Paying for Performance: Key Design Features and the Bigger Picture

Padhraig Ryan, Department of Health Care Policy, Harvard Medical School, Boston, MA, USA, & Centre for Health Policy and Management, Trinity College Dublin, Dublin, Ireland

ABSTRACT

Deficiencies in quality arise in health systems around the world. A common yet largely unproven response is “pay for performance” (P4P), which links performance targets to financial rewards. Of particular note among P4P’s myriad problems, specifying challenging targets may demotivate low-performers and reward existing high-performers without significant improvement, and low targets are likely cost-ineffective. Trade-offs arise also in defining reward magnitude, as larger rewards may be powerful but more susceptible to adversely impacting unrewarded measures. Quality improvement should prioritize patient outcomes, but difficulties in attribution impede linking rewards to clinical outcomes. Superior performance requires a multifaceted approach that nurtures clinical and managerial skills, while fostering professionalism and pride in high quality care. Some providers have improved outcomes substantially without P4P incentives for clinicians, by emphasizing a collaborative leadership culture and evidence-based practice. Unless intimately entwined with such factors, P4P may flounder.

Keywords: Chronic Disease, Economics, Incentives, Paying for Performance (P4P), Quality Improvement

1. INTRODUCTION

The problem of sub-optimal quality is common to health systems around the world (Fisher & Wennberg, 2003; Hussey et al., 2004; Wennberg, Fisher, Goodman, & Skinner, 2010). The clinical performance spectrum has been characterised as a bell curve, with a few providers delivering markedly sub-standard care, some others achieving markedly superior outcomes, and a large majority clustered around the median performance level (Gawande, 2007). In light of these deficiencies, health systems have pursued an array of quality improvement strategies.

The perceived key determinants of health system performance may vary according to one’s disciplinary perspective. On the one hand, clinicians are instilled with altruistic professional values and generally motivated to deliver high quality care, therefore the most important determinants of provider behaviour are ethics, professionalism, and clinical expertise. Yet tensions may exist between the ethical imperative to deliver high-quality care and the need to generate income, as clinician livelihoods and the viability of provider organisations depend on a steady income stream (Reinhardt, 2012). Accordingly an economic interpretation of care patterns is defensible, and the economic incentives instilled by provider payment mechanisms are recognised as another determinant
of performance (Roberts, Hsiao, Berman, & Reich, 2004). Motivated in part by the modest impact of quality improvement initiatives such as public performance reporting (Ryan, 2011), much policy attention now focuses on the use of financial incentives to improve quality.

Pay-for-performance (P4P) is a prominent tactic for quality improvement that modulates provider income in line with measured performance on specified indicators. Ideally, this sharpens incentives for high quality and compensates providers for the cost of investing in quality improvement. P4P can enjoy broad stakeholder support among payers, consumer groups, and political actors (Tanenbaum, 2009). Its popularity has grown, and a 2006 survey found that 126 out of 252 commercial HMOs in the United States (USA) were incorporating P4P into provider payment mechanisms (Rosenthal, Landon, Normand, Frank, & Epstein, 2006), while substantial investment in P4P is directed to low- and middle-income countries through the World Bank (Morgan, 2009).

But despite this burgeoning popularity, the evidence base for P4P’s clinical and cost-effectiveness is scant (Mullen, Frank, & Rosenthal, 2010; Van Herck et al., 2010; Scott et al., 2011). I briefly examine this evidence base before proceeding to the main focus of the paper, an analysis of P4P’s key design features. This literature review draws on four databases: PubMed, Embase, Google Scholar, and the Cochrane Library, until September 2012. Search terms included combinations of “pay for performance”, “financial incentives”, “quality and outcomes framework”, “performance based payment”, “health care”, and quality.

Relatively few P4P programmes have been robustly evaluated, and evaluations often uncovered no evidence of benefits. Some studies have found negative effects on the incentivized performance measures, implying performance inferior to a comparator group rather than deterioration in quality. In addition, valid inferences are often elusive due to the concurrent implementation of other quality improvement strategies (Mullen et al., 2010). A systematic review of P4P’s effects in low- and middle-income settings uncovered only nine studies, including non-randomized trials and interrupted time series studies. Results were mixed and interpretation difficult due to the high risk of bias (Witter, Fretheim, Kessey, & Lindahl, 2012).

In a systematic review of P4P’s effects in primary care, only seven studies met the eligibility criteria. Six of these reported modest improvements, but several caveats bear emphasis. Evaluation of multi-tasking was absent, meaning that any improvements in rewarded measures may have been offset by deterioration on other measures (Scott et al., 2011). Second, the durability of effectiveness is of concern. A P4P programme within Kaiser Permanente achieved modest increases in screening rates for cervical cancer and diabetic retinopathy, but after incentives were discontinued the screening rates dropped below their original levels (Lester et al., 2010). Consequently, although P4P may deliver some quality improvements it should not be viewed as a simple solution to quality deficiencies.

This paper examines six key design features of P4P, focusing mostly on the extensive experience of the USA, and I pinpoint some problems in P4P’s design and implementation. Illustrative examples of P4P programmes and other quality improvement initiatives serve to illuminate the challenges. Section 2.1 deals with the task of performance measurement, and in Section 2.2, I address the setting of targets and explore further difficulties in capturing the clinical complexities of performance. Sections 2.3 to 2.5 deal respectively with the magnitude of financial rewards, P4P’s implications for efficiency and value for money, and the question of whom to reward. Section 2.6 describes key characteristics of complementary quality improvement tools, and a concluding section draws some general inferences from P4P’s successes and failures to date.

A key lesson is that incentives are unlikely to be effective if providers lack the clinical and managerial skills to respond appropriately, therefore these skills must be nurtured while fostering professionalism and pride in high-
Land and Soil in the Religious Culture of Kôyasan in Medieval and Early Modern Japan