Author's response to reviews

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

Authors:

Sunny S Kim (sunny.kim@cgiar.org)
Disha Ali (disha_nan@yahoo.com)
Andrew Kennedy (a.l.kennedy@cgiar.org)
Roman T Gebremedhin (r.tesfaye@cgiar.org)
Amare W Tadesse (amareworku@addiscontinental.edu.et)
Teweldebrhan H Abrha (teweldebrhan.hailu@gmail.com)
Rahul Rawat (r.rawat@cgiar.org)
Purnima Menon (p.menon@cgiar.org)

Version: 4 Date: 16 March 2015

Author's response to reviews: see over
Dear Ms. Natalie Pafitis, Executive Editor:

Thank you for the peer review of our manuscript, “Assessing implementation fidelity of a community-based infant and young child feeding intervention in Ethiopia identifies delivery challenges that limit reach to communities: a mixed-method process evaluation study.” As per instructions, the point-by-point responses to the reviewers’ comments are enclosed as follows, and the revised manuscript (with all changes highlighted) is attached as a separate document.

Thank you for your continued attention and review of this manuscript.

Sincerely,

Sunny S. Kim, PhD, MPH
Research Fellow
Poverty, Health, and Nutrition Division
REFEREE 1’s COMMENTS:

1. Question well defined? Adequately.
3. Data sound? Yes, and absence of a comparison group is addressed.
5. Adhere to standards? Yes.

**RESPONSE:** Thank you for your positive assessments above.

6. Discussion and conclusions: They seem justified by the data. However, the conclusions could be more forceful. The main use of this study is to examine how the program could be more effectively implemented. The conclusions could largely have been guessed at before the study, and represent the common constraints for such programs. It would be useful to know what possible constraints (e.g. inappropriate messages) were not found, and what the priorities were among those identified, and how they could be dealt with. For instance, how would more intensive supervision be achieved (e.g. transport, more supervisors, change job descriptions, give better training on supervision, etc.)?

**RESPONSE:** In the conclusions, we assert that the volunteer cadre is underutilized and presents the greatest challenge to reaching mothers, and that linkages between HEWs and volunteers need to be strengthened. In order to achieve this, we proposed the following approaches: 1) more intense supervision of the HEWs and volunteers, particularly with explicit instructions and better guidance on supervision and a more feasible ratio of HEWs to volunteers, potentially through intermediary roles by leaders over teams of volunteers; 2) performance-based incentives to motivate these cadres of health workers; 3) a reassessment of the workloads of these cadres and task shifting to ensure that delivery of the expected IYCF interventions is manageable for both; and 4) simplifying the IYCF interventions, for example, by targeting fewer IYCF behaviors (or at least prioritizing those age-specific practices that require greater improvement) (lines 473-480).

7. Limitations are clearly stated.
10. Writing is acceptable.

**RESPONSE:** Thank you again for your positive assessment.

11. Recommended changes: Strengthen the conclusions/recommendations, as indicated above (Point 6). In addition, a comment on whether this type of extensive process evaluation produces results and conclusions that are worth it; or which usually could be reached and justified, with less extensive field work and analysis.

**RESPONSE:** A statement about the usefulness of the process evaluation in examining the implementation process and inform how the program achieves its intended impact, particularly in the absence of regular or comprehensive monitoring data, was included in the discussion. Given the longer duration and extensive fieldwork of the process evaluation, this study was not intended to replace routine monitoring for timely feedback and corrective actions. Still, the process evaluation helps to critically examine the implementation process and the conditions and factors influencing the process (lines 461-464).
**REFEREE 2’s COMMENTS:**

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.

   **RESPONSE:** Sub-headers have now been revised as suggested (lines 217, 248, 284, 313, 333, and 355).

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? FLW includes HEP supervisors, HEP, and community volunteers, and 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?)

   **RESPONSE:** Indeed, HEW supervisors used the tool only once a week or irregularly. However, the frequency of use (once to at least three times a week) and main reasons (i.e. time constraint) for not using the tool more often were similar for HEWs and volunteers. Therefore, HEW supervisors have now been removed from this statement (lines 314-318).

