

The Effect of Performance-Based Financial Incentives on Improving Patient Care Experiences: A Statewide Evaluation

Hector P. Rodriguez, PhD, MPH¹, Ted von Glahn, MS², Marc N. Elliott, PhD³, William H. Rogers, PhD⁴ and Dana Gelb Safran, ScD⁵

¹Department of Health Services, School of Public Health, University of California, Los Angeles, CA, USA; ²Pacific Business Group on Health, San Francisco, CA, USA; ³RAND Corporation, Santa Monica, CA, USA; ⁴Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA; ⁵Blue Cross Blue Shield of Massachusetts, Boston, MA, USA.

BACKGROUND: Patient experience measures are central to many pay-for-performance (P4P) programs nationally, but the effect of performance-based financial incentives on improving patient care experiences has not been assessed.

METHODS: The study uses Clinician & Group CAHPS data from commercially insured adult patients (n=124,021) who had visits with 1,444 primary care physicians from 25 California medical groups between 2003 and 2006. Medical directors were interviewed to assess the magnitude and nature of financial incentives directed at individual physicians and the patient experience improvement activities adopted by groups. Multilevel regression models were used to assess the relationship between performance change on patient care experience measures and medical group characteristics, financial incentives, and performance improvement activities.

RESULTS: Over the course of the study period, physicians improved performance on the physician-patient communication (0.62 point annual increase, $p < 0.001$), care coordination (0.48 point annual increase, $p < 0.001$), and office staff interaction (0.22 point annual increase, $p = 0.02$) measures. Physicians with lower baseline performance on patient experience measures experienced larger improvements ($p < 0.001$). Greater emphasis on clinical quality and patient experience criteria in individual physician incentive formulas was associated with larger improvements on the care coordination ($p < 0.01$) and office staff interaction ($p < 0.01$) measures. By contrast, greater emphasis on productivity and efficiency criteria was associated with declines in performance on the physician communication ($p < 0.01$) and office staff interaction ($p < 0.001$) composites.

CONCLUSIONS: In the context of statewide measurement, reporting, and performance-based financial

incentives, patient care experiences significantly improved. In order to promote patient-centered care in pay for performance and public reporting programs, the mechanisms by which program features influence performance improvement should be clarified.

KEY WORDS: financial incentives; pay-for-performance; public reporting; patients' experiences of care; quality improvement.

J Gen Intern Med 24(12):1281-8

DOI: 10.1007/s11606-009-1122-6

© Society of General Internal Medicine 2009

INTRODUCTION

Using financial incentives to induce meaningful and lasting improvements in health care quality is a strategy that is gaining rapid and widespread appeal in the United States¹⁻³. To date, pay-for-performance programs (P4P) have been employed in select markets in the US and most extensively implemented in the UK⁴⁻⁶. Recent efforts by the Centers for Medicare and Medicaid Services (CMS) to implement P4P in the Medicare program, however, have catapulted the approach into high prominence here⁷. In the US, the state of California has the largest and most long-standing experience with pay-for-performance. The Integrated Health Association's (IHA) statewide initiative in California, launched in 2004, addresses three areas of performance: clinical care processes, patient care experiences, and office-based information systems. Through the combined efforts of six insurers, California medical groups are offered financial incentives for achieving high performance in these areas. From 2004 through 2008, the annual incentive payments have ranged from \$38-60 million⁸. In addition to these financial incentives, medical group performance is motivated by the annual public reporting of group-level performance results in two of the three areas: clinical care processes and patient care experiences. Medical groups use incentive compensation in diverse ways. For example, some groups may choose to direct incentive compensation to individual physicians using group-defined financial incentives, while some groups use incentive compensation for other organizational priorities⁹.

The successes¹⁰⁻¹³ and challenges¹⁴⁻¹⁸ of improving clinical care processes have been well documented following the introduction of public reporting and/or financial incentives related to this area of measurement, but virtually nothing is known about the potential for patient care experiences to be improved through

Electronic supplementary material The online version of this article (doi:10.1007/s11606-009-1122-6) contains supplementary material, which is available to authorized users.

Financial Incentives and Patient Care Experiences

Received February 17, 2009

Revised August 3, 2009

Accepted September 3, 2009

Published online October 14, 2009

public reporting or financial incentive programs^{19,20}. Recent studies in the UK, however, indicate that pay-for-performance reduced continuity of patient care²¹, changed the dynamic between doctors and nurses and the nature of the physician-patient relationship²². Using longitudinal survey data from patients of 1,444 primary care physicians (PCPs) belonging to 27 California medical groups during 2004–2007, our study examines whether the magnitude and nature of medical group performance-based financial incentives are associated with improved patients' experiences of primary care.

