Title

Effects of Individual Physician-Level and Practice-Level Financial Incentives on Hypertension Care: A Randomized Trial

Journal

JAMA, 310(10)

ISSN

0098-7484

Authors

Petersen, Laura A
Simpson, Kate
Pietz, Kenneth
et al.

Publication Date

2013-09-11

DOI

10.1001/jama.2013.276303

Peer reviewed

Published in final edited form as:
JAMA. 2013 September 11; 310(10): 1042–1050. doi:10.1001/jama.2013.276303.

Effects of Individual Physician-Level and Practice-Level Financial Incentives on Hypertension Care: A Cluster Randomized Trial

Laura A. Petersen, MD, MPH[1], **Kate Simpson, MPH**[1,2], **Kenneth Pietz, PhD**[1], **Tracy H. Urech, MPH**[1], **Sylvia J. Hysong, PhD**[1], **Jochen Profit, MD, MPH**[1], **Douglas A. Conrad, PhD**[3], **R. Adams Dudley, MD, MBA**[4], and **LeChauncy D. Woodard, MD, MPH**[1]

[1]Health Policy and Quality Program, Michael E. DeBakey VA Medical Center Health Services Research and Development Center of Excellence, and Section for Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, TX

[2]American Heart Association, Southwest Affiliate, Houston, TX

[3]University of Washington, Department of Health Services, Seattle, WA

[4]Philip R. Lee Institute for Health Policy Studies and Department of Medicine, University of California, San Francisco

Abstract

Importance—Pay for performance is intended to align incentives to promote high quality care, but results have been contradictory.

Objective—To test the effect of explicit financial incentives to reward guideline-recommended hypertension care.

Address correspondence to: Laura Petersen, MD, MPH, Health Services Research and Development (152), Michael E. DeBakey Veterans Affairs Medical Center, 2002 Holcombe Blvd., Houston, TX, 77030; Phone: 713-794-8623; Fax: 713-748-7359; laurap@bcm.edu.

Additional Contributions

Oversight Committee:

Barry R. Davis, *University of Texas School of Public Health, Houston, TX*; Joseph Francis, *Department of Veterans Affairs, Washington, DC*; Gabriel Habib, *Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX*; Jeffrey Murawsky, *VA Great Lakes Health Care System, Westchester, IL*; Robert A. Petzel, *Department of Veterans Affairs, Washington, DC*; Brian Reed, *Baylor College of Medicine, Houston, TX*

VA Investigators:

Jan Basile, Rose Birkmeier, Warren D. Blackburn, Jr., Patricia Cioffi, Peter Friedmann, Saib Gappy, Stephen A. Geraci, Nicholas Haddad, Kent A. Kirchner, James Levenson, Steve Orwig, Alan Pawlow, Adam Powell, Shakaib Rehman, Amy W. Smith, Jessie M. Spencer, Vidya Sridharan, Donald Weinshenker

VA Primary Care Physician Participants:

Vijaya Ajjagottu, Jeffrey Austerlitz, Peter Baren, Andree Burnett, Eugene Constantinou, Donald Curran, Raminder Dhadli, Sean Ercan-Fang, Harold Fain, Swapna Gupta, Helen Han, Shireen Haque, Judith Hildebrand, Claudine Johnson, Yelena Kamenker-Orlov, Thomas Kumenda, Fengqi Liu, George Malatinszky, Minnie Martin, Nicholas Masozera, Ramon Matawaran, Raghuram Matta, Ivan Monserrate, Praveena Mungara, Kimberly Olson, Tuukka Ostenso, Radha Rao, Simona Retter, Terrence Shaneyfelt, Lubna Sheikh, Patricia Sullwold, Vanisree Suverna, Adeyinka Taiwo, Oanh Thai, Ishita Thakar, Rachel Wilson Peery

Dr. Petersen had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

The authors have no potential conflicts of interest to disclose.

Design, Setting, and Participants—Cluster randomized controlled trial of 12 Veterans Affairs hospital-based outpatient clinics with five performance periods and a 12-month washout. We enrolled 83 primary care physicians and 42 non-physician personnel (e.g., nurses, pharmacists) working with physicians to deliver hypertension care.

Interventions—Clinics randomized to one of four groups: physician-level (individual) incentives; practice-level incentives; individual- plus practice-level incentives (combined); or none. Intervention participants received up to five payments every four months; all participants could access feedback reports.

Main outcome measures—For each four-month period, the number of hypertensive patients among a random sample who achieved guideline-recommended blood pressure thresholds or received an appropriate response to uncontrolled blood pressure; and/or been prescribed guideline-recommended medications and the number who developed hypotension.

Results—Mean (standard deviation) total payments over the study were $4,270 ($459), $2,672 ($153), and $1,648 ($248) for the combined, individual, and practice-level interventions, respectively. The adjusted change over the study in patients meeting the combined blood pressure/appropriate response measure was 8.84 percentage points (95% confidence interval [CI], 4.20–11.80) for the individual-level, 3.70 (95% CI, 0.24–7.68) for the practice-level, 5.54 (95% CI, 1.92–9.52) for the combined, and 0.47 (95% CI, −3.12–4.04) for the control groups. For medications, the change was 9.07 (95% CI, 4.52–13.44), 4.98 (95% CI, 0.64–10.08), 7.26 (95% CI, 2.92–12.48), and 4.35 (95% CI, −0.28–9.28) percentage points, respectively. The adjusted estimated difference in the change between the proportion of patients with blood pressure control/appropriate response for individual incentive and control groups was 8.36 percentage points (95% CI, 2.40–13.00; P=.005). Use of guideline-recommended medications did not significantly change compared to controls, nor did the incidence of hypotension. The effect of the incentive was not sustained after a washout.

Conclusions and Relevance—Individual financial incentives, but not practice-level or combined incentives, resulted in greater blood pressure control or appropriate response to uncontrolled blood pressure; none of the incentives resulted in greater use of guideline-recommended medications or increased incidence of hypotension compared to controls. Further research is needed to understand the factors that contributed to our findings.

