Author’s response to reviews

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

Authors:

Oriane Bodson (oriane.bodson@ulg.ac.be)
Ahmed Barro (siebarro@gmail.com)
Anne-Marie Turcotte-Tremblay (anne-marie.turcotte-tremblay@umontreal.ca)
Nestor Zanté (nestorzante@hotmail.com)
André Somé (paulandre.some@gmail.com)
Valéry Ridde (valery.ridde@umontreal.ca)

Version: 1 Date: 04 Nov 2017

Author’s response to reviews:

Response to reviewer

We would like to thank the reviewer for careful reading of this manuscript and for the thoughtful and constructive comments, which help to improve the quality of this document. We have checked all the general and specific comments provided by the reviewer and have tried to do our best to respond to the points raised. Kindly find our response below.

General comments:

C1: 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.

R1: Thank you for you comment. Our intention was first to present the overall picture of the intervention implementation and then to develop the particularly under-implemented elements.
In other words, we have sought from the most general to the most particular which explains the presence in some cases of aggregations of scores. The choice to remain at the district level is justified by the fact that the health district represents in Burkina Faso the most important analysis entity where the implementation of the intervention is also played out. Indeed, it is clear that the implementation of a policy is also related to the leadership of the District Chief Doctor. The results by health facility are for information available in the attached file and are used in a complementary qualitative study (Ridde V, Yaogo M, Zongo S, Somé PA, Turcotte-Tremblay AM. Twelve months of implementation of health care performance-based financing in Burkina Faso : A qualitative multiple case study. Int J Health Plann Manage. 2017.).

C2: 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.

R2: Thank you. We have followed your recommendation and updated our references.

C3: 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.

R3: We tried to respond to your concern by clarifying our overall presentation (and therefore also response to C14 to C19).

Specific comments:

C4: 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.

R4: You are absolutely right on this, the correction has been made.

C5: 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.
R5: Thank you for suggestion. It has been considered and changes have been made to update the references used.

C6: 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?
R6: Yes, it was a typo. The correction has been made.

C7: P3, line 29-30 What is an interventions internal validity? Unclear.
R7: Thank you for flagging this. Internal validity is the approximate truth regarding cause-effect relationships. The point we are making here is to say that by identifying which elements were implemented and which were not, we are playing a part in the effort to determine the causal relationship. Indeed, evaluation of implementation fidelity is important because it may not only moderate the relationship between an intervention and its outcomes, but it may also prevent potentially false conclusions from being drawn about an intervention's effectiveness. We tried to clarify our point in the manuscript.

C8: 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?
R8: Good point, thank you. Change has been made to refer better to the internal validity mentioned above.

C9: 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?
R9: So-called "core" elements are perquisite if the intervention is to have its desired effects. An intervention may be implemented “successfully” and “meaningfully” even if every single component of the intervention is not implemented as long as the most essential and indispensable components are. The core elements are therefore specific to each intervention. As stressed by Carrol et al. (2007), the essential elements may be discovered by canvassing the designers of the intervention. And this is precisely what we have done: as developed in the methodology section, we have first compiled an exhaustive list of all the activities planned base on official documentation. Among this list, we have made a selection of activities considered to be core, which were fundamental to the intervention’s and submitted them to the authorities in charge of the RBF program which validated them as core activities.
C10: P4, line 12 When was this "first"? It's mentioned later but you can already mention it here.

R10: Change has been made.

C11: 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.

R11: With your comment, we understood that, it was not very clear that our collecting tools were build based on the intervention’s documentation (as action plans, implementation guide, meeting reports) and intervention theory. It is developed further (data collection section) but might need to be clarified at this point. Clarifications were made. Since we constructed the list of activities under study based on the intervention’s documents, we consider that the validated list represents the intervention as it was initially planned in 2014 - in sum, the “spirit” of the Burkina’s PBF scheme. In this context, the components that you are mentioning are missing in the documentation.

C12: 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.

R12: Thank you for your suggestion. We have taken it into account and rephrased as we could the section on data collection. We understand your concern and as we mentioned it in our manuscript, having only one CHR (and only two CMAs) surely represents a weakness our the sample. We are aware of it and stressed it as a limitation. However, we do not think any of the health facilities sampled should be excluded as one of the main strengths of our study lies in the originality of its subject and the diversity of the sampling (hospitals are rarely sampled), even if at the end it provides “only” a view of the situation that remains to be confirmed or countered in our next studies. Plus, as mentioned, this study is part of a larger research project designed as a multiple case study with several embedded levels of analysis. The selection of contrasted cases was not taken lightly but lies on criteria exposed in the (larger) research protocol (Ridde V, Turcotte-Tremblay A-M, Souares A, Lohmann J, Zombré D, Koulediati JL, Yaogo M, Hien H, Hunt M, Zongo S, et al. Protocol for the process evaluation of interventions combining
performance-based financing with health equity in Burkina Faso. Implementation Science. 2014; 9(1)).

