

Does pay-for-performance influence the quality of care?

Michael Schatz

Department of Allergy, Kaiser Permanente, San Diego, California, USA

Correspondence to Michael Schatz, MD, MS, Chief, Department of Allergy, Kaiser Permanente Medical Center, 7060 Clairemont Mesa Blvd, San Diego, CA 92111, USA

Tel: +1 858 573 5397; fax: +1 858 573 5343; e-mail: michael.x.schatz@kp.org

Current Opinion in Allergy and Clinical Immunology 2008, 8:213–221

Purpose of review

To summarize studies of pay-for-performance programmes designed to address clinical quality of care markers in ambulatory settings at the level of the physician group or individual physician.

Recent findings

Seven randomized controlled trials and 15 nonrandomized studies were reviewed. Less than half of the randomized controlled trials showed positive results, whereas all but one of the nonrandomized studies showed positive or mixed results. Characteristics of the quality measures, incentives, providers, patients, and concurrent interventions probably influenced the results. Study methodology problems such as small sample sizes, selection bias, and inadequate control for confounders were common. Asthma quality of care markers were included in three nonrandomized studies, two of which showed positive results and one of which was negative.

Summary

The data reviewed in this article suggest that pay-for-performance programmes can improve markers of quality, though not always. Even when studies suggest positive effects, the designs often do not permit assurance that the effects are due to the incentives as compared with other factors. More and better-designed studies are needed to determine the actual effectiveness of incentives themselves isolated from other factors, circumstances that promote effectiveness, effectiveness relative to other strategies, and cost-effectiveness.

Keywords

pay-for-performance, quality bonuses, quality improvement, quality incentives

Curr Opin Allergy Clin Immunol 8:213–221
© 2008 Wolters Kluwer Health | Lippincott Williams & Wilkins
1528-4050

Introduction

Pay-for-performance programmes are now firmly entrenched in the payment systems of American public and private insurers. A recent survey of 252 Health Maintenance Organizations (HMOs) drawn from 41 metropolitan areas in the United States reported that more than half of the HMOs, representing more than 80% of enrollees, use pay-for-performance in their provider contracts [1]. Evaluation of asthma medication was included in 71% of programmes. Most physician group leaders (89%) in a Massachusetts survey [2] reported pay-for-performance incentives in at least one commercial health plan contract. Incentives were tied to healthcare effectiveness data and information set (HEDIS) quality measures (89%), utilization measures (66%), use of information technology (52%), and patient satisfaction (37%).

Although pay-for-performance programmes are common, the evidence base for their use is less clear. The purpose of this article is to review recent literature on the effectiveness of pay-for-performance programmes in improv-

ing the quality of care, especially as it is related to the field of Allergy Immunology. Thus, the focus of this article will be on studies of pay-for-performance programmes designed to address clinical quality of care markers in ambulatory settings at the level of the physician group or individual physician.

Search methods

A Medline search was performed for the years 2006 and 2007 using the search terms 'Pay for Performance Outcomes' and 'Pay for Performance Results'. Relevant articles from the search were supplemented by articles identified from the bibliographies of identified articles, including the comprehensive systematic review of Petersen *et al.* [3].

Study methodology

Before reviewing the specific studies, it is important to consider methodological issues that may affect the validity and interpretation of the existing data. These

issues include the measures being assessed, advantages and disadvantages of randomized controlled trials (RCTs), selection bias, confounding, generalizability, and cost-effectiveness studies.

Performance measures

The results of a pay-for-performance outcome study will only be as meaningful as the measures being evaluated. Measures should be chosen such that improvement leads directly and substantially to improved health [4]. As outcome measures may be influenced by many factors beyond the control of the physician, they are less suitable for pay-for-performance programmes. Process measures with a proven relationship to outcomes that are important to patients are ideal, but many currently used process measures have not been shown to have such a relationship to health outcomes [4,5]. In addition, mechanisms must be in place to ensure the accuracy of the reported measures [5].

The ease of accomplishing the measure goal should also be considered. One important factor in this regard is the burden adherence puts on the patient [6^{••}]. Goals related to processes or outcomes that require greater patient cooperation and effort will likely to be more difficult to accomplish than goals that require little patient effort.

