



Should we expand the use of pay-for-performance in health care?

Dr Anne-marie Boxall
Social Policy Section

Executive summary

- Governments play a vital role in driving quality improvements in health care. They use various means, but publicly reporting performance data and linking performance and financial incentives—pay-for-performance (P4P)—are two common ones.
- P4P is used extensively in health care in some countries, for example the US and UK, even though it is controversial. There are only a few P4P schemes currently operating in Australia, but there have been recent proposals to expand its use.
- This paper examines the evidence on the impact of P4P schemes in health care. While the evidence is limited and yields mixed results, some experts argue that this should not spell the end for P4P. Instead, they suggest that policy-makers proceed with caution and use the existing evidence to highlight the challenges commonly associated with the use of P4P in health care. Some of the most important challenges are designing a system with:
 - substantial rewards and targets that are reasonably difficult to achieve;
 - minimal unintended consequences and incentives to ‘game’ the system, and;
 - incentives for both the best and worst performers to improve the quality of health care.
- The paper also outlines recommendations made by Professor Ian Scott on the necessary design features for successful P4P schemes in the Australian context.

Contents

Contents	2
Introduction	1
Pay for performance in health care in Australia	3
Strategies for improving the quality of health care without P4P	5
Can P4P schemes improve the quality of health care?	10
Expanding P4P in Australia: what are the most important lessons from overseas?	11
Designing incentives	11
Size of incentives	12
Difficulty of achieving targets	12
Unintended consequences	13
Gaming the system	13
Influencing the best and worst performers.	15
Prospects for Expanding P4P in Australia	16
Conclusion	17
Appendix 1: Recommendations from Professor Ian Scott for Australian P4P programs	18

Introduction

Pay-for-performance (P4P) schemes are increasingly being used in Australia and overseas as a way of driving quality improvement in health care. In essence, P4P schemes grant financial rewards to providers that meet specified quality improvement benchmarks.¹ P4P schemes can take various forms but examples include pay bonuses for individual health care professionals (performance pay), as well as funding bonuses or penalties for providers, such as general practices and hospitals.

P4P schemes are the most contentious of the range of measures used to drive quality improvement in health because they link performance and funding. Other measures, such as routine performance monitoring and public reporting of performance data, are still contentious but less so. Ongoing performance monitoring aims to improve the quality of care by giving providers timely and accurate feedback that can help them change their practices in problem areas. Public performance reporting is a mechanism for driving quality improvements in health care that does not link performance to remuneration. It is thought to improve quality by: helping consumers, or patients, shop around for the best health care provider or health service; giving consumers more confidence and power in negotiating their way through the health system; helping purchasers of health services to make better and more targeted choices about the services they purchase; and motivating providers and health care services to improve because of the kudos of receiving favourable reports.²

While P4P schemes are contentious in health care, they are used frequently in fields outside health.³ Their origins are in the ‘new public management’ approach to public administration that emerged in many developed countries during the 1990s.⁴ This approach saw a shift away from the more bureaucratic and hierarchical approach where it was assumed that the government should directly provide goods and services and that there was ‘one-best-way of working’.⁵ Under the new public management approach, more flexible management practices used in the private sector were adopted, better accountability mechanisms were put in place, and governments began to accept that the bureaucracy was not the only way to deliver goods and services.

-
1. A Scott, ‘Pay for performance programs in Australia: a need for guiding principles’, *Australian Health Review*, vol. 32, no. 4, 2008, pp. 740–749.
 2. R Hamblin and J Ganesh, ‘Measure for measure: using outcome measures to raise standards in the NHS’, 2007, viewed 16 September 2009, <http://www.policyexchange.org.uk/publications/publication.cgi?id=4>
 3. MB Rosenthal and RG Frank, ‘What is the empirical basis for paying for quality in health care’, *Medical Care and Research Review*, vol. 62, no. 2, 2006, pp. 135–57.
 4. AC Hyde, ‘Pay for performance’, *Public Manager*, vol. 34, no. 1, 2005, pp. 3–10.
 5. OE Hughes, *Public management and administration: an introduction* (third edition), 2003, Palgrave Macmillan, New York, pp.1–2.

Currently, P4P is used extensively in business both in Australia and internationally. It is common to use performance pay to partially remunerate company directors. In some countries, such as the US, UK and Australia, performance pay is also used in the public sector.⁶ Senior officials in the Commonwealth public service, for instance, can receive bonuses for reaching performance benchmarks. And recently, there has also been mounting pressure to introduce a performance pay system for public sector teachers.⁷

In other countries, P4P is also used routinely in health care. Some P4P schemes use incentive payments to encourage individual health care providers to achieve specified performance benchmarks, for example, screening a set proportion of eligible patients for diabetes. Some make bonus payments to providers that attain specified outcomes, for example budget savings. And in some schemes, penalties instead of rewards are imposed when providers fail to meet quality benchmarks. One example of such a scheme is refusing to pay providers when there are adverse health outcomes such as high levels of re-admission to hospital following discharge.⁸

So far, the United States (US) has been the most enthusiastic proponent of P4P in health care. In the public sector, the Centres for Medicare and Medicaid operate a number of P4P schemes including: the Hospital Quality Alliance Program, which provides financial rewards to hospitals that provide public reports on quality standards; the Physician Group Practice Demonstration program, which allows general practitioners to share in savings made under the Medicare program, and; the Premier Hospital Quality Incentive Demonstration program, where the best performing hospitals receive financial rewards and the poor performers the financial penalties.⁹ Many Health Maintenance Organisations (HMOs) operating in the

6. Hyde, 'Pay for performance', 2005.

7. Ministerial Council on Education, Employment and Youth Affairs, *Rewarding quality teaching: research report*, viewed 16 September 2009, http://www.mceecdya.edu.au/verve/resources/Rewarding_Teacher_Quality-Final_Report_-_GDA_2009.pdf

8. MB Rosenthal, 'Beyond pay for performance – emerging models of provider payment reform', *New England Journal of Medicine*, vol. 359, no. 12, 2008, pp. 1197–2000.

9. US Department of Health and Human Services, Centres for Medicare and Medicaid Services, 'Medicare demonstrations show paying for quality health care pays off', media release, 17 August 2009, viewed 16 September 2009, <http://www.cms.hhs.gov/apps/media/press/release.asp?Counter=3495&intNumPerPage=10&checkDate=&checkKey=&srchType=1&numDays=3500&srchOpt=0&srchData=&keywordType=All&chkNewsType=1%2C+2%2C+3%2C+4%2C+5&intPage=&showAll=&pYear=&year=&desc=&cboOrder=date> ; US Department of Health and Human Services, Centres for Medicare and Medicaid Services, 'Hospital Quality Alliance', Centres for Medicare and Medicaid website, viewed 16 September 2009, http://www.cms.hhs.gov/HospitalQualityInits/33_HospitalQualityAlliance.asp;

private sector also use P4P schemes as part of payment contracts with health care providers.¹⁰ However some, Kaiser Permanente for example, have begun to scale back their use of performance based financial incentives and now focus instead on transparent and systematic performance feedback.¹¹

In recent years, the United Kingdom (UK) has also embraced the use of P4P in health care. The most substantial scheme currently operating is part of the National Quality and Outcomes Framework for GPs, which was introduced in 2004. Under this scheme, GPs are eligible to receive financial rewards that amount to approximately 25% of total income if they meet certain quality benchmarks for patient care.¹² P4P schemes known as payment by result are also increasingly being incorporated into hospital purchasing contracts.¹³

This paper has several purposes. It provides an overview of P4P in the health sector overseas and in Australia. It examines recent proposals to expand P4P in Australia, as well as the existing evidence on alternative mechanisms for improving the quality of health care, such as public performance reporting and ongoing performance monitoring. It reviews the available international evidence on P4P schemes and identifies the key lessons for Australian policymakers. And, it summarises recommendations on the best design for P4P schemes in Australia, made previously by Queensland academic and clinician, Professor Ian Scott.¹⁴

Pay for performance in health care in Australia

There are currently only a few P4P schemes operating in Australia. Details are outlined in Table 1 below.

-
10. MB Rosenthal, BE Landon, ST Normand, RG Frank, AM Epstein, 'Pay for performance in commercial HMOs', *New England Journal of Medicine*, vol. 355, no. 18, 2006, pp. 1895–1902.
 11. Editorial, 'Pay for performance in primary care: proceed with caution, pitfalls ahead', *Healthcare Policy*, vol. 4, no. 1, 2008, pp10–15.
 12. UK Department of Health, 'Investing in general practice: the new general medical services contract', 2003, viewed 16 September 2009, http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4071966
 13. UK Department of Health, 'Payment by results', 2009, viewed 16 September 2009, <http://www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/NHSFinancialReforms/index.htm>
 14. Scott, 'Pay for performance programs in Australia: a need for guiding principles', *Australian Health Review*, 2008; see Appendix 1 for details of Professor Scott's recommendations for Australian P4P programs.

Table 1: P4P schemes operating in Australia

P4P scheme	Commencement date	Details
General Practice Immunisation Incentive Scheme (GPII) ¹⁵	1997	<ul style="list-style-type: none"> Rewards GPs that promote, monitor and provide childhood immunisation services with bonus payments
Practice Incentives Program (PIP) ¹⁶	1998	<ul style="list-style-type: none"> Operated by Medicare Australia for GPs Provides bonus payments to practices that meet or are working towards meeting practice accreditation standards
P4P scheme for hospitals ¹⁷	2006	<ul style="list-style-type: none"> Run through the Commonwealth Department of Veterans' Affairs Rewards hospitals providing high quality care to veterans Focuses on surgical outcomes, patient safety and satisfaction, and the management of chronic diseases
Clinical Practice Improvement Payment (CPIP) ¹⁸	2008	<ul style="list-style-type: none"> Run by Queensland Health Pays clinical units within public hospitals incentives for meeting specified clinical benchmarks
Council of Australian Governments (COAG)	Negotiated in 2008 and came into effect in	<ul style="list-style-type: none"> National Partnership Agreement on Preventive Health includes bonus

15. Australian Government, Medicare Australia, 'General Practice Immunisation Incentives Scheme', viewed 16 September 2009, <http://www.medicareaustralia.gov.au/provider/incentives/gpii/index.jsp>

16. Australian Government, Medicare Australia, 'Practice Incentives Program', viewed 16 September 2009, <http://www.medicareaustralia.gov.au/provider/incentives/pip/index.jsp>

17. Australian Government, Department of Veterans' Affairs, *Information for Tier 1 hospitals: veteran partnering information package*, February 2009, viewed 17 September 2009, http://74.125.153.132/search?q=cache:YLXmdhbeQdAJ:www.dva.gov.au/service_providers/hospitals/veteran_partnering/Documents/HospitalInfoGuideFebruary2009.pdf+Information+for+Tier+1+Hospitals:+Veteran+Partnering+Information+Package&cd=1&hl=en&ct=clnk&gl=au

18. M Ward, SA Daniels, GJ Walker, S Duckett, 'Connecting funds with quality outcomes in health care: a blueprint for a Clinical Practice Improvement Payment', *Australian Health Review*, vol. 31, Supp. 1 2007, pp. S54–S58.

P4P scheme	Commencement date	Details
National Partnership Agreements ¹⁹	January 2009	payments of 50% of total payments for states and territories that achieve specified performance benchmarks in prevention and health promotion <ul style="list-style-type: none">• The National Partnership Agreement on Hospital and Health Workforce Reform indicated that incentive payments would be developed in the future and used in conjunction with other hospital financing methods

In recent years, several proposals have been made to expand the use of P4P in Australia. The NSW Independent Pricing and Regulation Authority (IPART) recommended trialling a P4P scheme in NSW similar to Queensland's CPIP scheme (outlined in Table 1 above) in 2008.²⁰ And in July 2009, the National Health and Hospital Reform Commission proposed expanding the use of P4P in Australia.²¹ It suggested using financial incentives to reward health services that provide safe and timely access to care, and aged care providers that increased their use of e-health technologies.

Strategies for improving the quality of health care without P4P

Before examining the evidence on P4P schemes it is worth considering whether quality improvement in health could be achieved through the less costly approach of public performance reporting and ongoing performance monitoring.

The most common way of publicly reporting on the performance of various health care providers is to publish league tables or scorecards. They rank the performance using a range of indicators, such as mortality, hospital re-admission or infection rates.

Not long after the Labor Government came to power in November 2007, the Federal Minister for Health and Ageing, Nicola Roxon, began advocating a system for publicly reporting hospital performance data. Few details of the scheme were disclosed, but Roxon explained

19. Council of Australian Governments, *National Partnership Agreement on Hospital and Health Workforce Reform*, December 2008, viewed 16 September 2009, http://www.coag.gov.au/intergov_agreements/federal_financial_relations/docs/national_partnership/national_partnership_on_hospital_and_health_workforce_reform.pdf
20. Independent Pricing and Regulatory Tribunal NSW, *Framework for performance improvement in health: Other industries*, Final Report, September 2008.
21. National Health and Hospitals Reform Commission (NHHRC), *A healthier future for all Australians: final report*, NHHRC, Barton, ACT, 2009, viewed 16 September 2009, <http://www.nhhrc.org.au/internet/nhhrc/publishing.nsf/Content/nhhrc-report>

that the government was keen to make sure that ‘accountability targets’ were publicly available because it would mean the public would ‘know if there [were] very high infection rates from hospitals...[or] if there [were] more adverse events at a particular hospital than another’.²² She went on to say that ‘everyone understands that it’s in health service providers’ interests, and patients and consumers interests, to be able to compare across the whole system’.²³

At the time, the Government was unable to proceed with its plans to implement a national system for publicly reporting on hospital performance data. Some states, NSW in particular, strongly objected to the proposal. Reba Meagher, Minister for Health in NSW at the time, explained that her state had strong reservations because the system could be used to construct league tables that did not accurately reflect the quality of care delivered in hospitals.²⁴

One of the main objections critics raise about league tables is that they do not adequately take into consideration the complexity of cases treated or patient demographics (there is a strong link between socio-economic status and health outcomes for instance). A similar debate has erupted in the education sector after Julia Gillard, the Federal Minister for Education, outlined plans to publish schools’ performance on a range of national benchmarks in literacy and numeracy. Critics argued that public reporting would lead to the publication of crude league tables that ranked vastly different schools against one another.²⁵ Newspaper reports suggest that this has already occurred in Tasmania and Queensland.²⁶

-
22. N Roxon (Minister for Health and Ageing), ABC Radio, transcript, ABC, 1 December 2008, viewed 16 September 2009, <http://parlinfo.parlInfo/search/display/display.w3p;adv=yes;db=;group=;holdingType=;id=;orderBy=customrank;page=0;query=AuthorSpeakerReporter%3ARoxon%20Date%3A01%2F12%2F2008%20Dataset%3Aemms,radioprmtvprog;querytype=;rec=2;resCount=Default>
 23. Roxon, ABC Radio, 1 December 2008.
 24. ABC Radio, PM, transcript, ABC, 31 January 2008, viewed 16 September 2009, <http://parlinfo.parlInfo/search/display/display.w3p;adv=yes;db=;group=;holdingType=;id=;orderBy=customrank;page=0;query=AuthorSpeakerReporter%3ARoxon%20Date%3A31%2F01%2F2008%20Dataset%3Aemms,radioprmtvprog;querytype=;rec=0;resCount=Default>
 25. E Macdonald, ‘Principals hit out at school league tables’, *The Canberra Times*, 8 May, 2009, viewed 30 September 2009, <http://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;query=Id%3A%22media%2Fpressclp%2F7OHT6%22>
 26. C Bonner, ‘Using league tables to keep score of schools entrenches division’, *The Age*, 29 May 2009, viewed 30 September 2009, <http://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;query=Id%3A%22media%2Fpressclp%2FA7PT6%22>

League tables ranking hospitals and primary care trusts have been used in the UK since the early 1990s, but they were controversial initially.²⁷ Those in favour of league tables claimed they increase accountability, and stimulate providers to improve the quality of care.²⁸ However opponents argued that health care providers are generally hostile towards them, that they have a negative impact on public trust and professional morale, are harmful because the media overreact to low scores, and providers usually find ways to ‘stack the decks’ by selecting the healthiest patients to improve their scores.²⁹

The available evidence on the effectiveness of league tables and scorecards provides little certainty for policymakers considering whether or not to use them. According to some UK experts, most of the evidence on league tables suggests they do not benefit the public. They claim the public does not tend to seek out information published in league tables, does not understand, trust, or use it.³⁰ These experts also point out that data from focus groups in the UK suggests that some people think public reporting is ‘a punitive tool used by politicians to punish hard-working professionals’.³¹

According to research from the US, it is not just the public that are sceptical about league tables.³² Physicians also distrust league tables and put considerable effort into discrediting them. This was certainly the case in Britain when league tables were first introduced. Some members of the British Medical Association (BMA) argued that the proposed league tables were ‘perverse’, ‘suitable for sport but not for hospitals’, and were ‘actually harmful because they were diverting resources away from patient care’.³³ Other doctors were concerned that

