Pay for performance programs in Australia: a need for guiding principles

Ian A Scott

Abstract

Pay-for-performance (P4P) programs which reward clinical providers with incentive payments based on one or more measures of quality of care are now common in the United States and the United Kingdom and it is likely they will attract increasing interest in Australia. However, empirical evidence demonstrating effectiveness of such programs is limited and many existing programs have not had rigorous outcome evaluation. To maximise success, future P4P programs should incorporate the lessons and insights obtained from previous experience. Based on a review of published trials, program evaluations and position statements, the following principles that may guide future program design and implementation were synthesised: 1) formulate a rationale and a business case for P4P; 2) use established evidence-based performance measures; 3) use rigorous and verifiable methods of data collection and analysis; 4) define performance targets using absolute and relative thresholds; 5) use rewards that are sufficient, equitable and transparent; 6) address appropriateness of provider responses and avoid perverse incentives; 7) implement communication and feedback strategies; 8) use existing organisational structures to implement P4P programs; 9) attribute credit for performance to participants in ways that foster population-based perspectives; and 10) invest in outcomes and health service research. Recommendations flowing from these principles relevant to Australian settings are provided.

Aust Health Rev 2008: 32(4): 740–749

Ian A Scott, FRACP, MHA, MEd, Director
Internal Medicine and Clinical Epidemiology Department, Princess Alexandra Hospital, Brisbane, QLD.

Correspondence: Associate Professor Ian A Scott, Internal Medicine and Clinical Epidemiology Department, Princess Alexandra Hospital, Ipswich Road, Brisbane, QLD 4102.
ian_scott@health.qld.gov.au

What is known about the topic?

There is increasing interest in pay-for-performance as a mechanism to achieve health system outcomes. Yet there is limited empirical evidence of best practice.

What does this paper add?

This paper outlines 10 principles for success derived from a review of published trials, program evaluations and position statements.

What are the implications for practitioners?

While the principles outlined can assist in consistency of approach, the author stresses the need for the National Health and Medical Research Council and the federal health department to ensure the evaluation of P4P programs as they migrate from pilot projects to mainstream applications.

IN RECENT YEARS, pay-for-performance (P4P) programs have become popular in the United States[1] and the United Kingdom[2] as a tool for improving quality of health care, and other countries such as Canada[3] have expressed interest in adopting such schemes. In the US there are currently more than 150 P4P programs (alternatively termed quality incentive payment systems [QIPS]) in various stages of implementation[4] across hospitals and group clinics in both public and private sectors, led by the Centres for Medicare and Medicaid Services (CMMS)[5] and commercial health maintenance organisations (HMOs).[6] In 2004, the National Health Service (NHS) in the UK launched a P4P contract with 8105 family practices that allows practitioners to increase their income by up to 25% depending on their performance with respect to 146 quality indicators relating to clinical care of 10 chronic diseases, organisation of care and patient experience.[2] In Australia, the only significant P4P initiative to date has been Medicare Australia's Practice Incentive Payment scheme targeting general practice which, despite initiation in 1998, has not under-

taken formal outcome evaluation.[7] More recently, in Queensland, a phased program for paying incentives to public hospitals to participate in quality improvement activities and improve processes and outcomes has been initiated.[8]

The key attribute of P4P is a defined change in reimbursement to a clinical provider (individual clinician, clinician group or hospital) as a direct result of a change in one or more quality measures. Despite the speed with which P4P has been adopted, the empirical evidence remains preliminary and inconsistent. A recent systematic review of formal studies published up to November 2005 that required, as minimum inclusion criteria, quantitative measures of at least one quality aspect of care coupled with a baseline comparison, found only eight randomised trials and four controlled before–after trials.[9] While the majority of studies in this review were positive, the effects on quality of such programs were, in general, small and mostly limited to ambulatory preventive care services. Other reviews of published trials concluded that evidence is lacking: most reports are descriptive rather than evaluative, many studies target a narrow clinical area and do not assess long-term sustainability, and results are mixed, with effectiveness heavily dependent on design features, size of the incentives, and receptivity of local cultures.[10-14]

