PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Impact of using eHealth tools to extend health services to rural areas of Nigeria: Protocol for a mixed-methods, non-randomised cluster trial
AUTHORS	Ebenso, Bassey; Allsop, Matthew; Okusanya, BO; Akaba, Godwin; Tukur, Jamilu; Okunade, Kehinde; Akeju, David; Ajepe, Adegbenga; Dirisu, Osasuyi; Yalma, Ramsey; Sadeeq, Abubakar; Okuzu, Okey; Ors, Tolga; Jagger, Terence; Hicks, Joseph; Mirzoev, Tolib; Newell, James

VERSION 1 – REVIEW

REVIEWER	Soter Ameh Lecturer, Department of Community Medicine, Faculty of Medicine, College of Medical Sciences, University of Calabar, Calabar, Cross River State, Nigeria
REVIEW RETURNED	04-Jun-2018

GENERAL COMMENTS	This is a well-written study protocol, but the concerns below should be addressed before the protocol is published. It was not stated how the sample size of 300 patients was estimated/derived. The authors mentioned they would do an interrupted time-series analysis for the quantitative component of the study. This is great, but the authors need to be clear about the statistical analytical approach (e.g. segmented regression analysis using ARIMA/ARMA model) that will be used to account for autocorrelation inherent in longitudinal data of this nature in which measures are repeated. I also refer the authors to Diggle's sample size formula for repeated measures in an interrupted time-series data. In as much as the logical framework for the study has been clearly shown, it is not clear how the conceptual framework, which depicts Avedis Donabedian's theoretical framework for evaluating quality of health care, will be operationalised using data analysis to show the relationships between the factors categorised under input, process, output, outcome, and impact constructs. Read Donabedian's classic article on this approach. Although there will be no randomisation of the clusters (LGAs), I think, very strongly, that the Consolidated Standards of Reporting Trials (CONSORT) checklist for reporting trials is a more appropriate reporting document than the STROBE checklist for reporting observational studies. Therefore, in addition to the mERA checklist, the authors should consider the use of the CONSORT checklist because of the interventional nature of the study design indicated in this protocol.

INTRODUCTION SECTION

1. If the acronym “EXTEND” has a full meaning, the authors should define it the first time it is used in line 21 (Page 4 of 20).
2. Specific objective 1 is double-barrelled. The authors should split it into two objectives as suggested below:
    a. Strengthen service delivery through using video training (VTR) app to increase FHW knowledge and skills.
    b. Strengthen data management through the use of Clinical Patient Administration Kit (CliniPAK) app to promote efficiency in data management and use.
3. Specific objective 2 is a stand-alone objective which will become the third objective if objective 1 is split into two
4. Would it not be beneficial for the authors to have a robust dataset/result by including an objective that evaluates health output/outcome of pregnant women in this study? The hypothesis would be that eHealth improves clients’ health output/outcome.
5. Provide the full meaning of FHW the first time it is used in line 4 (page 5 of 20).
6. Although there will be no randomisation of the clusters (LGAs), I think, very strongly, that the Consolidated Standards of Reporting Trials (CONSORT) checklist for reporting trials is a more appropriate reporting document than the STROBE checklist for reporting observational studies. Therefore, in addition to the mERA checklist, the authors should consider the use of the CONSORT checklist because of the interventional nature of the study design indicated in this protocol.

STUDY DESIGN AND METHODS

7. There is no prior definition of the acronym PHCs in line 33 (Page 5 of 20). I guess the authors are referring to Primary Health Care (PHC) facilities. If I am correct, then the authors should consistently replace the acronym “PHCs” with “PHC facilities”.
8. Line 3 in page 6 of 20 should read thus: “The remaining 51 non-SatCom facilities….”
9. It may be necessary to mention in the conceptual framework or elsewhere in the protocol the possibility of study findings being influenced by competing interventions/alternatives in either the intervention or control arm of the study or both that were unknown or unanticipated at the time the protocol was developed.
10. It may also be necessary to state that there could be unintended positive or negative consequences of the intervention that are not mentioned in the protocol.
11. It was not stated how the sample size of 300 patients was estimated/derived. The authors mentioned they would do an interrupted time-series analysis for the quantitative component of the study. This is great, but the authors need