METHODS

Patient Sampling and Survey Administration

The study draws on commercially insured patients who had encounters with 1,444 adult PCPs from 27 medical groups in California between the years 2003–2006. During each of the four survey years (2004–2007), a random sample of approximately 100 patients per physician who had at least one visit with their PCP during the prior year were mailed a survey. All commercially insured patients visiting the physicians during the study year were eligible to receive a survey. Although an individual patient could respond to the survey in multiple years, the deidentified patient survey data could not be linked at the patient level.

The survey included core measures from the Clinician & Group CAHPS survey²³ and supplemental measures from the Ambulatory Care Experiences Survey (ACES), a validated survey that measures patients' experiences with a specific, named physician and that physician's practice²⁴. Mailings included an invitation letter, a printed survey, and a postage-paid return envelope. The survey invitation included a personal online code that gave respondents the option of completing the survey using the web. Previous analysis has demonstrated the absence of web survey mode effects for the patient survey questions²⁵. The survey invitation listed a toll-free number for patients to obtain surveys in Spanish. A second invitation and questionnaire were sent to non-respondents 2 weeks after the initial mailing. Each annual data collection effort spanned a period of approximately 8 weeks.

Patient Survey Content

For this analysis, we consider four survey composite measures: physician communication (six items), care coordination (two items), access to care (five items), and office staff interaction (two items) (see [Appendix](#)). The physician communication, access to care, and office staff interaction summary measures represent core item and composite content of the CAHPS Clinician & Group Survey²³, which was endorsed by the National Quality Forum (NQF) for use in evaluating ambulatory care received from individual physicians and their practices. All survey questions consist of a six-point response continuum ranging from "never" to "always". All questions reference care received from that particular physician and the physician's practice over the past 12 months. Composite measures all achieve physician-level reliability of 0.70 or higher with samples of approximately 30–40 established patients per physician^{24,26,27}.

Patient Survey Composite Scoring

As detailed elsewhere²⁴, survey questions were scored linearly from 0 ("never") to 100 ("always") points, with higher scores indicating more favorable performance. Composites were calculated as the unweighted average of responses to all items comprising the measure after applying the half-scale rule^{28,29}, which includes only respondents who complete at least half of the questions comprising a composite. Most measures had consistent item consistent and wording across years. For the access to care composite, we used an adjusted half-scale approach to calculate scores³⁰. This scoring method facilitated the comparison of composites across time because items were treated in a comparable way across years, and each individual item's contribution to the overall scale was factored into the composite calculation.

Analytic Sample

Of 399,392 outgoing patient surveys over the 4 study years, 14,226(3.6%) were undeliverable because of bad address information or patient death. Surveys were received from 145,522 respondents, yielding an adjusted response rate of 37.8%. The analytic sample included 135,401 respondents (average per physician=93.8) who confirmed having seen their PCP during the prior 12 months. Respondents who did not confirm the named physician as their PCP or indicated that they did not visit the physician during the prior 12 months (n=10,121) were excluded from the analysis. The commercially insured respondent sample was 35.1% male, 64.9% non-Hispanic White, 13.1% Hispanic, 12.2% Asian, 3.6% Black, 44.4% reported completing college, and 67.5% reported having at least one chronic medical condition; 47.1% reported being established with their PCP for 5 years or longer. Of the 1,444 physicians included in the analytic sample, 427 physicians participated in the survey initiative all 4 years (2004–2007), 283 physicians participated during 3 of the 4 years, and 734 participated during 2 of the 4 years.

Medical Director Interview

Medical director interviews were conducted via telephone between April and June 2007. The pool of eligible medical groups consisted of groups that participated in IHA's medical group performance-based financial incentive program and assessed patient care experiences at the individual physician-level in 2007. The interview assessed whether PCPs were eligible to receive performance-based financial incentives and the maximum possible magnitude of the incentives as a percentage of base compensation. In addition, directors were asked about the formulae used to calculate physician incentives, including the percent of the incentive that was based on productivity (e.g., average patients seen per day), efficiency (e.g., limiting referrals, effective panel management), patient experience measures, clinical quality measures, and "other" criteria (physician seniority, prescribing generic medications, and other contributions to the medical group). Finally, directors were asked whether their organization was presently engaged in various patient experience performance improvement activities, including sharing patients' experience measures with physicians in individual feedback sessions, interpersonal skills training, business practice redesign, and practice leader compensation. A composite measure reflecting

the total number of patient experience improvement activities adopted by groups was constructed by summing responses to the individual questions (range: 0–4, $\alpha=0.65$). Of the 27 medical groups eligible for this study (groups with 2 or more years of participation in the patient survey initiative), interviews were conducted with 25 medical group directors or designees, resulting in a 92.6% response rate.