-
27. UK Department of Health, *The NHS Performance Guide 1995–96*, viewed 16 September 2009, <http://www.performance.doh.gov.uk/tables96.htm>
 28. MN Marshall, PG Shekelle, HTO Davies, PC Smith, ‘Public reporting on quality in the United States and the United Kingdom’, *Health Affairs*, vol. 22, no. 3, 2003, pp. 134–148; RM Werner and DA Asch, ‘The unintended consequences of publicly reporting quality information’, *JAMA*, vol. 293, no. 10, 2005, pp. 1239–1244.
 29. P Adab, AM Rouse, MA Mohammed, T Marshall, ‘Performance league tables: the NHS deserves better’, *British Medical Journal*, vol. 324, 12 January, 2002, pp. 95–98; Marshall, Shekelle, Davies, Smith, ‘Public reporting on quality in the United States and the United Kingdom’, 2003.
 30. Marshall, Shekelle, Davies, Smith, ‘Public reporting on quality in the United States and the United Kingdom’, 2003.
 31. Marshall, Shekelle, Davies and Smith, ‘Public reporting on quality in the United States and the United Kingdom’, 2003, p. 142.
 32. J Kaplan, ‘Managed care report cards: evaluating those who evaluate physicians’, *Managed Care Interface*, vol. 13, 2000, pp. 88–94 cited in Marshall, Shekelle, Davies and Smith, ‘Public reporting on quality in the United States and the United Kingdom’, 2003.
 33. ‘Views and reviews’, *British Medical Journal*, vol. 309, 16 July, 1994, p. 200.

performance information might be incorrectly attributed to an individual thereby ‘potentially blighting their career’.³⁴

After league tables were implemented in the UK, a representative of the BMA stated that the organisation supported ‘measures intended to improve outcomes in hospital care’ but warned against the dangers of ‘simply publishing hospital death rates’. He went on to explain that league tables were not the best method of driving quality improvement because:

... crude mortality figures [were] grossly misleading as they are dependent on many factors. When these are taken into account the results become so complicated as to be almost incomprehensible, even to experts in that field.³⁵

While league tables and scorecards might not be popular in some circles, there is some evidence that they can improve the quality of care. A major review of the evidence published in 2008 concluded that public performance reporting could stimulate hospitals to improve the quality of care they provide. It cited several studies where this occurred.³⁶ Some studies in the US, for example, found that publicly releasing information on mortality rates following heart surgery caused hospitals with poor results to make changes that improved the quality of care.³⁷ Others, however, did not find that report cards led to quality improvements, but the evidence base in this field overall is quite limited.³⁸

Because the evidence to support publishing performance data, especially in the form of league tables, is limited and resistance from some in the health sector is strong, some health experts have suggested alternative methods of providing performance feedback to health care

-
34. G Clews, ‘Proposal to release doctor errors slated’, *BMA news*, 2005, January 5, p. 2.
 35. British Medical Association, Wales, *BMA Wales warns against crude mortality figures*, media release, 9 June 1998.
 36. CH Fung, Y-W Lin, S Mattke, C Damberg, PG Shekelle, ‘Systematic review: The evidence that publishing patient care performance data improves quality of care’, *Annals of Internal Medicine*, vol. 148, 2008, pp. 111–123.
 37. All studies cited in Fung, Lin, Mattke, Damberg, Shekelle, ‘Systematic review: The evidence that publishing patient care performance data improves quality of care’, 2008.
 38. See for example: WA Ghali, As Ash, RE Hall, MA Moskowitz, ‘Statewide quality improvement initiatives in and mortality after cardiac surgery’, *Journal of the American Medical Association*, 1997, vol. 277, pp. 379–382; M Moscussi, KA Eagle, D share, D Smith, AC De Franco, M O’Donnell et al, ‘Public reporting and case selection for percutaneous coronary interventions: an analysis from two large multicentre percutaneous coronary intervention databases’, *Journal of the American College of Cardiology*, 2005, vol. 45, pp. 1759–1765.

providers.³⁹ The BMA has developed a model where ‘control charts’ are used to monitor the safety and quality of care.⁴⁰

Control charts work by plotting data on expected outcomes for certain performance indicators against acceptable levels of variation and actual performance. One of the advantages of this approach is that it is easy for those responsible for monitoring performance to identify any outliers – that is unexpectedly high or low performers. These organisations are then required to investigate the reasons for poor performance and implement strategies to remedy problems. Because this form of performance reporting and monitoring is done routinely, it avoids many of the problems outlined earlier that are often associated with publishing league tables or scorecards.

According to the BMA, control charts also have other benefits.⁴¹ They are easier to interpret and more intuitive than league tables. Because they do not publicly report or rank performance, providers performing within acceptable standards are not penalised even if they are not at the top of the rankings. And, it is easier to learn from top performers using control charts. Providers exceeding expected performance standards are identified and the reason for their superior performance is identified. This may be difficult, at least initially, but it should be possible for other providers to adopt successful methods and procedures if they are suitable.

The Queensland Department of Health has already adopted a performance monitoring system along the lines of the one advocated by the BMA. In 2003, it began using the Variable Adjusted Life Year Display (VLAD) to monitor the quality of care in public and private hospitals in Queensland.⁴² The VLAD enables hospitals to see a graphical overview of their performance on a range of quality measures over time. The system flags differences between actual and expected performance and alerts those responsible for performance monitoring to investigate further at local level. In the first instance, it is important to ascertain whether flags are created because of poor data coding, or changes in demographical or practice patterns. After accounting for these factors, it is possible to identify potentially substandard systems of care or professional incompetence.

For some, the debate over different models of performance monitoring and reporting is peripheral because they believe it is most effective when used in conjunction with P4P schemes. When a group of American researchers tested this theory (using multivariate

39. Werner and Asch, ‘The unintended consequences of publicly reporting quality information’, 2005.

40. Adab, Rouse, Mohammed, Marshall, ‘Performance league tables: the NHS deserves better’, 2002.

41. Adab, Rouse, Mohammed, Marshall, ‘Performance league tables: the NHS deserves better’, 2002.

42. Queensland Government, ‘Quality and safety, Variable Life Adjusted Display (VLAD)’, viewed 16 September 2009, <http://www.health.qld.gov.au/quality/vlad.asp>

modelling), they found that P4P schemes were capable of stimulating quality improvement even in hospitals already monitoring and publicly reporting performance data.⁴³ Other researchers, however, have found the opposite.⁴⁴ The following section looks more broadly and reviews the evidence on the effectiveness of P4P schemes.

Can P4P schemes improve the quality of health care?