In contrast, recent evaluations of large-scale programs have yielded more promising results. First year results from the NHS experiment show a median performance achievement of 83% (ie, proportion of patients within each of 10 disease groups who satisfied all quality indicators).[15] In the CMMS Premier QIPS demonstration program involving 270 US hospitals, although incentives were small, program participation significantly improved quality of care (median quality score increases of between 3% and 12% for five clinical conditions/procedures) within a 12-month period compared with control hospitals.[5] In a large, heterogenous integrated delivery network, P4P contracts for physicians led to significant improvement in quality metrics relating to care of asthmatic and diabetic patients.[16] A program involving 45 000 physicians in California estimated that, in the first full year of P4P, 210 000 more patients received cervical and breast cancer screening, immunisations or diabetes tests.[17]

While the debate around efficacy continues, several commentators have raised concerns about unintended, adverse consequences of P4P and whether such a strategy can achieve sustained across-the-board improvements in care.[18-21] In particular, there is the risk that unless clinicians are firmly convinced of the validity of risk adjustment as applied to outcome data, or of process of care measures that take account of variation in patient preferences, they may avoid sick or challenging cases, or engage in other "gaming" strategies such as reclassifying patient conditions or "ticking the box" even when care has not been given or has been incompletely administered.[12] In the NHS experiment, the strongest predictor of reported achievement was a higher rate at which patients were excluded from the program (0.31% increase for every 1% increase in exception reporting).[15] Disadvantaged populations may suffer under P4P as clinicians or institutions serving such populations see revenues fall because quality thresholds used to award incentive payments are beyond their reach. Overemphasis on process-of-care measures may promote inappropriate over-treatment in patients with multiple diseases[22] and divert efforts away from co-ordinated care of chronic illness across different sectors. Financial incentives may further undermine morale and professional altruism[23] and erode holistic approaches to care as emphasis is given to "treating the measure" rather than the patient.[24]

In responding to the evidence vacuum, investigators have proposed a research agenda to fill the void[25] and policy makers are calling for more detailed evaluations of existing and future programs.[4] In the meantime, issues that need to be considered in designing effective P4P programs have been defined,[26] and professional organisations and government agencies have enunciated sets of principles to guide members and payers through the transitional period from

pilot programs to mainstream funding arrangements.[27-33]

As P4P strategies will attract increasing attention in Australian settings, this article enumerates guiding principles based on a review of the previously cited position statements, program evaluations and systematic reviews of P4P programs in developed countries. Each principle is accompanied by specific recommendations relevant to the Australian context.

Principle 1. Formulate a rationale and business case

The primary goal of any P4P program must be to promote quality care rather than simply to achieve monetary savings. The rationale underpinning such programs should be clearly stated and relate to existing evidence of unsafe or suboptimal care and the inadequacy of existing quality improvement efforts to fix the problem. A sustainable business case must recognise that implementing quality improvement systems may require substantial initial capital investment in information technology (for data collection and decision support), staff training, additional care delivery capacity and process change.[34] Lead clinicians and managers with expertise must be involved in design, implementation and evaluation of any program on a voluntary basis under a formal governance structure and be encouraged, by direct or indirect incentives, to participate in working groups, monitor data collection, participate in data analysis, and help create and implement action plans that result in continuous improvement. Shared accountability for quality should be matched with shared rewards that recognise efforts that extend beyond routine duties. Programs should be available to any clinician group, specialty or institution who wishes to participate and must not favour one group, specialty or institution over another. Before widespread roll-out, P4P programs should be pilot-tested for a sufficient duration to obtain valid data across a variety of settings and specialties, and to also test for potential for patient de-selection. Before local implementation, operational details must be made known to all relevant stakeholders and clinician "buy-in" and participation secured.