Trial registration—NCT00302718; www.clinicaltrials.gov

Background

As part of the Affordable Care Act, the U.S. government has introduced pay for performance to all hospitals paid by Medicare nationwide.[1] The New York City Health and Hospitals Corporation recently announced a performance pay plan for physicians.[2] These and other value-based purchasing systems are intended to align incentives to promote high quality health care.[3] Evaluations of the effectiveness of pay-for-performance programs directed at hospitals have shown contradictory results.[3–5] The Premier Hospital Quality Incentive Demonstration showed an increase of up to 4.1 percentage points in process-quality measures during the first two years,[6] but these modest gains were not sustained.[7] Moreover, risk-adjusted mortality in the same program showed no improvement.[8]

In contrast, recent studies assessing outcomes of hospital pay-for-performance programs implemented in the United Kingdom on a wider scale, with larger bonuses, and with different approaches to quality improvement, showed clinically significant mortality reductions.[9] Enhancing the face validity of these findings, these reductions were concentrated among hospitals that also showed the best process-measure performance.

Evaluations of incentives targeted at individual physicians and physician practice teams (i.e., clinicians, nurses, and support staff who deliver health care) also show variability.[3–5] A Cochrane review of incentives to improve the quality of primary care found that six of seven eligible studies showed a statistically significant positive effect, but the authors encouraged caution in interpreting findings due to design limitations and generalizability concerns.[4] Thus, many questions about pay for performance are unanswered.[4,5,10]

Given the implementation of the patient-centered medical home and models of accountable care,[11] the effects of financial arrangements that reward health care practice teams will become more interesting to payers and policy makers.[12] We are not aware of other multi-site randomized trials of pay for performance directed at both physicians and practice teams. Therefore, we designed a cluster randomized controlled trial to test the effect of explicit financial incentives to individual physicians and practice teams for the delivery of guideline-recommended care for hypertension in the primary care setting.

Methods

Study design and randomization

Characteristics of the study hospitals and detailed trial methods were published elsewhere.[13] Research assistants at the Houston coordinating center enrolled a minimum of five full-time primary care physicians from 12 hospital-based primary care clinics in five Veterans Affairs (VA) Networks. Then, the clinics were randomized to one of four study groups: (1) physician-level (individual) incentives; (2) practice-level incentives; (3) physician-level plus practice-level (combined) incentives; and (4) no incentives (control). We cluster-randomized by hospital to avoid contamination of the intervention; all participants at a hospital belonged to the same intervention group.[14] Randomization was constrained on teaching status, geographic and clinic location, and participation in the Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial (ALLHAT).[15] A data analyst assigned a uniform random number to each of the possible allocations using SAS version 9.1.3 (SAS Institute, Inc, Cary, NC) and selected the one with the highest random number. At the six hospitals randomized to receive a practice-level incentive, physicians could invite up to 15 non-physician colleagues (e.g., nurses, pharmacists) to participate as part of their practice team caring for hypertensive patients on their panel. Also, to meet the physician recruitment goal of seven per site (to account for attrition), we enrolled additional eligible physicians following randomization. The study was approved by the institutional review boards of all participating institutions.[16] All participants provided written informed consent.

Interventions

Education—Participants attended webinars beginning in February 2008 that reviewed the Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7) guidelines.[17] Participants were also informed of their study group assignments, the study measures and financial incentive amounts.

Financial incentives—The intervention phase included a four-month performance baseline period (August-November 2007), and four consecutive four-month periods, starting in April 2008. Following the end of each period, chart abstractors at the Houston coordinating center who were blinded to the study aims and study group assignments collected data from electronic medical records for 40 patients with hypertension randomly selected from each physician's panel. Participants who were absent for four or more weeks during a given four-month performance period were not eligible to earn rewards for that period. Practices could replace non-physician participants who withdrew during the study. Intervention group participants received up to five incentive payments in their paychecks approximately every four months and were notified each time a payment was posted..

Audit and feedback—Customized audit and feedback reports detailing performance for each period and the next period's performance goals were posted to the study's secure website. After the final feedback report in April 2010, we followed participants for a 12-month washout period in order to determine whether the effects of the intervention were sustained after the incentive ended. We then collected data for a four-month post-washout performance period, May-August 2011.

Study measures and rewards

Participants earned incentives for (1) achieving JNC 7 guideline-recommended blood pressure thresholds or appropriately responding to uncontrolled blood pressure (e.g., lifestyle recommendation for Stage 1 hypertension or guideline-recommended medication adjustment), and/or (2) prescribing guideline-recommended antihypertensive medications. The directors of the participating hospital regions contributed $250,000 for the incentives. Results from simulations using pilot data and accounting for estimated improvement rates showed a maximum per chart reward of $18.20; $9.10 for each successful measure.

For the practice-level payments, the aggregated earnings of the physician participants were equally distributed between the physician and non-physician participants in the practice team. Physicians in the combined incentive group received their individual-level performance payment plus their practice-level share. For example, there are seven physicians and three nurses at a site receiving the combined incentive. If we assume each physician meets both study measures for all 40 charts reviewed ($9.10 per measure * 2 measures * 40 charts * 7 physicians) in this hypothetical case, the practice's total incentive earning is $5,096. For the 10 member practice team, the per member share is $509.60. Each nurse would receive $509.60, and each physician would receive $1,237.60 ($509.60 team share plus $728 for a perfect individual performance).

Sample size

We adapted an algorithm by Donner and Klar to determine study power.[18] As is standard in such trials, the participants were contained within the units of randomization (clusters). In this study the participants were primary care personnel and the clusters were VA hospitals. We used estimates of the intraclass cluster correlation of 0.39 for appropriate medication and 0.14 for blood pressure control obtained from pilot data. Equal cluster sizes were assumed to obtain a minimum cluster size. The Donner and Klar algorithm calculates a minimum cluster size and number of clusters using an iterative algorithm based on the non-central t-distribution. Application of the algorithm yielded a result of three clusters per group and five physicians per cluster with a minimum of 40 patients per physician with a 95% significance level and 80% power to detect a difference of 17 percentage points between an intervention and control group for appropriate medication and 15 percentage points for blood pressure control. A target minimum of seven physicians per cluster was used to account for attrition.[19]

Statistical analysis

To evaluate the adequacy of randomization, we tested for differences in physician characteristics across the groups using chi-square tests for binomial variables (or Fisher's exact test when cell sizes were less than five) and the Kruskal-Wallis rank test for continuous variables. We performed a repeated measures longitudinal analysis using the hospital as a random effect. The unit of analysis was the physician. The analysis approach was intent-to-treat; physician performance was evaluated according to the group to which their hospital was randomly assigned. We evaluated the effect of incentive type for each outcome: (1) each incentive group versus the control group; (2) individual-level incentive groups versus the control group; and (3) practice-level incentive groups versus the control group. The goal of the analysis was to evaluate the rate of change in the proportion of patients who met the study measures over time for the intervention group physicians.[20] Models were developed independently for each incentive type and each study measure. Using an approach described by Cheng and colleagues,[21] we constructed a maximal model using scientifically relevant covariates selected *a priori*, and then performed backward elimination to delete variables with P values ≥.05, arriving at our final model. The maximal model allowed us to evaluate both the covariance structure and the list of covariates for inclusion in the final model. Maximal model covariates included characteristics of the hospital where the physician practiced (teaching hospital, ALLHAT site, located in northern versus southern region of US), physician demographic and practice characteristics (gender, race, and years practicing since completing residency), characteristics of the physicians' patients sampled for analysis (mean age, percent male, percent black, and percent diabetic) and whether the physician had reached the ceiling value for the study measure in the baseline period. Also, the models included the covariates of time, the effect of the intervention, and the rate of change of the effect of the intervention over time. In addition to the primary study measures, we evaluated whether the physician made any medication adjustment (i.e., not solely guideline-appropriate changes).

Because the patients for analysis were randomly selected from a physician's panel at each time point, time-dependent covariates were used. A ceiling value calculated using the

Achievable Benchmarks of Care methodology[22] was included in the maximal models because we expected performance to "top off" at some value before reaching 100%. Ceiling values were 95.0% for BP control or appropriate response to uncontrolled BP and 78.3% for guideline-recommended medications. We also determined whether any site-to-site variation existed and included the facility (cluster) as a random effect as necessary. We evaluated final models using a two-sided *t* statistic with 95% significance and adjusted for multiple comparisons using the Benjamini-Yekutiele method.[23] As a check against overfitting, we conducted a bootstrap analysis in which each physician's patients were resampled and the analysis was repeated 1,000 times.

We tracked participants' engagement with the study website. We compared viewing of the baseline or the first intervention period's audit and feedback report between intervention and control group participants using Fisher's exact test.

To evaluate performance following the washout, we performed a linear analysis with clustering by hospital with the post-washout performance rate as the dependent variable and the final intervention performance rate as a covariate. We evaluated the effect of each type of incentive and developed the models independently using backwards elimination.

Using data from automated processing of structured fields from electronic health records, we evaluated the incidence of hypotension among all patients with hypertension who had at least one primary care encounter between February and May 2009. We looked four months from the patient's encounter to identify low systolic blood pressure readings (defined as an outpatient systolic blood pressure < 90 mm Hg) and/or an outpatient diagnosis of hypotension.

All statistical analyses were performed using SAS software version 9.3 (SAS Institute, Inc, Cary, NC).

Results

Participant and site characteristics

Between February 2007 and April 2008, 83 VA primary care physicians and 42 non-physicians members of practice teams (e.g., nurses, pharmacists) were enrolled from 12 study sites (Figure 1). Feedback reports displaying participants' baseline performances were provided starting in October 2008. Eight participants withdrew before the first period's feedback report and payment; seven left the primary care setting or their facility, and one withdrew for a personal reason. After participants received intervention components for the first performance period, an additional physician left the primary care setting. Nine non-physicians (e.g., nurses, pharmacists) from four sites left during the intervention period and three new non-physicians (e.g., nurses, pharmacists) were enrolled. The 77 physicians who contributed at least two periods of performance data were included in the analysis. We completed post-washout data collection in April 2012 for the 55 physicians who remained enrolled by the end of the washout period.

Amongst physicians who participated in all five performance periods, the mean (standard deviation) total payment for physicians over the course of the study was $4,270 ($459) in the combined group, $2,672 ($153) in the individual group, and $1,648 ($248) in the practice group. There were no significant differences in the distributions of physician gender, race, years practicing since completing residency, or patient characteristics (Table 1). There were significant differences across groups in characteristics of the hospitals where the participants worked, including whether they were teaching hospitals ($P<.001$), whether they were ALLHAT sites ($P<.001$), and whether they were in the southern or northern US ($P=.04$).

Rewarded clinical measures

In unadjusted analyses, the proportion of patients either with controlled hypertension or receiving an appropriate response increased for each incentive group over the course of the trial (Figure 2a). The adjusted estimated change over the study of the patients meeting the combined blood pressure or appropriate response measure was 8.84 percentage points (95% confidence interval [CI], 4.20–11.80) for the individual group, 3.70 (95% CI, 0.24–7.68) for the practice group, 5.54 (95% CI, 1.92–9.52) for the combined group, and 0.47 (95% CI, −3.12–4.04) for the controls. The adjusted estimated difference over the study in the change between the proportion of the physician's patients achieving blood pressure control or receiving an appropriate response for the individual incentive group and the controls was 8.36 percentage points (95% CI, 2.40–13.00; $P=.005$; Table 2). Thus, a typical study physician in the individual group with a panel size of 1,000 patients with hypertension would be expected to have about 84 additional patients achieving blood pressure control or receiving an appropriate response after one year of exposure to the intervention. After accounting for multiple comparisons, this remained significant at the 0.05 level. Significance was confirmed by the bootstrap analysis. Site-to-site variation did not have a significant effect on the modeling results.

Over the course of the trial, unadjusted guideline-recommended medication management increased by the final intervention period compared to baseline in all study groups (Figure 2b). For the individual, practice, and combined incentive groups, and the controls, the adjusted estimated change over the study of the physicians' patients meeting the measure was 9.07 (95% CI, 4.52–13.44), 4.98 (95% CI, 0.64–10.08), 7.26 (95% CI, 2.92–12.48), and 4.35 (95% CI, −0.28–9.28) percentage points, respectively. While the use of guideline-recommended medication increased significantly over the course of the study in the intervention groups, there was no significant change compared to controls (Table 2). In adjusted *post-hoc* analyses assessing any medication adjustment (either to start a medication, add a medication, or make a dose adjustment) over the course of the study for the individual intervention group and controls group was 15.36 percentage points (95% CI, 0.20–28.41; $P=.05$). For those in the combined incentive group, the difference was 14.80 percentage points (95% CI, 0.00–27.11; $P=.07$) (see eTable).

Audit and feedback

Far more intervention than control group participants viewed their feedback reports on the website, [66 (67%) versus 5 (25%) respectively; *P*=.001], suggesting that participants were aware of the relationship between performance and rewards.

Post-washout performance

In a model adjusted for geographic region, physician race, and comparison of performance in the final intervention period to the post-washout performance period, there was a significant reduction in performance in the combined measure of blood pressure control or appropriate response to uncontrolled blood pressure in each intervention group compared to controls (Table 3). Therefore, the effect of the intervention declined significantly after the incentive was withdrawn.

Unintended consequences

In *post-hoc* analysis, patients cared for by intervention group participants were no more likely than controls to have hypotension (164 (1.2%) versus 54 (1.4%) patients respectively; *P*=.18).

Discussion

In this cluster randomized trial, we evaluated the effectiveness of pay for performance in primary care settings for a common, chronic condition. We tested incentives targeted at individual physicians, health care practice teams, and a combined incentive to both the individual physician and team. We found that physicians who were randomized to the individual incentive group were more likely than controls to improve their treatment of hypertension as measured by achievement of blood pressure control or appropriate response to uncontrolled blood pressure. Thus, a typical study physician in the individual arm with a panel size of 1,000 patients with hypertension would be expected to have about 84 additional patients achieving blood pressure control or receiving an appropriate response after one year of exposure to the intervention. While the use of guideline-recommended medications increased significantly over the course of the study in the intervention groups, there was no significant change compared to controls. We also showed that those in the individual incentive group were more likely to make antihypertensive medication adjustments in response to uncontrolled blood pressures. Enhancing the face validity of our findings, participants in intervention groups were far more likely than controls to sign into a secure website and view their performance reports. And similar to a report from Kaiser Permanente, we found that the effect of the intervention was not sustained after the incentive was withdrawn.[24] Although concerns about over-treatment have been cited in criticisms of pay-for-performance programs, we did not find a higher incidence of hypotension in the panels of physicians randomized to the incentive groups.

Some might consider the magnitude of the incentives small. The mean individual incentive earnings over the study represented approximately 1.6% of a physician’s salary, assuming a mean salary of $168,000.[25] However, the budget that VA administrative leaders allocated for the incentives was a reflection of their “real-world” constraints, enhancing the

generalizability of our findings. Also, the final amounts of the incentives were similar to those recently announced by the New York City Health and Hospitals Corporation for primary care physicians, for 13 measures, rather than a single, performance measure,[2] meaning that the incentive we used for a single condition (hypertension) was proportionately greater..

What aspects of the design and implementation of our study may have contributed to our findings? First, our measures are meaningful process measures to clinicians. Second, we measured and rewarded actions mostly under the control of physicians and their practice teams.[26] Because blood pressure is not completely under the clinician's control, we rewarded a combined measure of blood pressure control or an appropriate response to an uncontrolled blood pressure (a so-called "tightly-linked" measure).[27,28] Third, responding to an abnormal blood pressure is a discrete task, as opposed to complex problem-solving, such as diagnosing the etiology of abdominal pain. Fourth, we rewarded participants for their absolute rather than relative performance, avoiding a tournament or competition; participants received a pre-specified financial incentive each time they met a performance measure.[29,30] These aspects of our study enhanced the salience of the incentive rewards. Lastly, we combined clear audit and feedback with an incentive.[13] Monetary incentives might amplify the positive effects of performance feedback reports.[31] Bandura's self-efficacy theory states that incentives work by piquing an individual's interest in a task, leading to greater effort at performing the task and ultimately to an increased sense of self-efficacy.[32] We found that intervention group participants were much more likely than controls to view their feedback reports, suggesting that those who received financial incentives demonstrated greater interest in their performance than those who received audit and feedback alone.[33] This suggests that incentive-based interventions and audit and feedback interventions could be synergistic. The goal of the incentive is not to coerce an individual into performing the requested task but to increase their interest in their performance of the task, overcoming clinical inertia.[34]

Unexpectedly, performance gains did not hold after a 12-month washout period, during which we avoided prompting or interacting with participants. Although performance did not decline to pre-intervention levels, the decline was significant. While we speculate that the cessation of performance feedback information may have contributed to the performance decline, further research should elucidate why this phenomenon occurred.

The VA system has a nationwide quality monitoring and assessment program for primary care and chronic conditions, and a culture of performance improvement. As reported by Sutton et al.,[9] the cultural context of the performance rewards may be a significant contributor to their effectiveness. Although these contextual factors may have enhanced the likelihood that our intervention was effective, the high baseline performance of VA health care physicians with blood pressure control rates of approximately 75%[35] may have created a "ceiling effect," whereby gains in performance were more difficult to achieve than they might be in the non-VA setting. Therefore, the improvements might have been greater in a system where baseline performance was lower.

Given that health care organizations are restructuring to implement the patient-centered medical home,[11] we assessed the effect of rewards to health care practice teams. We hypothesized *a priori* that incentives to practice teams would be effective, but we did not find significant effects of either the practice team or the combined incentives. Our interviews with participants suggested that the integrity of the team construct may have been impaired at some sites, perhaps dampening the effectiveness of the practice-level as well as the combined incentives. At two of the practice-level incentive sites, participants noted that they did not know who else was in their practice. At one site, non-physician practice team members were moved to different clinic locations, preventing them from working as a team. Participants also noted the importance of a team in improving hypertension management, yet confirmed that the team structure they were under was rudimentary (prior to the implementation of the VA Patient Centered Medical Home known as the Patient Aligned Care Team (PACT)).[36] Thus, it is possible that, had the PACT structure already been in place, the practice-level and combined incentives would have had a greater impact.

Despite concerns that baseline performance and team cohesion might have dampened the effect of the interventions, several aspects of the VA health care delivery system made this an ideal setting to test the effectiveness of the incentives. First, because the VA uses a single payment approach, we eliminated the problem of multiple payers or varying performance measures diluting the effect of the incentive or the performance targets. Second, VA physicians are salaried, ensuring that the rewards were a clear addition to their expected pay. And while the VA is a uniquely well-suited laboratory for this study, because the structure of the VA health care system is similar to other large delivery systems such as Kaiser Permanente and the Department of Defense, our findings are generalizable. Although VA enrollees are overwhelmingly male, there is little reason to believe that would have systematically biased the study findings.

Hypertension is a common, chronic condition, affecting approximately 70% of those 65 years and older,[37] requiring careful follow-up, adjustments to medication and lifestyle, effective patient-doctor communication, and treatment plan adherence. Inadequate blood pressure control results in excess cases of coronary artery disease, congestive heart failure, renal insufficiency, peripheral arterial disease, and stroke.[17] Even small reductions in blood pressure translate into significant reduction in risk of morbidity and mortality[17] and in systemwide costs.[38] This trial addresses the needs of policy makers and payers for information about a clinically relevant payment intervention in routine practice. Payment-system interventions are attractive because of their potential scale and reach. However, payment-system interventions are only one piece of the solution to improve management of chronic diseases such as hypertension. More resource-intensive, tailored, patient-level self-management strategies may be needed to truly impact patient outcomes.

Individual financial incentives, but not practice-level or combined incentives, resulted in greater blood pressure control or appropriate response to uncontrolled blood pressure. None of the incentives resulted in greater use of guideline-recommended medications compared to controls. Further research is needed to understand the factors that contributed to our findings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding/Support

This work is supported in part by Veterans Affairs Health Services Research & Development (HSR&D) Investigator-Initiated Research (IIR) 04-349 (PI Laura A. Petersen, MD, MPH), NIH RO1 HL079173-01 (PI Laura A. Petersen, MD, MPH), the American Recovery and Reinvestment Act of 2010 (NHLBI 1R01HL079173-S2), and the Houston VA HSR&D Center of Excellence HFP90-020 (PI Laura A. Petersen, MD, MPH). Dr. Petersen was a recipient of the American Heart Association Established Investigator Award (Grant number 0540043N) and was a Robert Wood Johnson Foundation Generalist Physician Faculty Scholar (Grant number 045444) at the time that this study was planned and funded. Dr. Hysong was a recipient of an NHLBI Investigator Research Supplement to Promote Diversity in Health-Related Research (1R01HL079173-S1) during the early stages of the study and is currently a VA HSR&D Career Development Awardee (CDA 07-0181). Dr. Profit's contribution is supported in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development #1 K23 HD056298-01 (PI: Jochen Profit, MD, MPH). Dr. Conrad is an awardee of the Robert Wood Johnson Foundation Health Care Financing and Organization Program (Grant number 63214). Dr. Dudley is a Robert Wood Johnson Investigator Awardee in Health Policy. The views expressed are solely of the authors, and do not necessarily represent those of the authors' institutions.

Role of Funding Sources

None of the funding sources had a role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

We thank the following members of the study coordinating center at the Michael E. DeBakey VA Medical Center Health Services Research and Development Center of Excellence: Supicha S. Chitwood, MPH, for coordinating the incentive payments, Mark Kuebeler, MS, for managing the data, Meghan Z. Lutschg, BA, for coordinating data collection, and Richard SoRelle, BS, for managing the institutional review board requirements across the study sites. We also acknowledge the contributions of the trial's oversight committee, the VA study site investigators, and the VA primary care physician participants.

References

1. Centers for Medicare & Medicaid Services. [Accessed July 9, 2013] Roadmap for implementing value driven healthcare in the traditional Medicare fee-for-service program. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/downloads/vbproadmap_oea_1-16_508.pdf
2. Hartocollis, A. [Accessed July 9, 2013] New York City ties doctors' income to quality of care. New York Times. 2013 Jan 11. http://www.nytimes.com/2013/01/12/nyregion/new-york-city-hospitals-to-tie-doctors-performance-pay-to-quality-measures.html?pagewanted=all&_r=0
3. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? Ann Intern Med. 2006; 145(4):265–272. [PubMed: 16908917]
4. 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]
5. Houle SK, 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. [PubMed: 23247940]
6. Lindenauer PK, Remus D, Roman S, et al. Public reporting and pay for performance in hospital quality improvement. N Engl J Med. 2007; 356(5):486–496. [PubMed: 17259444]
7. Werner RM, Kolstad JT, Stuart EA, Polsky D. The effect of pay-for-performance in hospitals: lessons for quality improvement. Health Aff (Millwood). 2011; 30(4):690–698. [PubMed: 21471490]