	to be clear about the statistical analytical approach (e.g. segmented regression analysis using ARIMA/ARMA model) that will be used to account for autocorrelation inherent in longitudinal data of this nature in which measures are repeated. I also refer the authors to Diggle's sample size formula for repeated measures in an interrupted time-series data. 12. In as much as the logical framework for the study has been clearly shown, it is not clear how the conceptual framework, which depicts Avedis Donabedian's theoretical framework for evaluating quality of health care, will be operationalised using data analysis to show the relationships between the factors categorised under input, process, output, outcome, and impact constructs. Read Donabedian's classic article on this approach. DISCUSSION 13. There is confusion in the use of the terminologies such as prospective, non-randomised and case-control evaluation in line 54-55 (Page 11 of 20). Is there a design that uses a combination of prospective non-randomised and case-control designs? The authors should provide a referenceif such study design exists.

REVIEWER	Christine A. Hudak, Ph.D, CPHIMS, FHIMSS Kent State University School of Information Kent, Ohio 44242 USA
REVIEW RETURNED	23-Jul-2018

GENERAL COMMENTS	BMJ Review Question 1: The specific objectives of the study are not truly measurable. Objective one states that service delivery and data management will be strengthened. It is unclear what that means. While it can be inferred that it will get better because of the intervention, it would be useful to state a percentage of improvement or some other objective criteria to measure improvement. The second objective "Understand the acceptability to FHW"… is not measurable. Understanding is a vague concept that is relative to each person. Making this objective measurable would strengthen the protocol and make it easier to know if the objectives have been met. Question 5: There is no copy of the consent form within the protocol. While it is appropriate that FHW's be approached twice to gain consent for their participation, the only mention of consent is that "Participants will complete a consent form prior to participation. " If the study is to be replicated, a copy of the consent form should at least be available for review. Comments for Authors · Under strengths and Limitations of the study, the first point, should also include patient outcomes. You mentioned this in both the abstract and the body of the protocol. This is a minor point, but would provide consistency.

	· You will be collecting qualitative data in addition to the quantitative. How will the qualitative data be analyzed (InVivo, Atlas TI?). It was stated that the qualitative data would enable evaluations. A clearer description of how the data will be used would add to understanding. · There were multiple acronyms used in the paper. It became confusing to try to remember them all or to keep referring to the list at the end of the paper. Perhaps looking at the number of times a particular acronym is used and then simply writing it out in the text would be less confusing to the reader. (More frequently used acronyms can stay. Less frequently used can be written out.) · In the description of Phase 1: Baseline assessments and Phase 2. How did you arrive at the numbers of patients, policy makers and facility leads to interview? Were they deliberatelyselected or was it a convenience sample? It would be helpful to know this fact.

VERSION 1 – AUTHOR RESPONSE

Reviewer comments	**Author response to reviewer comments**
Soter Ameh	
It was not stated how the sample size of 300 patients was estimated/derived.	The sample size of 300 patients was chosen based on logistical and resource considerations, i.e. available funds, personnel and equipment. This has been clarified in the "Recruiting FHWs for CliniPAK and VTR interventions" section. Please note that although this sample size was not planned statistically (i.e. no power calculation was done), this is not the sample size for the primary outcome (% missing data in NHMIS indicator "total number of ANC visits"), which was based on a formal sample size calculation as explained.
The authors mentioned they would do an interrupted time-series analysis for the quantitative component of the study. This is great, but the authors need to be clear about the statistical analytical approach (e.g. segmented regression analysis using ARIMA/ARMA model) that will be used to account for autocorrelation inherent in longitudinal data of this nature in which measures are repeated. I also refer the authors to Diggle's sample size formula for repeated measures in an interrupted time-series data.	Thank you for highlighting the need to add further detail about our planned analysis. In relation to the interrupted time-series analysis we agree more detail is needed how we will deal with the possible issue of temporally autocorrelated errors. This has now been added to the "Data analysis" section (page 9). However, we have not provided any power calculation for two reasons. This is because the ITS is simply the method used to analyse the secondary outcomes, and we believe it is fair to say that typically secondary outcomes are not subject to formal sample size calculations, as they are of secondary importance and often exploratory (which certainly applies here). Hence we have only ensured that we have a (hopefully) sufficient level of power for our primary outcome subject to resources. In terms of the planned analytical approach, we will analyse the primary outcome (binary facility-level indicator measuring whether the monthly NHMIS indicator "total number of ANC visits" is

	complete), adjusted for covariates, using a two-stage method that accounts for between-cluster variation and is appropriate for cluster trials with relatively few clusters per arm. First, we will use a logistic regression model of the primary outcome including our covariates of interest, but excluding the treatment effect, to compute a difference residual for each cluster. Second, we will estimate the intervention effect as the absolute difference in the primary outcome (intervention minus control), and base our inference on the associated (t-statistic based) 95% confidence intervals and p-value (two-sided, 5% level of significance). For the interrupted time series analysis, we will use controlled interrupted time-series analysis to analyse the NHMIS indicators themselves, to understand whether there have been any changes in the levels and/or trends of all NHMIS indicators following implementation of the intervention. For all these monthly indicators, we will have 12-months' worth of pre-intervention data and 6-months' worth of post-intervention data for both intervention and control clusters. We will analyse the all NHMIS indicators, variables aggregated at the LGA level, using a linear regression model including a time x treatment x period (pre-intervention vs post-intervention) interaction to provide estimates of the changes in level and trend of outcomes before and after the intervention period. If model errors display non-negligible autocorrelation this will be accounted for using by fitting a generalised least squares model appropriate methods to deal with any problems observed in the models due to the time-series nature of the data adjusting for AR(1) errors. These additional details about our planned analysis have also been added to the manuscript.
In as much as the logical framework for the study has been clearly shown, it is not clear how the conceptual framework, which depicts Avedis Donabedian's theoretical framework for evaluating quality of health care, will be operationalised using data analysis to show the relationships between the factors categorised under input, process, output, outcome, and impact constructs. Read Donabedian's classic article on this approach.	An explanation to clarify how we will use data analysis to assess how we will operationalize the linear/simplified relationship between components of the project (i.e. inputs, processes, outputs, outcomes and impacts) has now been added (page 7). Our project is built on the assumption that the context within which a project is implemented, interacts with and influences project results. So, whilst our programme theory (Figure 1) lends structure from Donabedian's theoretical framework, to depict linear and simplified relationships between the project components, nonetheless, we will use insight from the analysis of documents review and qualitative interviews (see "methods of data collection" section of protocol) to assess whether and how the context of implementation of the project affects project results and the relationship between the various components of project. We have also provided examples (in response to a comment from

	Christine Hudak), of how context might affect project result or the relationship between components of the project.
Although there will be no randomisation of the clusters (LGAs), I think, very strongly, that the Consolidated Standards of Reporting Trials (CONSORT) checklist for reporting trials is a more appropriate reporting document than the STROBE checklist for reporting observational studies. Therefore, in addition to the mERA checklist, the authors should consider the use of the CONSORT checklist because of the interventional nature of the study design indicated in this protocol.	Thank you for this suggestion. We have now included (in addition to the mERA checklist, the Consolidated Standards of Reporting Trials (CONSORT) checklist as a checklist for reporting our study.
Christine Hudak	
Introduction: If the acronym "EXTEND" has a full meaning, the authors should define it the first time it is used in line 21 (Page 4 of 20).	The title, EXTEND, is not an abbreviation of a specific title, but was selected as a noun derived from the fact that the project is using Satellite Communication to host and extend e-Health interventions and basic health services to technologically disadvantaged areas of Nigeria.
Specific objective 1 is double-barrelled. The authors should split it into two objectives as suggested below: a. Strengthen service delivery through using video training (VTR) app to increase FHW knowledge and skills. b. Strengthen data management through the use of Clinical Patient Administration Kit (CliniPAK) app to promote efficiency in data management and use.	Thank you for this useful suggestion, we have split objective 1 into two objectives as advised, and have subsequently revised the way in which these objectives are framed to ensure they are measurable, in response to a comment from the BMJ Open editorial team (see first comment from BMJ Open below for further details).
Specific objective 2 is a stand-alone objective which will become the third objective if	Thank you for this reminder; following the split of objective 1 into two separate objectives, we have updated the previous