Randomized clinical trials

The gold standard medical intervention study is the RCT. The major strength of the RCT is that it minimizes the influence of measured and unmeasured confounders. The RCT would be the ideal study design to assess a pay-for-performance intervention [4]. A several arm RCT design would be the optimal way to compare the effectiveness of payment incentives to alternative approaches, such as regulatory or accrediting programmes or public reporting [4]. Nevertheless, RCTs are complex and expensive to design and conduct [4], and it may be difficult to enroll participants in RCTs that could affect their practice patterns or income [7]. Moreover, RCTs reduce the influence of confounding variables only when there are many participants in each group of the trial [7]. For pay-for-performance programmes aimed at physician organizations, it could be difficult for researchers to enroll sufficient numbers of physician organizations to ensure that randomization really balanced the confounding factors. In addition, small sample sizes increase the possibility of type 2 errors affecting the statistical significance of the findings.

Selection bias

Selection bias can occur in studies whenever the identification of individual participants for inclusion in the study depends on factors related to the outcome. In this case, the relationship between the intervention and the outcome observed among those who participated in the

study may be different from that for individuals who would have been eligible to participate but were unwilling or not selected. For example, a number of early pay-for-performance programmes involved volunteer physician groups [4]. As volunteers typically differ from nonvolunteers in most settings, simple comparisons between the experience of pay-for-performance volunteers and nonvolunteers may not provide convincing evidence of effectiveness due to the potential influence of selection bias [4].

Confounding

Confounding occurs when factors associated with the intervention in a study are also independently related to the outcome. Although observational studies can control for measured confounders through statistical techniques, this frequently is not done. Moreover, observational studies cannot control for unmeasured confounders. Confounders in pay-for-performance studies fall into two main categories [4]. First, participants and nonparticipants may differ in specific ways that directly affect their capacity and motivation to improve. For physician groups, these factors may include size, single compared with multispecialty composition, implementation of guidelines, educational activities, and the extent to which medical information is available electronically [4,6^{••},7]. For individual physicians, factors that might affect improvement include age, gender, specialty, board certification, country of graduation, workload, baseline income, and number of patients with the condition to which the incentive applies [6^{••},8]. In addition, patient characteristics, such as age and socioeconomic features, may differ between provider groups and affect performance on quality measures [6^{••},7,8,9]. Second, the performance outcomes may be influenced by factors other than the incentive programme itself. This is particularly a problem in before–after study designs. Even in controlled before and after studies [4], participants and nonparticipants may differ in their exposure to other influences that may independently produce improvement on the quality outcomes being measured. These influences include regulatory agency programmes and public reporting of the measures that are separate from the incentive programmes but may influence the same outcomes.

Generalizability

Even an internally valid study may not be substantially generalizable beyond the very specific incentives, measures, and/or study population utilized. The magnitude of the incentive, the proportion of each provider's patients to whom the incentive applies, and the provider cost of implementing the programme may all be expected to influence provider motivation and thus programme effectiveness [7]. Some measures are undoubtedly more difficult to affect than others [4], so that demonstrated effects on one measure cannot be easily generalized to

effects on another measure, even if the same incentive and population were involved. Finally, an effect demonstrated even in a RCT conducted in one group of physicians does not guarantee a similar effect in other physicians who may differ in many relevant ways, including their overall payment structure (fee-for-service compared with capitation) [7].

Cost-effectiveness studies

Although the effectiveness of a specific intervention such as pay-for-performance is important, the cost of the intervention must also be considered. One must try to determine if the degree of improved quality is worth the expenditures needed to achieve it. In particular, when there are potentially alternate ways to accomplish the same goal, comparative cost-effectiveness studies are necessary to identify the methods that provide the most value [10]. No cost-effectiveness studies of pay-for-performance interventions in the outpatient setting at the physician group or individual physician level were identified, although one study reported marginal cost-effectiveness results (see below).

Studies on the effectiveness of pay-for-performance programmes

Pay-for-performance programmes outside the health-care sector have produced mixed results regarding effectiveness [10]. Monetary incentives at the hospital level have been reported to improve quality indicators [11,12]. Studies evaluating the effectiveness of pay-for-performance incentives at the physician group or individual physician level in the ambulatory setting can be categorized into RCTs and studies using alternative designs.

Randomized controlled trials

Seven RCTs were identified [13–19] (Table 1). All of these trials dealt with preventive services, with four dealing with immunizations. All studies tested the provision of bonuses ranging from \$50 per provider [13] to \$7500 per provider [18]. Results were negative in three studies [13,15,16], positive in three studies [14,17,18], and mixed in one study [19].

Problems in all the three negative studies [13,15,16] included small sample sizes per group and relatively small bonuses per physician. In addition, two of the negative studies [15,16] were in the Medicaid population, in which reimbursement is typically low, practices are more constrained, and patients are harder to reach and enlist in quality improvement programmes [10]. In addition, in both of the studies [15,16] conducted in the Medicaid-managed care setting, the authors suggested that the study protocol may not have allowed enough time for the development of an effective intervention.

