PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Mitigating the mistreatment of childbearing women: Evaluation of respectful maternity care intervention in Ethiopian hospitals
AUTHORS	Asefa, Anteneh; Morgan, Alison; Gebremedhin, Samson; Lemango, Ephrem; Abebe, Sintayehu; Magge, Hema; Kermode, Michelle

VERSION 1 – REVIEW

REVIEWER	Patience Afulani University of California, San Francisco, USA
REVIEW RETURNED	27-Apr-2020

GENERAL COMMENTS	This paper contributes to the limited evidence on interventions to promote respectful maternity care. The paper is well written and results clearly presented. The methods are appropriate although there a number of limitations, which are noted. I just have minor comments: I am a little concerned about the use of multilevel modelling given there are only 3 clusters (3 facilities). Although there are no hard and fast rules, 3 clusters is considered too for multilevel models. There is probably no harm in using multilevel models when not necessary, but it would have been a simpler analysis to include the 3 facilities as predictors in a fixed effects model (with robust standard errors to account for the clustering). This would be helpful to show the difference between the three facilities in the multivariate analysis. If the authors have a good reason for the multilevel models with only 3 clusters, I would suggest running the fixed effects model with facilities as predictors in sensitivity analysis and noting if the findings are the same or different. It would also strengthen the paper to present bivariate results for the various predictors included the multivariable model showing the mistreatment scores pre and post intervention on these predictors. If this additional table was not presented because of limits to number of tables, the current tables 3 and 4 could be combined into table 3, and then table 4 will be the mistreatment scores by the various predictors. It would also be helpful to include the mistreatment scores by facility in this new table. The narrative on the experiences of mistreatment is too long. Given this is already presented in a table, you can just summarize the key findings in one paragraph, instead of the long 1.5-page narrative in the current version. The other narratives of results in table could also be shortened to highlight key findings and not repeat the tables. In the discussion (page 20 line 34 to 37), the authors note that “treating hospitals as random-effects in the statistical model controls for the impact of other interventions that may have happened around the same time as the study intervention.” This is

	not necessarily true as it only accounts for interventions within the facilities. I will edit to note that it "controls for the impact of other interventions that may have happened around the same time in those facilities." A key limitation of this paper is the use of number of mistreatment components as the outcome measure, which the authors acknowledge. Although this is an improvement over using a binary measure, it may still be underestimating the intervention effects as it does not capture changes within a component (e.g. verbal abuse may still be occurring, but at a lower frequency which is not captured by your measure). Using frequency response options instead of binary options could have helped address this, although that is too late to address. I would suggest noting the potential underestimation of the intervention effect size even with your count of mistreatment components measure in the limitations. Thank you for the opportunity to review this important work.

REVIEWER	Everlyn Waweru Institute of Tropical Medicine, Antwerp, Belgium Kenyan National
REVIEW RETURNED	03-May-2020

GENERAL COMMENTS	Thank you for the opportunity to review this thoughtful manuscript on a much needed area of research to improve the quality of care provided to child bearing women. I commend the authors for conducting this research in Ethiopia, a country, like many other low income African countries, where cases of mistreatment of child bearing women are under reported and there is little guidance on how to minimize these occurrences, deal with reported cases or resolve conflict. Questions Methods and materials Page 8 from line 1: In describing the study setting it is stated which hospitals were chosen but the criteria for their selection is unclear – or was it random? Was the intervention conducted in all the hospitals in the SNNPR, or in selected facilities of which three were included in the study? I feel this information is useful for a better description of the study setting. Page 11 line 39: is there any data on the response rate i.e. how many women were approached and number/reasons of refusals? Especially considering negative experiences might lead to refusal to participate in the study? Data management and analysis Explanation of the choice of the reference categories for comparison in model III Why was cadre giving care not included as part of the service characteristics? Results Page 17 line 1-10: the results described emphasize the necessity to consider the cadre of staff that women encountered. From experience in LMIC, women with complications during Pregnancy or delivery; and women who delivered by caesarean section after trial of vaginal or without trial of vaginal delivery might be treated by different cadres of staff – a parameter that is not included as a possible confounder in Model III results. This could also offer an additional perspective to the significant results on Table 5 Page 33 line 10-12

