Reviewer's report

Title: Design choices made by target users for a pay-for-performance program in primary care: an action research approach

Version: 1 Date: 9 December 2011

Reviewer: Evangelos Kontopantelis

Reviewer's report:

Thank you for giving me the opportunity to review this interesting paper. Before I comment on the paper let me disclose that I’ve been involved in QOF related research for a number of years so I apologise in advance if I sound cynical at times…

The authors ambitiously set out to develop a pay-for performance framework for the Dutch primary care. They argue that P4P programs should not necessarily be developed by the policy-makers (who pay the bill); but involvement of participants in the development process can have a positive impact and improve the scheme’s effectiveness (how do they define effectiveness?). General practices were very eager to be involved in the decisions and the payment scheme (shocking!) and 65 took part in the end. Through meetings, a consensus was reached on the indicators to be included (all ‘easy’ to meet process indicators and no intermediate outcome indicators), money was selected as the reward (!), but thought not to be high enough (!). It was also decided that penalising bad quality of care was not a good idea (!).

The whole project rests on the assumption that handing over most power to participants (it seemed more than involvement to me) will improve the participation rate and it’s ‘effectiveness’. In addition, the authors do not seem to have taken on board many of the lessons from the QOF, the largest and most expensive by far P4P health scheme in the world. It is true that the original 2004-05 QOF did not directly involve GPs as much – academic GPs drafted the (evidence based) indicators which were passed on to the Department of Health. However, GP representatives entered into negotiations with the DoH and changes were made to the proposal before it was accepted. Obviously, involving all GPs in England from the start of the project was unfeasible – and I fail to see how that will be possible for the Netherlands, unless the decisions reached by that small and biased sample are automatically enforced to all practices in the country. These days NICE deals with updating the QOF and indicators are added and removed on certain criteria. There exists a QOF advisory committee that examines the evidence and proposes changes to the QOF (with a large % of GP presence in the committee - http://www.nice.org.uk/aboutnice/qof/qof.jsp)

Coming back to the participation rate in the ‘low’ GP involvement QOF: over 99.9% of the practices in the UK are actively participating. So participation is not really an issue as long as the incentive is there.

The other issue is effectiveness, which the authors have not defined clearly. Do
they imply that if GPs decide to a large extent on the measures on which they will be assessed, they will perform better on them? That remains to be seen, but I would be more worried on the quality of the selected measures and the choice of indicators in the program seems to verify that worry. Only process indicators were selected and intermediate outcome indicators (the best proxies of health) were not included! Note that if the QOF continues past 2013 it is very likely that measurement process indicators will be removed. Also note that under the QOF, average scores across almost all process indicators are above 90% allowing the vast majority of practices to get 100% of the offered incentive (since the highest upper thresholds are at 90%). Therefore effectiveness does not seem to be an issue either, if we take the QOF as an example of a ‘low’ GP involvement P4P. Of course, the difficult to meet intermediate indicators are a different issue!

Some major points then

1) I don’t think the authors have learned much from the UK QOF lesson. A massive, extremely expensive program with unclear outcomes. More recent work has identified a potential issue with decline in non-incentivised aspects of care and another study examined intermediate outcomes and failed to find a QOF effect (both papers in the BMJ). There is a plethora of publications on QOF and the paper is not up to date. Just search for “QOF” in google scholar!

2) The paper is extremely abstract in terms of the remuneration. What is the accurate estimated cost per practice? Per patient (although mentioned in a table but not discussed at all in the paper)? What are the seven levels of reimbursement? What is the overall cost of the project? What would be the cost for the whole of the Netherlands if implemented? What would be the expected outcomes? A cost-effectiveness analysis would be ideal. In general the remuneration formula, if there is one, is absent. The authors criticise payment on absolute terms, which is fine – I’m pretty sure the decision for an absolute terms reward was practical since I can’t think how the authors propose to reward based on improvement. Until I see some formulas the choice is clear!

3) The disaster with the QOF was not taking into account baseline levels and trends. So we now have a project that has cost over 7 billion pounds and we are still uncertain about the value of its outcomes (and its cost-effectiveness) since performance was improving before the introduction of the incentive and would have eventually caught up within a few years. How do the authors suggest dealing with that?

4) As I explained above using QOF examples, I think that no practice would say no to a monetary incentive and therefore participation would not be affected. Performance is relative and I would prefer to see a performance on a hard quality measure than measurement of BP. Therefore I do not agree on the consensus procedure. Isn’t there a conflict of interest? How come this has not been addressed at all in the paper? The GPs chose the easiest measures, a monetary reward which they thought it wasn’t enough and no penalisation. I am not sure that is the best p4p (Dutch) money can buy.

Overall, considering the economic climate, the lessons from the QOF that have not been heeded (from which the NHS is desperately trying to backtrack while
keeping GPs happy, which is not easy) and the lack of mathematics (payment formulas, overall cost estimates, baseline values) I think there is much room for improvement.

**Level of interest:** An article of limited interest

**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 that I have no competing interests