objective 1 is split into two	standalone objective to become objective 3.
Would it not be beneficial for the authors to have a robust dataset/result by including an objective that evaluates health output/outcome of pregnant women in this study? The hypothesis would be that eHealth improves clients' health output/outcome.	We thank the reviewer for this suggestion. However, our interventions (SatCom, VTR and CliniPAK) are mainly "supply-side" interventions that are implemented at the health systems and health facility level, by health professionals, to directly influence outputs and short-term outcomes such as i) accuracy and completeness of data generated, ii) timeliness of generation and transmission of data to local government headquarters, iii) changes in test scores for health workers, iii) change in confidence and staff motivation to deliver their clinical tasks . So, as the project is neither implemented at the patient level, nor does it include "demand-side" activities such as stimulating service uptake by pregnant women, we believe that the SatCom, CliniPAK and VTR interventions are unlikely to significantly effect the health outcomes of pregnant women (although we are looking at, as secondary outcomes, change in i) numbers of pregnant women attending ANC, ii) proportion of pregnant women delivering in PHC facilities and attended by skilled birth attendants, iii) attending PNC), during the 2-year life span of the project). For the foregoing reasons, we have decided not to focus on measures of health outcomes of women in the study. Additionally, we inferred from contextual evidence available from qualitative interviews conducted during the mid-line evaluation of the project (see Phase 2), that general improvements noticed in some health indicators were largely due to other ongoing, concurrent MCH initiatives (such as the "Save One Million Lives initiative", and the "Midwives Service Scheme") that have been implemented in the project areas for over 4 years. In other words, while the EXTEND project is contributing to improved data management and indirectly to patient care at facility level, we believe the lifespan of VTR and CliniPAK implementation (<1year at the time of conducting the mid-line evaluation) was too short to make a significant change to health outcomes of patients.
Provide the full meaning of FHW the first time it is used in line 4 (page 5 of 20).	We have amended this manuscript in line with this comment (FHW = Frontline Health Workers).
Study design and methods: There is no prior definition of the acronym PHCs in line 33 (Page 5 of 20). I guess the authors are referring to Primary Health Care (PHC) facilities. If I am correct, then the authors should consistently replace the acronym	We have amended the manuscript in line with this comment. The acronym in line 33 (page 5 of 20) PHCs has been defined as Primary Health Care facilities. Furthermore, we replaced all acronyms "PHCs" with "PHC facilities".

"PHCs" with "PHC facilities".	
Line 3 in page 6 of 20 should read thus: "The remaining 51 non-SatCom facilities…."	We have amended the manuscript in line with this comment.
It may be necessary to mention in the conceptual framework or elsewhere in the protocol the possibility of study findings being influenced by competing interventions/ alternatives in either the intervention or control arm of the study or both that were unknown or unanticipated at the time the protocol was developed.	We have amended this manuscript in line with this comment. See page 7 of 23 under the subheading "Conceptual framework".
It may also be necessary to state that there could be unintended positive or negative consequences of the intervention that are not mentioned in the protocol.	We have amended the manuscript to acknowledge that there could be unintended positive or negative consequences of the intervention that are not mentioned in the protocol as part of our overview of the conceptual framework (page 7).
Discussion: There is confusion in the use of the terminologies such as prospective, non-randomised and case-control evaluation in line 54-55 (Page 11 of 20). Is there a design that uses a combination of prospective non-randomised and case-control designs? The authors should provide a reference if such study design exists.	The description of the study design has been altered from "prospective, non-randomised, case-control evaluation" to "mixed-methods, non-randomised cluster trial" which is reflected in the title and an accurate description of the study design.
BMJ review	
The specific objectives of the study are not truly measurable. Objective one states that service delivery and data management will be strengthened. It is unclear what that means. While it can be inferred that it will get better because of the intervention, it would be useful to state a percentage of improvement or some other objective criteria to measure improvement.	We have reviewed the objectives to enable changes resulting from eHealth interventions to be measured. For the first objective relating to the quantitative component, we first broke it down into two parts as advised by Christine Hudak (further details above). To ensure consistency with the logframe, the revised objectives now read, to: 1. Strengthen service delivery through enabling access to a video training (VTR) app that targets knowledge and skills, with >65% of FHWs showing improvements between pre- and post-test assessments 2. Strengthen data management using the Clinical Patient Administration Kit (CliniPAK) app to enable >90% participating PHC facilities to transmit accurate and timely data to LGA

	headquarters
The second objective "Understand the acceptability to FHW"… is not measurable. Understanding is a vague concept that is relative to each person. Making this objective measurable would strengthen the protocol and make it easier to know if the objectives have been met.	As per the previous comment, we have revised this objective to enable it to be clearly measurable, with the revised objective now reading: 3. Identify factors that influence the acceptability and use of VTR and CliniPAK at scale for Frontline Health Workers (FHWs)
Question 5: There is no copy of the consent form within the protocol. While it is appropriate that FHW's be approached twice to gain consent for their participation, the only mention of consent is that "Participants will complete a consent form prior to participation. " If the study is to be replicated, a copy of the consent form should at least be available for review.	We have included the consent form as Appendix 2, contained in the main document.
Under strengths and Limitations of the study, the first point, should also include patient outcomes. You mentioned this in both the abstract and the body of the protocol. This is a minor point but would provide consistency.	We have addressed this point and included patient outcomes under the strengths and limitations section (page 3).
You will be collecting qualitative data in addition to the quantitative. How will the qualitative data be analyzed (InVivo, Atlas TI?). It was stated that the qualitative data would enable evaluations. A clearer description of how the data will be used would add to understanding.	We have specified that we will use manual analysis and framework analysis. Quantitative and qualitative findings will be integrated and triangulated to answer the research questions. Furthermore, we will conduct a comparative analysis of variations in adoption and effectiveness of e-Health innovations in the three states to ascertain the influence of contextual factors on processes of implementation and project outcomes. These details have been outlined on page 10.
There were multiple acronyms used in the paper. It became confusing to try to remember them all or to keep referring to the list at the end of the paper. Perhaps looking at the number of	We have reviewed the use of acronyms, removing 7, with the aim of improving the readability of the manuscript.

times a particular acronym is used and then simply writing it out in the text would be less confusing to the reader. (More frequently used acronyms can stay. Less frequently used can be written out.)	
In the description of Phase 1: Baseline assessments and Phase 2. How did you arrive at the numbers of patients, policy makers and facility leads to interview? Were they deliberately selected or was it a convenience sample? It would be helpful to know this fact.	The number of participants was driven by discussion across the research team based on numbers that would be required to enable capture of necessary insights (informed by pilot testing undertaken in 2016), balanced against the study timeline, resources available and the study objectives. In terms of the sampling use, for qualitative work with policymakers and heads of facilities, purposive sampling was used. For service users, convenience sampling was used. We have added details of the sampling approach to the 'Methods of data collection and sampling' section of the manuscript.

VERSION 2 – REVIEW

REVIEWER	Soter Ameh Department of Community Medicine, Faculty of Medicine, College of Medical Sciences, University of Calabar, Cross River State, Nigeria.
REVIEW RETURNED	06-Sep-2018

GENERAL COMMENTS	The concerns raised in my first revision have been addressed

REVIEWER	Christine A. Hudak, Ph.D, CPHIMS, FHIMSS Professor & Director, Health Informatics Program Kent State University School of Information Kent, Ohio USA
REVIEW RETURNED	22-Aug-2018

GENERAL COMMENTS	The manuscript revisions have clarified the issues brought forward in the first review. One very minor issue: on page 8, line 5, there is a "to" missing in the sentence: "beyond the training to enable staff (to) use the tablets." I appreciate the authors' attending to the concerns in the previous submission.