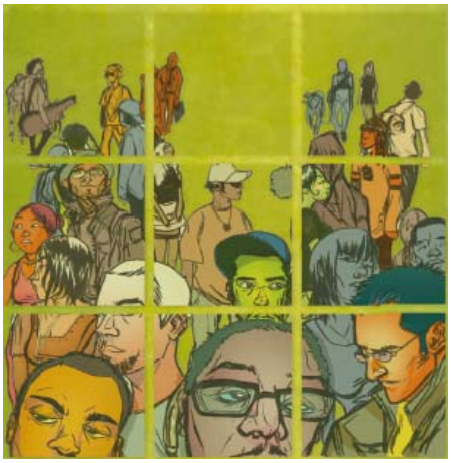
Pay For Performance (P4P) Programs in Health Services: What is the Evidence?

Richard M. Scheffler

World Health Report (2010)
Background Paper, 31

HEALTH SYSTEMS FINANCING
The path to universal coverage

World Health Organization



Pay For Performance (P4P) Programs in Health Services: What is the Evidence?

World Health Report (2010) Background Paper, No 31

Richard M. Scheffler[1]

[1] PhD; Director, The Nicholas C. Petris Center on Health Care Markets & Consumer Welfare, and The Global Center for Health & Economic Policy Research; Distinguished Professor of Health Economics and Public Policy University of California, Berkeley

Abstract

This paper examines the different kinds of efficiency in health including allocative, technical, and dynamic efficiency. A framework is provided to look at pay for performance (P4P) programs in health services delivery. The measures used and the basis of reward and types of rewards are discussed. This is followed by a discussion of the types of rewards and the payments of rewards. The resulting problems that need to be addressed are explained and an assessment of evidence is presented.

Introduction

It is widely recognized that health care systems in most countries are performing poorly. Resources are being used in inefficient ways and, in many instances, the health services that are needed most are not being produced. Additionally, the cost of health services is increasing at an alarming rate. Governments, private sectors, and health insurance companies cannot afford to pay for necessary health benefits. As out of pocket payments increase, the problem is compounded. These are among the major barriers in any attempt to expand health care coverage.

The purpose of this paper is to assess the literature on a new and growing trend of results-based finance in health called pay for performance (P4P). We begin this assessment with a broad discussion of three different efficiency concepts in healthcare: allocative efficiency, technical efficiency, and dynamic efficiency. This section is followed by the presentation of broad frame to view P4P programs. The framework shows the connection between the measures used, the basis for rewards, and the rewards given. The types of reward models are simply defined and the two key issues involved are attached.

Next, we review the experiments on P4P in selected countries. The literature on P4P in health is new, vast and rapidly growing. But most of the current research is quite preliminary since P4P models are only just being implemented. As a result, the research designs lack the rigor needed to evaluate their impact. For illustrative purposes, we have selected a few P4P program to give a sense of the kinds of P4P currently at work. Clearly, this is not intended to be exhaustive.

The final section of the paper assesses the evidence-based effectiveness of P4P. At best, we can find the evidence weak at this point. The paper concludes with some thoughts on the future of P4P and its importance in results-based financing and improving health care system.

Section 1

Efficiency as viewed by economists can have a variety of meanings. Allocative efficiency speaks to the marginal cost of producing health care and its marginal benefit or price. The concept desires to maximize the output of the healthcare system. It begs the questions: What is the best use of every dollar spent on the healthcare system? Should dollars be spent on inpatient care or outpatient care? Prevention versus treatment? Here, the benefits sought are those that increase health. These are the allocative efficiency questions that all healthcare systems need to address. Public spending on health systems must

make these allocations. When individuals pay for health services in a market with out-of-pocket funds, their willingness to pay is measured. In theory, the marginal cost of each output of the health system is weighted against the price the consumer is willing to pay.

Technical efficiency in the healthcare system deals with minimizing cost and maximizing quality, or both. It looks at the production function in the health systems. For example, what is the most efficient combination of doctors, nurses, midwifes and other health professionals in order to produce outpatient visits or inpatient stays? The idea is that technical efficiency is reached when the cost of the inputs is minimized for a given level of output or quality or both.

Health system change is extremely important and often overlooked. The dynamic efficiency of a health system asks the question: Is the rate of technological change in the healthcare system optimal? Technological change in healthcare is broadly defined. It includes new procedures, drugs and diagnostic tests. Many health economists attribute over half of the increase in health care costs in developed countries to an increase in technology. Part of the increase might be worth the cost if the new technology improves the productivity of the health system. The increased use of electronic medical records is an obvious example.

Results-based financing and P4P programs are an attempt to address all three of these efficiencies – allocative, technical, and dynamic. Equity goals are often considered as well. With this, P4P programs are being seen as a potential approach to improving and reforming health systems.

Section 2

Our review of the P4P literature on the supply side suggests a framework that can be used to evaluate P4P programs. Cash transfers and other consumer incentives on the demand side are also important, but are not included in this review. Figure A shows the structure of the P4P programs that we reviewed.

The measures used can be split into quality and efficiency measures. In developed countries P4P is largely used for quality improvements, though cost savings is certainly another element. The quality measures used can be viewed in the well-known paradigm: structure, progress and outcome. Structural measures of quality can be seen in the equipment and technology used, as well as health care facility spending. Process measures are those procedures that the facility routinely does, such as administering vaccinations, screening for diseases, and delivering the appropriate level and type of treatment.

The most valued measures are clearly outcomes, the delivery of a healthy baby, for example. Since many health outcomes are difficult to measure or to attribute to healthcare interventions,

intermediate outcomes are often used. An example of this would be chronic disease management, improved blood sugar and cholesterol levels. In most cases, P4P uses patient satisfaction as an outcome measure.

One low-income country and some middle-income countries valued the efficiency of P4P programs. Their health systems are seen as poor performing as they continue to focus on more output, as measured by more visits or more inpatient stays. In some cases quality is considered but it is clearly not the focus. Increasing hours worked, reducing absenteeism, improving the procedural practice of a facility is clearly the goal.

What are the major ways that improved performance, be it in quality or quantity, are best rewarded? It does appear that meeting a target is most often used. Examples include a target rate of breast exams or prenatal visits, though it is understood that targets do vary by type of practice and facility. In contrast to a target is the use of changes in measures or improvements in the rate of vaccination. A single version of this approach would be to have an increase in reward for an increase in a level of quality or quantity – perhaps both. Economic theory supports this basis of rewards since targets can produce distortions. Practices near the target rate are rewarded for little effort while those starting well below the target have little chance of a reward.

In industrialized countries one sees the use of rankings as a basis for reward. Being in top ten percent, for example, is one such ranking system. This method suffers from some of the same limitations as using targets. Those starting at the top have little incentive as do those at the bottom if they see their chances of improved ranking as small. What we observe is a multitude of different ways to reward performance often in combination with different measures and combinations of measures being used. At this time, no clear pattern of reward has emerged.

How are practices or institutions rewarded in P4P? The use of financial bonuses is clearly the most often reward given. It can be a flat bonus or a percent bonus and is most often given on a yearly basis. In industrialized countries, for example in the United States of America, the public acknowledgement of reaching a goal or the ranking achieved is used as a reward. The bonus can be given directly to providers but is most often given to the medical practice or group (see figure B). The practice then decides how to allocate or spend the reward. In the USA the reward is often given to the practice rather than directly to the doctors (80% of the time). In the United Kingdom the reward is most often given as a salary bonus directly to doctors.

In developing countries, there is considerable variation in who receives the reward. For example, in Rwanda the bonus is given to the practice who must then distribute. It is often paid as salary increases

but is sometimes used to improve the practice (e.g. advanced equipment). In Turkey the payment is given to the doctor if they increase their effort in the private sector.

Any method of reward has its difficulties. The payment of rewards needs to address case mix differences between practices and severity of illness of patients treated or reward systems will be distorted. The data required to do this properly, as well as the methodological changes, are significant. To date, there are no agreed upon methods of doing this is unlikely to emerge.

Another significant challenge is that of shirking. This happens if the cost of providers is too costly to measure, monitor, and attribute to the efforts of individual providers. If all providers are paid equal rewards then the incentive to improve performance is blunted. Another significant issue related to rewards comes about if there are different reward systems or levels of reward. This is more often the case in developed countries with public and private payers. Sorting out the impact of the performance in these cases may prove almost impossible.

Section 3

My assessment of the evidence concerning P4P at this time is that it is weak. This is due in large measure to the weaknesses of the evaluations. Most P4P programs have no research design that would enable one to determine how or if the program was working. Comparison groups are lacking. We cannot really attribute these changes to P4P without adequate control groups. In our cases, multiple changes are impacting the practice so it is really impossible to isolate the P4P impact or lack of impact. This doesn't mean that P4P is not working or it could be made to work. It does mean that the evidence may not be available or that the evidence is weak. Improved research designs are needed.

Then, there is the question that if P4P does work, will other issues naturally arise? For example, will the practice improve the things it is rewarded at the expense of the things it is not? How do we determine this without massive and expensive monitoring costs? More attention to this problem is clearly needed.

The impact of P4P on equity is also an issue worth considering. Without adequate risk adjustment for healthy patients who receive more care than those who are sick or perhaps difficult to treat. The measures used as a basis of reward may favor one type of patient over another. These concerns are real. On the other hand, P4P can be designed to improve equity. If we paid higher bonuses for very sick patients or poor patients, this might improve equity. In this case, it is not the P4P program that makes equity worse but the way it is implemented.

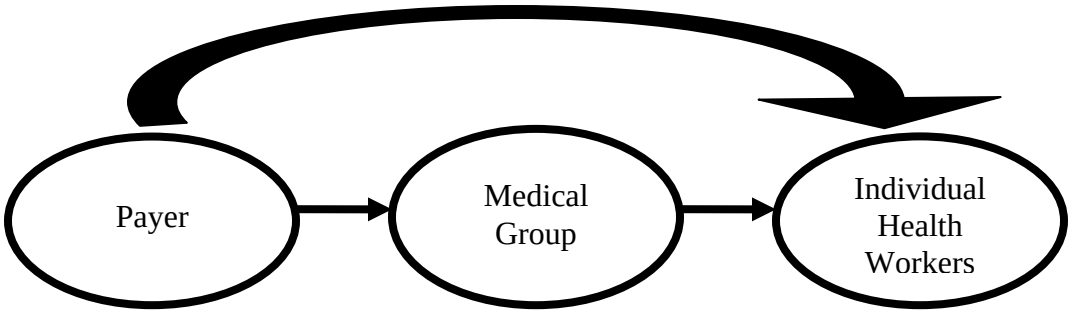
In theory, incentives matter in all types of behavior and certainly in the health system. They are clearly not the only thing that motivates provider behavior. We need much more experimentation in P4P and rigorous evaluation. There is little doubt that P4P is spreading rapidly around the world. If it is successful in improving efficiency then it can be used as an agent to reform health care systems. There is much truth to the saying "you get what you pay for" – caveat emptor.

Figure A: Structure of P4P Programs

Measures	Basis for Reward	Reward
Quality: – Structure: investment in technology, facilities, and equipment	Absolute level of measure: target or continuum	Bonus payment
	Change in measure	Publicize measures and ranking
Process: – vaccination rates, cancer screening, disease management	Relative ranking	
Outcomes: – chronic care measures, patient satisfaction		
Efficiency: – Cost savings or productivity improvements		

Source: Adopted from Scheffler RM: *Is There a Doctor in the House? Market Signals and Tomorrow's Supply of Doctors*, Stanford University Press, 2008.

Figure B: P4P Reward Payment Models



Implementation Issues

- Shirking
- Case mix
- Multiple payer-medical group relationships

Source: Adopted from Scheffler RM: *Is There a Doctor in the House? Market Signals and Tomorrow's Supply of Doctors*, Stanford University Press, 2008.

References

Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. "Effects of pay-for-performance on the quality of primary care in England," *The New England Journal of Medicine* 361, 2009:368-78.

Casalino LP, Alexander GC, Jin L, Konetzka LT. "General Internists' Views on Pay-for-Performance and Public Reporting of Quality Scores: A National Survey," *Health Affairs* 26(2), 2007: 492–499.

Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. "Pay-for-performance programs in family practices in the United Kingdom," *The New England Journal of Medicine* 355(4), 2006: 375-384.

Hibbard JH, Stockard J, Tusler M. "Does publicizing hospital performance stimulate quality improvement efforts?" *Health Affairs* 22(2), 2003:84-94.

Rebhun D, Williams T, *The California Pay For Performance Program: The Second Chapter Measurement Years 2006-2009*, Oakland, Calif: Integrated Healthcare Association, 2009.

Robinson JC, Williams T, Yanagihara T. "Measurement of and Reward for Efficiency in California's Pay-For-Performance Program," *Health Affairs* 28(5), 2009: 1438-1447.

Rosenthal MB, Dudley RA. "Pay-for-performance: will the latest payment trend improve care?" *JAMA* 297(7), 2007:740-744.

Rosenthal MB, Frank RG. "What is the empirical basis for paying for quality in health care?" *Med Care Res Rev.* 63, 2006:135-157.

Scheffler, RM. *Is There a Doctor in the House? Market Signals and Tomorrow's Supply of Doctors*. Palo Alto, Calif.: Stanford University Press, 2008.

Tanenbaum S. "Pay for Performance in Medicare: Evidentiary Irony and the Politics of Value," *Journal of Health Politics, Policy and Law* 34(5), 2009: 717-746.

Trisolini M, et al. *The Medicare Physician Group Demonstration: Lessons Learned on Improving Quality and Efficiency in Health Care*, New York: Commonwealth Fund, February 2008.

Eichler R, Auxila P, Antoine U, Desmangles B. *Performance-based Incentives for Health: Six Years of Results from Supply Side Programs in Haiti*. CGD Working Paper #121. Washington, DC: Center for Global Development, 2007.

Logie DE, Rowson M, Ndagije, F. "Innovations in Rwanda's health system: Looking to the future," *The Lancet 372*(9634), 2008: 256-261.

Meessen B, Kashala JPI, Musango, L. "Output-based payment to boost staff productivity in public health centres: contracting in Kabutare district, Rwanda," *Bulletin of the World Health Organization 85*, 2007: 108-115.

Meessen B, Musango L, Kashala JPI, Lemlin J. "Reviewing institutions of rural health centres: the Performance Initiative in Butare, Rwanda." *Name: Tropical Medicine and International Health* 11, 2006: 1303-1317.

Rusa L, Fritsche G. "Rwanda: Performance-Based Financing In Health," *Sourcebook on Emerging Good Practice in Managing for Development Results* Vol. 2: World Bank, 2006.

Soeters R, Habineza C, Peerenboom PB. "Performance-based financing and changing the district health system: experience from Rwanda," *Bulletin of the World Health Organization* 84, 2006: 884-889.

Vujicic M, Sparkes S, Mollahaliloglu S. *Health Workforce Policy in Turkey: Recent Reforms and Issues for the Future*. Washington, D.C.: World Bank, 2009.

Fulton BD, Scheffler RM. "Health Care Professional Shortages and Skill-Mix Options Using Community Health Workers: New Estimates for 2015," Berkeley, CA: The Global Center for Health Economics and Policy Research, 2009 (working paper).

Scheffler, RM. *Is There a Doctor in the House? Market Signals and Tomorrow's Supply of Doctors*. Palo Alto, Calif.: Stanford University Press, 2008.

Scheffler RM, Liu JX, Kinfu Y, Dal Poz MR. "Forecasting the Global Shortages of Physicians: An Economic- and Needs-based Approach." *The Bulletin of the World Health Organization* 86, 2008:516-523. Available at: http://www.who.int/bulletin/volumes/86/7/07-046474.pdf.

Scheffler RM, Mahoney CB, Fulton BD, Dal Poz MR, Preker AS. "Estimates of Sub-Saharan Africa Health Care Professional Shortages by 2015," *Health Affairs* 28, 2009:w849-w862.

Scheffler RM, Fulton BD, "Needs-Based Health Workforce Analysis: Methods and Empirical Estimates in Selected African Countries," book chapter for *Human Resources in Health in Africa: A New Look at the Crisis* (working paper).