J Ambulatory Care Manage
Vol. 32, No. 2, pp. 103–114

A Multimodal Strategy Based on Pay-Per-Performance to Improve Quality of Care of Family Practitioners in Argentina

Adolfo Rubinstein, MD, MSc, PhD; Fernando Rubinstein, MPH, MD; Marcela Botargues, MD; Mariela Barani, MD; Karin Kopitowski, MD

Abstract: Pay-for-performance has become increasingly common to complement physician reimbursement. We designed a quality framework to measure family physicians' performance in a managed care setting in Buenos Aires. We aimed to assess the effectiveness of a multimodal intervention based on pay-for-performance, teamwork, continuous education, and audit and feedback to improve quality. After 2 years, a significant improvement was observed in most of the indicators measuring clinical effectiveness and some improvements were observed in other domains. Despite these results, a better performance matrix is needed to capture not only specific conditions but also other aspects like integrating, prioritizing, and personalizing care. **Key words:** *financial incentives*, *pay-for-performance*, *primary healthcare*, *quality improvement*

THERE is widespread variability in the quality of care in all healthcare systems. The problem of inappropriate variation in the process and outcomes of healthcare has been described since the landmark article published by Wennberg in *Science*, 35 years ago (Wennberg & Gittelsohn, 1973). Interventions to improve professional performance are complex, and any interpretation of results requires sorting out the differences among professionals, the interventions studied, the targeted behaviors, and the study designs. The practice of paying physicians for performance (P4P) on quality targets has spread rapidly and internationally and has become one of the most prominent policy initiatives aimed at improving the quality of healthcare (Epstein, 2006). Therefore, P4P has become increasingly common as a reimbursement scheme for physicians with the underlying goal of improving health outcomes while reducing undue variation and enhancing appropriateness of service delivery and patient safety. The change in the professional context with the emergence of evidence-based medicine and information technology systems in the decade of 1990 made physicians more accountable for their clinical decisions. In some way, P4P schemes appear necessary because existing payment mechanisms do not reward providers for higher quality as do prices in most other markets (Robinson, 2001).

Author Affiliations: *Division of Family and Community Medicine, Hospital Italiano de Buenos Aires (Drs Adolfo Rubinstein, Fernando Rubinstein, Botargues, Barani, and Kopitowski), and Institute of Clinical Effectiveness and Health Policy (Drs Adolfo Rubinstein and Fernando Rubinstein), Buenos Aires, Argentina.*

The authors appreciate the assistance of Alejandro Lopez Osornio and Sergio Montenegro with the retrieval of data from the Hospital Italiano Health Information dataset. This research was partially funded by a scholarship to Mariela Barani from the National Commission of Health, Science and Technology (SACyT), Argentine Ministry of Health.

Corresponding Author: *Adolfo Rubinstein, MD, MSc, PhD, Division of Family and Community Medicine, Hospital Italiano de Buenos Aires, Gascón 450, Buenos Aires, C1181ACH, Argentina.*

However, it should be noted that although P4P programs create explicit incentives to reward improved performance, the underlying payment system exerts its own set of mostly implicit incentives. For example, fee-for-service payment creates an incentive to increase utilization of services while capitation payment may involve incentives to limit them. Therefore, if physicians are capitated for their services, the incentives to increase some practices would need to be larger than in fee-for-service system, in which providers already receive basic fees for their associated visits and procedures (Dudley & Rosenthal, 2006). One way to induce providers to improve healthcare quality would be to make their payments dependent to a certain extent on a set of indicators of process and health outcomes that acknowledges the provider effort to deliver high-quality care. Despite the growing enthusiasm for P4P mechanisms, there remains some dissent. Skeptics doubt the effectiveness of this approach in stimulating quality improvement and point to possible unintended consequences including redirecting physician time and attention away from their main purpose of caring for patients (Berwick, 1995; Shen, 2003). This behavior, which is generally called *gaming*, takes place when physicians find the way to maximize measured results without actually accomplishing the desired goal through risk selection practices or diversion of other dimension of care (ie, patient-centered aspects of care).

Process-of-care measures are generally more sensitive to quality improvement frameworks than are measures of outcomes, in part because a poor outcome does not necessarily mean poor quality (although the reverse might be true). Therefore, one way to change behavior may be to base the incentive on the combination of a process-of-care measure (ie, documentation of blood pressure [BP] control) and the outcome of interest (ie, BP control). This approach may avoid the pitfalls of process-of-care measures alone that might encourage gaming as well as the disadvantage of founding incentives solely on outcomes that may be sometimes beyond the control of the provider.