Recommendations: The Australian Institute of Health and Welfare (AIHW) and the National Institute of Clinical Studies should collaborate in building a profile of suboptimal care in Australia which prioritises areas of clinical practice to which P4P programs could be targeted. As chronic disease conditions and their acute complications account for most health expenditure, these should be targeted in the first instance. State and federal health departments (the latter in combination with the Health Insurance Commission) should develop tender templates to be used by public hospital and other health care sectors respectively to apply for grants for pilot P4P programs that focus on one or more of the priority areas and which incorporate the above-mentioned prerequisites for successful implementation.

Principle 2. Use established evidence-based performance measures

Central to any P4P program is the requirement for a set of performance measures that are valid, reliable, evidence-based, interpretable, feasibly ascertainable and actionable. Measures should pertain to key areas of care in which there is potential to improve and not be chosen simply because they can be documented at low cost. The ideal number of measures represents a balance between having too few (leading to over-emphasis on some aspects of care and neglect of others) and too many (causing confusion and administrative overload). Process-of-care indicators which, in the main, comprise proportions of eligible patients who receive specific interventions, must be rigorously developed and tested using accepted methods,[35] based on current evidence, and rapidly responsive to changes in guidelines and professional consensus. They should incorporate standard data definitions which are recognised by relevant professional organisations.[36]

Performance measures that assess outcomes such as death must be appropriately risk-adjusted for patient-level factors such as illness severity and co-morbidities using methods that are valid, transparent and conform to appropriate standards. Performance measures must be capable of providing reliable longitudinal comparisons among groups of providers such that measures are consistent across providers over time but remain current with advances in science and resistant to gaming. For purposes of P4P programs, performance measures should preferably relate only to clinician groups or institutions, as performance measures at the level of individual clinicians are constrained by problems of attribution (most patients with serious illness receive care from multiple clinicians) and inadequate sample sizes (few clinicians have enough patients for statistically meaningful analysis).[37]

Recommendations: P4P programs should focus on, or give more weighting to, process of care measures (proportion of diabetic patients who have their feet examined at least once a year) or tightly coupled intermediary outcome measures (such as the proportion of diabetic patients who achieve glycated haemoglobin [HbA_{1c}] <7%). While some of these measures will be aligned with existing quality and safety initiatives (such as the NICS venous thromboembolism prophylaxis program),[38] others will be new. More remote infrequent outcomes, such as deaths, and qualitative measures such as patient satisfaction should be de-emphasised. The most appropriate measures should have national coverage and be formulated by the relevant specialty colleges and their respective subspecialty societies or faculties under the auspices of the Australian Committee of Medical Colleges working in collaboration with the AIHW, NICS and the Australian Commission on Quality and Safety in Health Care.

Principle 3. Use rigorous and verifiable methods of data collection and analysis

The data collection and analysis must be scientifically valid and subject to periodic external audit. Programs should use an appropriate mix of accurate administrative data (for outcomes) and clinical data abstracted from medical records (for processes of care). Accessing routinely collected administrative data versus more difficult-to-obtain clinical data has attractions in regard to outcome measures, particularly if enhanced by adding a select few, readily accessible clinical and laboratory variables.[39] Data collection from clinical records should not be disruptive to clinical work practices. Administrative datasets or risk-adjustment models should be validated against high-quality clinical data. Physicians should be allowed to review, correct and supplement data, especially administrative data, in the absence of onerous appeals processes and before any data are released in publicly disclosed performance reports or used to determine levels of reward. Performance measures must be based on data collected over a significant period of time and, for process measures, relate care delivered (as numerator) to a statistically valid population of eligible patients (as denominator). Programs must have defined security measures for ensuring data confidentiality and protection from corruption or unauthorised use. Clinician groups or institutions need to be reimbursed for any added administrative costs incurred as a result of collecting and reporting valid data.