Statistical Analyses

Change over time for each patient survey composite was assessed using two sets of multilevel regression models. First, models were specified for each patient survey composite that used physician and medical group random effects to account for the clustering of patients within physicians and physicians and medical groups. These unadjusted models included a continuous measure of time (survey year). The second set of multilevel regression models was specified identically to the unadjusted models, but also controlled for patient age, gender, race/ethnicity, education, and self-rated physical health, which are commonly used to adjust patient experience measures^{31,32}. For the physician random effects, the intercept value represents the first year of participation in the survey initiative for each physician, and the slope represents the annual performance change over the physician's baseline year. The model also permitted calculation of the correlation between the physician intercept and slope³³ in order to assess the extent to which physician performance change over time was associated with baseline performance. Following Elliott et. al.³⁴, we standardized the change over time coefficients in terms of physician-level standard deviations (SDs) for each composite. This allowed us to describe the average annual change as the change in percentile points it would represent for a physician at the 50th percentile at baseline. We also examined the extent to which alternative explanations, including panel maturation, varying (and sometimes small) patient samples sizes per physician, and physicians' years of participation might account for change over time. These analyses are presented in the online appendix.

Physicians' baseline CAHPS composite scores were compared by medical group characteristics, financial incentives, and patient experience improvement activities. Differences in medical group baseline composite scores by medical group variables were assessed using multilevel models that accounted for the clustering of patients within physicians and physicians within medical groups. Finally, for each survey composite measure, we assessed the medical group-level predictors of physician performance change over time. We chose to introduce each of the medical group financial incentive and patient experience improvement variables independently because the modest number of medical groups involved ($n=25$) could result in overfitting of models. For example, 24 degrees of freedom at the medical group level could perfectly capture any arbitrary pattern of the medical group means, even if they were entirely unrelated to the specified predictors. Each model was specified using physician and medical group random effects and controlled for medical group type (staff model vs. IPA vs. hybrid). To account for differences in the patient case mix across individual physician practices and years, models also controlled for patient age, gender, race/ethnicity, education, and self-rated physical health. Each model included the main effect and time interactions for the medical group variable under study. Statistically significant medical group variables were then specified jointly in models. Joint Wald

tests were used to examine the extent to which the set of interaction terms (change over time predictors) jointly accounted for a statistically significant amount of the observed change over time on the composite measure. All continuous patient and medical group measures were standardized to a mean of 0 and a variance of 1 for the comparable interpretation of regression coefficients. All analyses were conducted using the XTMIXED module in STATA 10.0³³, which uses restricted maximum likelihood (REML) to estimate random effects.

RESULTS

Performance Change over Time

Table 1 presents unadjusted and adjusted annual change scores, by composite measure. During the course of the study period (measurement years 2004–2007), performance on the physician communication composite increased by 0.74 points per year [95% confidence interval (CI): 0.61, 0.87, $p<0.001$], indicating that physicians who participated in the initiative during all 4 years improved by an average of 2.22 points (95% CI: 1.83, 2.61). The magnitude of improvement decreased slightly in adjusted analyses (annual point change=0.62; 95% CI: 0.49, 0.75, $p<0.001$). A 0.62 point annual point improvement for a physician at the 50th percentile at baseline corresponds to a 10.8 percentile point increase in percentile rank during the first follow-up period, i.e., the equivalent of improving to the 61st percentile of the baseline distribution (Table 1, column 4). There were also statistically significant adjusted improvements on the care coordination (0.48; 95% CI: 0.26, 0.69) and office staff interaction (annual point change=0.22; 95% CI: 0.04, 0.40, $p=0.02$) composites. However, in adjusted analyses, performance on the access composite (annual point change=0.06; 95% CI: -0.19, 0.32) did not improve over time. The physician slope-intercept correlation was comparable across composite measures (between -0.34 and -0.51; Table 1, column 4). Negative correlations indicate that physicians with lower baseline performance on composite measures improved more over time compared to physicians with higher baseline performance.

Medical Group Activities and Baseline Performance

Physicians' baseline performance on patient survey composite scores did not differ by most medical group characteristics or the number of patient experience improvement activities adopted (Table 2). Baseline performance differed by medical groups' financial incentive formulae and financial incentive magnitude for the care coordination and office staff interactions composites, although the associations were not consistent across measures.

Medical Group Predictors of Performance Change

Some medical group financial incentive characteristics were significantly associated with change over time on patient experience measures when examined independently of other group-level predictors (Table 3). Greater emphasis on clinical quality and patient experience criteria in individual physician incentive formulas was associated with larger improvements on the care coordination ($p<0.01$) and office staff interaction (p

Table 1. Annual Change in Patients' Experiences of Primary Care (2004–2007)

Composite measure	Unadjusted annual score change	Adjusted annual score change ^a	Annual change in physician percentile points (adjusted model) ^b	Physician slope-intercept correlation (adjusted model)
Physician communication	0.74 (0.61, 0.87) **	0.62 (0.49, 0.75) **	10.8	-0.38 (-0.47, -0.28) **
Care coordination	0.76 (0.53, 0.98) **	0.48 (0.26, 0.69) **	6.7	-0.43 (-0.52, -0.34) **
Access to care	0.32 (0.06, 0.58) *	0.06 (-0.19, 0.32)	1.0	-0.34 (-0.42, -0.26) **
Office staff interactions	0.47 (0.28, 0.66) **	0.22 (0.04, 0.40) *	3.5	-0.51 (-0.57, -0.44) **

Notes:

Composite measures are scored on a 0–100-point scale. All change over time estimates account for the clustering of patients within physicians and physicians within medical groups

^aAdjusted estimates control for patient age, gender, race/ethnicity, education, and self-rated physical health

^bChanges in physician percentile points are calculated for the 50th percentile of physician performance at baseline

The physician slope-intercept correlation reflects the extent to which physician performance change is associated with baseline performance. A negative slope-intercept correlation indicates that physicians with lower baseline performance had larger improvements compared to physicians with higher baseline performance

* $p < 0.05$, ** $p < 0.001$

Table 2. Baseline Medical Group Performance on Patient Experience Measures by Group Characteristics

	Group n	Patient experience measure			
		Physician communication	Care coordination	Access to care	Office staff interactions
Overall	25	87.9	77.3	83.7	83.5
Group size—Number of primary care physicians (PCPs)					
≤20 PCPs	5	88.9	78.9	83.3	85.0
21–38 PCPs	6	87.0	76.7	81.4	82.1
39–86 PCPs	6	87.3	77.6	83.2	84.0
≥87 PCPs (reference)	8	88.1	77.2	84.2	83.4
Medical group type					
Independent practice association (reference)	13	87.1	77.8	85.3	84.2
Integrated medical group	10	88.5	77.4	82.8	82.9
Hybrid group	2	89.2	77.5	80.1	81.5
Financial incentive magnitude					
Primary care physicians not eligible	4	86.1	74.5	82.3	83.6
≤10% of base compensation	16	88.4	78.3*	82.8	82.9
>10% of base compensation (reference)	5	87.2	75.7	85.9	84.4
Financial incentive payment formulae					
Productivity and efficiency					
None	11	88.1	76.9	81.7	82.1
1–30% of weight	3	88.5	80.5	85.7	84.9
31% or more weight (reference)	7	87.4	76.3	85.3	84.2
Patient experience					
None	1	87.5	76.9	84.7	86.2**
1–30% of weight	15	87.8	77.7	85.1	84.4
31% or more weight (reference)	5	88.2	76.8	81.2	81.4
Clinical quality					
None	1	87.5	76.9	84.7	86.2**
1–30% of weight	12	87.9	78.0	85.8	84.6
31% or more weight (reference)	8	87.9	76.5	81.0	81.5
Other criteria					
None	7	87.7	75.4*	80.5	82.4
1–30% of weight	9	87.7	77.0	85.0	83.6
31% or more weight (reference)	5	88.7	80.5	85.4	85.0
Patient experience improvement activities (count) ^a					
None	6	86.9	77.9	84.4	83.3
Two	8	88.5	79.0	83.4	83.2
Three	7	87.6	76.5	81.9	83.0
Four (reference)	4	87.8	76.3	85.0	83.9

Notes: Patient experience measures are scored on a scale of 0–100, with higher values indicating better performance; * $p < 0.05$, ** $p < 0.01$ compared to reference group. Results account for the clustering of patients within physicians and physicians within medical groups using random effects.

^aNone of the medical directors reported adopting only one patient experience improvement activity

Table 3. Medical Group Characteristics and Performance Change over Time on Patient Experience Measures

	Patient experience measure			
	Physician communication	Care coordination	Access to care	Office staff interactions
Medical group predictors of change over time (time interactions)^a				
Financial incentive magnitude				
>10% of base compensation (reference)	–	–	–	–
≤10% of base compensation	0.36**	–0.83	0.24	0.74***
PCPs not eligible for incentives	–0.37	0.36	0.10	–0.24
Financial incentive payment formulae				
Patients' experiences (%)	0.01	0.03**	0.00	0.03***
Productivity and efficiency (%)	–0.19**	–0.23	0.01	–0.34***
Clinical quality (%)	0.20**	0.27*	0.08	0.30**
Patient experience improvement activities				
Improvement activities count	–0.11	0.04	–0.07	–0.18
Change over time joint tests^b				
Productivity and efficiency weight in formula	9.26 (2)**			
Clinical quality weight in formula				
Patient experiences weight in formula		8.93 (2)*		
Clinical quality weight in formula				
Productivity and efficiency weight in formula				17.56 (2)***
Patient experiences weight in formula				
Productivity and efficiency weight in formula	11.59 (4)*			
Clinical quality weight in formula				
Financial incentive magnitude (two terms)				
Productivity and efficiency weight in formula				20.28 (4)***
Patient experiences weight in formula				
Financial incentive magnitude (two terms)				

Notes:

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Change over time predictors were examined individually in models that controlled for patient age, gender, race/ethnicity, education, and self-rated physical health, medical group type, the main effect of the change over time variable, and accounted for the clustering of patients within physicians and physicians within medical group using random effects

^aPatient experience measures are scored on a scale of 0–100, with higher values indicating better performance. Coefficients represent the effect of the characteristic or activity on change on the composite score. Continuous measures, e.g., financial incentive payment formulae, are standardized to have a mean of 0 and a standard deviation of 1, so the coefficients represent the effect of a standard deviation change for the variable on performance change. Joint tests examine to the extent to which the interaction terms (change over time predictors) jointly account for a statistically significant amount of the observed change over time on the composite measure

^bThe χ^2 statistics and number of variables tested in the joint Wald tests are presented, along with statistical significance

<0.01) measures. By contrast, greater emphasis on productivity and efficiency criteria was associated with worse performance over time on the physician communication ($p < 0.01$) and office staff interaction ($p < 0.001$) measures. Contrary to our expectations, physicians belonging to groups that used smaller ($\leq 10\%$ of base compensation) incentives improved more over time on the communication ($p < 0.01$) and office staff interaction ($p < 0.001$) measures compared to physicians belonging to groups that used larger ($> 10\%$ of base compensation) incentives. This counterintuitive result likely stems from the fact that the groups with larger incentives used heavy productivity and efficiency criteria in their formulae (data not shown). The number of patient experience improvement activities undertaken by medical groups was not associated with performance change over time.

When significant medical group predictors of change over time were jointly tested in models that included one another, most joint tests indicated that the combinations remained statistically significant. This suggests that one or more of the variables included in the joint test explain change over time on the patient experience composite, but multilevel models that account for both interaction terms lack the statistical power to distinguish between them. For example, the magnitude and

formulae of individual physician financial incentives were independently associated with improvements on the physician communication composite, but not in models that included one another. Tests examining the joint significance of financial incentive magnitude and formulae ($\chi^2 = 11.59$, $p < 0.05$) suggest a substantial amount of shared association with improvements in physician communication.

DISCUSSION

This study assessing the relationship of medical group performance-based financial incentives and individual physician performance improvement on patient care experience measures has several important findings relevant to the design and implementation of pay-for-performance programs. First, performance on the physician communication, care coordination, and office staff interactions composite measures have significantly improved since the inception of California's pay-for-performance program in 2004. The magnitude of annual improvement on the physician communication and care coordination composites is associated with a 10.8 and 6.7 annual percentile point increase, respectively, for a physician

at the 50th percentile at baseline. Many stakeholders would consider a 5–10 point improvement in a physician's percentile rank over one year as a practically meaningful change. Moreover, previous research assessing the effect of communication interventions on improvements on patients' experiences measures³⁵ suggests that the observed changes in physician communication are clinically meaningful. Our findings underscore that even with fairly tight performance distributions commonly found for patient satisfaction and experience measures³⁶, significant improvement is possible.

The improvement is also noteworthy because previous analyses have found substantial declines in physician-patient relationship quality over time. For example, in the late 1990s, before the Institute of Medicine called attention to the importance of patient-centered care³⁷, there were no statewide or other large-scale efforts to measure patient care experiences. At this time, primary care performance reporting and incentives focused entirely on clinical quality measures, and substantial improvement was observed^{38–40}. Patients' experiences of primary care, however, deteriorated in both commercially insured adult and Medicare-insured elderly patient populations^{41,42}. The improvements in primary care patients' experiences observed in this study are a marked contrast to those earlier trends, and may be owed, in part, to the heightened salience of patient-centered care nationally during that period and the specific measurement and accountability activities related to patient care experiences in California. Our results were fairly robust to many sensitivity tests assessing alternative explanations, including small patient sample sizes for some physicians, physicians with fewer years of participation, and panel maturation. Future research should clarify the extent to which observed improvements stem from secular trends, performance-based financial incentives and other improvement activities.