There are several empirical studies in the United States on the effects of paying for quality in healthcare with mixed and controversial results (Amundson et al., 2003; Fairbrother et al., 1999; Hillman et al., 1998, 1999; Kouides et al., 1998; Lindenauer et al., 2007; Rosenthal et al., 2005, 2006; Rosenthal & Frank, 2006a; Roski et al., 2003; Schneider et al., 2001). A recent systematic review found that 5 of 6 studies evaluating financial incentives at the physician level showed partial or positive effects (Petersen et al., 2006). The British recently started one of the most impressive and boldest experiences in size and scope of P4P (Roland, 2004). In 2004, the UK National Health Services (NHS) committed additional funding ($1.8 billion) over a period of 3 years for a new P4P Quality of Outcomes Framework (QOF) program for family physicians (FPs). Under this program, intended to increase income of FPs up to 25%, responsibility moved from the individual physician to the practice, typically 3 to 6 FPs.

Despite the widespread expansion of P4P in developed countries, little is known about quality improvement strategies that include P4P in Latin America except for some very limited experiences (Cherchiglia et al., 1998; Escrivão & Koyama, 2007). Nevertheless, neither of these reports aimed to evaluate clinical performance of primary care providers.

In 1994, the Hospital Italiano Health Plan changed the reimbursement scheme to its FPs from a fixed salary to a capitation payment for a defined population of patients under their care. In 2004, our division implemented an annual bonus to individual physicians according to their accomplishment of some clinical and organizational targets but the impact was restricted to only a few physicians (Rubinstein, 2006). In 2005, we decided to create a Program of Quality Improvement (PQI) in healthcare shifting from an individual to a group program of quality incentives based on our 5 traditional primary care groups (clinical care and teaching units, UDA in its Spanish acronym). Since August 2005, a new scheme of P4P, inspired in the British experiment, started to operate as a complement

of the capitation reimbursement scheme, representing a potential increase of 8% of physicians' annual payment.

The objective of this article is to report the development of this multimodal intervention in our group and the first results after 2 years of this innovative plan to improve quality measures on FPs, based on P4P, teamwork, continuous education, and audit and feedback.

METHODS

Population and Setting

The Hospital Italiano of Buenos Aires is a large university hospital. Primary care services are offered by FPs, general internists, and pediatricians, who are in charge of a defined panel of patients, providing first-contact care in a capitated setting in 18 primary care centers distributed across the metropolitan area of Buenos Aires. The Division of Family and Community Medicine takes care of approximately 80 000 individuals. Each physician, who is in charge of a defined panel of approximately 1200 patients, belongs to 1 of 5 different primary care groups (UDAs), which are composed of 10 to 15 physicians responsible for the care of a population of 10 000 to 15 000 patients. The intended goal of analyzing quality by UDA was to limit the individual variability, to increase the pool of patients with less prevalent conditions, and to encourage group rather than individual commitment.

Definition of the quality improvement framework

Under this new scheme of P4P, each UDA can earn up to 1000 points if a complex set of indicators pertaining to 5 different dimensions is met. The domains that were agreed upon for assessment were clinical effectiveness, access and coordination of care for sentinel conditions, comprehensive practices, quality of documentation in the electronic medical record, and patient satisfaction with experiences of care.

At the end of the year, the PQI assigns to each UDA the points earned, after evaluating the degree of accomplishment on the different targets. No exclusions of patients are allowed from the numerator or the denominator of any individual indicator. Points are then translated into Argentine pesos (ARS $), which are paid annually to each UDA.

Assessment and weighing process of the domains

A first survey containing the selected dimensions and indicators was sent to all physicians along with some references of P4P experiences, including the QOF of the British. FPs were asked to weigh each dimension considering that all domains should add up to 1000 points and were asked to weigh each indicator within the domain in percentage points where all the indicators should add up to 100%.

A round of sessions was convened to show physicians the distribution of values, allowing discussion to elicit consensus on the domains and their specific indicators and targets by using modified Delphi techniques.

Then, a second survey was sent requesting to reassign values to the domains and indicators in the same way as described above. After excluding values below the 10th or above the 90th percentile, final weights for domains and indicators were agreed and established on the basis of the remaining values assigned by the group.

Final domains, indicators of performance, and targets

A description of scores (points) of the final domains, subdomains, and indicators of performance can be seen in Table 1. Domains were categorized as follows.