P4P schemes are used widely throughout the US and UK (the main schemes were outlined previously). There has been limited research on their effectiveness despite their widespread use, and the evidence that does exist shows mixed results.

The highest quality evidence in research comes from systematic reviews of published, peer-reviewed research studies in any field. There have been two systematic reviews of the research on P4P schemes in recent years that provide some useful insights, and another review that examines evidence from ‘real world’ P4P initiatives.⁴⁵ The first of these was published in 2006 and examined studies of P4P schemes operating across a wide range of health care areas, for example hospitals, general practice and health promotion.⁴⁶ The authors found that overall there was a lack of compelling evidence that P4P was effective. This was partly because there were so few studies done in each specific field of health care, but results were often mixed even within the one field. Despite this, the authors of the review concluded that there was some evidence that P4P schemes had a positive impact on the quality of care, but they were also keen to emphasise the need to monitor and prevent the many unintended adverse consequences seen in many P4P schemes (the following section deals with these in more detail).

-
43. PK Lindenauer, D Remus, S Roman, MB Rothberg, EM Benjamin, A Ma, DW Bratzler, ‘Public reporting and pay for performance in hospital quality improvement’, *New England Journal of Medicine*, vol. 356, no. 5, 2007, pp. 486–496.
 44. SW Glickman, F-S Ou, ER DeLong, MT Roe, BL Lytle, J Mulgund et al, ‘Pay for performance, quality of care, and outcomes in acute myocardial infarction’, *JAMA*, vol. 297, no. 21, 2007, pp. 2373–2380.
 45. LA Petersen, LD Woodard, T Urech, C Daw, S Sookanan, ‘Does pay for performance improve the quality of care?’, *Annals of Internal Medicine*, vol. 145, no. 4, 2006, pp. 265–272; A Mehrotra, CL Damberg, MES Sorbero, SS Teleki, ‘Pay for performance in the hospital setting: what is the state of the evidence?’, *American Journal of Medical Quality*, vol. 24, no. 19, 2009, pp. 19–28; JB Christianson, S Leatherman, K Sutherland, ‘Lessons for evaluations of purchaser Pay-for-Performance programs: a review of the evidence’, *Medical Care Research and Review*, 2008, vol. 65, no. 6, pp. 5S–35S.
 46. LA Petersen, LD Woodard, T Urech, C Daw, S Sookanan, ‘Does pay for performance improve the quality of care?’ 2006.

The most recent systematic review, which was published in late 2008, examined the impact of P4P schemes in hospitals.⁴⁷ It too commented on the lack of research in this area and poor quality of many studies. As a result, the authors focused on the three most rigorous studies of hospital P4P schemes; all of them evaluated the US Centres for Medicare and Medicaid Services Premier Hospital Quality Incentive Demonstration project. Two found that hospitals participating in P4P schemes had marginally better quality improvement than hospitals outside the scheme.⁴⁸ The other study found that P4P did not improve patient outcomes (mortality rates following a heart attack were the outcome of interest).⁴⁹ Like their colleagues, the authors of this review argued strenuously that more research was needed in this area, particularly because it was not clear whether the benefits of P4P outweighed the costs.

Although the final review, published in late 2008, included a range of studies other than randomised trials, it came to similar conclusions.⁵⁰ The authors found that most studies reported quality improvements after P4P schemes were introduced. However, in most studies it was impossible to attribute improvements to the P4P scheme alone because it was usually implemented along with other quality improvement initiatives.

While the available evidence does not provide conclusive evidence in favour of P4P schemes, it suggests that these schemes could be effective if some of the problems with them are addressed. The following section outlines the main issues policymakers would need to address if governments decided to expand the use of P4P schemes in Australia.

Expanding P4P in Australia: what are the most important lessons from overseas?

Designing incentives

One of the most important factors determining the ‘success’ with P4P schemes is the nature of the incentive itself. If incentives are to motivate behavioural change, health care providers must be aware of the incentive scheme, they must perceive it to be significant, and the

47. A Mehrotra, CL Damberg, MES Sorbero, SS Teleki, ‘Pay for performance in the hospital setting: what is the state of the evidence?’, 2009.

48. SR Grossbart, ‘What’s the return?: Assessing the effect of pay for performance initiatives on the quality of care delivery’, *Medical Care Research and Review*, vol. 63, no. 1, 2006, pp. 29S–49S; Lindenauer, Remus, Roman, Rothberg, Benjamin, Ma, Bratzler, ‘Public reporting and pay for performance in hospital quality improvement’, 2007;

49. Glickman, Ou, DeLong, MT Roe, BL Lytle, J Mulgund et al, ‘Pay for performance, quality of care, and outcomes in acute myocardial infarction’, 2007.

50. JB Christianson, S Leatherman, K Sutherland, ‘Lessons for evaluations of purchaser Pay-for-Performance programs: a review of the evidence’, 2008.

performance target or benchmark must be perceived as being achievable although not without some additional effort. This has not been the case with all P4P schemes.

Size of incentives

In one US P4P scheme, token rewards were given to physicians as an incentive to increase mammography referral rates.⁵¹ Rewards of approximately \$50 were provided if physicians reached a 50% referral rate for eligible patients. Researchers examined whether implementing a P4P scheme offering token rewards was more effective than simply educating physicians about the importance of making referrals and putting reminders into patient medical records. They found that token rewards were not as effective as education and reminders. As a result, they argued that financial rewards must be a significant part of the overall remuneration arrangements if they are to influence physicians' behaviour. Subsequent studies have confirmed these findings.⁵²

Difficulty of achieving targets

Offering large financial rewards does not necessarily guarantee success of a P4P scheme either. It is just as important to set appropriate performance targets or benchmarks and monitor them over time. Policymakers in the UK learnt this important lesson after implementing the Quality and Outcomes Framework (QOF) for general practices in 2004. This P4P scheme is a voluntary program providing financial incentives to practices that meet a range of clinical care, organisational, patient experience and service performance targets. Almost all GPs in England participate in the UK P4P scheme, and financial rewards representing up to 25% of total income are available if targets are met.