8. Jha AK, Joynt KE, Orav EJ, Epstein AM. The long-term effect of Premier pay for performance on patient outcomes. N Engl J Med. 2012; 366(17):1606–1615. [PubMed: 22455751]
9. Sutton M, Nikolova S, Boaden R, Lester H, McDonald R, Roland M. Reduced mortality with hospital pay for performance in England. N Engl J Med. 2012; 367(19):1821–1828. [PubMed: 23134382]
10. Epstein AM. Will pay for performance improve quality of care? The answer is in the details. New Engl J Med. 2012; 367(19):1852–1853. [PubMed: 23134388]
11. Fields D, Leshen E, Patel K. Analysis & commentary. Driving quality gains and cost savings through adoption of medical homes. Health Aff (Millwood). 2010; 29(5):819–826. [PubMed: 20439867]
12. Colla CH, Wennberg DE, Meara E, et al. Spending differences associated with the Medicare Physician Group Practice Demonstration. JAMA. 2012; 308(10):1015–1023. [PubMed: 22968890]
13. Petersen LA, Urech T, Simpson K, et al. Design, rationale, and baseline characteristics of a cluster randomized controlled trial of pay for performance for hypertension treatment: study protocol. Implement Sci. 2011; 6:114. [PubMed: 21967830]
14. Donner, A.; Klar, N. Design and Analysis of Cluster Randomization Trials in Health Research. London, England: Arnold; 2000.
15. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). JAMA. 2002; 288(23):2981–2997. [PubMed: 12479763]
16. Petersen LA, Simpson K, SoRelle R, Urech T, Chitwood SS. How variability in the institutional review board process affects minimal-risk multisite health services research. Ann Intern Med. 2012; 156(10):728–735. [PubMed: 22586010]
17. Chobanian AV, Bakris GL, Black HR, et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA. 2003; 289(19):2560–2572. [PubMed: 12748199]
18. Donner A, Klar N. Statistical considerations in the design and analysis of community intervention trials. J Clin Epidemiol. 1996; 49(4):435–439. [PubMed: 8621994]
19. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. Control Clin Trials. 1981; 2(2):93–113. [PubMed: 7273794]
20. Verbeke, G.; Molenberghs, G. Linear Mixed Models for Longitudinal Data. New York, NY: Springer; 2009.
21. Cheng J, Edwards LJ, Maldonado-Molina MM, Komro KA, Muller KE. Real longitudinal data analysis for real people: building a good enough mixed model. Stat Med. 2010; 29(4):504–520. [PubMed: 20013937]
22. Weissman, NW.; Kiefe, CI.; Allison, J., et al. Achievable Benchmarks of Care (ABC™) user manual. Birmingham, AL: University of Alabama at Birmingham, Center for Outcomes and Effectiveness Research and Education (COERE); 2001.
23. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Statist. 2001; 29(4):1165–1188.
24. Lester H, Schmittdiel J, Selby J, et al. The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators. BMJ. 2010; 340:c1898. [PubMed: 20460330]
25. Department of Veterans Affairs. Annual Pay Ranges for Physicians and Dentists of the Veterans Health Administration (VHA). Federal Register. 2009; 74(151):39734–39736.
26. Kuhn, M. Quality in Primary Care: Economic Approaches to Analysing Quality-Related Physician Behaviour. London, England: Office of Health Economics; 2003.
27. Kerr EA, Smith DM, Hogan MM. Building a better quality measure: are some patients with ‘poor quality’ actually getting good care? Med Care. 2003; 41(10):1173–1182. [PubMed: 14515113]
28. Woodard LD, Petersen LA. Improving the performance of performance measurement. J Gen Intern Med. 2009; 25(2):100–101. [PubMed: 19953336]

29. Conrad DA, Perry L. Quality-based financial incentives in health care: can we improve quality by paying for it? Annu Rev Public Health. 2009; 30:357–371. [PubMed: 19296779]
30. Frolich 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. [PubMed: 16624440]
31. Bonner SE, Sprinkle GB. The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. Accounting, Organizations, and Society. 2002; 27:303–345.
32. Bandura, A. Self-Efficacy: The Exercise of Control. New York, NY: W.H. Freeman and Company; 1997.
33. Hysong SJ, Simpson K, Pietz K, SoRelle R, Broussard Smitham K, Petersen LA. Financial incentives and physician commitment to guideline-recommended hypertension management. Am J Manag Care. 2012; 18(10):e378–391. [PubMed: 23145846]
34. O'Connor PJ. Overcome clinical inertia to control systolic blood pressure. Arch Intern Med. 2003; 163(22):2677–2678. [PubMed: 14662620]
35. Fletcher RD, Amdur RL, Kolodner R, et al. Blood pressure control among US veterans: a large multiyear analysis of blood pressure data from the Veterans Administration health data repository. Circulation. 2012; 125(20):2462–2468. [PubMed: 22515976]
36. Klein, S. The Veterans Health Administration: Implementing Patient-Centered Medical Homes in the Nation's Largest Integrated Delivery System. Washington, DC: The Commonwealth Fund; 2011.
37. Centers for Disease Control and Prevention (CDC). . Vital signs: prevalence, treatment, and control of hypertension—United States, 1999–2002 and 2005–2008. MMWR Morb Mortal Wkly Rep. 2011; 60(4):103–108. [PubMed: 21293325]
38. New England Healthcare Institute. [Accessed July 9, 2013] Thinking outside the pillbox: A system-wide approach to improving patient medication adherence for chronic disease. http://www.nehi.net/publications/44/thinking_outside_the_pillbox_a_systemwide_approach_to_improving_patient_medication_adherence_for_chronic_disease

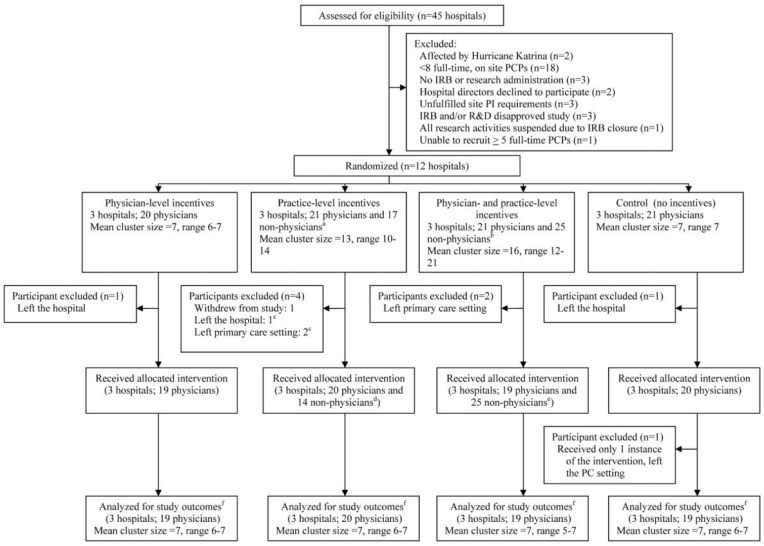

Figure 1.

