

Pay for performance in health care: strategic issues for Australian experiments

Ian A Scott

In recent years, pay-for-performance (P4P) strategies have attracted considerable interest in the United States,^{1–5} the United Kingdom,⁶ Australia,⁷ Canada⁸ and other Western countries (Box 1). Their key attribute is a defined change in reimbursement to a clinical provider (individual clinician, clinician group or hospital) in direct response to a change in one or more performance measures, as a result of one or more practice innovations (Box 2).

Currently, across the US, over 170 P4P programs (or quality incentive payment systems) are in various stages of implementation in hospitals and group clinics, in both public⁹ and private sectors,¹⁰ covering 50 million beneficiaries. In 2004, the UK National Health Service (NHS) launched the General Medical Services Contract – Quality and Outcomes Framework, which gives family practitioners up to a 25% increase in income if various quality indicators are met.⁶ In Australia, Medicare's Practice Incentives Program targets quality in general practice,⁷ and, in Queensland, a P4P program targeting public hospitals is being piloted from July 2007.¹¹

P4P is popular because of the propositions that:

- Traditional approaches to optimising care, including education and certification, do not appear to guarantee minimum standards;
- Current quality improvement efforts are slow at reforming systems of care; and
- Few financial incentives exist for clinicians and managers to modify the status quo and reward high performance.¹²

However, despite its rising popularity, how effective is P4P, and what are the key determinants of its success? Are its benefits sustainable over the longer term? Are there unintended adverse effects?¹³ What forms of P4P might be applicable to Australian settings? In this article, I attempt to provide some answers based on empirical evidence obtained preferentially from controlled evaluations,¹⁴ and propose a framework under which P4P might evolve in Australia.

How effective (and cost-effective) are financial incentives for quality?

In a review of 17 studies (including 12 controlled trials),¹⁵ of which 13 examined process-of-care measures, five of the six studies of doctor-level financial incentives and seven of the nine studies of provider group-level incentives showed positive (improvement in all measures) or partial (improvement in some measures) effects on measures of quality, while four studies suggested unintended effects. Quality gains for most studies were modest (eg, a 3.6% increase in rates of cervical cancer screening, a 6.8%–7.4% increase in immunisation rates, and a 7.9% increase in rates of smoking status determination).

The US Centers for Medicare and Medicaid Services (CMS)/Premier Hospital Quality Incentive Demonstration (HQID) program recently reported 2-year results, comparing changes in process-of-care measures for acute myocardial infarction, heart failure and pneumonia between 207 P4P hospitals and 406

ABSTRACT

- In response to persisting quality problems in clinical practice, policymakers in various countries, including Australia, are experimenting with pay-for-performance (P4P) schemes that tie a portion of provider payments to performance on measures of quality.
- Rigorous studies of P4P efficacy are relatively few, with many focused on preventive care in ambulatory settings and many suggesting only modest gains in performance.
- Several key issues need to be considered in determining the optimal design and implementation methods for P4P programs, including:
 - the choice of clinical practice area;
 - the size of financial incentives and who should receive them;
 - the selection of quality measures and performance thresholds that determine incentive eligibility;
 - data collection methods; and
 - the best mix of financial and non-financial incentives.
- A proposed framework to guide Australian initiatives in P4P emphasises early clinician involvement in development, a phased approach from "pay-for-participation" in performance measurement to P4P within several pilot demonstration programs, and investment in clinical information technology.

MJA 2007; 187: 31–35

matched control institutions.¹ After adjusting for differences between groups in baseline performance, condition-specific volumes and other hospital characteristics, P4P was associated with composite process improvements of 2.6% for acute myocardial infarction, 4.1% for heart failure and 3.4% for pneumonia.

Other uncontrolled, before–after studies suggest improvements in selected quality metrics for other large-scale programs.^{2,5,16} While initial reports on the NHS experiment appear encouraging (median of 83% achievement across six practice domains),⁶ no baseline analysis was undertaken, with anecdotal reports suggesting a pre-P4P mean performance improvement of 60%–80% that had been rising over previous years.¹⁷

Cost-effectiveness studies are few in number. An incentives package for improving nursing home care and access saved an estimated US\$3000 per stay.¹⁸ Another study of P4P for cardiac care across eight hospitals estimated that 24 418 patients received improved care, translating to between 733 and 1701 additional quality-adjusted life-years (QALYs), with cost per QALY ranging from US\$12 967 to US\$30 081.¹⁹ A study of physician-level payments for diabetes care within a large health maintenance organisation reported savings of US\$4.7 million due to better quality care over 2 years compared with a program