In the first year of operation, incentive payments made under the QOF vastly exceeded the budget allocation because the average GP practice scored 91% of the maximal quality points, a much higher rate than anticipated.⁵³ In subsequent years, fewer quality improvement gains were made, which suggests that the impact of the scheme had diminished over time. Analysts have argued that this might be because in the first year, GPs made all the improvements that were easy to make.⁵⁴ Subsequent improvements were much more difficult to achieve. In

51. KE Grady, JP Lemkau, NR Lee, C Caddell, 'Enhancing mammography referral in primary care', *Preventive Medicine*, vol. 26, 1997, pp. 791–800.

52. AL Hillman, K Ripley, N Goldfarb, I Nuamah, J Weiner, E Lusk, 'Physician financial incentives and feedback: failure to increase cancer screening in Medicaid managed care', *American Journal of Public Health*, vol. 88, no. 11, 1998, pp. 1699–1701; AL Hillman, K Ripley, N Goldfarb, I Nuamah, J Weiner, E Lusk, 'The use of physicians financial incentives and feedback to improve pediatric preventive care in Medicaid managed care', *Pediatrics*, vol. 104, no. 4, 1999, pp. 931–935.

53. Editorial, 'Pay for performance in primary care: proceed with caution, pitfalls ahead', 2008.

depth analyses confirmed that much of the incentive to improve was removed after the first year as most GPs had already received almost maximal incentive payments.

Unintended consequences

Perhaps the most concerning consequence of the QOF P4P scheme was that the quality of care for some non-targeted aspects of care declined after it was implemented.⁵⁵ This was particularly evident in the area of heart disease. Once the scheme began, quality scores declined for non-targeted aspects of care but increased for targeted aspects of care. A similar effect was seen on performance measures relating to asthma management but it did not occur until the scheme had been operating for a year. Researchers examining the impact of the QOF point out that although the gap in quality of care that resulted was unintentional, it was particularly concerning because quality was already lower in areas where there were no incentives.

The unintended consequences of rewards systems have been well documented in other fields. Those working in management and business were made aware of the problem in 1975 when an American academic, Steven Kerr, published his now classic piece 'On the folly of rewarding A, while hoping for B'.⁵⁶ In it, he argues that:

Whether dealing with monkeys, rats, or human beings, it is hardly controversial to state that most organisms seek information concerning what activities are rewarded, and then seek to do (or at least pretend to do) those things, often to the virtual exclusion of activities not rewarded.

Kerr explained that 'fouled up' reward systems operate in many areas – he provides examples from politics, academia, business consulting, and sport – because people tend to have a fascination with objective criteria, put too much emphasis on highly visible behaviours, or simply value other factors more highly. Kerr does not argue against reward systems but instead urges managers to thoroughly investigate which behaviours are actually being rewarded. Sometimes, the system in place rewards behaviours other than the desired ones.

Gaming the system

Another related problem seen often in P4P schemes is 'gaming'. It can take various forms. One common one is to neglect all problems or issues other than the ones being targeted.

54. SM Campbell, D Reeves, E Kontopantelis, B Sibbald, M Roland, 'Effects of pay for performance on the quality of primary care in England', *New England Journal of Medicine*, vol. 361, no. 4, 2009, pp. 368–378.

55. Campbell, Reeves, Kontopantelis, Sibbald, Roland, 'Effects of pay for performance on the quality of primary care in England', 2009.

56. S Kerr, 'An academic classic: On the folly of rewarding A, while hoping for B', *The Academy of Management Executive*, vol. 9, no. 1, 1995, pp. 7–14.

When this is intentional, it is considered gaming rather than an unintended consequence of P4P. Providers can game the system in a range of ways, but one relatively inconspicuous way is to ‘cherry-pick’ patients. This occurs when providers give preferential treatment or access to less complex patients because it is easier to reach performance benchmarks.

There is evidence of cherry-picking occurring in the US with mental health and drug and alcohol services. Researchers found that when a P4P scheme was introduced where rewards were largely determined by patient outcomes, providers encouraged more complex patients to use other services.⁵⁷ In another study, researchers found that providers treated fewer patients with severe substance abuse problems after a new performance-based contract system was implemented that rewarded high overall performance scores.⁵⁸

Experts also suspect that cherry-picking occurs in the UK’s QOF.⁵⁹ It can be done if GPs label people with borderline health conditions as actual cases. They are then eligible for incentive payments when they successfully achieve the relevant benchmarks, which is obviously relatively easy to do in borderline cases. GPs can also exclude the data on outcomes for more complex patients. In the first year the scheme was operating, a small number of GPs had exceptionally high exclusion rates – one excluded 86% of all eligible patients – suggesting that the practice occurs even though it might not be widespread.

Manipulating data in order to meet targets is another form of gaming that occurs under P4P schemes. It also occurs in public performance reporting systems. There is evidence that the introduction of more stringent performance reporting requirements in Australia has promoted some hospitals to ‘fudge the figures’ in order to meet benchmarks. According to the Victorian Auditor-General, at least one hospital had been manipulating waiting list data in order to hide the hospitals true performance.⁶⁰ There have been reports that cheating, dishonesty and ‘fudging the figures’ in other Australian states and in the UK’s NHS.⁶¹

-
57. Y Shen, ‘Selection incentives in a performance-based contracting system’, *Health Services Research*, vol. 38, no. 2, 2003, pp. 535–552.
 58. Shen, ‘Selection incentives in a performance-based contracting system’, *Health Services Research*, 2003.
 59. Editorial, ‘Pay for performance in primary care: proceed with caution, pitfalls ahead’, 2008.
 60. Victorian Auditor General, ‘Access to public hospitals: measuring performance’, PP no. 180, session 2008–09, viewed 16 September 2009, http://www.audit.vic.gov.au/reports_publications/reports_by_year/2009/20090401_hospital_indicators.aspx
 61. Editorial, ‘Reality catches up with Victoria’s Health Minister’, *The Age*, 3 April 2009; J Laurance, ‘Tricks and cheating rife to meet care targets, says BMA’, *The Guardian*, 1 July 2003.