Consistent with evidence assessing the effects of pay-for-performance programs on the technical quality of care¹¹, our slope-intercept correlation results indicate that physicians with lower baseline patient survey scores were more likely to improve performance compared to physicians with higher scores at baseline even with random effects controlling for regression to the mean. This suggests that targeting patient experience improvement activities at individual physician practices with lower baseline performance may result in the most cost-effective use of resources. Medical groups and health plans, however, might want to foster a quality culture by making improvement activities broadly available to all physician practices. For example, including high performing practices in improvement activities may facilitate organizational learning⁴³.

Finally, we found that the criteria used in individual physicians' financial incentive formula were associated with changes over time on most composite measures. Specifically, greater emphases on clinical quality and patient experience performance and fewer emphases on productivity and efficiency criteria were associated with larger improvements. Our results are consistent with evidence that strong use of productivity incentives may result in unintended consequences, including physician⁴⁴ and patient⁴⁵ dissatisfaction. Productivity incentives might not effectively cultivate the working relationships of physicians, advanced practice clinicians, and office staff⁴⁶, and a weak relationship emphasis may spill over to patient care. In addition to clarifying the precise mechanisms by which medical group financial incentive characteristics affect performance on patient experience measures, other organizational influences,

including group culture⁴⁷, should be examined in larger samples of medical groups.

Our findings should be considered in light of important limitations. First, while detailed information about the magnitude and nature of financial incentives directed at individual physicians was analyzed, we could not account for secular trends in performance improvement analytically because a control group was not available. However, the study is the first to find large-scale improvements in patients' experiences of primary care over time, suggesting that the public reporting and/or financial incentive program may have induced improvement. Second, response rates across years were modest, and information about non-respondents was not available. As a result, it is not possible to assess the extent to which differential patient non-response by physician may affect the reliable measurement of physician performance. Previous analyses indicate that the *nature* of non-response on the patient survey measures does not differ significantly across physicians and that differences in the *extent* of non-response across physicians are too small to meaningfully affect overall results²⁴. Consequently, non-response seems unlikely to threaten the integrity of the physician-level analyses presented. Third, although the study's physician sample is large, our results might not generalize to other states with different demographic distributions or market conditions. However, the medical groups studied are a mix of integrated, salaried practices and independent practice associations and are likely representative of groups that are actively engaged in performance improvement. Finally, we did not assess physician influences on performance change over time. Individual physicians account for the largest proportion of explainable variation on ambulatory care experience measures^{24,48}, so physician factors are likely to explain performance change on patient experience measures. For example, data on physician characteristics and attitudes concerning public reporting and pay-for-performance implementation⁴⁹ might explain performance improvement.

In conclusion, our study, the first to assess the relationship between the use of performance-based financial incentives and changes in patients' experiences over time in the US²¹, suggests public reporting and pay-for-performance can potentially improve physician communication, access to care, and office staff interactions as experienced and reported by patients. In the context of statewide measurement, reporting, and performance-based financial incentives, patient care experiences significantly improved. The study is the first to document large-scale improvements in patient care experiences over time^{41,42}. The magnitude and nature of financial incentives directed at individual physicians explained a significant proportion of the observed improvements, indicating that the incentives used by groups or unmeasured characteristics of medical groups associated with the use of incentives, e.g., group culture, account for changes in physician performance on patient experience measures over time. In order to promote patient-centered care in pay-for-performance and public reporting programs, the mechanisms by which program features affect performance improvement should be clarified.

Acknowledgements: This project was supported by a research grant from the Commonwealth Fund (no. 20070067). We would like to thank Serwar Ahmed, Rachel Brodie, and Amy Rassbach for contributions to the collection of medical director interview data.

Angie Mae C. Rodday, Angela Li, Sean M. Devlin, and Hong Chang for assistance with data management and analytic support, and Gary Chan for statistical consultation. Special thanks to Doug Conrad for helpful suggestions on an early draft of the paper. We also are grateful to Jeff Burkeen and his staff at the Center for the Study of Services (CSS) for their expertise in the sampling and data collection activities that supported this work.

Conflict of Interest: None disclosed.

Corresponding Author: Hector P. Rodriguez, PhD, MPH; Department of Health Services, School of Public Health, University of California, Box 951772, Los Angeles, CA 90095-1772, USA (e-mail: hrod@ucla.edu).