	Page 29 line 7: the age of women who participated in the study ranges from 15 to 44. It might be wise to include information on how was informed consent obtained, especially for women under 18 years of age. Discussion Well discussed with suggestions on a few areas requiring clarification Page 19 line 31: one could argue that the poor understanding of pregnant women's rights to information, privacy and confidentiality and preferences, applies to both health workers' lack of understanding or how to practice; but also the women's lack of knowledge that they should demand those rights. Consequently, this knowledge also needs to be communicated not only to professionals but to expectant mothers and the communities they live with. Page 20 line 1: 'hospitals included in this study do not have a private ward which means that several women are labouring in the same room' - this information comes in the discussion without mention in the study setting or results section i.e. is it possible to include this details in the description of the three hospitals section? Page 20 line 30-44: are your results in line with these arguments or contrary? The studies you cite suggests that younger and less experienced providers may be less supportive during labour; in the same vein one could argue that it may be easier to change behavior in younger providers who have had less exposure to negative normalized behavior – I feel this section would benefit from a more robust discussion of what impacts practice but also change in behavior since the article is a pre-post analysis. Page 21 line 22 and 23: information on political violence should have been reported in the results section (who / what was the source of this information) or what form of violence should be presented in the results section if it is to be included in the discussion section, otherwise it leaves the reader with more questions of the degree of its influence on the intervention (was it considered in analysis?) Page 21 line 44: how many women who participated in the study were admitted in a shared ward? This information is not presented in the women's obstetric and maternal health characteristics in the results (neither in Table 3 nor in Table 5) Curiosity questions i.e. may not require changes to the manuscript Was the utilisation of the hospitals in March similar to July/August? Sometimes the number of pregnant women who have come to deliver can affect workload, amount of time or attention awarded to each mother, staff attitudes etc. Regarding staff turnover: was it always the staff who were trained who offered care to the women and children post intervention? Once again many thanks for the opportunity to review this insightful article. I believe these suggestions will contribute to its technical soundness and ease of understanding. I look forward to reading the published article.

REVIEWER	Christie Pettitt-Schieber University of North Carolina - Chapel Hill, USA
REVIEW RETURNED	25-May-2020

GENERAL COMMENTS	Page 7 row 1 - should be "in the Lancet" not "on the lancet" Page 16 - what percentage of patients experienced failure of providers to obtain consent for procedures in pre- and post-intervention groups? No % cited. Would recommend having someone proofread, there are minor writing issues throughout that could use revision or tightening.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

This paper contributes to the limited evidence on interventions to promote respectful maternity care. The paper is well written and results clearly presented. The methods are appropriate although there a number of limitations, which are noted. I just have minor comments:
Thank you.

1. I am a little concerned about the use of multilevel modelling given there are only 3 clusters (3 facilities). Although there are no hard and fast rules, 3 clusters is considered too for multilevel models. There is probably no harm in using multilevel models when not necessary, but it would have been a simpler analysis to include the 3 facilities as predictors in a fixed effects model (with robust standard errors to account for the clustering). This would be helpful to show the difference between the three facilities in the multivariate analysis. If the authors have a good reason for the multilevel models with only 3 clusters, I would suggest running the fixed effects model with facilities as predictors in sensitivity analysis and noting if the findings are the same or different.

   Response:
   Thank you for the insightful comment. As you have well-described, we also believe that there are no hard and fast rules on the minimum number of clusters to consider multilevel analysis. We chose the multilevel analysis to have a conservative analysis and to overcome the inherent limitations of traditional regression analysis techniques and consider hospital-level residuals. One of the hospitals (Adare) is located in Hawassa city, capital of the regional city, and is the base for the regional health bureau, partner organisations, medical and health sciences teaching colleges which are closely tied with the hospital's operation. This consideration led us to ensure we identified hospital-level residuals in a multilevel model to account for prior level of mistreatment or RMC practice.

   Furthermore, multilevel models treat clusters (in our case hospitals) as a random sample from a set or population of clusters, and helps to make inference to a population of clusters. As we are aiming to take the lessons learned from this intervention to inform the scale up of similar RMC intervention in other hospitals in the country, we believed that the use of multilevel model would be preferable.

As you have nicely suggested, for comparison purpose, we have also run a fixed effects model with robust standard errors which included hospitals along with other variables of model III as fixed effects (the output table is appended to this response). Accordingly, the Aβ for the intervention remained the same; however, the CI was narrower (Aβ = 0.82, 95%CI: 0.76-0.89). Therefore, using multilevel model would make more benefits due to the above-mentioned points, and we prefer to remain with the multilevel model. We have included this information both in the Methods (Page 13 Line 3-5) and Results sections (Page 17 Line 20-21).

2. It would also strengthen the paper to present bivariate results for the various predictors included the multivariable model showing the mistreatment scores pre and post intervention on these predictors. If this additional table was not presented because of limits to number of tables, the current tables 3 and 4 could be combined into table 3, and then table 4 will be the mistreatment scores by the various predictors. It would also be helpful to include the mistreatment scores by facility in this new table.

   Response:
   Thank you for this comment. We have now added a column in Table 5 (now Table 4) which shows the bivariate results and included a narrative in the Results section, including mistreatment scores by hospitals (Page 32, Table 4). Additionally, we have combined Table 3 and 4, as suggested (Page 30-31, Table 3). That has helped also to address the next comment.

3. The narrative on the experiences of mistreatment is too long. Given this is already presented in a table, you can just summarize the key findings in one paragraph, instead of the long 1.5-page narrative in the current version. The other narratives of results in table could also be shortened to highlight key findings and not repeat the tables.

   Response:
   We have now succinctly presented the subsection as per the comment (Page 15 Line 7 – Page 16 Line 23). We have also made revisions with the entire Results sections.

4. In the discussion (page 20 line 34 to 37), the authors note that “treating hospitals as random-effects in the statistical model controls for the impact of other interventions that may have happened around the same time as the study intervention.” This is not necessarily true as it only accounts for interventions within the facilities. I will edit to note that it “controls for the impact of other interventions that may have happened around the same time in those facilities.”

   Response:
   Thank you for the insightful comment. We have now corrected the sentence accordingly (Page 21 Line 16).

5. A key limitation of this paper is the use of number of mistreatment components as the outcome measure, which the authors acknowledge. Although this is an improvement over using a binary measure, it may still be underestimating the intervention effects as it does not capture changes within a component (e.g. verbal abuse may still be occurring, but at a lower frequency which is not captured by your measure). Using frequency response options instead of binary options could

have helped address this, although that is too late to address. I would suggest noting the potential underestimation of the intervention effect size even with your count of mistreatment components measure in the limitations.

Response:

We have added that as a limitation (Page 22 Line 1-5).

Reviewer: 2

Thank you for the opportunity to review this thoughtful manuscript on a much needed area of research to improve the quality of care provided to child bearing women. I commend the authors for conducting this research in Ethiopia, a country, like many other low income African countries, where cases of mistreatment of child bearing women are under reported and there is little guidance on how to minimize these occurrences, deal with reported cases or resolve conflict.

Thank you

Methods and materials

1. Page 8 from line 1: In describing the study setting it is stated which hospitals were chosen but the criteria for their selection is unclear – or was it random? Was the intervention conducted in all the hospitals in the SNNPR, or in selected facilities of which three were included in the study? I feel this information is useful for a better description of the study setting.

   Response:

   Thank you. The hospitals were selected purposively, and the intervention was implemented only in these three hospitals; we have added a description in the Methods (Page 7 Line 11-14).

2. Page 11 line 39: is there any data on the response rate i.e. how many women were approached and number/reasons of refusals? Especially considering negative experiences might lead to refusal to participate in the study?

   Response:

   Yes, there is. Only four women declined to participate in the survey. We have now included information on response rate in the Methods (Page 10 Line 14-17).

Data management and analysis

3. Explanation of the choice of the reference categories for comparison in model III

   Response:

   Thank you! Reference categories were chosen using two strategies: normative, using the largest category, and aligned with existing evidence of factors associated with the mistreatment of women. As there are no hard and fast rules on these, we believe that readers would easily interpret the reported measures of associations.

4. Why was cadre giving care not included as part of the service characteristics?

   Response:

   The question soliciting handling cadres was removed from the survey tool because there is no practice of using badges in the study hospitals, and consequently women respondents may not

be able to accurately identify their providers' cadre, and more than one provider could assist women during labour and childbirth. Therefore, that variable was not included in the analysis. The primary author has conducted similar studies before in the same setting and the information generated through women's survey on such variables is of poor quality.

Results

5. Page 17 line 1-10: the results described emphasize the necessity to consider the cadre of staff that women encountered. From experience in LMIC, women with complications during Pregnancy or delivery; and women who delivered by caesarean section after trial of vaginal or without trial of vaginal delivery might be treated by different cadres of staff – a parameter that is not included as a possible confounder in Model III results. This could also offer an additional perspective to the significant results on Table 5 Page 33 line 10-12

   Response:

   That is logical insight. We have added a sentence in the Discussion to reflect on that (Page 19 Line 14-16).

6. Page 29 line 7: the age of women who participated in the study ranges from 15 to 44. It might be wise to include information on how was informed consent obtained, especially for women under 18 years of age.

