Reviewer's report

**Title:** Systematic review: Effects, development choices, and context of pay-for-performance in health care

**Version:** 2  **Date:** 23 June 2010

**Reviewer:** Frank Eijkenaar

**Reviewer's report:**

Reviewer’s report

Title: Systematic review: effects, development choices, and context of pay-for-performance in health care

Authors:

Pieter van Herck
Delphine De Smedt
Lieven Annemans
Roy Remmen
Meredith B Rosenthal
Walter Sermeus

Version: 2
Date: 23 June 2010

Reviewer’s response to authors:

1. Title and abstract
   1.2.4: agreed.

Other comments (all discretionary revisions):

- 3rd sentence under ‘background’: maybe the authors can briefly specify which effects of P4P are analyzed in the paper. In other words: effects of P4P on what?
- Wouldn’t ‘design choices’ fit better than ‘development choices’ (used throughout the text)?
- 1st sentence under ‘results’: I would suggest to start a new sentence regarding the other quality domains after ‘equity of care’. e.g.: ‘However, less evidence on the impact on….was found’.

2. Introduction and background
   2.1-2.1.2: agreed.
   2.3: agreed.
Other comment (discretionary revision):
- After ‘(3)’ on page 5: I’d suggest replacing ‘what is’ by ‘to analyze’.

3. Methods
3.1.1-3.1.3: agreed, it is much clearer now.
3.2: agreed.
3.3: you are right; I only added it because is important to keep such overlap in mind. Clearly, you did.
3.4: agreed.
3.5: agreed.
3.6: this still seems inappropriate because generally review papers are understood as papers reviewing the literature (not meant here). Leaving out the word ‘review’ would solve this.
3.7: agreed.
3.8: agreed.

4. Results: effect findings
4.2-4.3: agreed, it's more transparent like this.
4.5: agreed.
4.6: issue clarified.
4.7.1.1-4.7.1.2: agreed!
4.7.1.3: In the first table row: replace ‘outcome’ by ‘effect’. Minor essential revision: regarding the issue on reporting number of positive, neutral, and negative studies: agreed. Instead, you could report only the number of studies in which (in general) negative results were found (‘minority of cases’) and then (if applicable) state that some of these studies also found positive results on some other measures. This would offer the reader at least some idea of the distribution of results. For details, he is referred to the table. Also, I still think a sentence should be included on how positive the rest of the results were in general (e.g. about 5% improvement, but a lot of variation, depending on the measure and program). So not necessarily a number of studies, but an overall idea of the effect size. Just see if this is possible.
4.7.1.5: agreed.
4.8-4.10: agreed.

5. Results: development choices
5.2-5.3: ok.
5.3.1: I agree that if this issue is elaborated on, it should be dealt with in the discussion. Obviously risk-adjustment is an important element of a P4P scheme and becomes more important when outcome measures are included. Just see if you have the space; it is not crucial to include, but nonetheless important. In
some respects, this insight may help in interpreting the results. Just keep it in mind and include it if necessary.

5.3.2: also agreed. I understand that because of scope and space limitations it is impossible to discuss all these aspects fully. I just added them for you to keep in mind. However, I do think it is important to provide some (crude) information on the (mean) number and type of indicators predominantly evaluated in the studies because ‘treating to the test’ is such an important and widely cited caveat in P4P. This would be an interesting addition to the paper (discretionary revision)

5.7: this would be your own choice. I do think that risk-adjustment is a crucial element and determines to a great extent the success of the program and the prevalence of undesired consequences. It’s thus a very important design issue. However, if the studies do not report it properly, a one-sentence addition to the discussion regarding this lack of information seems warranted.

5.10: agreed.

Comment: page 13: replace ‘lesser’ by ‘smaller’ (discretionary revision).

5.11: If this document provides this information, I think it could be a valuable citation.

5.12: this is definitely not a key priority. It would just be an interesting observation, as is the case with a lot of other things. It would indeed go too far to discuss these issues here. If this document provides some information on this, you could, if possible, dedicate one overall sentence on this issue and refer for details to Annemans et al.

5.13: ok.

5.15: ok. When reviewing your paper in March, I was not aware of the fact that table 1 only pertained to clinical effectiveness.

5.17: ok. I was surprised to see this little information on evaluation, but if the information is not provided, there is indeed nothing to add.

Results: context findings

6.2: agreed.

6.5: of course it is impossible to draw hard conclusions, but you could present is as facts/observations so the reader can determine for himself. So it may be possible to notice some likely trends/patterns. I do think that the level of decision-making, the one that you included, it a very important one; it implicitly already includes to some extent the question ‘how is healthcare purchased and by whom’. I raised this point because health care system characteristics is quite a broad category and you only shared included one element to it.

6.7.1: you are right. This is a matter of definition; in a broad sense, over- and underuse indeed capture most elements that P4P attempts to correct. I accidentally labeled this a major compulsory revision (it should have been a discretionary one).

6.7.2: agreed.
6.8: ok.
6.13: ok.
6.14-6.14.1: Under patient characteristics, I would add a sentence like I suggested and then refer to the section on quality measurement (page 12), where, I would recommend, you could elaborate on this a bit more and state that evidence on (the prevalence of) risk-selection is minimal, but that some evidence is available. You could refer to the equity section if applicable.

7.5: yes
7.5.1: I understand your point now and agree that is does not belong in the results section. I understood these sentences as facts, e.g. as follows: 'programs in which the goal was specified as supporting minimal standards were generally effective; programs that were directed at boosting improvement across all providers were less effective'.
7.6: then I would rephrase this sentence a little bit.
7.10.2: this would definitely also be a problem in a lot of other Western countries, including the Netherlands and Germany (EU competition law is applicable). You could just state that one should be cautious for anti-trust issues. I don’t understand what you mean by ‘rebalancing’ though.

Level of interest: An article of importance in its field
Quality of written English: Acceptable
Statistical review needed: No
Declaration of competing interests: I declare that I have no competing interests

Level of interest: An article of importance in its field
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