Clinical effectiveness

Four conditions with 13 indicators were included in this domain, including cancer-screening practices (mammograms every 2 years in women older than 49 years, Papanicolaou tests every 3 years in women 18 to 65 years of age, annual fecal occult blood with guaiac in individuals older than 50 years); high BP documentation and control; type 2 diabetes mellitus documentation and control;

Table 1. Scoring of indicators from all domains and subdomains assessed

Domain	Subdomain	No. of indicators in each item	Points (minimum-maximum)
Clinical effectiveness	Screening practices	3	0–111
	Hypertension	2	0–95
	Diabetes	4	0–111
	High-risk cardiovascular disease	2	0–95
Comprehensive practices	Psychosocial interventions	3	0–58
	Management of particular women health problems	2	0–59
	Well-child visits	2	0–48
	Coordination of care	4	0–30
	Joint infiltrations	1	0–25
	Smoking cessation interventions	1	0–30
Continuous medical education activities		4	0–147
Documentation in electronic medical records		3	0–176

and treatment of patients with cardiovascular disease.

Comprehensive practices

Four procedures were included: well-child care visits, Papanicolaou tests, psychosocial interventions (ie, detection and treatment of depression, domestic violence, family crises, etc), and joint infiltrations performed by FPs.

Access and coordination of care for sentinel conditions

We intended to measure the referral rate to specialists in certain high-prevalence conditions that are usually managed by general practitioners (GPs) as a way to assess coordination of care. Sentinel conditions taken into account were hypothyroidism, irritable bowel syndrome, and headache. Physicians were not aware of which sentinel conditions had been selected.

Quality of documentation in the electronic medical record

Eight clinical records per physician were reviewed every year. A special instrument to audit medical records was designed.

Satisfaction with experience of care

We used an instrument adapted from the "CAHPS Clinician & Group Survey 4.0" that evaluated access, coordination of care, communication skills, managing decisions about patient care, and an overall score of the GP. Trained interviewers administered a telephone survey to a random sample of 30 patients per GP.

In addition to evaluating its face validity, we included an internal question for criterion validation purposes: "Would you recommend your physician to a person you care for?"

The response to this question was highly consistent with the scores of the brief satisfaction questionnaire.

Development of targets and scoring for each dimension

Standards and targets are developed for each of the 5 domains and UDAs earn points on the basis of the achievement of the different targets of the indicators measured. For instance, one of the targets in the domain of clinical effectiveness was the proportion of diabetic patients in each UDA who had

Table 2. Examples of targets for indicators in some domains

Colorectal cancer screening with faecal occult blood		**BP control in hypertensive patients**	
% of patients with the practice done	**Points**	**% of patients with the practice done**	**Points**
≥30%	50	≥56%	60
24%–29%	40	46%–55%	45
17%–23%	30	35%–45%	30
Hemoglobin A_{1c} control in diabetics		**Paps performed by FPs**	
% of patients with Hb A_{1c} <8%	**Points**	**% of FPs performing Paps**	**Points**
≥60%	30	≥60%	40
51%–60%	20	30%–59%	30
40%–50%	10	10%–29%	20

Abbreviations: BP, blood pressure; FP, family physician; Hb A_{1c}, hemoglobin A_{1c}; Pap, Papanicolaou test.

a glycated hemoglobin below 8%. In this case, if less than 40% of patients are under the target, the UDA does not earn points, whereas if more than 60% of patients are on target, the UDA may earn 30 points. Another target in the domain of comprehensive practices was the proportion of FPs in each UDA who perform at least 15 Papanicolaou tests per year. In this case, if less than 10% of physicians meet the requirements, the UDA does not earn points. On the other hand, if more than 60% of physicians perform more than 15 Papanicolaou tests a year, the UDA earns the maximum, 27 points (Table 2).

Information system

The data required for the indicators of clinical effectiveness are extracted periodically from the Hospital Information System. Access to this tool is regulated by institutional-level privacy policies. The PQI obtains information on the performance of the clinical indicators for the whole group of physicians, each UDA or each physician within the UDA.

Data analysis

Clinical indicators were obtained from the hospital information system predefined consults to the database and the other indicators from the collection of specific primary data (survey to patients) or from the auditing of electronic medical records.

As a preliminary evaluation of the first experience with the PQI in our setting, we decided to compare the indicators in each reported domain for 2007 with the baseline measurements obtained in 2005. Depending on the indicator measured, results are reported as the proportion of patients or the proportion of physicians achieving the different targets. The results are evaluated as a before-after study, using χ^2 and Z tests for comparison of proportions.