FOR DEBATE

1 Descriptions of selected large-scale pay-for-performance programs

Program and scope	Quality measures	Financial incentives	Program evaluation
CMS/Premier Hospital Quality Incentive Demonstration¹ <ul style="list-style-type: none">• 270 hospitals• Initiated 2003	34 quality measures relating to acute myocardial infarction (AMI), heart failure (HF), pneumonia, coronary artery bypass surgery, and hip and knee replacement	Composite process scores for each condition/procedure calculated annually; hospitals categorised into performance deciles: those in top decile receive 2% bonus, those in second decile receive 1% bonus, those not achieving improvement above baseline penalised 1%–2%	Composite process scores over 2 years from intervention hospitals (adjusted for baseline performance, condition-specific volume and hospital characteristics) improved compared with control hospitals by 2.6% for AMI, 4.1% for HF, and 3.4% for pneumonia
CMS Physician Group Practice Demonstration² <ul style="list-style-type: none">• 10 doctor groups (total 5000 physicians)• Initiated 2005	32 quality measures relating to diabetes, HF, coronary artery disease, and preventive care services	Performance payments, allocated between efficiency and quality, earned if actual Medicare spending for assigned populations is below the annual average	None reported
Bridges to Excellence³ <ul style="list-style-type: none">• Multilateral coalition backed by the Leapfrog group of employers (170)⁴ and others; targeting doctors in Boston, Cincinnati/Louisville, Albany/Schenectady• Initiated 2003	Three initiatives comprising Physician Office Link (POL), Cardiac Care Link (CCL) and Diabetes Care Link (DCL). Each "link" comprises a broad set of measures, each accorded points towards total overall score	Up to US\$50 per patient for meeting POL goals, US\$200 per patient for CCL, and US\$100 per patient for DCL	None reported, as insufficient follow-up data. Broad measures of self-reported patient satisfaction, knowledge and confidence. High-performing participating doctors said to provide care at 15%–20% lower cost than non-participating doctors
Integrated Healthcare Association California Pay for Performance Program⁵ <ul style="list-style-type: none">• 90 preferred provider organisation groups (total 45 000 doctors)• Initiated 2003	Measures identified in three areas: clinical quality (covering seven conditions); patient experience; and investment and adoption of information technology to support patient care. Efficiency measure added for 2007	Each health plan uses its own methodology and formula to calculate bonus payments	Comparing 2005 data with 2004 data, an estimated 150 000 and 60 000 more women received cervical and breast cancer screening, respectively; 30 000 more children received immunisations; and 12 000 more patients received a diabetes test
NHS GMS Contract – Quality and Outcomes Framework⁶ <ul style="list-style-type: none">• 8105 family practices• Initiated 2003	146 measures identified in four areas: clinical quality (76) covering 10 disease groups; organisational quality (56); additional services related to women's and children's health (10); and patient experience (4)	Points assigned to each indicator based on importance, workload and costs; 1050 total possible points, each worth on average £120 per practice; during first 3 years, incentive payments paid regardless of performance to help practices implement required infrastructure	No baseline analysis In first 12 months practices scored a median of 95.5% of available points, 96.7% for clinical quality. Quality targets for eligible patients and for all patients achieved in 83.4% and 82.9% of practices, respectively, overall
Medicare Australia Practice Incentives Program⁷ <ul style="list-style-type: none">• General practitioners throughout Australia• Initiated 1998	Measures identified in seven areas: information systems (3); after-hours care (3); teaching (1); participation in NPS quality use of medicines program (1); care of diabetes, asthma, cervical screening or mental health (10); practice nurses (1); rural location (1)	\$3 per standard patient per year for providing data to Medicare Australia; information systems/after-hours care \$2 ppp; teaching \$100 per session; quality prescribing \$1 ppp; diabetes, asthma and cervical screening 25c to \$1 pp; service incentives \$35 to \$100 pp; outcomes payments \$2 to \$20 pp; mental health \$250 pp; practice nurses \$7 to \$8 pp; rurality 15% to 50% loading of total payment	None reported

CMS = Centers for Medicare and Medicaid Services. NHS = National Health Service. GMS = General Medical Services. NPS = National Prescribing Service. ppp = per patient per annum. pp = per patient.

cost of US\$2.2 million.²⁰ However, on the negative side, program expenditure within the first year of the NHS experiment vastly exceeded (by US\$700 million) the US\$3.2 billion additional funding allocated over 3 years, suggesting that quality improvements may not be commensurate with the rapid rise in costs.¹⁷

What types of clinical conditions or health care should be the target of financial incentives?

Most controlled trials have assessed preventive care and care of chronic diseases within primary care settings, with fewer targeting acute hospital care.¹⁵ High-volume, high-cost conditions with

demonstrated practice variation and a strong evidence base to guide practice improvement are preferred as initial targets. However, others argue that minority populations with special needs and demonstrably suboptimal care should not be neglected.²¹

To whom should incentives be directed?

Should incentives be given to individual clinicians, clinician groups (multi-clinician associations, from single clinic or hospital department to multi-site collaborations or networks), hospitals, or all three? Currently, 14% of US physician P4P programs target individual clinicians alone, 25% target both individuals and groups, and 61% target groups alone.¹⁶ Results of the NHS trial,⁶ and of studies reporting the largest absolute increases in quality measures,¹⁵ suggest that paying individuals or small clinician groups, rather than large groups or hospitals, may be the better strategy. However, deciding which of several clinicians caring for a patient with multiple illnesses should be assigned primary responsibility for care is problematic,²² and detracts from using P4P to entice large groups to share risk and invest in population-level systems of care improvement.

What is the optimal mix of financial and non-financial incentives?

Non-financial strategies for improving care quality (such as clinical guidelines, audits and feedback) yield median improvements of 6%–13% in process-of-care measures.²³ Do they provide an equal or better return on investment compared with financial incentives when used alone, or do both strategies in combination yield synergistic effects? In one randomised trial involving 60 physicians, bonus payments combined with feedback led to a 25% increase in documentation of immunisation status compared with no change with feedback alone.²⁴ In addition, non-cash incentives such as professional pride, recognition, opportunities to do more challenging work and the removal of administrative burdens may, in some circumstances, be equally motivating and should not be underrated.

What level of incentive is sufficient?

The “dose–response” relationship for incentives in terms of size, frequency and duration remains uncertain. Some P4P programs pay as little as \$2 per patient and exert impact, while others offer bonuses of up to \$US10 000 per practice and have no effect.¹⁶ The 25% income increase possible for NHS family doctors is postulated as the reason for the observed high levels of performance.⁶ In contrast, data from US health plans, which on average pay 9% of base income as bonus payments, show no consistent relationship between incentive size and response.¹⁶ One private-sector survey found incentives had to average 5%–20% of income for individual physicians and 1%–4% for hospitals.¹⁶ Payments in the CMS/Premier HQID, for example, can be up to a

2 Performance measures and practice innovations

Performance measure	Practice innovation
Clinical quality/effectiveness	Clinical guidelines, pathways, protocols; clinical audit and review procedures
• Process of care (eg, proportion of eligible patients with acute myocardial infarction who are prescribed a β-blocker at discharge)	
• Clinical outcomes (eg, deaths, complications, readmissions)	
Efficiency (eg, length of stay)	Discharge planning; criteria-led discharge; patient flow; community aftercare services; preadmission processing
Utilisation (eg, average number of laboratory requests per patient)	Test-ordering systems incorporating forcing functions, alerts for duplicate requests, audit and feedback of test request patterns; education in test utility and cost-effectiveness
Access (eg, clinic waiting time for new patient referrals)	Triage systems for filtering new referrals; discharge procedures for patients being reviewed whose condition is stable; preprocessing of patients prior to clinic visit
Safety (eg, numbers of medication errors per patient)	Incident reporting and review systems, pharmacist outreach services, standardised medication charts, drug safety alerts and restrictions
Patient satisfaction (eg, proportion of patients who report feeling satisfied with a consultation or episode of in-hospital care)	Patient education and consent procedures; self-management programs; patient-held management plans; patient-initiated clinical reviews
Practice/hospital management (eg, the level of information technology in a practice/hospital)	Electronic health records, clinical registries, order entry and reporting systems; personal digital assistants; wireless laptops for bedside access to clinical information

2% increase for high performance versus up to a 2% loss for poor performance.¹

It is possible that changes to the base funding formula, rather than add-on incentives, could facilitate higher-quality care,¹² especially within capitated or salaried environments. Under such circumstances, care improvements, which may incur additional “downstream” costs but no increased revenue, could financially negate incentives (which tend to be smaller than those under fee-for-service).²⁵ Another debate centres on whether incentives should involve reallocation of fixed funds from low performers to high performers or elicit additional (“new”) funding. Clearly, payments must compensate for the incremental net costs of undertaking all actions required of P4P programs.