   Response:

   Women between the age of 15 to 17 and who already have a child/children are considered as fit to provide informed consent in Ethiopia; we used two different forms one for literate and the other for illiterate women. Data collectors read the plain language statement (information sheet) for women. There was also a separate consent sheet with a checklist for illiterate women which was signed by data collectors to confirm that they have read all required information for a woman and the woman has agreed to participate. We have included an ethics statement as per the journal's requirement (Page 24 Line 2-8).

Discussion

Well discussed with suggestions on a few areas requiring clarification

7. Page 19 line 31: one could argue that the poor understanding of pregnant women's rights to information, privacy and confidentiality and preferences, applies to both health workers' lack of understanding or how to practice; but also the women's lack of knowledge that they should demand those rights. Consequently, this knowledge also needs to be communicated not only to professionals but to expectant mothers and the communities they live with.

   Response:

   Thank you. "Among providers" was missing from the sentence and we have now added that (Page 19 Line 11-12). In the sentence, we aimed to focus on why there was no change despite the intervention and what could be done. As women in both the pre- and post-intervention groups did not receive any intervention, it would be difficult to make conclusions on their knowledge of RMC in this study.

8. Page 20 line 1: ‘hospitals included in this study do not have a private ward which means that several women are labouring in the same room’ - this information comes in the discussion without mention in the study setting or results section i.e. is it possible to include this details in the description of the three hospitals section?

   Response:
   We have removed the information from the Discussion (to avoid repetition) and included it in the Methods section (Page 7 Line 14-15).

9. Page 20 line 30-44: are your results in line with these arguments or contrary? The studies you cite suggests that younger and less experienced providers may be less supportive during labour; in the same vein one could argue that it may be easier to change behavior in younger providers who have had less exposure to negative normalized behavior – I feel this section would benefit from a more robust discussion of what impacts practice but also change in behavior since the article is a pre-post analysis.

   Response:
   Thank you. Arguments could be made on both directions. If young graduates receive adequate behavioural change intervention before they are deployed, they will become powerful change agents; we have referred to a lesson from a Tanzanian study and cited it in the Discussion. Accordingly, we have inserted a new sentence in the paragraph (Page 20 Line 13-15).

10. Page 21 line 22 and 23: information on political violence should have been reported in the results section (who / what was the source of this information) or what form of violence should be presented in the results section if it is to be included in the discussion section, otherwise it leaves the reader with more questions of the degree of its influence on the intervention (was it considered in analysis?)

    Response:
    There are various sources of information for the crisis in the study area (Sidama Zone, SNNPR region) https://www.theafricareport.com/18565/abiy-ahmed-and-the-struggle-to-keep-ethiopia-together/. Not only that, there have been nationwide instabilities due to a political transition in the country which are still ongoing. While the paragraph is supported by evidence, we have now recognised that this issue is very sensitive, and we would like to remove the paragraph from the Discussion (Page 21 Line 5-11).

11. Page 21 line 44: how many women who participated in the study were admitted in a shared ward? This information is not presented in the women’s obstetric and maternal health characteristics in the results (neither in Table 3 nor in Table 5)

    Response:
    All women in the study hospitals were admitted in shared wards as none of the study hospitals have private wards. Consequently, we did not include this in the results tables. We have added a description in the Methods (Page 7 Line 13-15) and a phrase in the Discussion (Page 19 Line 24).

Curiosity questions i.e. may not require changes to the manuscript

12. Was the utilisation of the hospitals in March similar to July/August? Sometimes the number of pregnant women who have come to deliver can affect workload, amount of time or attention awarded to each mother, staff attitudes etc.

    Response:

    Yes, it is almost consistent. We have checked the client flow of the preceding year (2017) of all hospitals during the design of the study.

13. Regarding staff turnover: was it always the staff who were trained who offered care to the women and children post intervention?

    Response:

    None of the trained staff left the hospitals until this study was concluded. However, there were five providers who did not attend the face-to-face training from one of the hospitals. We have mentioned that in the methods. However, these providers took part in the supportive supervision visits and were provided with the training manuals and orientation by their supervisor.

Reviewer: 3

1. Page 7 row 1 - should be "in the Lancet" not "on the lancet"

   Response:

   We have made changes accordingly (Page 5 Line 26).

2. Page 16 - what percentage of patients experienced failure of providers to obtain consent for procedures in pre- and post-intervention groups? No % cited.

   Response:

   As part of succinctly presenting the Results section and not repeating figures in the tables, as per the comments of Reviewer 1, we have revised this section to briefly describe the findings and remove the proportions of most of the narrated variables. Instead, readers are referred to Table 3. (Page 15 Line 12-24).

3. Would recommend having someone proofread, there are minor writing issues throughout that could use revision or tightening.