Since we did not find significant differences in the characteristics of each UDA, we compared both periods pooling the information obtained for each indicator per year.

RESULTS

Baseline characteristics of patients and FPs and satisfaction with care for each UDA are shown in Table 3. We assessed patients' satisfaction with care provided by their FP once, during the first 3 months of this program. A random sample of 1784 individuals (approximately 30 patients per GP) was surveyed by

Table 3. Baseline characteristics of patients and general practitioners of each UDA

Patients	UDA 1	UDA 2	UDA 3	UDA 4	UDA 5
Number (*n*) (% total)	8224 (18%)	9930 (22%)	9697 (21%)	9339 (20%)	8510 (19%)
Age <40 y	2127 (25.8%)	2794 (28.2%)	2646 (27.3%)	2674 (28.6%)	2732 (32.1%)
Age 40–65 y	4641 (56.5%)	5728 (57.6%)	5566 (57.4%)	5337 (57.2%)	4802 (56.4%)
Age >65 y	1456 (17.7%)	1408 (14.2%)	1485 (15.3%)	1328 (14.2%)	976 (11.5%)
Men	2860 (34.8%)	4204 (42.3%)	4241 (43.7%)	4132 (44.2%)	3814 (44.8%)
Women	5364 (65.2%)	5726 (57.7%)	5456 (56.3%)	5207 (55.8%)	4696 (55.2%)
Diabetic patients	434 (5.3%)	490 (4.9%)	450 (4.6%)	440 (4.7%)	422 (4.9%)
Hypertensive patients	2410 (29.3%)	2527 (25.4%)	2640 (27.2%)	2484 (26.6%)	2434 (28.6%)
Patients with cardiovascular disease[a]	600 (7.3%)	665 (6.7%)	727 (7.5%)	747 (8.0%)	570 (6.7%)
Obese patients body mass index >30	1406 (17.1%)	1559 (15.7%)	1813 (18.7%)	1606 (17.2%)	1667 (19.6%)
Satisfaction with care (out of 100)	91.3	90.4	91.1	92.6	90.1
Physicians					
5–15 y since graduation	45%	41%	51%	44%	49%
>15 y since graduation	55%	59%	49%	56%	51%

[a]Total patients (45 700). Patients with cardiovascular disease are subjects with a previous cardiovascular event such as acute myocardial infarction, stroke, abdominal aortic aneurism, cardiovascular revascularization, or percutaneous transluminal coronary angioplasty). Patient characteristics and physician distribution in the UDAs were balanced.

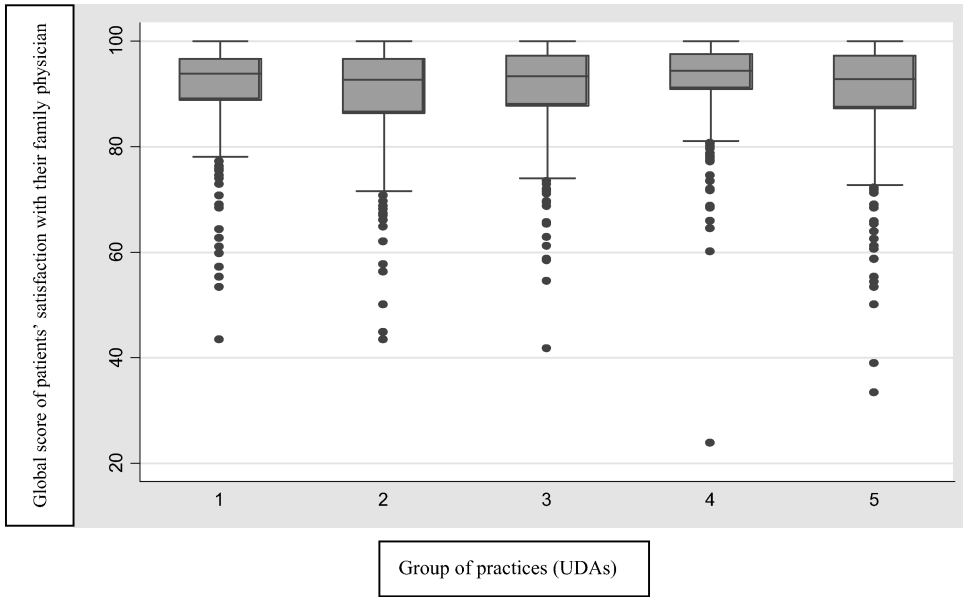

Figure 1. Global score by group of practices (UDAs).

phone. In addition to the responses to the different questions, 95% of them responded that they would definitely recommend his/her physician. The global score of their FPs in a scale from 1 to 10 was 8.86 (SD 1.2). A "ceiling effect" was observed in all the individual items of the survey and in the global score also. The global score was 91 (95% CI, 88–97) of 100 points score possible. The patients who would recommend their GPs had a significantly different global score than did those who would not, 92.1 versus 71.4; $P < .001$ (Fig 1).

