Reviewer’s report

Title: A study on the implementation fidelity of the performance-based financing policy in Burkina Faso after 12 months

Version: 0 Date: 22 Sep 2017

Reviewer: D Renmans

Reviewer's report:

General comments:

Research on the actual implementation of an intervention is the first phase in every good evaluation. However, too many evaluations assume an implementation conform the intervention on paper. However, what has been implemented in the field often differs in many ways from what has been planned. This has strong ramifications for the attributability of effects to certain components of the project or the intervention as a whole. This study is thus a very important contribution to the evaluation of the PBF intervention in Burkina Faso. It also takes a very interesting approach and gives relevant insight in how implementations may differ from the foreseen implementation.

However, I think that the aggregation of the findings in general scores leads to a loss of very relevant information on how the intervention was eventually implemented. More than knowing whether an intervention scored 40%, 60% or 90% on a fidelity scale, it is important to know what eventually got implemented in order to attribute effects to the correct components. It is therefore advisable to give more information on the activities besides the information on the different dimensions. This can give more depth to the analysis. To further increase the depth of the analysis it would also be interesting to look at the differences between facilities and the why of the lack of implementation.

I think the references can be updated. Although we are almost at the end of 2017, very few articles from 2017 or 2016 are being mentioned or discussed in the paper. Updating the paper with findings and insights from more recent papers from the published academic literature will increase its relevance.

The denominators used in the paper are not always clear. Including more information on the different activities may help to understand it better. Also being more explicit about the denominator and the units of observation/analysis can help.
Specific comments:

P3, Line 3-10 According to Musgrove's terminology, RBF is the generic term for all results-based approaches (supply and demand side) while PBF is the specific term for what is being discussed in the paper. You might want to change this in order to contribute to conceptual homogeneity across the literature.

P3, Line 14-17 Given the many studies published on PBF, I'm sure there are more adequate and more recent references to give from the academic literature. Also, it's not all bad and recently there have been some interesting studies that aim to look at the change mechanisms of PBF. It might be nice to give credit to them.

P3, Line 17-20 Typo: Still needed are "more theoretical and qualitative that address the 'how and why'..." (1, p. 698): more theoretical and qualitative..what?

P3, Line 29-30 What is an interventions internal validity? Unclear.

P3, Line 32-34 "this makes it easier to interpret the link between an intervention's implementation and its results (8; 9)"

Is it really easier or just more valid?

P3, Line 54-59 Can you sum up some of those elements? Or are they specific to each intervention? What are they for a PBF scheme?

P4, Line 12 When was this "first"? It's mentioned later but you can already mention it here.

P5 Many PBF schemes have other important components besides the financial incentives (e.g. more autonomy, changed way of delivery of drugs, extra support supervision, etc.) is this the case here as well? If so please also explain.

P7, Line 41- P8, Line 10 This part is at first glance a bit confusing. Unclear that these 21 interviews are both individual interviews and FGD. Bring forward the claim that if possible FGD were held with all the relevant HWs. Also an overview table of the cadres interviewed can be useful.

Also, how many interviews were performed between in the CHR? Given that only one secondary care facility was included I propose not to include it in the analysis as it distorts the analysis and it would increase the internal validity and coherence.

P8, Line 12-26 I wonder whether the health workers have sufficient knowledge about the implementation and about what needed to be implemented. Were there differences between the facilities where you only interviewed one person and those where you interviewed a group? How do the ethnographic notes compare to the statements of the health workers? Was there a difference? Did you sometimes have to correct statements from the HWs?
I understand that it is sometimes interesting to transform "qualitative data" into quantitative data. Especially when doing comparisons. However, aggregating the information received in one score makes you lose a lot of relevant qualitative information. In the light of the objectives to "discern the intervention's strength and weaknesses" and "to easier interpret the link between an intervention's implementation and its results", it would be better to also include the dis-aggregated information, i.e. of what the different scores are composed (see the supplementary data file). Hence, a summary of the table in the supplementary file that includes the activities instead of the dimensions should not be too big to include in the paper.

It is clear how the score for a certain activity is being calculated, however how is it being calculated for a dimension? This is not made clear? Again, using the bigger table that includes the activities may make this more clear.

No explanation on how coverage was assessed.

Results section

In this section there are a lot of percentages with different denominators and it is not always clear where these denominators come from? Is the unit of observation the program or the health facility or the separate activities in each of the separate facilities?

Tables and figures

Not sure what the journals policy is, but I would prefer to see the tables and figures below the paragraphs they are mentioned in.

Table 1

Maybe add a line with 'added activities' ?

Make clear that "training" is part of the planning component or create a larger table as proposed earlier. It is also not clear what is meant with training, because in the data file trainings are only relevant for the CHR which distorts the analysis at this point (see earlier comment). Or do you also include the knowledge building sessions?
This heterogeneity is another argument to give a more detailed table with less aggregated data.

In the methodology section coverage is being defined as 'the public affected by the public'. Like an earlier comment, there is no information of how the authors operationalized this and used this in their analysis. It is thus difficult to understand what is meant with coverage statistics. (I assume that coverage here relates to how the fidelity is assessed across districts, however this is not in tune with the definition in the methodology section)

Are the differences significant?

What is the 'pyramid of complexity'?

Is this being observed in this case as well?

I am sure better references exist in the published academic literature to make this point.

What are these core elements for this intervention?

The denominators differ. If the denominators are indeed different, you cannot just add them to become a new 'adaptation rate'.

"it would seem that its not much the adaptation rate that should be examined, as the nature of these adjustments" => Indeed! Why isn't this done? No clear reason is given why the choice was made to focus solely on the quantitative aspects of the fidelity.

This comes to the core of your methodology of course. How did you manage this, how did you counter this? Because this means that your main data source is untrustworthy.

Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

Please indicate the quality of language in the manuscript:

Acceptable

Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests'

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

Yes