Recommendations: As preference is to be given to process of care performance measures, data collection systems will need to be developed and funded that capture such data in real time in clinical settings. At the very least, hybrid systems which blend administrative or claims data with abstracted medical records data may require start-up investment in information technology, especially in group practices of private medical specialists. Where they currently exist, automated electronic systems such as pharmacy databases, clinical registries, structured discharge summaries, and medical record and prescribing software may be more efficient in minimising time and resources in retrieving required data. Existing data dictionaries should be consulted as reference standards. The fraction of total available cases that should be sampled to ensure statistical

reliability for specific sets of performance measures should be pre-determined by the AIHW, together with methods for verifying data quality which allow for independent and confidential review. Independent audits of the validity and reliability of collected data will be required periodically to ensure sufficient accuracy of data for P4P purposes.

Principle 4. Define performance targets using absolute and relative thresholds

P4P programs require agreed and achievable targets or thresholds of performance to determine eligibility for rewards. Such thresholds can be of three types: 1) *absolute* — achievement of a predefined absolute threshold of performance (such as 75% of eligible patients receiving specific interventions); 2) *relative* — improvement over baseline performance by a specified margin (such as a 30% increase), or rankings (often as percentiles) relative to some external benchmark; or 3) *all cases* — payment for each instance of high-quality care regardless of overall performance (as in the NHS experiment).

Also, consideration will need to be given to determining if absolute or relative thresholds pertain to *individual instances of care* (% of eligible patients who receive specific single interventions), *composite care* (% of total instances in which eligible patients receive all of several interventions), or to *all-or-nothing care* (% of patients who each receive all the interventions they are eligible to receive).[40] Clearly all-or-nothing thresholds constitute the most rigorous standard but may be setting the bar too high.

Evidence suggests that P4P programs which use both absolute and relative thresholds have more consistently positive results and promote a larger improvement in those individuals or groups with the lowest baseline performance.[41] Programs which reward both the size of change in quality and sustainability of that improvement over time are also more likely to achieve lasting effects.[42] Therefore, performance thresholds should be pre-specified as both absolute thresholds (which should be set fairly high) and relative thresholds (which should be seen as reasonably achievable), and, where appropriate, include an adjustment for sustainability over a given time period. Rewards tied to outcomes may depend on the extent that observed outcomes equal or exceed, by a pre-defined margin, those that would be expected based on appropriate risk-adjustment methods.

Where analyses show wide gaps between actual performance and target thresholds, an assessment should be made, in consultation with participating clinicians, of how much this relates to constraints regarding demographic characteristics of patient populations being serviced, local service and expertise availability, referral patterns, and managerial perspectives. UK experience suggests that servicing socioeconomically disadvantaged populations and having small general practices with low practitioner to population ratios can hinder achievement of optimal performance levels.[15] Programs must also take account of "knock-on" or "downstream" resource implications of participants requesting more services as part of incentive-driven good-quality care, but where such services are either non-existent or insufficient to meet demand.[43]

Recommendations: Relative performance thresholds applied to process of care measures should be prioritised according to how the majority of eligible patients will benefit, in absolute terms, from receiving specific clinical interventions. For example, reducing mean glycated haemoglobin (HbA_{1c}) levels by 2% in diabetic patients with early-age onset of disease and a baseline value of 11% will reduce the absolute risk of blindness considerably more than the same change in patients with later onset disease and a baseline value of 8%.[44] In the first instance, the provision of specific interventions to individual patients should be the performance measure as opposed to composite or all-or-nothing measures which may actually discourage some providers from trying to achieve target levels. Finally, the units to which performance measures are applied should be hospitals or large private group practices or divisions, rather than to individual clinicians for several reasons: little practice variation

can be attributed to individual clinicians; [45] much of modern care is provided by teams and production of quality is thus an exercise in teamwork; organisation-level and group-level incentives play an important role in moulding and reinforcing clinician-level reward structures; and variation in care is greatest at facility or large-group level compared with the level of the individual[46] which, in turn, calls for remedial strategies that can be applied to large-scale patient populations. Having said this, it will be important to ensure that funds received by these intermediaries include sufficient incentive to front-line clinicians to personally engage in the quality effort rather than see the whole exercise as favouring investment in infrastructure as the "quick fix" for all quality problems.