By the end of year 2007, the UDAs achieved 60% to 80% of all possible points for that year. In terms of financial incentive, for each member of the UDA who obtained the best score, the reward in pesos would represent the equivalent of 4% of the mean year income for direct patient care in our division. Since there was no heterogeneity among the 5 groups, we are reporting the pooled results.

After the first 2 years, a significant improvement in all indicators related to clinical effectiveness (screening and control of prevalent or important conditions in terms of burden of disease) was achieved (Table 4).

When evaluating comprehensive practices, management of psychosocial problems and participation in continuous medical education activities showed a significant improvement. The other indicators, except for well-child visits, showed a nonsignificant trend toward improvement in performance, although the power for statistical tests was limited (Table 5).

A nonsignificant trend toward improvement was also observed when evaluating quality of documentation in the e-records (57% vs 62%).

DISCUSSION

After the first 2 years, FPs performed very well as a response of the implementation of this program of P4P, attaining 71.2% of the available points. The set of indicators evaluated by the program was sensitive enough to show some variability in the performance across different UDAs (UDA No. 3 = 79.8% vs UDA No. 5 = 62.5%), an important fact considering that the composition of the groups was similar.

Table 4. Clinical effectiveness indicators measured in 2005 and 2007

Clinical effectiveness indicators	2005	2007	Diff	*P*
Colorectal cancer screening	7%	29%	22%	.01
Breast cancer screening	59%	64%	5%	.01
Cervical cancer screening	46%	63%	17%	.01
Measurement of HbA1c in diabetic patients during the previous 12 mo	64%	91%	27%	.01
HbA1c <8% in diabetic patients	64%	79%	14%	.01
Measurement of LDL-C in diabetic patients during the previous 12 mo	77%	95%	18%	.01
LDL-C <130 mg/% in diabetic patients	42%	74%	32%	.01
Measurement of Ldl-c in patients with CVD during the previous 12 mo	82%	93%	11%	.01
LDL-C <100 mg/% in patients with CVD	25%	50%	25%	.01
BP record in hypertensive patients	67%	90%	32%	.01
BP <140/90 mm Hg in hypertensive patients	35%	64%	29%	.01

Abbreviations: BP, blood pressure; CVD, cardiovascular disease; Hb A_{1c}, glycosilated hemoglobin A_{1c}; LDL-C, low-density lipoprotein cholesterol.

Attainment of quality indicators across the different dimensions showed some discrepancies. Performance on quality indicators of clinical effectiveness showed improvement in both cancer-screening practices and documentation and control of chronic conditions, but performance on achievement of comprehensive practices showed contradictory results. While targets for psychosocial interventions, including the detection and management of depression, showed a significant 32% of improvement, well-child visits targets showed a decrease of 6%. A small but significant improvement was observed in the documentation of relevant data in the electronic medical record as well as in the coordination of care of FP.

We evaluated satisfaction with care as one important outcome indicator of the program, and the results were very satisfactory across all groups. The ceiling effect observed in the scores, consistent with previous evaluations of satisfaction, might be explained by the fact that patients can choose their physicians freely and may change without any restraint just by requesting an appointment with a different one. On the other hand, some patients may grade higher scores for courtesy or for concern. As a result of this initial evaluation, we decided not to measure patients'

Table 5. Results for comprehensive practices

Comprehensive practices	2005	2007	Difference	*P*
Psychosocial interventions	15%	48%	32%	<.01
Joint infiltrations	18.2%	28.6%	10.4%	NS
Pap tests performed by GPs	36%	38%	2%	NS
Well-child visits	39.7%	33.3%	−6.4%	NS
% of GPs involved in continuous medical education activities	27%	47.6%	20.6%	<.05

Abbreviations: GPs, general practitioners; NS, not significant; Pap, Papanicolaou test.

satisfaction again and redistributed the points assigned to this item among the other domains assessed.