A more subtle form of gaming seen frequently in P4P schemes is ‘hitting the target but missing the point’.⁶² According to UK health expert James Gubb, this happens when providers become so preoccupied with hitting targets that patient experiences and outcomes become secondary. Again, it is not restricted to P4P schemes but appears to be a consequence of any performance monitoring scheme. A recent example from NSW illustrates the point. In order to meet performance benchmarks, one hospital set up a new unit within the Emergency Department using the concept of ‘virtual beds’.⁶³ These beds were fictitious hospital beds not real ones. Employees were asked to admit patients to these virtual beds when there were no actual beds available because it meant they could still meet performance targets for waiting times. It made no tangible difference for patients because they were still waiting for treatment if they were admitted to a ‘virtual’ bed.

Influencing the best and worst performers

One of the other key challenges with P4P schemes is working out how to design a system that encourages the poorest as well as the best performers to improve the quality of care. There is evidence that in P4P schemes with fixed performance targets, the poorest performers make greatest improvements but the best performers receive the highest financial rewards.⁶⁴ Under these circumstances, the substantial cost of achieving quality gains is difficult to justify.

There are a number of reasons why P4P schemes struggle to improve quality in both the top and bottom performers.⁶⁵ It is difficult for high performing hospitals or health services to make gains because they tend to already have sophisticated methods for ensuring quality in place. They often have nursing and support staff, staff dedicated to performance data collection and monitoring, electronic health records and reminder systems. Financial rewards, therefore, are not needed to develop and implement quality improvement systems. Doctors working in top performing hospitals with P4P schemes also explained that for them, the desire to be the best – to provide the highest quality care – was the main motivating factor to improve the quality of care. While it is difficult to quantify, there is much anecdotal evidence to suggest that altruism rather than financial rewards is a strong motivating factor for doctors and other health professionals.

62. J Gubb and G Bevan, ‘Have targets done more harm than good in the English NHS?’, *British Medical Journal*, vol. 338, 21 February, 2009, pp. 442–443.

63. J Robotham, ‘Bullying, bottlenecks and death by a thousand paper cuts’, *Sydney Morning Herald*, 28 March 2009.

64. MB Rosenthal, RG Frank, Z Li, AM Epstein, ‘Early experience with pay for performance: from concept to practice’, *JAMA*, vol. 294, no. 14, 2005, pp. 1788–1793.

65. KM Sautter, BG Bokhour, B White, D Min, GJ Young, JF Burgess, D Berlowitz, JRC Wheeler, ‘The early experience of a hospital-based pay-for-performance program’, *Journal of Healthcare Management*, vol. 52, no. 2, 2007, pp. 95–108.

In poorer performing hospitals or health services, it is difficult to improve the quality of care using P4P schemes because they lack the necessary resources and infrastructure to take advantage of any financial rewards they do receive, which tend to be lower than those in high performing organisations anyway.⁶⁶ Small hospitals, for example, do not have the resources to monitor quality in the same way as large hospitals. They find it difficult to employ additional staff or purchase expensive technology that will help improve the quality of care. Smaller hospitals are also often located in rural or remote areas. They tend to service people from lower socio-economic backgrounds that also have poorer health outcomes.

In an effort to improve the reward structure for P4P schemes, researchers in the US trialled a scheme that rewarded or penalised GPs in direct proportion to their relative performance.⁶⁷ In this scheme, participating GPs put a set proportion of their income into a funding pool, which was then used to reward high quality performance. The main performance measure used to assess performance was compliance with guidelines for chronic disease management. GPs that performed poorly risked losing part or all of the money they had put into the pool. Those that performed well receive a proportion of the pooled funds, the amount determined by their relative performance.

While overall GPs performance increased after the P4P scheme was implemented, the rate of improvement was no greater than it was beforehand. Improvements in quality of care, therefore, were attributed to secular trends and not the P4P scheme. The researchers suggested a number of reasons for the lack of effect, all of them proffered reasons why other schemes have failed – the scheme had not been in place long enough to change behaviour, the rewards were not substantial enough, and GPs lacked the necessary infrastructure to make quality improvements.

Prospects for Expanding P4P in Australia

Policymakers seeking to justify the widespread implementation of P4P schemes in the health care sector based on the strength of the available evidence will find it a difficult task. The evidence that P4P works in other sectors, such as the business sector, is also underwhelming. For instance, a recent report on executive remuneration by the Productivity Commission found that the increasing complex financial reward systems put in place for executives were not necessarily the best means of stimulating them to improve company outcomes.⁶⁸

66. Sautter, Bokhour, White, Min, Young, Burgess, Berlowitz, Wheeler, 'The early experience of a hospital-based pay-for-performance program', *Journal of Healthcare Management*, 2007.

67. GJ Young, M Meterko, H Beckman, E Baker, B White, KM Sautter et al, 'Effects of paying physicians based on their relative performance for quality', *Journal of General Internal Medicine*, vol. 22, 2007, pp. 872–876.

68. Australian Government, Productivity Commission, *Executive remuneration in Australia: Productivity Commission discussion draft*, Canberra, Productivity Commission, 2009, viewed 30 September 2009, <http://www.pc.gov.au/projects/inquiry/executive-remuneration/draft>

According to some experts, however, the ‘evidence vacuum’ need not spell the end for P4P. After reviewing the international evidence on P4P schemes in health care, Queensland academic and clinician, Professor Ian Scott, argues that ‘P4P programs have the potential to improve quality of care’ but accepts that ‘they are not without potential to harm if poorly designed and implemented’.⁶⁹ Some British experts also agree that merits of P4P schemes need to be critically evaluated. Professor Brian Hutchison, editor of the *Healthcare Policy* journal, for example, argues that ‘depending on the context and design features, pay for performance can yield small gains at large cost’.⁷⁰ He advises policy-makers determined to proceed with P4P that ‘making haste slowly may be the best course of action’ and to consider introducing a phased and carefully monitored introduction.⁷¹

Conclusion

There are indications that the current federal government will put measures in place that encourage state and territory governments to expand the use of P4P in public hospitals. It is impossible to design and operate a perfect scheme, one that: pinpoints the most appropriate size of financial rewards; sets benchmarks that are difficult to achieve but still attainable; avoids any negative consequences for non-targeted aspects of care; totally prevents cherry-picking and gaming; and facilitates quality improvement in both the best and worst performers. There is, however, much to learn from international evidence on P4P schemes that will help Australian policymakers avoid some of the potential pitfalls. The principal lesson is that P4P is not a magic bullet. There would be substantial costs and risks involved with expanding the use of P4P that must be weighed up against the potential benefits in the quality and safety of health care. The potential benefits and risks of alternative systems for driving quality improvement, for example public performance reporting and routine, ongoing performance monitoring, also need to be given serious consideration. Ultimately though, the decision whether or not to expand the use of P4P in Australia will largely be a political one because there is not enough high-quality evidence in this field to guide decision-making.

69. Scott, ‘Pay for performance programs in Australia: a need for guiding principles’, *Australian Health Review*, 2008; In order to ensure any Australian P4P schemes takes heed of international evidence and experience, Professor Scott makes a number of recommendations on how to best design and implement them. They are too numerous to list here, but some of the main ones are outlined in Appendix 1. More detail on Scott’s recommendations can be found in the publication cited above. While many of these recommendations need to be developed further, they provide a useful framework, informed by evidence, which could be used to guide policy development in this area. It is beyond the scope of this paper critically evaluate each of these recommendations.

70. Editorial, ‘Pay for Performance in Primary Care: Proceed with Caution, Pitfalls Ahead’, 2008.

71. Editorial, ‘Pay for Performance in Primary Care: Proceed with Caution, Pitfalls Ahead’, 2008, pp. 13–14.

Appendix 1: Recommendations from Professor Ian Scott for Australian P4P programs

Design

1. The rationale, objectives and design of any P4P schemes should be clearly articulated
2. Capital investment requirements, for example information technology, staff training and the capacity to deliver additional capacity, need to be identified in advance
3. Lead clinicians and managers should be involved in the design, implementation and evaluation of schemes
4. Schemes should be pilot tested before widespread roll-out
5. The Australian Institute of Health and Welfare and the National Institute of Clinical Studies should collaborate to build a profile of areas of suboptimal care where P4P could be used to improve performance

Performance measures

1. Performance measure should pertain to key areas of care where there is potential to improve quality
2. Process of care measures that are flexible and evidence-based must be developed and rigorously tested
3. Outcome measures should be risk-adjusted for patient level factors (for example the severity of illness and number of other co-morbid conditions)
4. More weight should be given to process of care measures (for example the number of diabetic patients that have their feet checked each year) or closely related intermediary outcome measures (for example the number of diabetic patients that sustain target blood glucose levels)
5. Performance measures should apply nationally and be developed with the help of experts from professional associations and organisations

Evaluating performance data

1. Performance data should be collected in a way that is scientifically valid and subject to periodic external audit
2. Performance data should include routinely collected administrative data (for outcome measures) and clinical data from medical records (for process of care measures)
3. Physicians should be allowed to review, correct and supplement data, especially administrative data, before it is publicly released or used to determine financial rewards
4. Performance data used to determine rewards should be collected over a significant period of time
5. Automated systems that already exist (for example pharmacy databases and clinical registries) should be used for data collection where possible

Meeting performance targets

1. Performance benchmarks should include a combination of targets, some specifying absolute thresholds (for example 75% of eligible patients must receive an intervention), some specifying relative improvement thresholds (for example 30% improvement over baseline). Absolute thresholds should be set so they are reasonably difficult to achieve. Relative thresholds should be seen as reasonably achievable).
2. It should be possible to adjust targets over time to ensure the sustainability of the scheme
3. Comprehensive assessments should be undertaken of cases where there are wide gaps between performance and targets

Paying for performance

1. P4P schemes should be based on rewards not penalties
2. Incentive payments should be bonus payments over and above base funding
3. Incentive payments must be sufficient to offset the costs of establishing information systems, training programs and better systems for service delivery
4. P4P schemes should include some kind of 'trickle down' process to ensure that individuals and clinical units receive a share of the incentive as a reward for improving the quality of care

Dealing with perverse incentives

1. P4P schemes should develop vigilant procedures for detecting perverse incentives such as cherry-picking and gaming
2. Schemes should have the capacity to quickly correct any flaws detected
3. Expert advisers (for example specialists colleges) should be involved in developing criteria that allow outcomes for certain patient populations to be exempt from inclusion in P4P schemes

Governance and communicating with participants

1. P4P schemes should have explicit mechanisms in place to ensure participants receive detailed information on performance results
2. Information on strategies used to achieve exceptional results should be disseminated widely by a central agency
3. Reports from P4P schemes should be periodically submitted to a central agency, such as the Australian Institute of Health and Welfare or Health Insurance Commission

Implementation

1. Where possible, P4P programs should be implemented through existing local structures such as clinical-led networks or Divisions of General Practice. They should also be responsible for operationalising and evaluating any schemes implemented.

Should we expand the use of pay-for-performance in health care?

© Copyright Commonwealth of Australia

This work is copyright. Except to the extent of uses permitted by the *Copyright Act 1968*, no person may reproduce or transmit any part of this work by any process without the prior written consent of the Parliamentary Librarian. This requirement does not apply to members of the Parliament of Australia acting in the course of their official duties.

This work has been prepared to support the work of the Australian Parliament using information available at the time of production. The views expressed do not reflect an official position of the Parliamentary Library, nor do they constitute professional legal opinion.

Feedback is welcome and may be provided to: web.library@aph.gov.au. Any concerns or complaints should be directed to the Parliamentary Librarian. Parliamentary Library staff are available to discuss the contents of publications with Senators and Members and their staff. To access this service, clients may contact the author or the Library's Central Entry Point for referral.