Principle 5. Use rewards that are sufficient, equitable and transparent

P4P programs must be based on rewards and not penalties, and rewards should be based on performance related to pre-defined thresholds (as discussed above), not on rankings arising from benchmarking exercises involving other clinician groups or institutions. Incentives need to be bonus payments over and above base funding using a central supplemental funding pool which is estimated in advance. Programs must provide participants with full explanations as to the methods and performance measures used to determine incentive eligibility and incentive amounts. While the dose–response relationship remains uncertain, incentives should equal, on average, a 10% increase over base funding for clinician groups and a 2% increase for institutions.[47] Incentives must be sufficient to offset opportunity costs of instituting information systems, training programs and new forms of service delivery. In terms of timing, while annual incentive payments are most common, semi-annual or concurrent payments appear to be more effective in changing physician behaviour, presumably by preventing decline in awareness of the intervention and of performance feedback. [48]

Programs must reward all participants who actively participate and who achieve pre-specified performance targets, while ensuring that incentives are designed not to minimise patient access or financially disadvantage those participants who serve minority patient populations. Programs should include a "trickle down" process that ensures individual clinicians or units respectively receive some of the incentive directly as reward for effort expended. Program participants should be able to review and appeal the process by which incentives are awarded, and be considered, on a case-by-case basis, for a moiety of "in good faith" funding if achieving improvement that approaches but does not equal target thresholds.

Recommendations: Incentives should follow an evolutionary process that rewards providers, in the first instance, for voluntarily reporting performance measures and/or establishing information systems that will furnish the necessary performance data. The next phase is paying providers for meeting performance benchmarks. Initially, incentives would be given as a bonus beyond usual reimbursement but over time would transform to a "zero-sum" or "budget-neutral" stance whereby a portion of aggregate funds is set aside before allocation and then distributed to programs which have passed performance thresholds. As large group practices and hospitals in both the public and private sectors are the preferred units of accountability, between 2% and 5% increase in baseline Medicare or state government levels of remuneration should constitute sufficient incentive for improvement supplemented, if necessary, by the same percentage increase in gap private rebates from health insurance funds.

Principle 6. Address appropriateness of provider responses and avoid perverse incentives

P4P programs should explicitly consider the behaviours to be discouraged and those to be encouraged. Programs must implement vigilant procedures for detecting perverse incentives (such as adverse patient selection, gaming, and treating the metric rather than the patient) and be

prepared to quickly correct any design flaws that have unintended consequences. Where problems have been detected, these must be quickly made known, in writing, to all program participants.

Recommendations: Explicit rules need to be formulated by specialty colleges around criteria that allow certain patient populations to be exempt from P4P programs as a result of socio-economic disadvantage, special ethnic or racial needs, or endemic non-adherence to clinician advice. Both public and private programs would need to be regularly audited for their adherence to actually providing the care they say they have, with demands for return of payments in cases of demonstrably serious breaches.

Principle 7. Implement governance procedures which incorporate communication and feedback strategies

Programs should have explicit and transparent systems of governance and means for communicating program details and results to all participants and for receiving their feedback on how the program is progressing. Care should be taken to identify and, where appropriate, proactively respond to contentious issues that have repeatedly surfaced in questionnaire surveys and informant interviews involving other programs:[49-51] role of institutional culture, structure and stability, community context, quality measurement issues, nature and size of the incentives, administrative burdens of customised programs, absence of standardised performance measures and accurate and timely performance data, potential for conflicting financial incentives, and inadequate patient time, staff support and access to colleagues.

Where programs identify clinicians with exceptional performance, or identify strategies that accelerate practice improvement, this information should be widely promulgated. Newsletters and email bulletins should be disseminated on a regular basis from a central agency which collates performance data from different programs (de-identified as appropriate). This should be combined with open forums on at least an annual basis at which activities of different programs can be discussed with a broad group of participants.

Recommendations: Each P4P program should have a governing board or committee with representation of key stakeholders that oversees program policies and activities under a transparent and agreed charter of operations and terms of reference. Meeting schedules for reporting program outcomes and revising program activities should be made known to all committee members and annual reports circulated widely to all interested parties. Evaluative reports from each program, either public or private, should be periodically submitted to a central agency which may be the Health Insurance Commission in the first instance, working in collaboration with the AIHW.