Participant Enrollment

Abbreviations: IRB, Institutional Review Board; PC, primary care; PCP, primary care practitioners; PI, principal investigator; R&D, Research & Development

[a]Non-physician participants included 7 registered nurses, 4 nurse practitioners, 4 registered nurse case managers, 1 physician assistant, and 1 pharmacist.

[b]Non-physician participants included 10 licensed practical nurses, 9 registered nurses, 3 medical support assistants, 2 registered nurse care coordinators, and 1 pharmacist.

[c]Non-physician participant

[d]Two additional non-physicians left the practice-level incentives group during the intervention. One new non-physician was enrolled.

[e]Seven non-physicians left the physician- and practice-level incentives group during the intervention. Two new non-physicians were enrolled.

[f]Data from physicians only were used to assess both physician-level and practice-level performances; each physician participant received >/= 2 more instances of the intervention.

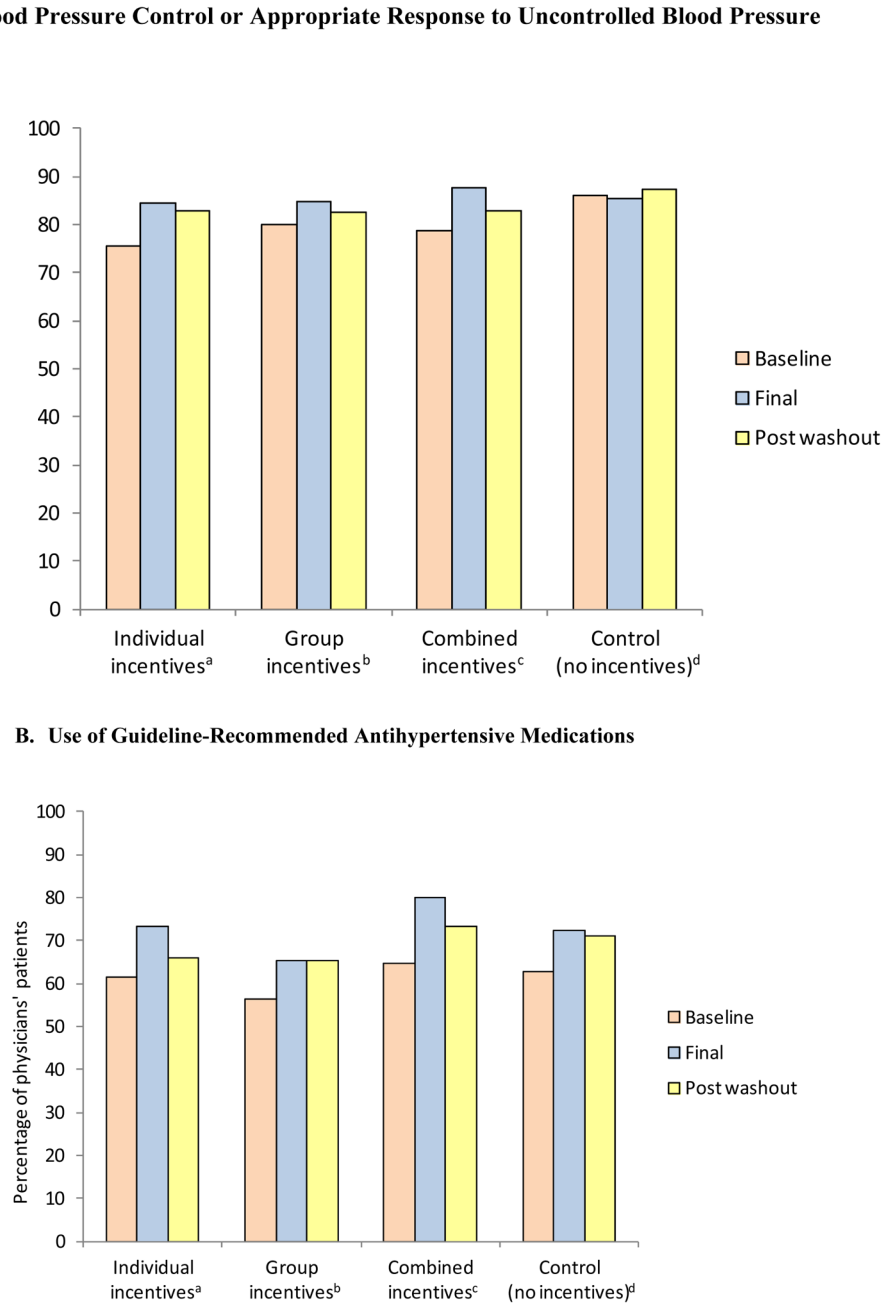

Figure 2.

Figure 2(A and B). Unadjusted Proportion of Physicians' Patients Meeting the Rewarded Clinical Measures by Study Group

A. Blood Pressure Control or Appropriate Response to Uncontrolled Blood Pressure

B. Use of Guideline-Recommended Antihypertensive Medications

[a]19 physicians participated in the baseline period; 760 patient charts reviewed. 18 physicians participated in the final intervention period; 720 patient charts reviewed. 13 physicians in the post-washout period; 520 patient charts reviewed.

[b]18 physicians participated in the baseline period; 720 patient charts reviewed. 19 physicians participated in the final intervention period; 760 patient charts reviewed. 17 physicians in the post-washout period; 680 patient charts reviewed.

