6–19.0)	17.1% (6.4–27.8)	27.9% (13.5–42.2)	7.9% (–2.7–18.5%)	1.35 (0.74–2.47)	0.32	

The absolute between-group differences in performance change between the Incentive and Control groups indicate whether the incentive overall changed the performance on each measure. IVD: ischemic vascular disease; HTN: hypertension; DM: diabetes mellitus.

^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. The statistical model creates odds ratios and we change this to more easily interpreted adjusted percentages using the predicted estimates from the model.

^b Baseline rates did not differ between control vs. intervention practices (p>0.05 for all).

^c Odds ratios and p-values are for the interaction term of study group and study year, with positive values indicating that the incentive group had greater improvement compared to baseline than the control group and negative values indicating that the control group had greater improvement compared to baseline than the incentive group. The p-values in the last column indicate whether the effects are statistically significant.

^d Smoking cessation interventions: For patients ages 18 years or older identified as current smokers, receipt of cessation counseling, referral for counseling, or prescription or increase dose of a cessation aid, documented in the EHR.