3. Lines starting 322 on Supervision and feedback: It is 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.

   **RESPONSE:** The numbers of supervisees by FLW type were included in lines 337-339, as suggested. These personnel ratios were not defined by the program, as the HEP structure was pre-established by the government. On average, each HEW supervisor supervises 6-8 HEWs, and each pair of HEWs are responsible for 50 community volunteers in their kebele/ward. A feasible ratio of supervisors to HEWs is in place, yet monthly supervision exposure among HEWs is quite low. Supervision of community volunteers by HEWs faces greater difficulty, given the greater numbers of volunteers and the context of HEWs’ heavy workload.

4. Lines 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.).

   **RESPONSE:** The paragraph on the qualitative findings have now been summarized at a higher level as suggested (lines 364-368), in order to reduce text.
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.

RESPONSE: Our study findings were reported back to the program, but given that the program was phasing out by early 2014, our understanding is that the results were considered more in the design of a second phase of the program, to be initiated in 2015. A brief statement reflecting on this has been included in the discussion (lines 458-461).

6. In the discussion, the points about variable levels of capacity and resources in different geographic areas and the (positive and negative) influences on fidelity and quality are issues that affect all interventions 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.

RESPONSE: There were challenges in measuring and interpreting actual work time, given the wide portfolio of activities and responsibilities of FLWs and frequent task shifting, and few perceptions of time and workload were gathered in the qualitative work. Given the limited results, we have decided to keep the discussion about FLWs’ work context within the discussion section.

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. Lines 404-409). Carroll et al.’s implementation fidelity framework paper refers to the “essential components” need 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.

RESPONSE: For our present study, we considered the “essential components” of the interventions as the key messages that were embedded as the program tracers. We did not draw out this element explicitly for our implementation fidelity study, but discussed them as part of our results of dissemination of program tools and messages. We intend to address these specific components further as well as the success of the program in addressing the various IYCF practices, when later assessing the effect of the intervention on outcomes as part of our ongoing impact evaluation. However, we have now included mention of this fifth aspect from Carroll et al.’s framework in the discussion section (lines 412-420).

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.

RESPONSE: Acronyms that were used less frequently were removed (i.e. EA for enumeration area, MCH for maternal and child health, and SS for supportive supervision); HEP as an acronym was
also removed as suggested. FLW has now also been spelled out in its first appearance in the main body of the paper (line 99).

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

   RESPONSE: Revised as suggested (line 25).

3. Lines 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).”

   RESPONSE: Revised as suggested (lines 153-155).

4. Lines 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.

   RESPONSE: In each enumeration area/kebele, there is usually one health post, which consists of two HEWs. Both HEWs, their supervisor, and approximately four volunteers from a list of community volunteers identified by HEWs in each enumeration area were selected to participate in the FLW survey. This process of selecting frontline workers were clarified in lines 155-159.

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

   RESPONSE: In Table 3 (previously Table 2), bolded text has now been removed, and a footnote has been added to explain the 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?

   RESPONSE: Maintenance of stocks of program tools by regional and woreda-level trainers were included in the PIP, in order to reflect the importance of logistics management functions. In our study, only few HEWs reported shortage of materials as the reason for not disseminating, which was mentioned briefly in the results (line 296-297). The numbers of copies received and sufficiency of materials were assessed during the qualitative study, and most HEWs reported receiving more than one copy but used the tools directly without disseminating them.

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.

RESPONSE: In order to address any confusion about the various data sources, a summary table of all data collection methods and sample has now been added (Table 1). Also, a brief description of the baseline survey methods has been included in the sampling and data collection section (lines 168-171). Although similar sampling methods and instruments were used, there were differences in sample sizes between the two cross-sectional surveys (e.g. baseline included households with children under five years of age, thus more households, and a smaller sample of FLWs).

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 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. Even in discussion Lines 377-380 talks about increased IYCF knowledge and efficacy, etc., which suggests quantitative comparisons. Please clarify analysis methods.

RESPONSE: “Compared qualitatively” means that all survey results (between the 2010 baseline and 2013 PE surveys) were compared by describing their trends and patterns; no statistical tests were applied. This explanation has now been included in the analysis section (lines 181-183).