[c]17 physicians participated in the baseline period; 680 patient charts reviewed. 15 physicians participated in the final intervention period; 600 patient charts reviewed. 11 physicians in the post-washout period; 440 patient charts reviewed.

[d]18 physicians participated in the baseline period; 720 patient charts reviewed. 19 physicians participated in the final intervention period; 760 patient charts reviewed. 14 physicians in the post-washout period; 560 patient charts reviewed.

Table 1

Distribution of Physician Characteristics Across Study Groups[a]

Characteristic	Individual incentives (N=19)	Practice incentives (N=20)	Combined incentives (N=19)	Control (no incentives) (N=19)	*P* Value[b]
Male sex – no. (%)	8 (42)	12 (60)	10 (53)	13 (68)	0.41
Race – no. (%)					
White	7 (37)	7 (35)	9 (47)	8 (42)	
Black	3 (16)	2 (10)	1 (5)	0	0.86[c]
Asian Indian	5 (26)	5 (25)	6 (32)	5 (26)	
Other	4 (21)	6 (30)	3 (16)	6 (32)	
Years practicing	12.9±8.4	13.3±8.7	11.4±6.8	13.4±7.4	0.88
Work at teaching hospital – no. (%)	19 (100)	20 (100)	12 (63)	13 (68)	<.001
Work at ALLHAT site – no. (%)	6 (32)	20 (100)	12 (63)	13 (68)	<.001
Work in southern US – no. (%)	13 (68)	7 (35)	12 (63)	6 (32)	0.04
Percent of male patients[d]	91±23	98±3	96±4	93±23	0.63
Percent of black patients, median (IQR)[d]	33 (5–43)	41 (9–48)	10 (5–37)	5 (3–43)	0.07
Average age of patients, y	65±5	66±3	66±3	67±5	0.69
Percent of patients who are diabetic[d]	42±11	39±8	41±7	38±9	0.39

Abbreviations: ALLHAT, Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial; IQR, interquartile range

[a] Plus-minus values are means ± SD. Because the patients for analysis were randomly selected at each time point, characteristics of physicians' patient samples in the first intervention period are displayed.

[b] For binomial variables, *P* values are from a chi-square test for difference in distribution across study groups (or Fisher's exact test if any cell sizes were less than 5). For continuous variables, *P* values are from the Kruskal-Wallis rank test.

[c] Because of cell counts <5, racial categories were collapsed into white versus non-white for a chi-square test of differences across study groups (df=3).

[d] Denominator by study group: 760 patients (individual); 800 patients (practice); 764 patients (combined); and 760 patients (control).

Table 2

Longitudinal Modeling Results of the Effect of the Intervention on Rewarded Clinical Measures

Rewarded Clinical Measure	Study Group	Estimated Change in the Percentage of Physicians' Patients Meeting the Measure (95% Bootstrap CI)	Estimated Difference in the Change Between Each Incentive Group and Control Group (95% Bootstrap CI)	*P* Value for Estimated Difference[a]
Blood pressure control or an appropriate response to uncontrolled blood pressure[b]				
	Individual incentives	8.84 (4.20, 11.80)	8.36 (2.40, 13.00)	.005
	Practice incentives	3.70 (0.24, 7.68)	3.24 (−1.48, 8.92)	.26
	Combined incentives	5.54 (1.92, 9.52)	5.08 (−0.04, 10.56)	.09
	Control (no incentives)	0.47 (−3.12, 4.04)	Reference	
Use of guideline-recommended antihypertensive medications[c]				
	Individual incentives	9.07 (4.52, 13.44)	4.72 (−1.44, 10.92)	.09
	Practice incentives	4.98 (0.64, 10.08)	0.64 (−5.32, 7.32)	.81
	Combined incentives	7.26 (2.92, 12.48)	2.92 (−2.76, 9.64)	.30
	Control (no incentives)	4.35 (−0.28, 9.28)	Reference	

Abbreviation: CI, confidence interval

[a] *P* values compare the change in the control group with the change in each incentive intervention group.

[b] Model adjusted for proportion of diabetic patients, whether physician works at a teaching hospital, and whether physician works in the northern or southern region of the US.

[c] Model adjusted for proportion of diabetic patients, mean age of patients, whether physician works at a teaching hospital, and ceiling effects.

Table 3

Linear Regression Analyses of Post-Washout Physician Performance

Rewarded Clinical Measure	Study Group	Unadjusted Difference in Physician Performance, Estimated Median (IQR)[a]	Difference in the Adjusted Mean Performance between Each Incentive Group and Control Group (β Coefficient (95% CI))
Blood pressure control or an appropriate response to uncontrolled blood pressure			
Overall		−2.5 (−7.5, 5.0)	
	Individual incentives	−2.5 (−7.5, 2.5)	−9.34 (−14.86, −3.82)
	Practice incentives	−1.3 (−7.5, 2.5)	−7.14 (−11.87, −2.42)
	Combined incentives	−5.0 (−12.5, 5.0)	−8.16 (−13.20, −3.12)
	Control (no incentives)	0 (−5.0, 7.5)	Reference[b]
Use of guideline-recommended antihypertensive medications			
Overall		−5.0 (−10.0, 2.5)	
	Individual incentives	−10.0 (−12.5, −5.0)	−10.03 (−16.44, −3.61)
	Practice incentives	−1.3 (−6.3, 5.0)	−5.60 (−11.25, 0.04)
	Combined incentives	−7.5 (−15.0, −2.5)	−4.65 (−11.13, 1.82)
	Control (no incentives)	0 (−7.5, 2.5)	Reference[c]

[a] Change in performance from the final intervention period to the post-washout period. Fifty-four of the 55 physicians examined for the post-washout period had performance data for both study periods.

[b] Model adjusted for differences in physician performance on the measure in the final intervention period, whether physician is white or black, and whether physician works in the northern or southern region of the US.

[c] Model adjusted for differences in physician performance on the measure in the final intervention period, proportion of male patients, whether physician works in the northern or southern region of the US, and ceiling effects.