In contrast, two of the positive studies [17,18] of pediatric immunizations provided much larger bonuses. The authors of these studies, however, point out that the apparent increased immunization rate may have been primarily the result of better documentation rather than an actual increase in immunizations given. The third positive study [14], which focused on influenza immunization rates, apparently achieved positive results with a relatively small bonus and a small cost per extra immunization. However, these results may not be generalizable to the general population of physicians because all study participants had previously volunteered for an influenza Medicare demonstration project.

The study with mixed results [19], which tested the effects of incentives on smoking processes and outcomes, also provided relatively large bonuses per physician (\$700–\$5000). In this study, however, positive results were obtained in the easier process task (tobacco use status identification), but not in the more difficult to accomplish process (providing advice to quit) or outcome (documented quitting rate) measures.

Studies using designs other than randomized controlled trials

Fifteen studies using nonrandomized designs were identified, including seven uncontrolled before–after studies [20–23,24•,25,26••] (Table 2), six studies [27–32] using a controlled before–after design (Table 3), and two studies [33,34] using cross-sectional survey methodology (Table 4). These studies evaluated the effect of pay-for-performance programmes on a wider spectrum of diseases and measures than in the RCTs. All but one of the studies showed positive or mixed results.

Three studies evaluated outcomes before and after new general practitioner contracts in the United Kingdom [20,25,26••] (Table 2). The first of these studies [20] evaluated outcomes after the 1990 contract that included targeted payments for higher immunization rates, and the latter two studies evaluated diabetes [25] or multiple measures [26••] after the \$3.2 billion pay-for-performance contract of 2004 that rewarded quality targets in managing 10 chronic diseases. Control groups were not possible in these studies. Although increasing trends were already documented before the contract for the evaluated measures in all the three studies, the rate of increase after the contract was greater for at least asthma and diabetes measures in two of the studies [25,26••]. The most comprehensive study [26••] compared results for measures for which there was a specific rewarded target with results for measures for the same diseases but for which there was not a specific financial reward provided. No significant differences in performance were found between these two categories of measures. Although this could be interpreted to mean that the incentives were themselves ineffective,

Table 1 Randomized controlled trials

Ref.	Disease/Measure	Setting/Sample	Incentive	Results	Comment
Grady <i>et al.</i> [13]	Referral for mammography	61 primary care practices (23 education only, 18 education + reminder cues, 20 education + reminder cues + feedback and incentive)	\$50.00 bonus/provider for 50% referral rate	Negative: Both cue groups demonstrated increased referral rate compared to education only, but cue + reward group was not significantly greater than cue group without reward Positive: Median improvement in incentive group significantly greater than that in control group; marginal cost-effectiveness reported to be \$3.02 per extra immunization	Very small monetary incentive; small sample size
Kouides <i>et al.</i> [14]	Influenza immunization rate	54 solo or group practices (27 incentive group, 27 control group)	Bonus: 10% above standard fee (\$8.00) for >70% rate; 20% above standard fee for rate >85%; mean bonus \$162/group	Negative: No significant difference between groups in mean increase in compliance scores (pap test, colorectal screening, mammography, breast exam) Positive: Bonus group improved significantly in documented up-to-date immunization status; no significant changes occurred in the other groups	Selection bias a potential problem – participants had to have volunteered to participate in previous Medicare demonstration project
Hillman <i>et al.</i> [15]	Cancer screening in women > age 50	52 primary care provider groups (26 incentive group, 26 control group)	Full and partial bonuses (20%, 10% of capitation); \$570 – \$1260/site	Negative: No significant differences between groups in the increase in total compliance scores (reflecting immunizations and other indicators) Positive: Bonus group improved significantly in documented up-to-date immunization status; no significant changes occurred in the other groups	Possibly not enough time to develop effective intervention and too small a bonus; Medicaid managed care
Hillman <i>et al.</i> [16]	Pediatric preventive care guidelines	49 primary care provider groups (19 feedback + incentive; 15 feedback only, 15 control group)	Full and partial bonuses (20%, 10% of sites total 6 month capitation for members ≤ age 6); \$772 – \$4682/site	Negative: No significant differences between groups in the increase in total compliance scores (reflecting immunizations and other indicators) Positive: Bonus group improved significantly in documented up-to-date immunization status; no significant changes occurred in the other groups	Possibly not enough time to develop effective intervention and too small a bonus; Medicaid managed care
Fairbrother <i>et al.</i> [17]	Pediatric immunizations	60 physicians (15 bonus, 15 enhanced fee, 15 feedback, 15 control groups)	Bonus: \$1000 – 20% improved, \$2500 – 40% improved, \$5000 – 80% up-to-date; enhanced fee: \$5 per vaccine given when due; \$15/visit at which >1 due and all given Bonus: \$1000 – 30% improved, \$2500 – 45% improved, \$5000 – 80% up-to-date, \$7500 – 90% up-to-date; enhanced fee: as in study above	Positive: Both bonus and enhanced fee group improved significantly in up-to-date immunization status versus controls Bonus group, but not other groups had significant decrease in missed opportunities to immunize (MOI) during sick visits; no significant change in any group for MOI during well visits	Good incentive; feedback given to both intervention groups; increases primarily due to better documentation
Fairbrother <i>et al.</i> [18]	Pediatric immunizations	57 physicians (24 bonus, 12 enhanced fee, 21 control group)	Bonus: \$1000 – 30% improved, \$2500 – 45% improved, \$5000 – 80% up-to-date, \$7500 – 90% up-to-date; enhanced fee: as in study above	Positive: Both bonus and enhanced fee group improved significantly in up-to-date immunization status versus controls Bonus group, but not other groups had significant decrease in missed opportunities to immunize (MOI) during sick visits; no significant change in any group for MOI during well visits	Good incentive; feedback given to both intervention groups; increases primarily due to better documentation
Roski <i>et al.</i> [19]	Smoking	37 primary care provider groups (13 incentive group, 9 incentive group + registry, 15 control group)	Bonus for hitting targets: \$5000 for sites with 1–7 providers; \$10 000 for sites with ≥ 8 providers	Mixed: Change in tobacco use status identification; incentive group significantly greater than other groups; no significant differences between groups in providing advice to quit or in the documented quitting rate	Small sample sizes; easier task affected more than harder tasks; why was incentive group more effective than incentive group + registry?

Table 2 Studies with uncontrolled before–after designs

Ref.	Disease/Measure	Setting/Sample	Incentive	Results	Comment
Ritchie <i>et al.</i> [20]	Childhood immunizations	95 GP practices (313 GPs) in Grampian region (UK)	1990 GP contract in UK, which included target payments for higher immunization rates	Positive: Significant increase in immunization rates after the contract	Continuation of increasing immunization rate documented for 4 years before the contract
Morrow <i>et al.</i> [21]	Rates of MMR immunizations and cholesterol screening	IPA-HMO in four NE US states; MMR: 418 offices; cholesterol: 271 offices; 10–20 patients/office	Increased reimbursement based on quality elements (no additional details provided)	Positive: Offices meeting standards for MMR vaccination and screening for cholesterol increased significantly over the 3-year course of the study	Financial incentive accompanied by feedback and peer review
Amundson <i>et al.</i> [22]	Smoking identification and advice to quit	Network-model health plan in Minnesota: 14 489 records from 20 medical groups	Bonus payment of \$6 650–\$43 750 per medical group	Positive: Significant increase in smoking identification and documentation of advice to quit	Performance feedback received; some public reporting also included
Greene <i>et al.</i> [23]	Acute sinusitis care patterns	500 internists, 200 GP, 200 pediatricians in a NE US communitywide IPA	5% decreased withhold for top 5% of performers and 5% increased withhold for bottom 5% of performers	Positive: Exceptions to recommended care, including antibiotic and radiology exceptions, decreased significantly in the year following the intervention compared to the baseline year	Intervention also included physician education and feedback
Mandel and Kotagal [24*]	Flu shot, controller medication, and providing action plans in asthmatic children	44 pediatric practices in Ohio	Up to 7% fee schedule increase	Positive: Perfect care (severity classified, written self-management plan, controller for persistent asthma) increased from 4 to 88%; flu shot increased from 22 to 62%	Practices were required to develop asthma registry and reliable data collection processes, for which there was some payment
Millett <i>et al.</i> [25]	Diabetes outcomes (HgbA1c, BP, cholesterol)	36 general practices in London (population of 243 519)	2004 GP contract in UK with 25% of income derived through quality targets in managing 10 chronic diseases, including diabetes	Positive: Significant increase in all three outcomes after contract; rate of improvement greater than previously reported trends in UK; ethnic disparities not improved	Adjusted for age, gender, deprivation (postcode), and practice
Campbell <i>et al.</i> [26**]	CHD (15 indicators), asthma (12 indicators), diabetes (21 indicators); 30 indicators for which financial incentive were provided versus 17 for which financial incentives were not given	42 primary care practices in UK with quality indicators before (1998, 2003) and after (2005) P4P contracts initiated; representative of patient demographics in UK, but solo practices underrepresented	2004 GP contract in UK with 25% of income derived through quality targets in managing 10 chronic diseases, including CHD, asthma, and diabetes	Mixed: Rate of improvement increased between 2003 and 2005 versus predicted improvement based on 1998–2003 for asthma and diabetes but not for CHD; rate of improvement 2003–2005 did not differ between indicators for which financial incentives were provided compared to those for which incentives were not provided	Other initiatives affecting these indicators had been initiated earlier and were ongoing, especially for CHD; no difference between indicators with versus without incentive may indicate lack of effectiveness of incentive or indirect effects of incentives for those diseases

BP, blood pressure; CHD, coronary heart disease; GP, general or family practitioner; HMO, Health Maintenance Organizations; IPA, Independent Practice Association; MMR, measles-mumps-rubella; NE, northeast; P4P, Pay-for-performance.

Table 3 Studies with controlled before–after designs

Ref.	Disease/Measure	Setting/Sample	Incentive	Results	Comment
Kouides <i>et al.</i> [27]	Influenza immunization rates in the elderly	53 primary care physicians admitting to one hospital in New York versus 82 admitting to other hospitals in New York	Bonus: 10% above standard fee (\$8.00) for >70% rate; 20% above for rate >85%	Positive: Significantly increased immunization rate in incentive versus control groups	Potential for selection bias since all physicians previously participated in influenza immunization Medicare demonstration project; adjusted for confounders, including characteristics of physicians, practices, and patients
Young <i>et al.</i> [28]	Diabetes (four process indicators: HgbA1c, urinalysis, LDL, eye exams)	Rochester, New York IPA: 334 physicians, mean 22 patients/physicians; compared to national and state trends for these indicators	Return of 5% withheld (estimated \$1500)	Mixed: Postintervention trends only significantly increased compared with preintervention trends for one measure (eye exams); national and state trends for eye exams flat; eye exam results lowest at baseline.	Incentive may not have been large enough; may need longer time frame to show an effect; no adjustment for potential confounders
Armour <i>et al.</i> [29]	Colorectal cancer screening	6749, 50-year-old patients of physicians in a SE U.S. managed care health plan who were 'eligible' for the bonus programme compared with patients of physicians ineligible for the bonus programme.	Year-end bonus (additional details not provided)	Positive: Significant increase in screening during the bonus year compared with the prebonus year, adjusting for various intervention group patient and physician characteristics; no increase in screening in patients of the comparison physicians	Potentially confounding differences between physicians (and their patients) who were 'eligible' versus ineligible for the bonus were not explored
Rosenthal <i>et al.</i> [30]	Cervical cancer screening, mammography, HgbA1c tests	134 California physician groups (Incentive) versus 33 Pacific Northwest Physician groups (Control)	Bonus: 5% of capitation payments (0.8% of overall revenue) for performance targets met or exceeded; \$3.4 million in bonus payments	Mixed: % improvement significantly increased only for cervical cancer screening	Groups with highest baseline performance improved the least but received the largest share of the bonus payments; no adjustment for potential confounders
Beaulieu and Horrigan [31]	Diabetes – 10 process and outcome measures	21 New York primary care physicians (476 patients) received incentives versus 600 patients managed by other physicians in health plan (Control)	PMPM bonus based on composite score (hitting target or 50% improvement); \$3000 – \$12 000 per physician	Positive: Significant improvement in 9/10 process and outcome measures	Incentive group also received guidelines, a chart assessment tool, a list of patients, and feedback; no adjustment for potential confounders
Levin-Scherz <i>et al.</i> [32]	Diabetes (adults), asthma (children) HEDIS measures	Patients of Partners Community Healthcare Inc (PCHI) provider network (with pay-for-performance contract) versus patients of non-PCHI physicians in Massachusetts	Withhold for HEDIS measures at the network level that translated to \$3000 – \$5000 per physician if network met its HEDIS target	Mixed: Three diabetes measures improved significantly more in PCHI versus others; PCHI improved more for the asthma measure (4.1%) than others (2.3%), but not statistically significant	Confounded by the fact that PCHI pay-for-performance contracts were accompanied by a disease registry database, outreach by nonphysician staff, and internal feedback

HEDIS, healthcare effectiveness data and information set; IPA, Independent Practice Association; LDL, low density lipoproteins; PMPM, per member per month; SE, southeastern.

Table 4 Studies with cross-sectional survey designs

Ref.	Disease/Measure	Setting/Sample	Incentive	Results	Comment
McMenamin <i>et al.</i> [33]	Smoking cessation interventions	1104 national physician organizations in US	Report of receipt of: financial incentives from HMOs to promote smoking cessation interventions, additional income for scoring well on quality measures, or better contracts for scoring well on quality measures	Positive: Receipt of financial incentives and additional income but not better contracts significantly increased the odds of having support for smoking cessation interventions	Adjusted for other organizational characteristics, including public recognition for quality measures, requirement to report HEDIS scores, and awareness of guidelines, all of which were independently associated with support for smoking cessation interventions
Pourat <i>et al.</i> [34]	Guideline concordant delivery of STD care (5 markers)	Primary care physicians contracted with Medicaid HMOs in California	Reimbursement for quality of care dimensions in contracts with health plans	Negative: PCPs with reimbursement under salary or capitation and quality of care system did not report adherence to guidelines significantly more often than primary care physicians with other methods of compensation	Control for multiple confounders, including receiving guidelines and feedback

HEDIS, healthcare effectiveness data and information set; HMO, Health Maintenance Organizations; STD, sexually transmitted diseases.

an alternate explanation is that the nonrewarded disease measures were indirectly affected by improving the rewarded measures through a 'halo effect' [26^{••}]. This study did include 12 asthma process indicators.

Four additional studies [21–23,24[•]] employed an uncontrolled before–after design (Table 2). One of the studies [24[•]] demonstrating positive results dealt specifically with asthma process measures in asthmatic children. A major problem with these types of uncontrolled studies is being able to attribute the changes that are observed after intervention to the incentive compared with the many other factors discussed above that could lead to improvement in the types of process measures assessed in these studies. For example, nonmonetary interventions employed in four of the studies that could have contributed to improved performance included enhanced data collection processes, physician education, provider feedback, peer review, and public reporting [21–23,24[•]].

Six studies [27–31,32] used a controlled before–after design that employed a comparison group (Table 3), although the comparison group in one of the studies [28] was state and national trends for the same measures. The major problem with this design is ensuring the comparability of the intervention and control groups. The results of only two of these studies [27,29] were adjusted for potential confounder. Nevertheless, selection bias may have been a problem in the first of these studies [27] (Table 3), and the second study [29] that adjusted for confounders did not adjust for potential differences between the intervention and comparison groups (Table 3). Levin-Scherz *et al.* [32] evaluated diabetes and asthma HEDIS measures and demonstrated significant positive results only for the diabetes measures. However, this study was confounded by the use of a number of potentially effective nonmonetary interventions in the pay-for-performance group but not the control group (Table 3). Similarly, in the positive study of Beaulieu and Horrigan [31], the incentive group also received guidelines, a chart assessment tool, and feedback.

Finally, two studies [33,34] used a cross-sectional survey design (Table 4). Both were adjusted for potential confounders. The study assessing reported smoking cessation interventions was positive, whereas the study assessing reported adherence to sexually transmitted diseases (STD) guidelines was negative. Associations made in cross-sectional surveys are particularly difficult to interpret regarding cause and effect relationships.

Discussion

Pay-for-performance has the potential to benefit all participants in the healthcare system, including health plans, physicians, employers, and consumers [35]. However, as

pay-for-performance programmes have monetary costs and potential unintended consequences [35], a firm evidence base for their effectiveness must be sought. This article suggests that positive or mixed results have been obtained in most studies of pay-for-performance effectiveness at the physician or physician group level in the ambulatory care setting. Factors apparently associated with success in these studies include large enough incentives, measures that are more amenable to change, responsive populations, and possibly combination with other strategies such as information system enhancements, guidelines, feedback, and public reporting.

As discussed initially, a number of methodological issues must be considered in the interpretation of the data. Although RCTs are probably the optimal design to ensure internal validity, only a minority of the identified studies were RCTs, and many of those RCTs were small. A major problem with the nonrandomized studies is the ability to attribute positive effects to the intervention, as opposed to the many other physician factors, system factors, or concurrent interventions, both from within and external to the pay-for-performance programme, which could be responsible for or at least contribute to improvements in quality indicators observed. For example, the publication of performance data itself has been associated with an improvement in health outcomes, although in a limited number of studies [36]. In addition, the performance feedback that may be associated with incentive programmes may itself positively affect performance outcomes [10]. Furthermore, a recent RCT has shown that internal medicine recertification with an asthma Practice Improvement Module (PIM) was associated with lower patient self-reported asthma severity measures [37].