Two studies of Hilman et al. on the use of performance feedback and financial bonuses based on performance on quality-of-care measures in a Medicaid health maintenance organization (Hillman et al., 1998) and another study in the same Medicaid health maintenance organization focused on pediatric immunization rates (Hillman et al., 1999) showed no effect of the incentives on measured outcomes. A study of Kouides et al. (1998) examined the effect of targeted financial incentives for increasing influenza immunization rates, showing that despite the small size of the incentive (<1% of physician revenues), this resulted in a statistically significant 7% improvement in immunization rates. Fairbrother et al. (1999) studied the impact of financial incentives paired with performance feedback on childhood immunization rates in a low-income urban population. Neither feedback alone nor the enhanced fees improved the likelihood of childhood immunization in the study population. Roski et al. (2003) studied the impact of financial incentives to support adherence to smoking cessation guidelines. Although financial incentives improved both documentation and advice to quit compared with the control group, there was no significant impact on smoking cessation rates. Rosenthal et al. (2005) evaluated physician group quality from a large health plan with P4P as compared with other without P4P, showing only slight improvement of one indicator in the first group. Interestingly, few groups reached a majority of targets, consistent with the low correlation in performance across clinical areas that has been observed in other studies (Schneider et al., 2001). Pearson et al. (2008) recently evaluated the impact on quality of all P4P programs introduced into physician group contracts during 2001–2003 by the 5 major commercial health plans operating in Massachusetts. Overall, P4P contracts were not associated with greater improvement in quality compared to a rising secular trend. Because the US healthcare system is characterized by a large number of overlapping contracts between payers (ie, health plans and government programs) and providers, financial incentives introduced by any one payer must be a relatively large percentage of total reimbursement by that payer to justify any quality improvement effort with substantial fixed costs (Pearson et al., 2008).

As mentioned, our framework was adapted from the QOF of the British with respect to the attainment of points in relation to the accomplishment of desired goals. The British program of P4P was intended to increase FPs' income by up to 25% depending on their performance on 146 quality indicators related to clinical care for 10 chronic diseases, practice organization, and patient satisfaction with experiences with care. After 1 year, FPs performed extremely well. For clinical indicators, targets were met for 83% of eligible patients and practices earned nearly 97% of the possible points available. This resulted in higher payments than those predicted by the NHS, which had anticipated only 75%, contributing to a NHS deficit (Doran et al., 2006).

Our results, albeit the important improvement observed in some areas, were not as extreme as those obtained by British FPs. However, some factors might help explain these differences. First, as opposed to the British experiment in which some improvements could have been due to an existing secular trend not related to the incentives, we had baseline results which were used as benchmarks to set the indicator targets with the desired improvement so as to avoid setting "easy to achieve" results. Second, exceptions for patients to be discretionally excluded from the measurement because of different reasons were not allowed in our program. Although the exception rate reported by British FPs was low and consequently its effect was small, the possibility of some "gaming" cannot be excluded.

Economic theory suggests that the reward should be commensurate with the incremental cost of the quality improvement required, including the lost revenue that the provider could generate in other activities, such as seeing more patients (Rosenthal & Adams Dudley, 2007). In this regard, it could be

argued that some differences in the results of P4P programs in the United Kingdom and the United States might be related to the magnitude of the financial incentives at stake. In effect, while British QOF implied up to 25% of income increase for FPs, financial incentives to improve quality in the United States were in the order of 1% to 5% of the gross income of providers, for reported failures in the latter could be due to unilateral, small-scale bonus arrangements that were insufficient to change behaviors on the part of physicians and hospitals. Although the magnitude of incentives for FPs in our program was very similar to that of the United States, our results were much better, approaching those of the United Kingdom, which, in turn, has much more powerful incentives.

We believe that financial incentives for quality in our program were only one component of a broader strategy that included audit and feedback, education, teamwork, and a peer pressure for improvement, especially competition among the different UDAs. Henceforth, a key feature of the healthcare setting that is likely to affect the impact of payment incentives is the role that professionalism and trust play in physician-patient interactions. Physician behavior has been shown to be influenced by professional norms. These may alter the impact of financial incentives related to quality (Rosenthal & Frank, 2006b). Although it is very difficult to assess the relative contribution of each component to the whole result, we think that the financial incentive just triggered complex behaviors on the part of our FPs that led outcomes toward the desired goals.