   Response:

   Thank you. We have proofread the paper for minor issues as per the comment.

Fixed effects model with robust standard errors

```
Iteration 0:   log pseudolikelihood =  -753.4971
Iteration 1:   log pseudolikelihood = -745.81772
Iteration 2:   log pseudolikelihood = -745.79521
Iteration 3:   log pseudolikelihood = -745.79521

Poisson regression                              Number of obs     =        388

                                                Wald chi2(35)     =     336.97
Log pseudolikelihood = -745.79521               Prob > chi2       =     0.0000
```

NEWCountsOfDisrespectAndAbuseExp	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
QHospital						
Leku	.9290555	.0557592	-1.23	0.220	.8259524	1.045029
Yirgalem	1.222096	.0641635	3.82	0.000	1.102592	1.354553
PreVsPostIntervention						
Post-Intervention	.819171	.033438	-4.89	0.000	.7561868	.8874014
Q__101_Residence						
Rural kebele	1.062235	.0485434	1.32	0.186	.9712282	1.161769
AgeRecoded						
25-34	.9554583	.0550583	-0.79	0.429	.853417	1.069701
35-44	.8079257	.0816873	-2.11	0.035	.6626872	.9849954
___103_Age_at_first_pregnancy_in	1.013433	.0073011	1.85	0.064	.9992237	1.027844
___104_Marital_status						
Married	.797094	.2103299	-0.86	0.390	.4752283	1.336955
Separated	1.047075	.4411449	0.11	0.913	.4585182	2.391106
___105_ReligionN						
Orthodox, christian	.9384883	.0754304	-0.79	0.430	.801704	1.09861
Muslim	1.070404	.0788207	0.92	0.356	.9265486	1.236594
Christian Catholic	1.001655	.1085283	0.02	0.988	.8100117	1.238639
Others	.8088834	.0657676	-2.61	0.009	.6897276	.9486242
___106_EthnicityN						
Amhara	.973585	.1043193	-0.25	0.803	.7891652	1.201102
Oromo	.9315676	.0945635	-0.70	0.485	.7634996	1.136632
Wolayita	1.162562	.0949083	1.85	0.065	.9906643	1.364288
Others	1.008379	.0746753	0.11	0.910	.8721437	1.165894
EducationalStatusMerged						
Primary education	.9555612	.0570413	-0.76	0.446	.8500546	1.074163
Secondary education	.9495579	.074086	-0.66	0.507	.8149095	1.106455
College and above	1.035474	.0961237	0.38	0.707	.8632206	1.242101
___108_OccupationN						
Private employee	1.081835	.1390853	0.61	0.541	.8408663	1.391859
Government employee	.9665975	.1083552	-0.30	0.762	.7759367	1.204107
Private business	1.013767	.0929002	0.15	0.881	.8471007	1.213224
Others	.9823208	.0869227	-0.20	0.840	.8259105	1.168352
__110_Regular_Income						
Yes	.9279531	.0841517	-0.82	0.410	.7768453	1.108453
NumberofDeliveriesRecoded						
Two and more	.8572468	.0577147	-2.29	0.022	.7512736	.9781683
___204_antenatalVisit						
Yes	.9508803	.0762446	-0.63	0.530	.8125944	1.112699
___206_Complic_Durin_Curr_Preg						

Yes	1.15485	.0649418	2.56	0.010	1.03433	1.289413
___207_Complic_Durin_Curr_Labor						
Yes	1.154048	.0554858	2.98	0.003	1.050265	1.268087
___301_Referred_Or_NonReferred						
Non-referred	1.059845	.0518125	1.19	0.234	.9630078	1.16642
HoursOfStay	1.001129	.0010195	1.11	0.268	.9991332	1.00313
___305_Gender_Of_Main_Provider						
Male	1.025867	.0439068	0.60	0.551	.9433217	1.115635
___306_Type_of_Birth						
Caesarean birth after labour trial	.7728305	.0581345	-3.43	0.001	.6668905	.8955998
Caesarean birth without labour trial	.6806892	.0977486	-2.68	0.007	.5137057	.9019519
AssistedDeliveryRegression						
Yes	1.040015	.0607132	0.67	0.502	.9275744	1.166085
_cons	4.715746	1.337949	5.47	0.000	2.704258	8.223426

Note: **_cons** estimates baseline incidence rate (conditional on zero random effects).

VERSION 2 – REVIEW

REVIEWER	Patience Afulani University of California, San Francisco, USA
REVIEW RETURNED	03-Jul-2020

GENERAL COMMENTS	The authors have adequately revised the manuscript in response to reviewer comments