A number of potentially important factors are often not explored in pay-for-performance studies [6^{••}], including the ones reviewed herein. These include financial factors (incentive as a proportion of total income, cost of complying, payment structure), environmental factors (number of other incentives in place, community-level professional initiatives in place to improve quality), provider factors (workload, proportion of patients for whom incentive is relevant), and organizational capabilities (information systems, use of guidelines, leadership, and culture) [6^{••}]. In addition, the reviewed studies do not assess the effect of financial incentives on care for conditions for which no financial incentives were provided to assess one important type of unintended consequence of specifically targeted pay-for-performance programmes [26^{••}].

Asthma was specifically included in three nonrandomized studies. Asthma HEDIS measures in children were not significantly improved compared with control children in a study [32] that was associated with improved diabetes

measures in adults. In contrast, several asthma process measures substantially improved in another study [24[•]] of asthmatic children, although this study was uncontrolled. Finally, a number of asthma process measures improved more rapidly after the 2004 pay-for-performance contract was instituted in the United Kingdom than before it was implemented, although no control group was possible in this study either [26^{••}]. As with the studies reviewed in this article in general, the results for asthma measures in particular are encouraging but not definitive.

Conclusion

Do pay-for-performance programmes that reward physician groups or individual physicians improve quality of care in ambulatory practice settings? The data reviewed in this article suggest they can improve markers of quality, but not always. Even when studies suggest positive effects, the designs often do not permit assurance that the effects are due to the incentives compared with other factors. Many parameters probably contribute to the success of pay-for-performance programmes or lack of it, including characteristics of the quality targets, the incentives, the physicians, and the patients. There are also alternative strategies for accomplishing the same quality improvement goals, such as regulatory initiatives, recertification processes, education programmes, direct subsidies for infrastructure improvements, computerized decision aids, provider feedback, public reporting of quality measure results, and patient incentives [6^{••},10,37,38]. More and better-designed studies are clearly needed to determine the actual effectiveness of incentives themselves isolated from other factors, circumstances that promote effectiveness, effectiveness relative to other strategies, and cost-effectiveness.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

Additional references related to this topic can also be found in the Current World Literature section in this issue (pp. 276–277).

- 1 Rosenthal MB, Landon BE, Norman ST, *et al.* Pay for performance in commercial HMOs. *N Eng J Med* 2006; 355:1895–1902.
- 2 Mehrotra A, Pearson SD, Coltin KL, *et al.* The response of physician groups to pay for performance incentives. *Am J Man Care* 2007; 13:249–255.
- 3 Petersen LA, Woodard LD, Urech T, *et al.* Does pay-for-performance improve the quality of healthcare? *Ann Intern Med* 2006; 145:265–272.
- 4 Chassin MR. Does paying for performance improve the quality of care? *Med Care Res Rev* 2006; 63:122S–125S.
- 5 Pronovost PJ, Miller M, Wachter RM. The GAAP in quality measurement and reporting. *JAMA* 2007; 298:1800–1802.
- 6 Frølich A, Talavera JA, Broadhead P, Dudley RA. A behavioral model of clinician responses to incentives to improve quality. *Health Policy* 2007; 80:179–193.
- Review of the literature on the effectiveness of pay-for-performance and public reporting and presentation of a conceptual model of factors that influence provider behavior.
- 7 Dudley RA. Pay-for-performance research: How to learn what clinicians and policy makers need to know. *JAMA* 2005; 294:1821–1823.

