Reviewer's report

Title: Systematic review: outcomes, mechanisms, and context of pay-for-performance in health care

Version: 1 Date: 3 April 2010

Reviewer: Sukyung Chung

Reviewer's report:

• Major Compulsory Revisions

1. Health care systems differ substantially and physicians are paid very differently across study settings. I expect that implementation and anticipated effects of P4P would be highly dependent on the underlying payment scheme and delivery system. The first two subsections of "Context Findings" ((1) health system characteristics and (2) payer characteristics) in current form are not informative in providing the influences of these factors on the P4P outcomes. These two sections could be revised (perhaps combined). I would be interested in how the effect of P4P differs by underlying physician payment system (e.g. capitation, FFS, salaried, other type of bundled payment, or combination of two+) (classification by nation is not very intuitive given the variation in payment schemes in the US, where the majority of studies have been conducted), and by organization (i.e. separately for interventions targeted for ambulatory care - primary care vs. specialty care - and inpatient care settings - I assume most studies focus on metrics relevant to primary care). Dominant payment system characteristics may not be reported in the publications, but the authors may be able to identify it based on other context information, e.g. country, organization?

2. Define "mechanism" and "context" in more specific ways. E.g. mechanism of incentive distribution? organizational context in which the program is implemented?

3. Under "mechanism findings", it would be interesting to know the effect by the proportion of panel (% patients) or by the scope of metrics affected by the P4P. The more extensive the program is, the bigger the effect might be?

4. While the influence of "(3) Provider characteristics" in p.13 is very informative, it could be more convincing and coherent if some theoretical background to support the findings is provided. In addition, second paragraph of this section should be moved to "(3) P4P incentives" section in p.10.

5. The main contribution of this review would be to inform the influences of mechanism of P4P incentive administration and context in which P4P is implemented, with some revised work as suggested above, in addition to updating the 2 years gap in the literature. This should be emphasized in the Intro and Discussion.

6. In the discussion (p.19), the two categories: "absence" of evidence and evidence indicating "conflicting" findings should be distinguished. In that section,
"dose-response relationship" could be inferred from multiple studies (if there is no evidence on this in one study). For example, bonuses in UK is about 10 times more than bonuses in the US studies, and is the effect size much larger in the UK studies than US studies?

7. Published effects of P4P could be overstated or understated due to various reasons, including (1) potential ceiling effects because in many studies participation was voluntary, (2) P4P is combined with other QI efforts – an isolated effect of P4P itself may be hard to estimate and if possible it might be underestimation while combined effect may be overestimation of P4P effect, (3) potential publication bias. etc. A summarization in the discussion would be helpful.

• Minor Essential Revisions

8. 

9. In p.7, the first paragraph of “Description of studies” section, information on the rate of publication in years “2000-2006” is not given.

10. (p.8) Absence of effect in some conditions, e.g. CHD, may be in part due to small number of target population incentivized.

11. Several sentences, including those below, could be rewritten or expanded for better explanation/clarification:

a. p.8. “Finally, one study found positive effects for P4P targets when …”

b. p.11. “a diluting effect for incentive size due to payer fragmentation likely affected the P4P results”.

c. p.19. “Conflicting evidence, likely obscured by other mediating factors, does not justify the use of any incentive size, while still expecting P4P programs to deliver results”

12. In p.9, it says “intermediate outcome measures yielding in-between rates.” Why would it be?

13. p.10, what are the examples of “acute care”?  

14. p.10, "(3) P4P incentives" # types/amount of incentives?

15. p.10, "no clear relationship between incentive size and the reported P4P results" # see my comment #6.

16. p.12. isn’t the substantial incentive amount that mattered most for the UK’s success? Comparison by “nation” itself is not very informative; variation in the payment system and incentive characteristics across nations might be the underlying reasons for the difference. See my comment #1.

17. p.19. I do not see in the paper how the authors could draw this recommendation: “Provide quality improvement support to participants through staff, …”

18. In p.19, influence of patient behavior seems to be a mediator of the P4P effect rather than a contextual factor"
• Discretionary Revisions

19. There are also studies published immediately after the review time frame, including the two below, if the authors would like to mention in the discussion.


**Level of interest:** An article of importance in its field

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests