Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. Knowledge of principles of research ethics and guidelines (Page 12).

Only two principles are listed in the pre-test results. Is this because participants could only identify two? If not, what was the third?

Response: The first process in analyzing the answers to the question on principles of research ethics was to enumerate all the responses mentioned by trainees. Codes were then assigned to the answers indicating whether the response corresponds with any of the three principles. At baseline trainees provided answers directly related to beneficence and justice but none provided answers related to respect for persons. Some of the terms listed by trainees which were not directly related to any of the principles are were ‘risks associated with research’, ‘conduct ethically sound research’, non-personalizing issues, ‘non-exploitation’, ‘notification of side effects’. The answers that are not directly related to any of the principles of ethics were regarded as incorrect.

2. In the same section, it is reported: ‘The mean scores for knowledge of the principles of research ethics rose from 0.67 out of 3 points to 2.25 at post-test; this dropped to 2.19 at follow-up (p<0.05).’ This is the most impressive increase in knowledge reported in the paper, but it is not clear how this was tested, nor to which question in the questionnaire these results relate as separate results are reported for questions 5 and 6. Please can the authors clarify this?

Response: As stated in the appendix, question 5 assessed trainees’ knowledge
of the three principles of ethics. We analyzed the results of this question by assigning one mark for each of the correct answer given to each of the principles of ethics, namely respect for persons, beneficence and justice. We generated a three-point mean knowledge score for this item. We analyzed knowledge about international guidelines separately as reported on page 12.

3. Application of the principles of research ethics (page 13)

The results for this section are presented as percentages in Table 2, so it is not clear how the authors calculated the increase in mean scores from 2.64 at pre-test to 3.05 and 3.24 at post-test and follow-up respectively (p<0.05). Please can this be explained. This also applies to the following section which also reports scores out of 7, rather than reflecting the percentages in table 3.

Response: The proportion shown on Table 2 refers to those who gave correct answers to the application of the principles of research ethics; we summarized the findings for the 7 statements by assigning one point to those who provided correct answers leading to a 7-point mean knowledge score. The results of this are presented as mean scores. The same applies to the data on Table 3 (see page 11-12).

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions (which the author can choose to ignore)

1. In Table 2 (which somewhat confusingly reports the answers to questions 8-14 but is not labelled as such), it would help the reader to know which principles are considered to be the correct applications for the statements, rather than labelling each statement true or false.

Response: The correct applications of each of the statements are now indicated in Table 2 (page 24).

2. In the discussion, it may be worth commenting further on the difficulties of testing the effectiveness of research ethics training using a questionnaire. I'd be interested in the authors' views about alternative methods, for example, interviews with a selection of participants to identify what they found most valuable about the course, or repeat interviews with members of the IRB at 12 months to see if the quality of applications had changed following the workshop.

Response: We have discussed many of the limitations of use of questionnaires to determine impact of training on research ethics (see discussion section pages 17-18).
Reviewer's report
Title: Outcome of a training program on research ethics among clinicians and scientists in a Nigerian university
Reviewer: Michael Kalichman
Reviewer's report:

General

-------------------------------------------------------------------------------

Major Compulsory Revisions
1. Data analysis: The authors note using "Chi square test", but it isn't clear how this was done. Were these 2 X 2 analyses, or were multiple factors included in a single analysis? And how were variables categorized to allow for the analysis?
Response: We analyzed the results of this question by assigning one mark for each of the correct answer given to each of the principles of ethics, namely respect for persons, beneficence and justice. We generated a three-point mean knowledge score for this item and compared the findings by time (before, after and follow-up). We analyzed knowledge about international guidelines separately as reported on page 12.

2. Response Rate: Apparently 97 out of 133 persons completed both the pre and post-test survey. If this is correct, then is it also correct to assume that some people completed only the pre-test and some only the post-test? If so, then those numbers should be provided and it should be made clear how those results fit in with the other results already reported. Also, if the authors know that particular individuals completed both tests, then is it correct to assume that they can identify pre- and post-tests by the individual? If so, then a more powerful analysis might have assessed the change in individuals (e.g., with a paired t test or with a Wilcoxon signed rank test)

Response: A total of 133 persons participated in four rounds of training programs. However, only 97 completed the pre-post tests while only 59 completed the questionnaire at follow-up. The results of the data presented in this paper refer to these three categories of persons i.e. 97 at both pre-test and post tests and 59 at follow-up. Thirty-six trainees did not complete the pre-test because they either arrived late (i.e. after the pre-test had been conducted) or did not provided informed dissent. Those who did not complete the pre-test were told not to take part in the post-test or follow-up survey. The authors did not know the exact identity of the particular individuals who completed the questionnaire because it was anonymous. All the 97 trainees who completed the pre-post test were given the questionnaire with instructions that this should be completed one month later. These persons were contacted later for collection of the
questionnaire; we assume a return of the questionnaire as informed consent.

3. Case studies: The statements to be ranked in the case studies did not always seem clear. In several instances, it seems that the participants would not have a reasonable basis for judging the statement (e.g., "The study was not reviewed by an ethics review committee" – the fact that a review committee is not mentioned does not mean that it hasn't been reviewed. Is it possible that many/most of those who answered this question incorrectly had marked it "Disagree" simply because they had no way of knowing that a review had not actually occurred?)

Response: We agree with the reviewer that it is possible that individuals who disagreed to the statement "the study was not reviewed by an ethics committee" in the case study may have done so because this information was not provided in the case study. We have pointed out this limitation in the discussion section (see page 17).

Minor Essential Revisions

4. Endpoints: It appears that big changes were documented in endpoints such as the ability to identify relevant documents. However, more important endpoints (e.g., understanding of ethical principles or knowledge of IRB operations) improved only minimally, if at all. Perhaps this discrepancy should be clearly noted, with an emphasis on the need for particular types of educational approaches different from those used here?

Response: We have discussed the implications of this problem in the discussion section (see paragraph 4, page 16)

5. Inadequate citation: Ref. [14] is important, but the information given is inadequate for someone who wishes to review the work.

Response: We have provided a full citation for this reference (page 21).

6. Needs Assessment, first paragraph: The authors mention a list of "11 items." It is not clear how this list was created nor is it clear what items are on this list. On the assumption that the next sentence lists the items, it seems that only 10 items are included. What is the correct list and how was the list created?

Response: We acknowledge this mix-up. This problem has been addressed: there are nine items on the list and these were derived from the FHI training manual (page 7).

7. Ethical Challenges (Needs Assessment, item #1): What did the authors mean by "ethical challenges"? Are these supposed to be barriers to good ethical
consideration (e.g., "delay in review") or true ethical dilemmas (e.g., how much can you pay a research subject before it becomes coercive?). It seems that the meaning of ethical challenge is not clear here.
Response: We mean barriers to good ethical consideration; we have modified this sentence as suggested by the reviewer (page 7-8).

8. Ethical Challenges (Needs Assessment, item #3): The content listed here takes many different forms or qualities. It isn't clear how these items can be ranked when some have to do with overriding principles (e.g., "principles of research ethics"), others are specific kinds of principles (e.g., "justice and obligation," which is itself a subset of "principles of research ethics"), and others either overlap or contain the topics (e.g., "informed consent" and IRB).
Response: We have clarified this sentence; it now reads "Each FGD participant was requested to rank three contents in order of their priority." (page 7).

9. Ethical Challenges (Needs Assessment, item #3): It is disturbing that no "participants considered standard of care, obligations of researchers to study participants and conflict of interest to be relevant contents for training." This is worthy of some comment by the authors. Is this because of a lack of understanding by the participants? Because these issues aren't relevant in these communities (and, if so, why not?) Or for some other reason?
Response: We thank the reviewer for making this important point. We have raised this issue in the discussion section (see paragraph two on page 16).

10. Recruitment of trainees: It's worth noting that this recruitment plan looks excellent. However, it isn't clear why the numbers don't add up (32 Departments/Units; 28 responded and 3 did not, what happened to the other Department/Unit?).
Response: We have noted this error and made the correction. The sentence now reads "Of the 32 Departments/Units invited to nominate staff, 29 (91%) responded 3 (9%) did not." (see page 9).

11. Implementation of the Workshops and Measures: It isn't explicitly stated, but presumably these studies were conducted in English. Assuming that English is not the first language for the participants, this raises questions about the linguistic and cultural appropriateness of the training and the assessment tools. And, more specifically, it appears that the training heavily emphasized the Belmont Report principles. While those are established as standard in the U.S., is it appropriate or necessary to use those same principles in other countries...
(e.g., the principle of autonomy may be viewed very differently in other countries). Also, the listed statements in the section on Measures are worded in ways that might be misinterpreted (or "over" interpreted) even by someone for whom English is a first language.

Response: The workshops and all the assessments were conducted in English, the official language in Nigeria. The training emphasized the importance of the principles of research ethics as described in the Belmont Report; however the resource persons provided several examples on the local application of these principles (pages 9-10).

12. Knowledge of principles of research ethics and guidelines: As listed, it appears that trainees were very consistent in their choice of wording before taking the course. Is this because they had some other common training or because the authors have used some method for categorizing their responses? If the latter, then this should be explained.

Response: The first process in analyzing the answers to this question was to enumerate all the responses mentioned by trainees. Codes were then assigned to the answers indicating whether the response corresponds with any of the three principles (see page 11). At baseline trainees provided answers directly related to beneficence and respect for persons but none provided answers related to justice. Some of the terms mentioned which were not directly related to any of the principles were ¿risks associated with research¿, ¿non-exploitation¿, ¿non-personalizing issues¿, ¿notification of side effects¿. The answers that are not directly related to any of the principles of ethics were regarded as incorrect.

13. Knowledge about operations of IRB: In this paragraph, the authors have italicized both the quotations (intentionally) and their own text (unintentionally).

Response: We thank the reviewer for making this observation. We have made the correction; only direct quotations are now italicized.

14. Ethics reasoning in case studies: This first sentence is not clear. What is meant by "this" in the statement "but this dropped to 88% at follow-up"? Presumably, the point is that 95% of post-test respondents recognized Madral's conflict, but only 88% did so one month later.

Response: we have corrected this mistake and revised the sentence as suggested by the reviewer.

15. Discussion, paragraph 3: On what basis do the author's conclude that the participants' understanding was increased?

Response: We have modified this and the sentence now reads ¿improvements were found in participant¿s knowledge of the principles of research, their application, international regulations, and operations of IRB after the training¿.
page 17.