Principle 8. Use existing organisational structures to implement P4P programs

Rather than set up new governance structures for P4P programs, it is more efficient and politically strategic to use existing structures. In particular, as P4P needs to be clinician-led, clinician networks, collaborations or large group practices may be the preferred vehicles to operationalise P4P, given their potential to enable rapid program implementation, access to large numbers of clinicians working in different settings, consistency of measurement and economies of scale, and testing and revision of program methods. Clinician-led service networks (CSNs) and Divisions of General Practice networks (DGPNs) have become popular in recent times, both in Australia[52,53] and elsewhere,[54] as agents of service change and improvement across wide jurisdictions, with some acting as budget-holders and assuming purchaser–provider functions. Recent studies suggest that tightly structured CSNs can reduce hospital costs[55] and improve quality of care.[56]

Recommendations: As the responsibilities of CSNs and DGPNs include the promotion and

monitoring of high standards of care, having such networks assume responsibility for operationalising P4P programs seems logical. Tenders for grants should give preference to programs that will use pre-existing collaborative CSNs and DGPNs as the vehicles for implementing both the program and its governance structure.

Principle 9. Attribute credit for performance to participants in ways that foster population-based perspectives

The use of aggregate performance data from multiple clinical units or hospitals engenders collaborative team approaches involving different clinical disciplines and helps create shared accountability along the total patient care journey. Participants in P4P programs should be rewarded on the basis of patterns of care across their serviced populations, not for care on a case-by-case basis. Improving overall quality of care, including efficiency, should be the goal rather than attempting to eliminate outliers. Special, one-off rewards to care teams, multidisciplinary disease management initiatives, and programs that target needy populations should be considered in amplifying the potential for marginal gains.

Recommendations: As mentioned, payment of incentives at the facility or large group practice level facilitates the building of integrated systems of care that would not be possible with individual providers, and which can create reputational incentives that promote population-centred systems of care. Federal and state health departments, aided by the Health Insurance Commission, should publicise facilities and practices that satisfy criteria in meeting the needs of whole populations, not just of individuals.

Principle 10. Invest in outcomes and health service research

In areas of practice where the evidence base for defining care standards is inadequate or where accurate performance measurement using current methods is suboptimal or not feasible, P4P programs should encourage investment and participation in data collection and analysis efforts and clinical studies that shed light on how to fill these gaps. In addition, research should be conducted into the impacts (both intended and unintended) of P4P programs on access, costs, quality, health outcomes and clinician and patient satisfaction, and how their efficacy compares with other quality improvement approaches (such as clinical guidelines, audit and feedback, and academic detailing).

Recommendations: The National Health and Medical Research Council and the federal health department should tender for a consortium of independent researchers to investigate the design, implementation and evaluation of P4P programs as they migrate from pilot projects to mainstream applications. Any application for funding arising from such programs should be accompanied by mandatory registration of the program with the research consortium and consent of the program's governing body to provide detailed, de-identified data for investigative purposes.

Conclusion

While P4P programs have potential to improve quality of care and appear to have been taken up with speed in various jurisdictions around the developed world, they are not without potential to harm if poorly designed or implemented. A set of guiding principles combined with context-specific recommendations derived from current research and experience will likely increase the chances that new P4P programs are supported by clinicians and managers and are successful in improving quality of care.

Competing interests

The author declares that he has no competing interests.

References

1 Epstein AM, Lee TH, Hammel MB. Paying physicians for high-quality care. *N Engl J Med* 2004; 350: 406-410.

2 Roland M. Linking physicians' pay to the quality of care – a major experiment in the United Kingdom. *N Engl J Med* 2004; 351: 1448-54.

3 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.

4 Baker G, Delbanco S. Pay for Performance: Nationwide Perspectives. 2006 Longitudinal Survey Results with 2007 Market Updates. Medvantage and The Leapfrog Group, 2007. Available at: http://www.medvantage.com/Pdf/2006NationalP4PStudy.pdf (accessed Feb 2008).

5 Grossbart SR. What's the return? Assessing the effect of "pay-for-performance" initiatives on the quality of care delivery. *Med Care Res Rev* 2006; 63 (1 Suppl): 29S-48S.

6 Rosenthal MB, Landon BE, Normand S-LT, et al. Pay for performance in commercial HMOs. *N Engl J Med* 2006; 355: 1895-902.

7 Medicare Australia. Formula for calculating PIP payments. www.medicareaustralia.gov.au/providers/incentives_allowances/pip/calculating_payment.htm (accessed Sep 2005).

8 Ward M, Daniels S, Walker G, Duckett S. Reconnecting funds with outcomes in healthcare: a blueprint for a 'Clinical Practice Improvement Payment.' *Aust Health Rev* 2007; 31 (Suppl 1): S54-8. Available at: http://www.aushealthreview.com.au/publications/articles/issues/ahr_31_1_0407/ahr_31_1_s054.asp (accessed Sep 2008).

9 Petersen LA, Woodard LD, Urech T, et al. Does pay-for-performance improve the quality of care? *Ann Intern Med* 2006; 145: 265-72.

10 Chaix-Couturier C, Durand-Zaleski I, Jolly D, Durieux P. Effects of financial incentives on medical practice: results from a systematic review of the literature and methodological issues. *Int J Qual Health Care* 2000; 12: 133-42.

11 Gosden T, Farland F, Kristiansen IS, et al. Impact of payment method on behaviour of primary care physicians: a systematic review. *J Health Serv Res Policy* 2001; 6: 44-55.

12 Armour BS, Pitts MM, Maclean R, et al. The effect of explicit financial incentives on physician behaviour. *Arch Intern Med* 2001; 161: 1261-6.

13 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 04-0057.)

14 Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in health care? *Med Care Res Rev* 2006; 63: 135-57.

15 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: 375-84.

16 Levin-Scherz J, DeVita N, Timble 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.

17 Integrated Healthcare Association. Advancing quality through collaboration: the California Pay for Performance Program. McDermott S, Williams T eds. Oakland, Calif: IHA, February 2006.

18 Berwick DM. The toxicity of pay for performance. *Qual Manag Health Care* 1995; 4: 27-33.

19 Kirsch M. Pay-for-performance – quality or quagmire. *Am J Gastroenterol* 2006; 101: 2453-5.

20 Fisher ES. Paying for performance – risks and recommendations. *N Engl J Med* 2006; 355: 1845-7.

21 Bonis PAL. Quality incentive payment systems. Promise and problems. *J Clin Gastroenterol* 2005; 39: S176-82.

22 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: 716-24.

23 Frey BS. On the relationship between intrinsic and extrinsic work motivation. *Int J Indust Organis* 1997; 15: 427-39.

24 Eggleston K. Economic modelling of methods to stimulate quality improvement. *Int J Qual Health Care* 2005; 17: 521-31.

25 Dudley RA. Pay-for-performance research. How to learn what clinicians and policy makers need to know. *JAMA* 2005; 294: 1821-3.

26 Scott IA. Pay for performance in health care - strategic issues for Australian experiments. *Med J Aust* 2007; 187: 31-5.

27 American Medical Association. Guidelines for Pay-for-Performance Programs. Chicago, IL: AMA, 2005.

28 American College of Physicians. Linking physician payments to quality of care. Position Paper. Philadelphia, PA: ACP, 2005.

29 American Academy of Family Physicians. Shaping the future of pay for performance programs. *Ann Fam Med* 2005; 3: 562-4.

30 Bufalino V, Peterson ED, Burke GL, et al. Payment for quality: guiding principles and recommendations: principles and recommendations from the American

Heart Association's Reimbursement, Coverage, and Access Policy Development Workgroup. *Circulation* 2006; 113: 1151-4.

31 Brush JE, Krumholz HM, Wright JS, et al. American College of Cardiology 2006 principles to guide physician pay-for-performance programs: a report from the American College of Cardiology Work Group on Pay for Performance. *J Am Coll Cardiol* 2006; 48: 2603-9.

32 Milgate K, Cheng SB. Pay-for-performance: the MedPAC perspective. *Health Aff (Millwood)* 2006; 25: 413-19.

33 Institute of Medicine. Rewarding provider performance: aligning incentives in Medicare. September 2006. Available at: http://www.iom.edu/CMS/3809/19805/37232.aspx (accessed December 29, 2006).

34 Casalino L, Gillies RR, Shortell SM, et al. External incentives, information technology, and organised processes to improve health care quality for patients with chronic diseases. *JAMA* 2003; 289: 434-41.

35 Hickey A, Scott I, Denaro C, et al. Using clinical indicators in a quality improvement programme targeting cardiac care. *Int J Qual Health Care* 2004; 16 (Suppl1): i11-25.

36 Chew DPB, Allan RM, Aroney CN, Sheerin NJ. National data elements for the clinical management of acute coronary syndromes. *Med J Aust* 2005; 182 (Suppl): S1-16.

37 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: 2098-105.

38 The National Institute of Clinical Studies Venous Thromboembolism Program. Available at: www.nhmrc.gov.au/nics/asp/index.asp?page=programs/programs_article&cid=5263&gid=191&id=499 (accessed Feb 2008).

39 Pine M, Jordan HS, Elixhauser A, et al. Enhancement of claims data to improve risk adjustment of hospital mortality. *JAMA* 2007; 297: 71-6.

40 Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. *JAMA* 2006; 295: 1168-70.

41 Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance. From concept to practice. *JAMA* 2005; 294: 1788-93.

42 Birkmeyer NJ, Birkmeyer JD. Strategies for improving surgical quality-should payers reward excellence or effort? *N Engl J Med* 2006; 354: 864-70.

43 Sanderson AAF. Pay-for-performance programs in the United Kingdom [letter]. *N Engl J Med* 2006; 355: 1832; author reply 1833.

44 Vijan S, Hofer TP, Hayward RA. Estimating benefits of glycemic control in microvascular complications in type 2 diabetes. *Ann Intern Med* 1997; 127: 788-95.

45 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: 2098-105.

46 Turenne MN, Hirth RA, Pan Q, et al. Using knowledge of multiple levels of variation in care to target performance incentives to providers. *Med Care* 2008; 46: 120-6.

47 MedVantage Inc. Provider pay-for-performance incentive programs: 2004 National Study Results, 2005. Available at: http://www.medvantage.com/resources.htm (accessed Dec 2006).

48 Roski J, Jeddeloh 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: 291-99.

49 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: 443-51.

50 Trude S, Au M, Christianson JB. Health plan pay-for-performance strategies. *Am J Manag Care* 2006; 12: 537-42.

51 Teleki SS, Damberg CL, Pham C, Berry SH. Will financial incentives stimulate quality improvement? Reactions from frontline physicians. *Am J Med Qual* 2006; 21: 367-74.

52 Queensland Health. Clinical Networks Policy, February 2006. Available at: http://www.health.qld.gov.au/cpic/documents/Networks_policy.pdf (accessed Jan 2007).

53 Southon G, Perkins R, Galler D. Networks: a key to the future of health services. *Aust Health Rev* 2005; 29: 317-326. Available at: http://www.aushealthreview.com.au/publications/articles/issues/ahr_29_3_0805/ahr_29_3_317.asp (accessed Sep 2008).

54 Edwards N. Clinical networks. *BMJ* 2002; 324: 63.

55 Proenca EJ, Rosko MD, Dismuke CE. Service collaboration and hospital cost performance. Direct and moderating effects. *Med Care* 2005; 43: 1250-8.

56 Fleury MJ. Integrated service networks: the Quebec case. *Health Serv Manage Res* 2006; 19: 153-65.

(Received 30/01/07, revised 12/02/08, accepted 3/08/08) □