Author’s response to reviews

Title: Effect of Integrated Infectious Disease Training and On-site Support on the Management of Childhood Illnesses in Uganda: A Cluster Randomized Trial

Authors:

Peace D Imani (imani9p@yahoo.com)
Brian Jakech (bjakech@idi.org)
Ibrahim Kirunda (ibrakiru@yahoo.co.uk)
Martin K Mbonye (mbonye@yahoo.com)
Sarah Naikoba (naikobasarae@gmail.com)
Marcia R Weaver (mweaver@uw.edu)

Version: 3
Date: 5 June 2015

Author’s response to reviews: see over
Dear Dr. Torrey,

Thank you for the opportunity to resubmit a manuscript entitled “Effect of Integrated Infectious Disease Training and On-site Support on the Management of Childhood Illnesses in Uganda: A Cluster Randomized Trial” to BioMed Central Pediatrics. Below are responses to each of the reviewers’ insightful comments.

As a reminder, IDCAP results have been previously reported for two other outcomes: 1) Trainee performance on clinical vignettes, (Weaver et al PLOS One, 2012) and 2) Facility performance, (Mbonye et al. PLOS One 2014 and Weaver et al. PLOS One, 2014). We are resubmitting the manuscript that focuses on the clinical assessments.

It is a pleasure to resubmit this manuscript and response to reviewers to you.

Thank you for your consideration.

Peace Imani, MBChB, MMED, MPH

Dr. Ayieko’s Reviewer’s Report

1. The introduction is well written and the final paragraphs provide convincing basis for a separate analysis of these data from the IDCAP project. The introduction can be strengthened by providing a more convincing link between paragraphs on how human resource for health can contribute both towards attainment of the objective of universal health coverage and reduction in child mortality (outlined in paragraph 1). As currently presented only a single sentence tries to link these issues: “Shortage of health worker in the country could be a contributing factor to the slow decline in the under-five mortality.”

The sentence has been replaced with two sentences and two new references:

“Cross-country comparisons show that health outcomes improved and coverage of cost-effective interventions increased with more health workers. Infant and under five mortality were inversely related to the total number of doctors, nurses, and midwives per 10,000 population [Anand, 2004], and vaccination coverage was positively related to it [Anand, 2007].”


2. Regarding the cluster RCT design there is no explicit statement in the paper justifying use of this design. The authors should consider inserting a statement to this effect.
Two sentences in the subsection on Trial Design have been revised as follows:

“Thirty-six eligible health facilities were selected from all major regions of Uganda and randomized as clusters (26) and randomized as clusters, because many of the facility performance indicators depended on a team of clinicians, laboratory professionals and data entry staff rather than individuals.”

3. Further the authors state that “The clusters were health facilities drawn from all major regions of Uganda”. In the analysis section the authors go on to state: “All regression analyses were clustered on the trainee” and then in the sensitivity analysis the following statement is included: “Estimates were also obtained with clustering on the site instead of the trainee”. Considering these three statements the data structure is not clear. I would suggest that the authors use the terms “cluster and clustering” more cautiously to reduce confusion describing data structure and the analysis models. For the models it will be useful to state the number of levels in the model and the variables used to define each level.

The statements on clusters were clarified in the revised text. In the section on Trial Design, the revised sentence is, “Thirty-six eligible health facilities were selected from all major regions of Uganda (26) and randomized as clusters, because many of the facility performance indicators depended on a team of clinicians, laboratory professionals and other staff rather than individuals.” In the section on Data Management and Statistical Methods, the revised sentence is, “Although patients were observed with random effects for trainee nested within health facility, preliminary analyses suggested that the random effect for facility did not affect the results, and was not included in the primary model reported below.”

To avoid confusion, we deleted the statements about the sensitivity analysis with cluster on facility instead of the trainee from the sections on Data Management and Statistical Methods, and Sensitivity Analysis.

4. From the presentation of sample size section it is apparent that this analysis did not form the basis of the sample size calculation for the IDCAP project. It is also clear that the sample size calculation for the IDCAP study that is referenced (Naikoba et al) cannot support comparisons done in this analysis for the following reason: whilst the IDCAP project used facilities as the unit of analysis, the current analysis reports that “The patients was the unit of analysis”. The reasons presented for the absence of formal sample size calculations are admissible but the authors could consider conducting and reporting power analysis to demonstrate that the data collected for purposes of the IDCAP project can be used to answer the question related to childhood illnesses with reasonable power to detect important differences.

The following sentences about power calculations were added to the revised manuscript:

“Given 36 trainees per arm and five observations per trainee, we planned to have a sample of 180 observations per arm each time period. In a simple comparison of proportions at time 1, a sample size of 180 would detect an increase from 60% to 75% of tasks performed correctly with a power of 0.84 and an increase from 70% to 85% of tasks performed correctly with a power of 0.91.”
5. Under patient selection, the authors state that “patients were a convenience sample who reported to the clinic on the day of assessment”. In randomized trials it is preferable to use probability sampling rather than non-probability sampling approaches like convenience sampling for including individual participants in clusters for the purposes of the trial. This deviation has not been highlighted in the discussion and its methodological impact is uncertain. I would suggest that as minimum the authors should comment on this methodological issue.

The following sentences on patient selection were added to the subsection on Limitations:

“MLP were observed managing a convenience sample of patients rather than a random sample. It would be difficult to select patients at random in the absence of an appointment system. It’s unclear however, what how the selection process may have biased the results.”

6. The authors should avoid use of terms like “personal” bias and “general” bias to refer to well established types of biases and instead use correct terms to refer to the potential types of bias that occur when an observer classifies a group more favorably due to knowledge of allocation to intervention or due to the observer’s involvement in activities in a given health facility.

We changed the terms “personal bias” and “general bias” to “observer bias” as defined in Hróbjartsson et al (2012), which is also referred to as detection bias by Juni et al (2001).


7. The participants’ characteristics are adequately summarized and presented in table 2. The trainee characteristics and health facility characteristics are not summarized with the same level of clarity. The authors could consider including these data in an appendix.

We presented some information about the trainees and facilities in the original submission at the beginning of the subsection on Facility and Trainee Flow. To provide complete information on their characteristics, we added one sentence to that paragraph (“Four of the five hospitals were randomized to arm B.”) and revised the description of the arms in Figure 1 as follows:

“Intervention arm
Health facilities randomized to arm A (n=18): 35 HCIV and 1 hospital
2 MLP/site (n=36): 24 clinical officers and 12 registered nurses

“Control arm
Health facilities randomized to arm B (n=18): 32 HCIV and 4 hospitals
2 MLP /site (n=36): 22 clinical officers, 10 registered nurses and 4 registered midwives”
8. In paragraph 1 on page 20 the text reporting “a 42% increase in the proportion of physical system examined correctly (aRR = 1.40 (95% CI = 1.16- 1.68))” should be changed to agree with the aRR of 1.40 reported in the Table 3.

Thank you for noting this inconsistency. The text has been corrected.

9. Two sections of text appear to be misplaced within the body of the manuscript. The authors appear to be aware that the text under the section “changes after trial commencement” are out of place when they write “These revisions will be easier to understand after the outcomes are introduced and explained in the outcome section below”. I suggest that this section be moved to the appropriate point in the text. Second, is the section on “recruitment of sites and trainees” presented in the result sections after the” flow of trainees” has already been presented. Please consider whether the recruitment details could fit better in the methodology section.

These are very good suggestion. The subsection on Changes After Trial Commencement has been moved to after the subsections on Outcomes and Construction of Outcome Variables, and revised as follows:

“Changes after trial commencement
“The patient sample and the assessment tool changed. When some facilities did not have patients aged under five years on assessment days, a few trainees were assessed on children over 14 years. The observations were excluded from the analysis.

“The assessment tool was also revised prior to conducting the endline assessment to address problems with the patient history and physical examination sections identified at baseline. The patient history questions about HIV status and immunization were revised. At baseline, it was unclear whether “yes”, meant that the child or mother was positive or that the HIV status was known. Immunization status of the child was similarly unclear as to whether “yes” meant that the immunization status of the child was up-to-date or that the child’s immunization card was presented. For both tasks, a “yes” at baseline was interpreted to mean that the trainee spoke with the caregiver about the topic. At endline there were three separate tasks for asking about mother’s HIV status, PMTCT and child’s HIV status. A “yes” for the child’s status meant that the child was HIV-positive. For immunization status at endline, there were two distinct tasks for asking: 1) Whether or not the child’s immunization status was up to date and 2) Whether or not the immunization status was verified with an immunization card. Physical examination questions about the mouth were added to the form. The endline version of the assessment form is available for researchers as Supplementary File 1.

Data on HIV status, immunization and examination of the mouth were only included in the sensitivity analysis performed with the end line sample. (See Sensitivity Analysis section below).”

The subsection on Recruitment of Sites and Trainees was moved after the subsection on Human Subjects Approval, which presented the informed consent process.

10. The authors do a good job in the sensitivity analysis to demonstrate the potential impact of changes that occurred in the clinical faculty teams between baseline and endline, and the potential impact of missing data. However, given the recent advances in the field of missing data imputation it
might be worth stating in text the obvious limitations of best and worst case scenario analysis that was employed in the sensitivity analysis.

The following sentence was added to the Limitations section about the sensitivity analysis of missing data, “The missing data analysis demonstrated the range of results with extreme values for the tasks that were used to construct each dependent variable, multiple imputation of missing values would have been challenging because the number of appropriate tasks varied across patients for four of the dependent variables.”

11. There is no indication of either apriori or observed within cluster correlation values. While the authors do not have to report such measures these values will enable comparisons with other trials and help inform conduct of future Cluster RCTs.

This is a very good suggestion, and as the reviewer notes, beyond the scope of this manuscript.

12. Because of the proven incremental effect of OSS, the authors need to give a clear and careful description of OSS and what was unique about OSS that led to a significant interaction with IMID training. Factors that can be highlighted include the effect modifiers of feedback highlighted in the Cochrane review on audit and feedback by Ivers and colleagues.

The description of OSS in the revised manuscript has been clarified as follows:

The OSS intervention combined educational outreach and continuous quality improvement (CQI). A mobile faculty team visited each facility in arm A for two days, once a month for a period of nine months. Each mobile faculty team was comprised of a medical officer with CQI expertise, a clinical officer, a nursing officer, and a laboratory technologist. OSS sought to improve individual practice through clinical mentoring. At each of the facilities in arm A, eight clinicians spending 80% of their time on patient care were selected for one-to-one mentoring by two clinical faculty (medical officer and clinical officer) while the laboratory staff were mentored by the laboratory technologist. OSS also sought to build and foster team work, improve facility performance, and support the use of data to monitor facility performance through multidisciplinary team (MDT) training, cadre-specific clinical breakout sessions, and CQI activities. All clinical staff were invited to participate in the MDT training sessions and their respective cadre specific break-out sessions. The break-out session for clinicians may have also contributed to improving clinical practice. IMID trainees in arm A were required to attend seven of nine OSS sessions to receive a certificate. During each OSS visit, the MDT and cadre-specific sessions were focused on a selected topic from a predefined list of priority areas. A new topic was addressed each visit, and follow-up support on topics covered in the previous visits was also provided. OSS is described in more detail in Miceli et al. (28) Naikoba et al. (29) and Mbonye et al. (24).

O’Brien et al.’s Cochrane Review (2007) on educational outreach did not report effect modifiers that we could highlight in this manuscript.

Paul Mullan Reviewer’s report

• Discretionary Revisions
1. You mentioned a few task shifting articles (references 10-12) related to inpatient HIV testing, pharmacy refills, and once on HIV primary care roles. References #10 and 11 are less related to your article’s content than #12. Consider adding an article in addition to #12, from a more resource constrained setting (than South Africa) that relates to your article’s topic, such as Monyatsi, 2011 (PMID 22273135).

References 10 and 11 were replaced by Huicho et al, 2008, which was previously cited, and Monyatsi, et al. 2011.


Revision made.

3. Methods line 296 – this limitation could be in the limitations section.

The following sentences were added to the limitation section,

“Trainees and observers did not know the allocation of sites to arm during most of the baseline data collection, but they knew during the intervention and endline data collection. The mobile team members did not observe their mentees at endline, but they could have been biased in favor of intervention arm.”

4. Results. Line 383 – consider adding a line that states something about “The proportions of each type of provider per treatment arm were not significantly different (p<0.05)” to allay reader fears that the samples were not equal (by eyeball calculation, they look approximately the same).

Our colleagues who are biostatisticians discouraged us from performing statistical tests on anything other than our hypotheses. We are concerned that there were more registered midwives in arm B than arm A, and added the following sentence, “As noted above, we controlled for trainee cadre in the analysis to adjust for the potential of residual confounding.”

5. Line 409 – “described”

Thank you. Correction made.

6. Line 490 : into # in

Thank you. Correction made.

7. Line 504 – unclear sentence.

The sentence was deleted. Please see the revised paragraph in response to Comment 8 below.
8. Line 497 – Consider adding a reason as to why you believe treatment and pt/caregiver education did not improve significantly.

The paragraph on the four sets of tasks for laboratory, diagnosis, treatment and patient/caregiver education was rewritten as follows:

“Our results that IMID and OSS improved patient history and physical examination are consistent with the results from the IMCI multi-country evaluation (15-18). We did not find strong evidence that IMID and OSS improved laboratory tests, diagnosis, treatment, and patient/caregiver education, whereas several evaluations of IMCI have shown effects on treatment and patient/caregiver education (16-18) (22). There are three potential explanations for the absence of effects on these sets of tasks: 1) the intervention was not effective, 2) trainee practice on these tasks was higher at baseline than on patient history and physical examination, and had less room for improvement, and 3) the full effects were not measured because of the structure of the clinical assessment. Concerning the third explanation, the IDCAP clinical assessment was structured with an interruption after the patient history and physical examination when the faculty completed or corrected them. From this point forward, the trainee had the results of a complete patient history and physical examination and may have been able to prescribe correct treatment and provide appropriate information to patients and caregivers. The intervention may have had larger and significant effects on these tasks if the clinical faculty did not intervene to complete the patient history and physical examination.”

9. Line 521 – insured # ensured

Thank you. Correction made.

• Minor Essential Revisions

1. Introduction p 5, line 79-80 – please expand on how the supervision visits were “effective” and how supervision “may help maintain health worker skills.” As described here, it leaves the reader wondering just how effective these interventions were, and these interventions form the basis of this manuscript.

The introductory paragraph on the effect of IMCI training and supervision has been revised as follows:

“Evidence from the Integrated Management of Childhood Illness (IMCI) Multi-Country Evaluation showed that training improved the quality of care, but there was room for further improvement (15-19). Among health workers with IMCI training, the quality of care was better with at least one supervision visit every six months in Uganda (15), and with study supports including supervision in Benin (22). Horwood et al. (2009) recommended further research on the role of supervision to maintain IMCI skills and on different models of supervision (21). The Joint Uganda Malaria Training Program (JUMP) combined classroom sessions, practice and supervision visits, and was effective at improving case management of children with fever (20). Several of these facilities achieved high levels of performance after four years of ongoing site visits with data surveillance and feedback.”
2. Methods – line 324. Six sets of categories were tested at 0.05 significance, giving approximately a 0.30 chance of a type I error. Consider either adjusting for the fact that you had six categories (e.g. Bonferroni method: 0.05/6 = 0.008; if no a priori hypothesis existed for which of the 6 would be significant) or choosing one as your primary outcome for your hypothesis (if one existed a priori to your experiment). Or – expand/explain this “caveat” and its effect on significantly increasing the chance of a type 1 error.

In the Data Management and Statistical Methods section, we expanded on the caveat as follows: “A 5% level of significance was used with the caveat that there were multiple comparisons, which increased the chances of a Type I error, i.e. the probability of erroneously concluding that there was an effect of the IDCAP interventions.”

The sentence on multiple comparisons in the Limitations section has been revised to state, “Given that a 5% level of significance was used despite multiple comparisons, it is possible that we erroneously concluded that the effects of the interventions were statistically significant.”

3. Methods – lines 112-116 – This 3 sentence paragraph is confusing and its content either need to be incorporated further down in the methods section after the outcomes are described, or integrated elsewhere. Revisions to a tool are being discussed before the tool is discussed.

Please see the response to Dr. Ayieko’s comment 9 above.

4. Methods – line 118 – Please clarify that these trainees are the MLPs that were discussed earlier.

The sentence was revised as follows, “The MLP trainees included clinical officers (CO) and registered nurses (RN) who managed patients in the outpatient clinics, devoted at least 80% of their time to patient care, and were available to participate in the evaluation for 21 months.”

5. Methods line 252 – remove “the”

Thank you. Correction made.

6. Results line 386 – this was not an exclusion criteria mentioned earlier. Either add this exclusion criteria to the methods or include this MLP in the analysis.

The following exclusion criteria was added to the subsection on Trainees, “For the clinical assessments, trainees were excluded from the analysis if they were observed managing patients 14 years of more of age.”

7. Line 454 – Non significance is not “weak evidence.” It is “not significant.” Likewise line 473, if the CI truly includes 1.0, then one can not say it is statistically significant (if this is a “round upwards” from 0.9958, then you can’t; if it is a round downwards, from 1.0024 then you can).

The sentence was revised as follows, “The improvements in physical examination (aRRR=1.27 (95% CI=0.99-1.63)) and laboratory tests (aRRR=1.24 (95% CI=0.99-1.54)) were not statistically significant.”
The confidence interval is (1.002924, 1.599385), and we rounded downwards. In checking the results however, we noted that the result is for missing = 0 rather than missing=1, and the statement has been correct in the revised manuscript.

8. Line 491 – Repetition is also what engrains bad habits of practice. Remove this sentence or the first half of it as the data from this study don’t necessarily prove how the art of learning is achieved.

The sentence was removed.

• Major Compulsory Revisions

1. Intro – line 89-90; Please expand this sentence in 1-3 sentences to describe the effects on these variables by the intervention. This helps the reader understand the clinical impact of these interventions without needing to read 3 more articles.

The introductory paragraph on IDCAP has been revised as follows:

“The objectives of the Integrated Infectious Disease Capacity Building Evaluation (IDCAP) were to design two integrated training interventions for mid-level practitioners (MLP), evaluate their effectiveness, and estimate their cost-effectiveness at 36 facilities in Uganda. The two interventions were an Integrated Management of Infectious Diseases (IMID) training program, and on-site support (OSS). The effects of the interventions on the clinical competence and clinical practice of individual MLP, 23 facility performance indicators, and mortality among children under five years of ages were tested. Clinical competence measured by vignettes, which are sometimes referred to as case scenarios, increased significantly after the 3-week core IMID course and persisted for 24 weeks; no incremental effect of OSS was observed (23). Similarly IMID was associated with statistically significant improvements in two facility performance indicators, and the combination of IMID and OSS were associated with statistically significant improvements in five.(24, 25). Despite large incremental effects of OSS on 10 indicators, none were statistically significant. The results for clinical practice in management of common childhood illnesses are reported below.”

2. Methods, line 97; describe what a mid-level practitioner’s training entails (e.g. 4 years of nursing school, etc.).

The following sentences about the training of the mid-level practitioners were added:

“The clinical officers had a secondary school education, three years of pre-service training, and two years of internship. The registered nurses and registered midwives had a secondary school education, and three years of pre-service training.”

3. Results – line 423 – This sentence sounds contradictory. “no difference in practice... except lower appropriate lab tests ordered and higher % of correct diagnoses.” Please correct this sentence; and, please put some data (percentages, p values, something...) to let the reader know what “lower” and “higher” really mean. The validity of this intervention is largely based on the question “Were these two groups the same at baseline?” This must be described in detail where data are available. Clarifying this will also help line 458’s validity.
The sentence about the baseline results was revised as follows:

“There was no statistically significant difference in the practice of MLP in arm A compared to arm B at baseline (time 0), but the percentage of appropriate laboratory tests ordered was lower (aRR=0.85 (95% CI=0.72-1.01)) and the percentage of correct diagnoses was higher (aRR 1.14 (95% CI=0.98-1.33)) in arm A than arm B. For the primary model, the multivariate analysis controls for differences across arms at baseline."

The sensitivity analysis of endline data does not control for differences across arms at baseline, but differences across arms were not statistically significant for laboratory tests and correct diagnoses.

4. This was described as a continuous quality improvement project, but traditional statistics were used for all of the analyses. I was surprised not to see any of standard QI-related statistical process control (SPC) charts or run charts to describe any of the QI findings from the study. SPC charts would account for some of the changes over the course of the intervention rather than looking at the combined statistics for each time periods and drawing conclusions. With over 300 observations per group, you could easily draw proportion charts (p charts) over each of the three month periods for arms a and b so that the reader could see whether the effect of IMID on A and B was significant in and of itself over the three month period, and whether the endline period had any similar trends. This would be much more revealing to what actually was happening patient to patient. I took a picture (might not transfer in BMC upload version but this book is likely owned by authors or easily accessible) from p. 16 and 17 of Provost,Murray “The Health Care Data Guide, 2011” to illustrate this issue. The results from this study mimic figure 1.5 showing a before and after change standard model. Even though you don’t have continuous data from Jan 2010 to Feb 2011, you do have two continuous time periods (Jan 20 to Mar 10) and (Dec 10 to Feb 11) from which you could look at the phenomena of the training itself, and importantly how differences were seen from the beginning to end of each time period. Were the findings from arm A similar to Figure 1.6 Case 5 where the improvement in Arm A happened before the OSS intervention (in the latter half of the first period), but such an effect didn’t happen in Arm B? Or one of many other scenarios. If you don’t have the capability to do an SPC chart for these data, then it should at least be mentioned in the limitations that an SPC chart would account for the time based nature of the data more predictively for the “next patient.” The first author is at Baylor College of Medicine (where I trained as well), where SPC charts are routinely used in QI projects – the Chief Quality Officer there, Dr. Eric Williams, could either help or refer the first author to a QI specialist who could determine if p charts could be successfully used for this paper’s analysis.

Thank you for the reference to Provost and Murray, which was published 3 years after our study was initiated and 6 months after our data collection was complete. Figures 1.5 and 1.6 were available on Amazon.com. We understand the role of interrupted time series data in quality improvement research. For example, we have monthly data on the 23 facility performance data and have visually inspected run charts for each of them.

Although the clinical assessment data were collected over a period of three-months, they were limited to two time points at each facility. The observers spent two days at each facility in each time period, one day devoted to the observing the management of childhood illness and the other devoted to observing
HIV clinical practice. Each mobile team could visit at most two facilities per week, including travel time, and it took about four weeks to visit all 36 facilities. Add in holidays and scheduling return visits to the facilities when trainees were not available for observation or changed at the last minute, and the process extended for three months.

The following sentences were added to the limitations section:

“The clinical assessment data were limited to two time points at each facility, and don’t provide details of the monthly progress in clinical practice. To the extent that there were temporal trends in clinical practice arm B would control for them, so that the effect of OSS is measured accurately. The pre/post change in arm B however, could reflect these temporal changes as well as the effects of IMID.”