- 8 Doran T, Fullwood C, Gravelle H, *et al.* Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006; 355:375–384.
- 9 Zaslavsky AM, Hochheimer JN, Schneider EC, *et al.* Impact of sociodemographic case mix on the HEDIS measures of health plan quality. *Med Care* 2000; 38:981–992.
- 10 Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in healthcare? *Med Care Res Rev* 2006; 63:135–157.
- 11 Grossbart SR. What's the return? Assessing the effect of "Pay for Performance" initiatives on the quality of care delivery. *MCR&R* 2006; 63 (suppl): 29S–48S.
- 12 Bratzler DW, Nsa W, Houck PM. Performance measures for pneumonia: are they valuable, and are process measures adequate? *Curr Opin Infect Dis* 2007; 20:182–189.
- 13 Grady KE, Lemkau JP, Lee NR, Caddell C. Enhanced mammography referral in primary care. *Prev Med* 1997; 26:791–800.
- 14 Kouides RW, Bennett NM, Lewis B, *et al.* Performance-based physician reimbursement and influenza immunization rates in the elderly. *Am J Prev Med* 1998; 14:89–95.
- 15 Hillman AL, Ripley K, Goldfarb N, *et al.* Physician financial incentives and feedback: Failure to increase cancer screening in Medicaid managed care. *Am J Pub Health* 1998; 88:1699–1701.
- 16 Hillman AL, Ripley K, Goldfarb N, *et al.* The use of physician financial incentives and feedback to improve pediatric preventive care in Medicaid managed care. *Pediatrics* 1999; 104:931–935.
- 17 Fairbrother G, Hanson KI, Friedman S, Butts GC. The impact of physician bonuses, enhanced fees, and feedback on childhood immunization coverage rates. *Am J Pub Health* 1999; 89:171–175.
- 18 Fairbrother G, Siegel MJ, Friedman S, *et al.* Impact of financial incentives on documented immunization rates in the inner city: results of a randomized controlled trial. *Ambul Ped* 2001; 1:206–212.
- 19 Roski J, Jeddell R, An L, *et al.* The impact of financial incentives and a patient registry on preventive care quality: increasing provider adherence to evidence-based smoking cessation practice guidelines. *Prev Med* 2003; 36:291–299.
- 20 Ritchie LD, Bisset AF, Russell D, *et al.* Primary and preschool immunizations in Grampian: progress and the 1990 contract. *BMJ* 1992; 304:816–819.
- 21 Morrow RW, Gooding AD, Clark C. Improving physicians' preventive health-care behavior through peer review and financial incentives. *Arch Fam Med* 1995; 4:165–169.
- 22 Amundson G, Solberg LI, Reed M, *et al.* Paying for quality improvement. Compliance with smoking cessation guidelines. *Jt Comm J Qual Saf* 2003; 29:59–65.
- 23 Greene RA, Beckman H, Chamberlain J, *et al.* Increasing adherence to a community-based guideline for acute sinusitis through education, physician profiling, and financial incentives. *Am J Manag Care* 2004; 10:670–678.
- 24 Mandel KE, Kotagal UR. Pay for performance alone cannot drive quality. *Arch • Pediatr Adolesc Med* 2007; 161:650–655.
Uncontrolled study showing a dramatic increase in 'perfect' asthma care after the institution of an asthma improvement collaborative that included financial incentives
- 25 Millett C, Gray J, Saxena S, *et al.* Ethnic disparities in diabetes management and pay-for-performance in the UK: The Wadsworth Prospective Diabetes Study. *PLoS Med* 2007; 4:1087–1093.
- 26 Campbell S, Reeves D, Kontopantelis E, *et al.* Quality of primary care in • England with the introduction of pay for performance. *N Engl J Med* 2007; 357:181–190.
Comprehensive report of changes in multiple quality of care markers after the institution of the 2004 pay-for-performance contract for family practitioners in the UK
- 27 Kouides RW, Lewis B, Bennett NM, *et al.* A performance-based incentive program for influenza immunization in the elderly. *Am J Prev Med* 1993; 9:250–254.
- 28 Young GJ, Meterko M, Beckman H, *et al.* Effects of paying physicians based on their relative performance for quality. *J Gen Intern Med* 2007; 22:872–876.
- 29 Armour BS, Freidman C, Pitts Mm, *et al.* The influence of year-end bonuses on colorectal cancer screening. *Am J Manag Care* 2004; 10:617–624.
- 30 Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance. From concept to practice. *JAMA* 2005; 294:1788–1793.
- 31 Beaulieu ND, Horrigan DR. Putting smart money to work for quality improvement. *HSR: Health Serv Res* 2005; 40:1318–1334.
- 32 Levin-Scherz J, DeVita N, Timbie J. Impact of pay-for-performance contracts and network registry on diabetes and asthma HEDIS measures in an integrated delivery network. *Med Care Res Rev* 2006; 63:14S–28S.
- 33 McMenamin SB, Schauffler HH, Shortell SM, *et al.* Support for smoking cessation interventions in physician organizations: results from a national study. *Med Care* 2003; 41:1396–1406.
- 34 Pourat N, Rice T, Tai-Seale M, *et al.* Association between physician compensation methods and delivery of guideline-concordant STD care: is there a link? *Am J Manag Care* 2005; 11:426–432.
- 35 Schatz M, Blaiss M, Green G, Aaronson D. Pay for performance for the allergist-immunologist: potential promise and problems. *J Allergy Clin Immunol* 2007; 120:769–775.
- 36 Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data. What do we expect to gain? A review of the evidence. *JAMA* 2000; 283:1866–1874.
- 37 Simpkins J, Divine G, Wang M, *et al.* Improving asthma care through recertification. A cluster randomized trial. *Arch Intern Med* 2007; 167: 2240–2248.
- 38 Epstein AM, Lee TH, Hamel MB. Paying physicians for high-quality care. *N Engl J Med* 2004; 350:406–410.