What quality measures should be rewarded and how should they be ascertained?

Quality measures should ideally be clinically relevant, reasonably stable over periods of 2–3 years to allow serial comparisons (yet responsive to important new changes in practice), feasibly ascertainable, accurate, and actionable with high impact.

What types of measure should be accorded priority?

Ninety-one per cent of US programs target clinical quality measures, 50% cost-efficiency, 42% information technology, and 37% patient satisfaction (Box 2).⁶ The NHS experiment targets six care

domains using 146 indicators (Box 1).¹ The predominance of process-of-care measures across P4P programs relates to their sensitivity in detecting, in a timely manner, suboptimal care that is under direct clinician control. Their downside is the data collection burden required to determine patient eligibility for specific treatments. Outcome measures, such as mortality, complications and symptom relief, reflect patient-centred aggregate effects of health care, but must be risk-adjusted for variation in disease severity and other confounders unrelated to quality of care. The compromise of selecting process measures tightly linked in clinical trials with outcomes is also problematic, as absolute differences between hospitals in risk-adjusted mortality for several diagnoses, despite quite large variations in key process measures, are often small (no more than 2%) and the process and outcome measures are weakly correlated.²⁶

How many measures should be chosen?

Too few measures may lead to overemphasis on certain aspects of care and neglect of others; too many may cause confusion and administrative overload. Among existing P4P programs, measures number between 10¹⁶ and 146.⁶

Should measures be calculated at the level of individual clinician, clinical group or hospital?

Quality measures at the level of individual clinicians raise issues of attribution (most patients with a serious illness receive care from multiple clinicians) and analytic power (few clinicians have enough patients for statistically meaningful diagnosis-specific measurements).²⁷ Most P4P programs use measures pertaining to clinician groups, institutions or health plans, with some promoting population health perspectives that reward coordinated use of resources across different health care sectors.²⁸

Which data source — administrative or clinical — is preferred in generating quality measures?

In using risk-adjusted outcome measures, accessing routinely collected administrative data versus more difficult-to-obtain clinical data has its attractions, particularly if accuracy is enhanced by adding a select few, readily accessible clinical and laboratory variables.²⁹ However, for process measures, more resource-intense data abstraction from case records will continue to be required in the absence of electronic health records or clinical registries.

What performance thresholds should be used to determine reward eligibility?

Eligibility for incentive payment may rest on one or more performance thresholds:

- *absolute* — achievement of a predefined absolute threshold (eg, 75% of eligible patients receiving specific interventions);
- *relative* — improvement over baseline performance by a specified margin (eg, a 30% increase), or ranking measures (often as percentiles) relative to some external benchmark;
- *all cases* — payment for each instance of high-quality care regardless of overall performance (as in the NHS experiment); or
- some combination of these.

If absolute thresholds are to be used, should they be set at 75%, 85%, or 95%?

Also, will thresholds pertain to:

- *individual instances of care* — percentage of eligible patients who receive specific interventions;
- *composite care* — percentage of total instances in which eligible patients receive all of several interventions; or
- *all-or-nothing care* — percentage of patients who each receive all the interventions they are eligible to receive?³⁰

Most programs to date use a combination of individual and composite care measures, and those that use both relative and absolute thresholds³¹ appear to show more consistently positive results and exhibit the largest improvements in individuals or groups with the lowest baseline performance. The danger of ceiling effects with use of absolute thresholds was seen in the small gains in cervical cancer screening, mammography, and glycated haemoglobin testing within one program involving 172 physician groups — 75% of the bonus payments went to those already at or above the target 75th percentile.³²

What might be unintended adverse consequences?

The clinical dictum “first, do no harm” underlies the need for comprehensive evaluation frameworks to accompany P4P programs. Unless clinicians are convinced that performance measures are adequately risk-adjusted and take account of patient preferences, they may avoid very sick or challenging patients, or engage in other “gaming” strategies, such as reclassifying patient conditions or “ticking the box” even when care has not been provided or has been incompletely administered.¹³ In the NHS experiment, the strongest predictor of increases in reported achievement was a higher rate at which patients were excluded from the program (0.31% increase for every 1% increase in exception reporting).⁶ Clinicians or institutions serving disadvantaged populations, in whom performance thresholds may be difficult to achieve, might see falls in revenue. A predominant focus on process-of-care measures may promote inappropriate over-treatment in patients with multiple diseases.³³ Financial incentives may further undermine morale and professional altruism and erode holistic patient care.

What are the attitudes of clinicians and health managers towards incentives?

In one study of 22 clinician groups and nine hospitals, views on incentives varied widely depending on institutional culture, structure and stability, community context, quality measurement issues, nature and size of the incentives, and sustainability of interventions.³⁴ Among 35 health plans which had initiated P4P, providers expressed concerns about administrative burdens of customised programs, absence of standardised performance measures, and potential for conflicting financial incentives.³⁵ Physicians in California, while supporting P4P in general, wanted accurate and timely performance data, more patient time and staff support, and greater access to colleagues.³⁶

A proposed approach to P4P in Australia

In light of the above discussion, I propose the following framework under which P4P programs might evolve in this country, as opposed to the employer-subsidised managed care and competitive market environment of the US.

- A phased approach is suggested, which starts with “pay-for-participation” schemes in which participants focus on the development and testing of robust, standardised and preferably nationally consistent performance measures and systems of measurement that are integral to P4P.^{11,16} Only then should there be moves

towards "pay-for-performance" which, initially, might be a bonus to base funding for care improvement (which may include one-off investments in information technology upgrades), but which eventually becomes embedded into funding formulas (such as casemix funding for hospitals and fee-for-service items in private care), with financial penalties in cases of consistent and clearly evident poor performance.

- Pilot demonstration programs adopting this phased approach could be funded, by state or federal government, or private health funds (as appropriate), over a 3–4-year period, in different geographical (metropolitan, urban, rural/remote) and clinical (primary/community care, hospital care) settings in both public and private sectors. Funding could be focused on several well defined, common clinical conditions to provide proof of concept and assess the relative impacts of different design elements.³⁷
- Program design should: (i) target hospitals and clinician groups (arranged as clinical service teams, networks or collaborations), not individual clinicians; (ii) reward all high-quality care, not just that reaching predefined thresholds; (iii) include predominantly process measures with a select few unambiguous outcome measures (such as mortality) that, where feasible, not only relate to discrete episodes of care of individuals, but to cycles of care over time for whole populations (utilising linked national databases such as hospital discharge abstracts, the Pharmaceutical Benefits Scheme and the Medicare Benefits Schedule); and (iv) provide flexible payments for both capital purchases, particularly clinical information technology, and provider incentives linked to future performance guarantees (and associated penalties if these are not met).
- All P4P programs should have an appropriate governance structure comprising clinicians, health managers, quality improvement methodologists, and data managers/analysts, with advice being sought from health economists and epidemiologists.

Conclusion

While P4P programs can potentially improve care, they are not without problems if poorly designed or implemented in the absence of clinician involvement, or not subjected to ongoing evaluation and modification as circumstances demand. Key design elements discussed here deserve consideration if P4P programs are to be supported by clinicians, successful in meeting their objectives, and capable of providing useful answers to research questions.

Competing interests

None identified.

Author details

Ian A Scott, FRACP, MHA, MEd, Director

Department of Internal Medicine and Clinical Epidemiology, Princess Alexandra Hospital, Brisbane, QLD.

Correspondence: ian_scott@health.qld.gov.au

References

- 1 Lindenauer PK, Remus D, Roman S, et al. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med* 2007; 356: 486-496.
- 2 Trisolini M, Pope G, Kautter J, Aggarwal J. Medicare Physician Group practices: innovations in quality and efficiency. New York: The Commonwealth Fund, December 2006.
- 3 Bridges to Excellence. Rewarding quality across the healthcare system [website]. <http://www.bridgestoexcellence.org/> (accessed May 2007).
- 4 The Leapfrog Group [website]. <http://www.leapfroggroup.org/home> (accessed May 2007).
- 5 McDermott S, Williams T, editors. Advancing quality through collaboration: the California Pay for Performance Program. Oakland, Calif: Integrated Healthcare Association, February 2006. <http://www.oha.org/wp020606.pdf> (accessed May 2007).
- 6 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.
- 7 Medicare Australia. Formula for calculating PIP payments. http://www.medicareaustralia.gov.au/providers/incentives_allowances/pip/calculating_payments.shtml (accessed May 2007).
- 8 Pink GH, Brown AD, Studer ML, et al. Pay-for-performance in publicly financed healthcare: some international experience and considerations for Canada. *Healthc Pap* 2006; 6: 8-26.
- 9 Guterman S, Serber MP. Enhancing value in Medicare: demonstrations and other initiatives to improve the program. New York: The Commonwealth Fund, January 2007.
- 10 Rosenthal MB, Landon BE, Normand S-LT, et al. Pay for performance in commercial HMOs. *N Engl J Med* 2006; 355: 1895-1902.
- 11 Ward M, Daniels SA, Walker GJ, Duckett S. Connecting funds with quality outcomes in health care: a blueprint for a Clinical Practice Improvement Payment. *Aust Health Rev* 2007; 31 (Suppl 1): S54-S58.
- 12 Berwick DM, DeParle NA, Eddy DM, et al. Paying for performance: Medicare should lead. *Health Aff (Millwood)* 2003; 22: 8-10.
- 13 Fisher ES. Paying for performance — risks and recommendations. *N Engl J Med* 2006; 355: 1845-1847.
- 14 Campbell NC, Murray E, Darbyshire J, et al. Designing and evaluating complex interventions to improve health care. *BMJ* 2007; 334: 455-459.
- 15 Petersen LA, Woodard LD, Urech T, et al. Does pay-for-performance improve the quality of health care? *Ann Intern Med* 2006; 145: 265-272.
- 16 Baker G, Carter B. Provider Pay-for-Performance Programs: 2004 national study results. San Francisco, Calif: Medvantage, 2005. http://www.medvantage.com/Pdf/MV_2004_P4P_National_Study_Results-Exec_Summary.pdf (accessed Dec 2006).
- 17 Galvin R. Pay-for-performance: too much of a good thing? A conversation with Martin Roland. *Health Aff (Millwood)* 2006; 25: w412-w419.
- 18 Norton EC. Incentives regulation of nursing homes. *J Health Econ* 1992; 11: 105-128.
- 19 Nahra TA, Reiter KL, Hirth RA, et al. Cost-effectiveness of hospital pay-for-performance incentives. *Med Care Res Rev* 2006; 63 (1 Suppl): 49S-72S.
- 20 Curtin K, Beckman H, Pankow G, et al. Return on investment in pay for performance: a diabetes case study. *J Health Manag* 2006; 51: 365-374.
- 21 Institute of Medicine. Unequal treatment confronting racial and ethnic disparities in health care. Washington, DC: National Academy Press, 2002.
- 22 Pham HH, Schrag D, O'Malley AS, et al. Care patterns in Medicare and their implications for pay for performance. *N Engl J Med* 2007; 356: 1130-1139.
- 23 Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004; 8: iii-iv, 1-72.
- 24 Goldfarb S. The utility of decision support, clinical guidelines, and financial incentives as tools to achieve improved clinical performance. *Jt Comm J Qual Improv* 1999; 25: 137-144.
- 25 Dudley RA, Frolich A, Robinowitz DL, et al. Strategies to support quality-based purchasing: a review of the evidence. Rockville, Md: Agency for Healthcare Research and Quality, 2004. (Technical Review 10, AHRQ publication No. 04-0057.)
- 26 Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA* 2006; 296: 2694-2702.
- 27 Hofer TP, Hayward RA, Greenfield S, et al. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA* 1999; 281: 2098-2105.
- 28 Kindig DA. A pay-for-population health performance system. *JAMA* 2006; 296: 2611-2613.
- 29 Pine M, Jordan HS, Elixhauser A, et al. Enhancement of claims data to improve risk adjustment of hospital mortality. *JAMA* 2007; 297: 71-76.
- 30 Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. *JAMA* 2006; 295: 1168-1170.
- 31 Fairbrother G, Hanson KL, Friedman S, Butts GC. The impact of physician bonuses, enhanced fees, and feedback on childhood immunisation rates. *Am J Pub Health* 1999; 89: 171-175.
- 32 Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance. From concept to practice. *JAMA* 2005; 294: 1788-1793.
- 33 Boyd CM, Darer J, Boult C, et al. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. *JAMA* 2005; 294: 716-724.
- 34 Conrad DA, Saver BG, Court B, Heath S. Paying physicians for quality: evidence and themes from the field. *Jt Comm J Qual Patient Saf* 2006; 32: 443-451.
- 35 Trude S, Au M, Christianson JB. Health plan pay-for-performance strategies. *Am J Manag Care* 2006; 12: 537-542.
- 36 Teleki SS, Damberg CL, Pham C, Berry SH. Will financial incentives stimulate quality improvement? Reactions from frontline physicians. *Am J Med Qual* 2006; 21: 367-374.
- 37 Rosenthal MB, Dudley RA. Pay-for-performance. Will the latest payment trend improve care? *JAMA* 2007; 297: 740-744.

(Received 29 Jan 2007, accepted 17 Apr 2007)