One of the main limitations of our framework (as well as the British one) is that it evaluates performance of FPs through some domain-specific measures mostly devised by specialists to assess disease control of particular conditions. Nevertheless, most patients presenting in primary care have multiple, interacting problems where comorbidity is the rule rather than the exception. Moreover, most scientific evidence on which disease-specific quality measures are based explicitly exclude people with comorbid conditions (Starfield, 2003). Therefore, primary healthcare (PHC) quality measures are oftentimes embedded within the specialist-fragmented paradigm of physicians treating multiple conditions not related to one another. This paradigm has been reinforced lately by chronic disease-management strategies promoted by managed care organizations where each condition (e.g., type 2 diabetes mellitus, congestive heart failure, or asthma) is treated separately by disease specialists who provide integrated care and follow-up through a vertical approach that usually overlooks PHC as the horizontal component. Recent evidence points out to the contradictory findings as well as lack of concordance of intermediate outcome measures and patient outcomes in which the reduction of a particular measure such as BP or glycated hemoglobin not necessarily leads to a better outcome and in fact may be associated with a worse patient outcome (Nissen & Wolski, 2008), highlighting the importance of taking into account not only the reduction of the risk factor or measure but also the strategy used to accomplish this result (Krumholz & Lee, 2008).

Meanwhile, the accomplishment of some disease control standards might conflict with each other in the achievement of a better quality of life for her/his patient. If we are to measure the performance of FPs, instead of looking only at disease-specific indicators, we should concentrate our efforts in developing and validating indicators representing essential attributes of PHC practice associated with better patient outcomes such as first contact, access, comprehensiveness, longitudinality, coordination of care, and family and community approach (Starfield, 1998). This requires integrating dimensions not easily captured by the traditional quantitative approach and then turned into a measurable quantity. The everyday job of a physician, particularly in primary care, entails much more than a set of quantitative measures to evaluate important aspects of practice. The challenge remains to develop different indicators using mixed methods, integrating qualitative techniques to evaluate physician-patient interaction with other indicators that can reflect

coordination of care, problem resolution, and comprehensive approach.

The results reported should be considered as preliminary since 2 years is too short a period to test whether P4P in the context of the other factors included in the intervention has a sustainable effect throughout time. In addition, despite the fact that some of the results are remarkable, we cannot rule out a secular trend that might explain quality improvements regardless of the intervention implemented.

In conclusion, despite the significant improvement in the quality of care obtained in most of the selected indicators, especially those measuring clinical effectiveness, we should be careful not to address only specific condition targets, provision of selective preventive services, or special practices but also find out the appropriate indicators that measure higher-order functions of primary care providers such as first contact, comprehensive and personalized care across multiple illnesses, and domains.

REFERENCES

Amundson, G., Solberg, L. I., Reed, M., Martini, E. M., & Carlson, R. (2003). Paying for quality improvement: Compliance with tobacco cessation guidelines. *Joint Commission Journal on Quality and Safety, 29*(2), 59–65.

Berwick, D. M. (1995). The toxicity of pay for performance. *Quality Management in Health Care, 4*(1), 27–33.

Cherchiglia, M. L., Girardi, S. N., de Castro Vieira, R., Bibiani, R., de Aguiar, M., Mendes, P, et al. (1998). Remuneración y productividad: el caso de la Fundación Hospitalaria del Estado de Minas Gerais, Brasil, 1992–1995. *Rev Panam Salud Publica/Pan American Journal of Public Health, 4*, 112–120.

Doran, T., Fullwood, C., Gravelle, H., Reeves, D., Kontopantelis, E., Hiroeh, U., et al. (2006). Pay-for-performance programs in family practices in the United Kingdom. *The New England Journal of Medicine, 355*, 375–384.

Dudley, R. A., & Rosenthal, M. B. (2006). *Pay for performance: A decision guide for purchasers*. Rockville, MD: Agency for Healthcare Research and Quality. AHRQ Pub. 06-0047.

Epstein, A. M. (2006). Paying for performance in the United States and abroad. *The New England Journal of Medicine, 355*, 406–408.

Escrivão, A., & Koyama, M. F. (2007). O relacionamento entre hospitais e operadoras de planos de saúde no âmbito do Programa de Qualificação da Saúde Suplementar da ANS. *Ciência & Saúde Coletiva, 12*, 903–914.

Fairbrother, G., Hanson, K. L., Friedman, S., & Butts, G. C. (1999). The impact of physician bonuses, enhanced fees, and feedback on childhood immunization coverage rates. *American Journal of Public Health, 89*(2), 171–175.

