Reviewer's report

Title: Factors associated with data quality of the Routine Health Information System in Benin

Version: 1
Date: 9 February 2014

Reviewer: Magdalena Raban

Reviewer's report:

This is an interesting paper using a novel approach to assess the quality of RHIS data in Benin, and attempts to explore factors which may be associated with the data quality. As such, it has the potential to inform further development of RHIS data quality in Benin. However, some major revisions are needed.

Major compulsory revisions

Methods

The Methods would benefit from a more detailed description of the reports filled in by the health staff. The authors state these are periodic and are activity summary tables. However, I would be interested in how often these are filled in, the type of information contained in the reports, the number of data items per report etc. Is there only one type of report, or are there a number of different ones for different health programmes? Perhaps the authors can even provide details of how long it takes to complete a report. Also, are these reports used at the health facility level for any purpose? Or is there any feedback from the district level to the health facilities regarding performance or other issues?

Additionally, whose responsibility is it to fill in the reports? The authors mention that some staff who fill in reports are facility managers as well. How is the responsibility assigned? Is there only one such person per health facility? Is it always the same person who fills in the forms?

I also think it’s very important to clarify that the private health centres are also part of the RHIS (if that’s the case) and to what extent, at the start of the Methods. Often, in LMIC, the private health sector is not part of the RHIS, and the authors do mention in the Discussion that the private sector is “not greatly involved in the RHIS”. However, private health centres were sampled in the study.

The sampling description (pages 4-5) should probably include the actual numbers of health centres selected, and then within those the numbers of workers (if not just one per centre), and the number of data items in each batch. Perhaps a figure illustrating the sampling strategy may be helpful.

How were the data items sampled from a batch? Were all the reports sent in the last 12 months compiled, and then the data items sampled from all those reports. Or was a report randomly selected and then data items sample from within that
report? If the latter is true, then there would be some reports which were of lesser
quality than others, biasing the results.

The authors state in the Methods (page 4) that “computerisation” is still
rudimentary in the health centres. Do the authors have details of the break-up of
health centres with and without computers? Did the authors consider whether
this may have had an effect on data quality?

Results
In the first part of the results (page 8), the authors present the percentage of
batches which were rejected due to poor data quality, as well as percentage of
batches rejected based on the comprehensiveness criterion. What about the
other criteria of reliability and accuracy?

Page 8: it would be useful to know the median number of samples prior to batch
rejection for the sample overall. Also, is it possible to present these data for each
of the criteria (comprehensiveness, reliability, accuracy)?

Figures 1 & 2: Why have the authors chosen only these two independent
variables for the figures? Could perhaps all the results for each independent
variable be presented in one figure (comprised of several small graphs) or table?
This would allow the reader to a quick visual summary of all the results
(significant and not significant). Also, p-value would be useful on the figures.

Focus group results, Page 10: it is stated that the health workers use
“flowcharts”, however it is unclear what these are.

Discussion
The authors state that “In our study, perceived self-efficacy was association with
the quality of the data, although the results should be viewed with caution from a
statistical point of view”. I assume the authors are referring to the p-value being
of borderline significance at 0.052. I think the language needs to be changed
here as the phrase “caution from a statistical point of view” is vague and casts
doubt on the statistical analysis.

The authors state that the “private sector in Benin is not greatly involved in the
RHIS” (page 12). This statement needs to be clarified. It seems from the study
that the private facilities also report data. So why are they “not greatly involved”? This
needs to be explained in the methods so the reader understands how the
RHIS in Benin works and can read the results with this context in mind.

In the sentence (page 12): “Moreover, since in our sample three-quarters of the
health workers were from the public sector and 8 out of 10 of them were women,
the relationship with data quality could be the result of a mistake”, I think the
authors mean that the relationship could be due to confounding. Please revise
this sentence, as saying that there could be a mistake casts doubt on the
statistical analysis and results. The sample distribution should perhaps be
presented with greater clarity and detail in Table 1, so that such differences are
visible from the outset (e.g. have two columns, one for public and one for private
Page 11: the study results are compared to other studies of RHIS. Did these studies use similar methods to this study to make them comparable?

Conclusion, second sentence, page 13: this sentence is unclear – could it please be revised.

Minor essential revisions
Introduction, page 3: Need the full form of JSI.

Introduction, page 3: are the references cited for the PRISM framework correct (references 2 & 3)? The main PRISM reference by Aquil 2004 should be cited here (reference number 4 in the paper).

Methods, page 6, line 4: LQAS is used without previously clarifying what the full-form of it is.

Focus group results, page 10: it is stated that “health workers also seemed to be insufficiently qualified, as few had been trained and the training received had not always been appropriate”. I assume this means they did not receive RHIS training, but this need to be clarified in this sentence.

Discussion, page 11, second para, first sentence: replace “routine health information system” with RHIS to be consistent throughout the paper.

Discussion, page 11: the term “healthcare pyramid level” is confusing, consider revising e.g. health facility level.

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests.