Reviewer's report

Title: Assessing implementation fidelity of a community-based infant and young child feeding intervention in Ethiopia identifies delivery challenges that limit reach to communities: a mix-method process evaluation study

Version: 3 Date: 2 February 2015

Reviewer: Maria Elena Jefferds

Reviewer's report:

Understanding factors that influence the performance of nutrition interventions is an important gap limiting their effectiveness. Framing this analysis with an emphasis on implementation fidelity is quite useful; not enough attention is paid to implementation fidelity in nutrition whether it’s taken into consideration during the design (is it feasible, will delivery staff be able to carry out expected activities with fidelity) or during implementation.

Discretionary revisions:

1. Line 208: “Adherence and exposure” - consider renaming this header as “Training adherence and exposure” so it’s more specific. Similarly, consider renaming Quality and responsiveness as “Training Quality and responsiveness” in line 239 etc. because the same sub-headers were used for the different sections it was easy to get confused between sections/pages.

2. Lines 307-8: “…FLWs perceived that either they or the mothers had insufficient time to use the tool to discuss its content.” Did this vary by FLW type? The FLW includes HEP supervisors, HEP, and community volunteers, one would expect each of these staffing categories to have different amounts of time to dedicate to using the tools (e.g., I would expect HEP supervisors who presumably mostly supervise and don’t really deliver the intervention to be mostly irrelevant for this statement, no?)

3. Lines starting 322 on Supervision and Feedback. It’s difficult to interpret these paragraphs without knowing how many people each supervisor is expected to supervise monthly (e.g., 5 or 100). Was this defined when developing PIP? The number of people each supervisor is assigned to supervise will greatly influence the number they can observe each month and whether they’re providing SS every month (or if that’s even feasible). Providing a bit more info on that beyond the PIP figure would be helpful.

4. Line 355-368: Consider summarizing these qualitative findings at a higher level and reducing text because they essentially are reinforcing the quantitative findings and not providing much additional information in these additional paragraphs (e.g., Qualitative findings reinforced that the SS didn’t cover all components as expected and they routinely did W but not X, Y, and Z).

5. The conclusions nicely framed potential changes to consider in the future to
address problems identified. It would be good to know if any changes were made to the ongoing program to address problems identified when the data were collected.

6. In the discussion, the points about variable levels of capacity and resources in different geographic areas and the (positive or negative) influences on fidelity and quality are issues that affect all interventions and are very relevant and often not given enough attention, so I’m glad to see it discussed here. I did expect to see data related to lines 398-400 and perceived burden by staff presented in the results as their motivation is fundamental for quality, so consider whether that could be briefly included in the results.

7. The discussion also mentions the complexity of the intervention and inherent challenges related to implementation fidelity with increasing intervention complexity, especially because these are demanding in time and skill (e.g., Line 404-409). Carroll et al’s implementation fidelity framework paper refers to the “essential components” needed to effectively implement an intervention and identifying the “essential components” should be the priority, which I think you’re getting at so might want to mention “essential components” specifically. I also think it’s worth reflecting on the fact that even with a PIP, it’s hard to estimate the burden and there’s a push to address all IYCN behaviors, which is a lot of behaviors and hard to do in any context.

Minor Essential Revisions

1. Please avoid the use of acronyms as it limits the ability of readers to follow the text easily (and acronyms are used a lot in this paper). Suggest not using HEP as an acronym because it’s very similar to HEW. Also, when I did a search for FLW (because I couldn’t remember what it meant), I only found it spelled out in the abstract and not in the text.

2. Line 25 “need” should be past tense (or sentence rewritten to be grammatically correct)

3. Line 156/157. Double check this text because “EA as first cluster” doesn’t make sense. Seems it should say something like “…using a two-stage cluster sampling method (EA selected PPS in the first stage of sampling and random selection of households with children under two years of age as the second stage of sampling.)”

4. Line 153-155 – clarify whether all HEW supervisor, HEWs and community volunteers were interviewed per cluster (EA), or if not all in the cluster then how they were selected

5. Tables – explain in footnotes why some cells with results (not just describing n’s) are bolded, e.g., Table 2 training on ENN-BCC for 2013 results. I think it’s because the bolded values refer to the bolded training name?

6. The PIP seems to assume that logistics management functions appropriately, but poorly run logistics and stock out of supply (whether products or teaching aids etc) are a common problem. Was stock out an issue for the tools or any other program materials, especially since few community volunteers received tools?
Major Compulsory Revisions

1. I appreciate that with various data sources it’s complicated to describe all the methods for the results presented. While it is a strength of this paper to have multiple sources and be able to triangulate the findings and/or try to confirm internal validity, I was confused in a few cases.

a. In the methods, I don’t think the baseline survey methods are described, it just says that 2010 results compared to the 2013 findings in lines 162-164. Based on the methods for the PE, it sounds like both the baseline and the PE are population based representative surveys? Presumably the same or similar instruments and questions were asked? Please clarify.

b. Lines 162-164: it says 2010 and 2013 results are “compared qualitatively, given their different sample sizes”. I don’t know what is meant that quantitative data will be compared qualitatively. Also, is this only for the results stated in that sentence or for all results between the two surveys? The sentence is not in the analysis section, but at the end of the sampling which is also confusing. The language in the results text frames some results as larger or smaller, substantial increase etc, which connotes there are statistically significant differences (so doesn’t seem like a qualitative comparison). Even in discussion Lines 377-380 talks about increased IYCF knowledge and efficacy etc., which suggests quantitative comparisons. Please clarify analysis methods.

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

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.