Hillman, A. L., Ripley, K., Goldfarb, N., Nuamah, I., Weiner, J., & Lusk E. (1998). Physician financial incentives and feedback: Failure to increase cancer screening in Medicaid managed care. *American Journal of Public Health, 88*(11), 1699–1701.

Hillman, A. L., Ripley, K., Goldfarb, N., Weiner, J., Nuamah, I., & Lusk, E. (1999). The use of physician financial incentives and feedback to improve pediatric preventive care in Medicaid managed care. *Pediatrics, 104*(4), 931–935.

Kouides, R. W., Bennett, N. M., Lewis, B., & Cappuccio, J. D. (1998). Performance-based physician reimbursement and influenza immunization rates in the elderly. The Primary-Care Physicians of Monroe County. *American Journal of Preventive Medicine, 14*(2), 89–95.

Krumholz, H. M., & Lee, T H. (2008). Redefining quality: Implications of recent clinical trials. *The New England Journal of Medicine, 358*, 2357–2359.

Lindenauer, P. K., Remus, D., Roman, S., Rothberg, M. B., Benjamin, E. M., Ma, A., et al. (2007). Public reporting and pay for performance in hospital quality improvement. *The New England Journal of Medicine, 356*, 486–496.

Nissen, S. E., & Wolski, K. (2008). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. The Action to Control Cardiovascular Risk in Diabetes Study Group: Effects of intensive glucose lowering in type 2 diabetes. *The New England Journal of Medicine, 358*, 2545–2559.

Pearson, S. D., Schneider, E. C., Kieinman, K. P., Coitin, K. L., & Singer, J. A. (2008). The impact of pay-for-performance on health care quality in Massachusetts, 2001–2003. *Health Affairs (Millwood), 27*, 1167–1176.

Petersen, L. A., Woodard, L. D., Urech, T., Daw, C., & Sookanan, S. (2006) Does pay-for-performance improve the quality of health care? *Annals of Internal Medicine, 145*, 265–272.

Robinson, J. C. (2001). Theory and practice in the design of physician payment incentives. *The Milbank Quarterly, 79*(2), 149–177.

Roland, M. (2004). Linking physicians' pay to the quality of care—A major experiment in the United Kingdom. *The New England Journal of Medicine, 351*, 1448–1454.

Rosenthal, M. B., & Adams Dudley, R. A. (2007). Pay-for-performance: Will the latest payment trend improve care? *JAMA, 297*, 740–744.

Rosenthal, M. B., & Frank, R. G. (2006a). What is the empirical basis for paying for quality in health care? *Medical Care Research and Review, 63*, 135–157.

Rosenthal, M. B., & Frank, R. G. (2006b). What is the empirical basis for paying for quality in health care? *Medical Care Research and Review, 63*, 135–157.

Rosenthal, M. B, Frank, R. G., Li, Z., & Epstein, A. M. (2005). Early experience with pay-for-performance: From concept to practice. *JAMA, 294*, 1788–1793.

Rosenthal, M. B., Landon, B. E., Normand, S. L., Frank, R. G., & Epstein, A. M. (2006). Pay for performance in commercial HMOs. *The New England Journal of Medicine, 355*, 1895–1902.

Roski, J., Jeddeloh, R., An, L., Lando, H., Hannan, P., Hall, C., et al. (2003). The impact of financial incentives and a patient registry on preventive care quality: Increasing provider adherence to evidence-based smoking cessation practice guidelines. *Preventive Medicine, 36*(3), 291–299.

Rubinstein, A. (2006). Atención Primaria y Servicios de Salud in "Medicina Familiar y Práctica Ambulatoria." In *Medicina Panamericana* (2nd ed., pp. 40–55). Ed. Med. Panamericana: Buenos Aires Médica Panamericana.

Schneider, E. C., Zaslavsky, A. M., Landon, B. E., Lied, T. R., Sheingold, S., & Cleary, P. D. (2001). National quality monitoring of Medicare health plans: The relationship between enrollees' reports and the quality of clinical care. *Medical Care, 39*, 1313–1325.

Shen, Y. (2003). Selection incentives in a performance-based contracting system. *Health Services Research, 38*, 535–552.

Starfield, B. (1998). *Primary care: Balancing health needs, services and technology* (p. 448). New York: Oxford University Press.

Starfield, B. (2003). Primary and specialty care interfaces: The imperative of disease continuity. *British Journal of General Practice, 53*, 723–729.

Wennberg, J. E., & Gittelsohn, A. (1973). Variations in medical care among small areas. *Science, 246*(4), 100–111.