C13: 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?

R13: You are raising a good point about health workers possibly not knowing what needed to be implemented. We also thought this could have been a potential obstacle but our methodology helped us to minimize the risk as we compiled thoroughly all the activities planned and could therefore, based on our list, ask precisely the informants (based on what we knew was supposed to be implemented). We did not observe strong data differences between facilities were we conducted individual interviews and focus groups. Yes, it happened a very few times to observe an incoherence between the notes taken (based on what the informant said to the survey team) and the status chosen (implemented as planned, not implemented as planned, or modified). In these very few cases, we corrected the status based on the notes taken as we assume it did not represent the health worker’s statement. We made some clarification in the manuscript.

C14: P8, line 36-44 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.

R14: I think there is maybe a problem of understanding on this point: The activities analyzed are the activities that were part of the list under study, no more, no less. So, there was no actual loss of information. Moreover, while it is true that we first presented our results dimension by dimension (and therefore offered a general viewpoint), we later considered in more details some particular activities when it seemed relevant, and thus we dug first the most general to reach more particular results. Generally speaking, we tried to respond to your presentation concern (C14 to C19) by clarifying our overall presentation.
C15: P8, Line 46-58 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.

R15: Since a dimension is composed by several activities, its implementation fidelity is the calculated on the basis of fidelity of implementation score for the activities which compose it. We tried to respond to your concern by clarifying our overall presentation (and therefore response to C14 to C19).

C16: P8, Line 31-61 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'?

R16: As suggested, we tried to clarify our overall presentation (and therefore response to C14 to C19). Regarding Table 1, the added activities are already included (denominator 380) according to their status (implemented or not).

C17: P9, Line 45 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?

R17: As suggested, we tried to clarify our overall presentation (and therefore response to C14 to C19).

C18: P10, line 24 This heterogeneity is another argument to give a more detailed table with less aggregated data.

R18: Please, see R14. As suggested, we tried to clarify our overall presentation (and therefore response to C14 to C19).

C19: P11, 'Coverage' 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)

R19: As suggested, we tried to clarify our overall presentation (and therefore response to C14 to C19).

C20: P13, 'temporality' Are the differences significant?

R20: Yes, in general the differences are significant even if it is quite difficult to talk about significance here given the low implementation fidelity of some RBF dimensions. As we stress it in the manuscript, we decided to suggest our temporality results to be taken as nuances.

C21: P14, line 37 What is the 'pyramid of complexity'?

R21: It refers to the complexity of activities studied. We wanted to stress the fact that we observed that more complex an activity is, less organized/implemented it is. We tried to clarify our point by rephrasing it.

C22: P14, line 44-51 Is this being observed in this case as well?

R22: Considering the methodology of our study, we did not observe these effects (providers’ trust undermined and frustrations). Not observing them does not mean they are automatically absent but the collecting tools we used do not allow us to make such conclusions. An ongoing qualitative study will help us understand better if the fears flagged are founded.

C23: P14, Line 49-51 I am sure better references exist in the published academic literature to make this point.

R23: We tried to update our references following this commentary and your previous commentary (C2).

C24: P15, line 24 What are these core elements for this intervention?

R24: See C9 and response. The activities under study are understood as being the core elements of the intervention.
C25: P15, line 43-46 The denominators differ. If the denominators are indeed different, you cannot just add them to become a new 'adaptation rate'.

R25: Thank you for flagging the typo. The denominators are not different: 28 activities out of 380 have been modified (and not out of 360 as previously written).

C26: P15 Line 46-48 "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.

R26: We totally agree with on the importance of looking not only at the adaptation rate but also at the nature of the adaptations. Although interesting, we chose not to answer this concern in our article. We think that the main goal of our study was to give a very first shot to implementation fidelity assessment of a RBF scheme and offers an overall implementation picture. To date, this is first such study. Our article demonstrates interesting conclusions and paves the way for some complementary lines of research such as studying the nature of adaptations. Be assured that we will be attentive to responding to this particular and interesting question in a future study.

C27: P16, line 51-56 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.

R 27: Thank you for flagging it. We probably have not expressed ourselves well in the manuscript. In no case, we have not been able to meet a resource person. It was sometimes complicated though but we have been able to meet at least one informant person for each facility under study. We have rephrase the sentence not to be confusing anymore.