Reviewer's report

Title: Selecting effective incentive structures in health care: a decision framework to support health care purchasers in finding the right incentives to drive performance.

Version: 1 Date: 4 November 2007

Reviewer: Bruce Guthrie

Reviewer's report:

General

Thank you for asking me to review this paper which addresses an issue which is currently of great importance to policy makers internationally, namely the design of pay-for-performance programs to improve the quality of healthcare. However, I am not convinced that this is a research paper. However, that partly reflects that it sits in the grey area between research and policy, rather than being 'traditional' research, and I am uncertain whether it fits with the journals' policy for article selection.

Please note that I was unable to access the supplementary files, and did not receive a response to a query to the editorial office before completing the review.

The paper describes a potentially useful framework to help a policymaker make a rational choice between the various types of pay-for-performance program. By rational, I mean reasoned and evidence informed, rather than rational in the clinical 'evidence-based' sense (to create a guideline applicable in most or all situations). My reading of the literature is that there is no strong evidence for pay-for-performance in healthcare delivering better care (and even less evaluation of adverse effects eg in crowding out of non-incentivised care), although there are of course good theoretical reasons and evidence by extension from other fields to believe that pay-for-performance changes behaviour, even if not always in the intended way.

The key weakness of the paper from a research perspective is that it starts from the assumption that pay-for-performance will be implemented (ie the belief that it will work) and then examines the evidence to inform that decision. My problem with that is that the initial decision to implement is not evidence based, because there is little research or evaluation evidence of effectiveness in healthcare (as the authors themselves conclude in the paragraph between 'insert table 1' and 'insert table 2' on p7 of the manuscript). So the claim in the final sentence of the background on p4 is unsustainable ("A key contribution of the decision framework is its focus on offering explicit selection criteria that helps health policy-makers and purchasers identify the most appropriate incentive models to achieve the desired performance improvement.")

A second weakness is that the way it informs policy is at a relatively high level
(whether to implement a bonus, a withhold etc. There are a set of more detailed, but equally crucial questions. For example, what kind of bonus (competitive - winners [however defined] take all; or non-competitive - everyone can get a reward of some sort)? How much 'pay' (eg the proportion of compensation at risk; reward per measure incentivised etc)? Linear pay-for-performance or reward triggered by thresholds? There is little good research evidence for any of this in healthcare, but it would have been very interesting to understand how these issues are conceptualised in the wider, non-healthcare literature, and the methods used by existing pay-for-performance programs (if available). There is no absolute reason why the paper should have addressed this level of detail given its aim of informing an earlier stage of the policy making process, but I personally believe that the detail of the scheme implemented is at least as important as the higher level question of bonuses vs thresholds etc.

Overall therefore, I found this an interesting read (albeit one where I didn't always agree with the interpretation of the literature), but I wasn't convinced that it is a piece of research per se (which could reflect the writing rather than the reality). Rather it is a report of a rapid literature review to inform a firm policy decision to implement a pay-for-performance program.

-------------------------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. The paper does not clearly distinguish between empirical evidence from studies of pay-for-performance in healthcare, empirical evidence from other fields, and theoretical evidence. For example, in the second paragraph on p8, the authors say that witholds are more effective than bonuses, referencing two papers which appear to be from outside healthcare, although it isn't clear if they are empirically based. All three kinds of evidence are relevant, but conflating them in the results is potentially misleading. It isn't at all clear to me that there is therefore "obvious benefit" from withholds in healthcare.

2. Overall, I thought the paper was overly positive about the strength of the evidence that pay-for-performance works (eg p12, para 2 where choice of models is restrited to those "that have proven to be effective"). For example, the references for the effectiveness of bonuses (25-29) include several which are weak (or not evidence for improvement at all). eg ref 27 reports quality in the first year of the UK GP contract, but there is no 'before' data nor any comparison group. eg ref 29 - to my knowledge IHA California P4P demonstrates improvement over time, but such a finding is commonplace when quality is routinely measured even without P4P, so again not good evidence that the P4P is that important.

3. Linked to this, it was disappointing that little attention was paid to potential adverse effects of pay for performance (which I presume reflects the framing of the work, which is to inform as-rational-as-possible implementation after the decision to implement something has already been taken). This is probably because the empirical healthcare literature is relatively silent on this issue, but if
that is the case, then that should be explicit, and it would be interesting to at least get a sketch of how negative effects are understood in the theoretical literature or studies in other fields).

A key implication is that evaluation of implemented programs is essential, and such evaluation has to examine both expected/intended effects on incentivised measures, and unexpected/unintended effects on non-incentivised aspects of care. Disappointingly, the importance of evaluation of implemented programs gets one paragraph on p18, which given the uncertainty about effectiveness seems rather too little (although I accept that this may have been beyond the remit of the commissioned work).

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. Table 1: an additional incentive model used in the US related to 'cost differentials for beneficiaries' is the quasi-financial one of 'auto-assignment' as used in Medicaid to steer patients to higher quality health plans or providers (ie where the State makes the decision rather than the patient prompted by differential premiums or other signals of quality).

2. Table 1 describes a set of incentive models using a list which is then not used in table 2. It would be better if the two tables used the same headings.

3. p17 para 3 makes an unreferenced claim that "improving diabetes care can lead to savings for the healthcare purchaser of up to 10-15% per patient per year." The research I'm aware of on this topic comes from US healthcare, where savings accrued over 10 years, but savings only outweighed costs towards the end of that period. The authors may have other evidence in mind, but should reference it, and acknowledge that it may be context specific.
http://www.economics.harvard.edu/faculty/dcutler/papers/diabetes_case_2-3-03.pdf
http://content.healthaffairs.org/cgi/reprint/22/2/17.pdf

4. There are some minor textual repetitions and errors that need correcting (some examples below)
p8, para 2, ‘than bonuses’ repeated
p8 para 2, 'must go through to major changes’
p9 para 1, 'over time' repeated
p9 para 2, 'haven't find' used instead of 'haven't found'
p9 para 3, 'generated from the funding' doesn't read well

Discretionary Revisions (which the author can choose to ignore)
What next?: Reject because scientifically unsound

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

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

Declaration of competing interests:
I declare I have no competing interests