REFERENCES

- Conrad DA, Christianson JB. Penetrating the "black box": financial incentives for enhancing the quality of physician services. *Med Care Res Rev.* 2004;61(3 Suppl):375-68S.
- Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *JAMA.* 2007;297(7):740-4.
- Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *N Engl J Med.* 2006;355(18):1895-902.
- Doran T, Fullwood C, Gravelle H, et al. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med.* 2006;355(4):375-84.
- Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. *N Engl J Med.* 2007;357(2):181-90.
- Ashworth M, Millett C. Quality improvement in UK primary care: the role of financial incentives. *J Ambul Care Manage.* 2008;31(3):220-5.
- Institute of Medicine. Rewarding Provider Performance: Aligning Incentives in Medicare. Washington, DC: Board on Health Care Services (HCS), Institute of Medicine (IOM); 2007.
- Damberg CL, Raube K, Williams T, Shortell SM. Paying for performance: implementing a statewide project in California. *Qual Manag Health Care.* 2005;14(2):66-79.
- Damberg CL, Raube K, Teleki SS, Dela Cruz E. Taking stock of pay-for-performance: a candid assessment from the front lines. *Health Aff (Millwood).* 2009;28(2):517-25.
- Beaulieu ND, Horrigan DR. Putting smart money to work for quality improvement. *Health Serv Res.* 2005;40(5 Pt 1):1318-34.
- Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. *JAMA.* 2005;294(14):1788-93.
- 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):486-96.
- Young GJ, Meterko M, Beckman H, et al. Effects of paying physicians based on their relative performance for quality. *J Gen Intern Med.* 2007;22(6):872-6.
- Dudley RA. Pay-for-performance research: how to learn what clinicians and policy makers need to know. *JAMA.* 2005;294(14):1821-3.
- Conrad DA, Saver BG, Court B, Heath S. Paying physicians for quality: evidence from the field. *Jt Comm J Qual Patient Saf.* 2006;32(8):443-51.
- Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in health care? *Med Care Res Rev.* 2006;63(2):135-57.
- 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):265-72.
- Glickman SW, Ou FS, DeLong ER, et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *Jama.* 2007;297(21):2373-80.
- Snyder L, Neubauer RL. Pay-for-performance principles that promote patient-centered care: an ethics manifesto. *Ann Intern Med.* 2007;147(11):792-4.
- Rowe JW. Pay-for-performance and accountability: related themes in improving health care. *Ann Intern Med.* 2006;145(9):695-9.
- Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of Pay for Performance on the Quality of Primary Care in England. *N Engl J Med.* 2009;361(4):368-78.
- Campbell SM, McDonald R, Lester H. The experience of pay for performance in English family practice: a qualitative study. *Ann Fam Med.* 2008;6(3):228-34.
- Agency for Healthcare Research Quality, American Institutes for Research, Harvard Medical School, Rand Corporation. The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Clinician and Group Survey: Submission to National Quality Forum. <http://www.aqaalliance.org/October24Meeting/PerformanceMeasurement/CAHPSCtext.doc>. Accessed August 26, 2009.
- Safran DG, Karp M, Coltin K, et al. Measuring patients' experiences with individual primary care physicians. Results of a statewide demonstration project. *J Gen Intern Med.* 2006;21(1):13-21.
- Rodriguez HP, von Glahn T, Rogers WH, Chang H, Fanjiang G, Safran DG. Evaluating patients' experiences with individual physicians: A randomized trial of mail, internet, and interactive voice response telephone administration of surveys. *Med Care.* 2006;44(2):167-74.
- Rodriguez HP, von Glahn T, Chang H, Rogers WH, Safran DG. Patient samples for measuring primary care physician performance: Who should be included? *Med Care.* 2007;45(10):989-96.
- Rodriguez HP, von Glahn T, Chang H, Rogers WH, Safran DG. Measuring patients' experiences with individual specialist physicians and their practices. *Am J Med Qual.* 2009;24(1):35-44.
- Nunnally J, Bernstein I. *Psychometric Theory.* New York: McGraw-Hill; 1994.
- Rogers WH, Chang H. Patient Assessment Survey (PAS) 2008: Consumer Reporting Methods for Office of the Patient Advocate. http://www.opa.ca.gov/report_card/pdfs/Scoring%20Documentation_PAS%20Medical%20Group%20Measures_August%202008.pdf. Accessed August 26, 2009.
- Rogers WH, Chang H. Scoring Documentation for Consumer Reporting for IHA 2008 Effectiveness of Care Measures. http://www.opa.ca.gov/report_card/pdfs/Scoring%20Documentation_IHA%20Clinical%20Measures_August%202008.pdf. Accessed August 26, 2009.
- Zaslavsky AM, Zaboriski L, Cleary PD. Does the effect of respondent characteristics on consumer assessments vary across health plans? *Med Care Res Rev.* 2000;57(3):379-94.
- Zaslavsky AM, Zaboriski LB, Ding L, Shaul JA, Cioffi MJ, Cleary PD. Adjusting performance measures to ensure equitable plan comparisons. *Health Care Financ Rev.* 2001;22(3):109-26.
- Longitudinal/Panel Data Reference Manual* [computer program]. Version. College Station. Stata Press, TX; 2007.
- Elliott MN, Zaslavsky AM, Goldstein E, et al. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res.* 2009;44(2 Pt 1):501-18.
- Rao JK, Anderson LA, Inui TS, Frankel RM. Communication interventions make a difference in conversations between physicians and patients: a systematic review of the evidence. *Med Care.* 2007;45(4):340-9.
- Rosenthal GE, Shannon SE. The use of patient perceptions in the evaluation of health-care delivery systems. *Med Care.* 1997;35(11 Suppl):NS58-68.
- Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*, Vol 6. Washington, DC: National Academy Press; 2001.
- Lied TR, Sheingold S. HEDIS performance trends in Medicare managed care. *Health Care Financ Rev.* 2001;23(1):149-60.
- Bost JE. Managed care organizations publicly reporting three years of HEDIS measures. *Manag Care Interface.* 2001;14(9):50-4.
- Health plans bear down on quality, HEDIS scores improve dramatically. *Manag Care.* 2001;10(10):34-5.
- Montgomery JE, Irish JT, Wilson IB, et al. Primary care experiences of medicare beneficiaries, 1998 to 2000. *J Gen Intern Med.* 2004;19(10):991-8.
- Murphy J, Chang H, Montgomery JE, Rogers WH, Safran DG. The quality of physician-patient relationships. Patients' experiences 1996-1999. *J Fam Pract.* 2001;50(2):123-9.
- Garvin DA, Edmondson AC, Gino F. Is yours a learning organization? *Harv Bus Rev.* 2008;86(3):109-+.
- Grumbach K, Osmond D, Vranizan K, Jaffe D, Bindman AB. Primary care physicians' experience of financial incentives in managed-care systems. *N Engl J Med.* 1998;339(21):1516-21.
- Rodriguez HP, Von Glahn T, Rogers WH, Safran DG. Organizational and market influences on physician performance on patient experience measures. *Health Serv Res.* 2009;44(3):880-901.

46. **Safran DG, Miller W, Beckman H.** Organizational dimensions of relationship-centered care. Theory, evidence, and practice. *J Gen Intern Med.* 2006;21(Suppl 1):S9-15.
47. **Helfrich CD, Li YF, Mohr DC, Meterko M, Sales AE.** Assessing an organizational culture instrument based on the Competing Values Framework: Exploratory and confirmatory factor analyses. *Implement Sci.* 2007;2:13.
48. **Rodriguez HP, Scoggins JF, von Glahn T, Zaslavsky AM, Safran DG.** Attributing sources of variation in patients' experiences of ambulatory care. *Med Care.* 2009;47(8):835-41.
49. **Meterko M, Young GJ, White B, et al.** Provider attitudes toward pay-for-performance programs: development and validation of a measurement instrument. *Health Serv Res.* 2006;41(5):1959-78.

APPENDIX

Table 4. Patient Survey Composite Measures

Composite measure	Item content
Physician communication ($\alpha=0.93$)	In the last 12 months... ...how often did this doctor <i>explain things</i> in a way that was easy to understand? ...how often did this doctor <i>listen carefully</i> to you? ...how often did this doctor give you easy-to-understand instructions about <i>what to do to take care of the health problems or concerns</i> that were bothering you? ...how often did this doctor seem to know the important information about your medical history? ¹ ...how often did this doctor spend enough time with you? ...how often did this doctor show respect for what you had to say?
Care coordination ($\alpha=0.61$)	In the last 12 months... ...how often did this doctor seem <i>informed and up-to-date</i> about the care you got from specialist doctors? ...when this doctor sent you for a blood test, x-ray, or other test, how often did someone from the doctor's office follow up to give you the test results?
Access to care ($\alpha=0.85$)	In the last 12 months... ...when you called this doctor's office to get an appointment for care <i>you needed right away</i> , how often did you get an appointment as soon as you thought you needed it? ...when you made an appointment for a <i>check-up or routine care</i> with this doctor, how often did you get an appointment as soon as you thought you needed it? ...when you called this doctor's office with a medical question <i>during regular office hours</i> , how often did you get an answer to your question that same day? ...when you called this doctor's office <i>after regular office hours</i> , how often did you get the medical help or advice you needed? Wait time includes times spent in the waiting room and exam room. In the last 12 months, how often did your visits at this doctor's office start within 15 min of your appointment?
Office staff interactions ($\alpha=0.89$)	In the last 12 months... ...how often were clerks and receptionists at this doctor's office as helpful as you thought they should be? ...how often did clerks and receptionists at this doctor's office treat you with courtesy and respect?

Notes: Response scales and scoring for all questions: never (0 points), almost never (20 points), sometimes (40 points), usually (60 points), almost always (80 points), always (100 points)