PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Women's Satisfaction with the Quality of Antenatal Care Services Rendered at Public Health Facilities in Northwest Ethiopia: the application of partial proportional odds model
AUTHORS	Emiru, Amanu; Alene, GD; Debelew, Gurmesa

VERSION 1 – REVIEW

REVIEWER	Ashley Sheffel Johns Hopkins Bloomberg School of Public Health USA
REVIEW RETURNED	06-Mar-2020

GENERAL COMMENTS	I appreciate the opportunity to review the manuscript, Women's Satisfaction with the Quality of Antenatal Care Services Rendered at Governmental Health Facilities in Northwest Ethiopia: the application of partial proportional odds model. The authors conducted a health facility assessment comprised of a facility audit, provider interview, direct observation of ANC consultations, and exit interviews with ANC clients in order to holistically measure ANC quality of care (structure, process, and outcomes) and assess the relationship between client satisfaction and structure/process quality. The study has some important results relating to quality of ANC in Northwest Ethiopia. While I think the overall findings are interesting and important, I there are some areas that could be more thoroughly presented to build on the strength of this study. Please see my detailed comments in the attached document. **MAJOR COMMENTS:** **Research objective:** • The research objective is not clearly defined and differs between the abstract and introduction. It would be helpful to perhaps present two research objectives: 1) To describe the quality of ANC at health facilities in Northwest Ethiopia, including dimensions of structure, process, and outcome; and 2)To assess the relationship between ANC client satisfaction with structure and process quality or To assess factors associated with ANC client satisfaction – it's not quite clear to me which if these was the goal. **Methods:**

	• There are some weaknesses in describing the sample size calculation and sampling procedures.
	  ○ What is the sample meant to be representative of? Generally, there is a sample size calculation for facilities, providers, and clients in order to ensure structural/readiness indicators are also representative of the population of interest. Can you further explain the selection of 15 facilities per district and five districts?
	  ○ How were health workers sampled? What was the criteria to be considered a "health provider"? There appears to be a provider interview, but no information provided on how providers were sampled.
	  ○ Can you provide more information on how women were selected at the facilities? Was it the first xx number of ANC clients over a certain number of days? What is the range of number of women per facility (i.e. some facilities had 5 clients others had 25?)
	• More detail is needed on the items included in the structure and process indices.
	  ○ Association between structure/process and outcome can vary depending on what is included in the index. Please include as a supplementary table details on the items included in the indices. This is also important for contextualizing these findings with other surveys. For example, how were diagnostics measured – the high score is surprising and may be a result of the indicator definitions.
	• The criteria for categorizing facilities and providers as poor, fair, and good needs further explanation.
	  ○ The paper cited does not present WHO criteria. These are criteria used in a paper published in the WHO bulletin. In addition, this paper doesn't use poor, fair, and good as thresholds. Please explain further how these cut-points were created.
	• Methods for creation of the outcome variable are unclear
	  ○ How PCA was used to create the client satisfaction score/outcome variable is unclear. Was the first component resulting from the principal component analysis used to create a score? If so, how was that converted to a percentage between 0 and 100?
	  ○ Why was a different approach taken for the construction of the outcome variable as compared to the structure and process indices?
	• Data analysis lacks information on model building
	  ○ What are the other co-variates? How were they selected?
	  ○ Was bivariate analysis conducted prior to the final multi-variate model?
	  ○ Did you check for co-linearity of predictor variables?
	  ○ Why aren't all the structural indices included in the model?
	  ○ How did you account for the nested structure of the data (i.e. women within providers and providers within facilities)?
	• How was the survey design accounted for in the analysis?
	  ○ Was the complex survey design accounted for in the analysis (i.e. weights, stratification, clustering)?

Results:

- The availability of individual structural and process items should be presented including disaggregation by facility type. These are part of the main findings of the results section and thus should be presented. I recognize that this makes the paper quite dense, but these are really important findings!
- Table 6 needs information on the reference categories for each variable that is categorical. Explanation of the construction of the covariates in the methods section would help clarify which are categorical and which are continuous.

Discussion and Conclusion:

- The discussion could use some revisions to clearly state the main findings and secondary findings and link those findings more deeply to the country context. While there are a LOT of findings from this paper, it was a bit difficult to figure out which were the most important and why they are important in the Ethiopia context.
- The discussion would benefit from a paragraph on the public health implications of this work. How might this information be useful to regional government? Are there are actionable recommendations? Did this research generate additional future research questions?
- The limitations paragraph could use mention of other limitations, for example:
    - Unmeasured factors influencing satisfaction
    - Client expectations varying by social and cultural beliefs
    - Social and courtesy bias
    - Generalizability of findings
-

MINOR COMMENTS:

Introduction:

- Pg. 4 says “According to this model, the ultimate goal for quality is to produce client satisfaction …”. I think the goal of the model is to improve health outcomes, which includes but is not limited to client satisfaction. Many people perceive changes in health status as the ultimate goal, so I think it’s important not to lose sight of this.
- The justification for the study only relates to the first objective and not much to the second. I think this can be expanded to cover both research objectives and well as to make an argument of how this information would be useful in the context of Northwest Ethiopia (i.e. improve service delivery, health systems strengthening planning, fill current information gap, etc.). In addition to research projects assessing quality of ANC in Ethiopia, what health facility assessments have been implemented nationally in Ethiopia? What was the scope of content and geographic representativeness? (There was a SARA in 2016 and 2018 in Ethiopia – what the information gap from this assessment?)

	Methods: • The tools used for data collection are referred to inconsistently in the paper – for example facility audit, facility survey; client observation, observational study. In addition, in the strengths and limitation bullet points four tools are mentioned (facility audit, client exit interview, provider interview, and direct observation) while in the methods section only three tools are mentioned (facility survey, direct observation, exit interview). Would be helpful to be consistent. • No mention of IRB approval. • Pg. 7 – "52.7% of ANC clients booked between 3 and 6 months after conception"; this value doesn't appear in Table 1. Suggest aligning result and table. • You mention that the use of the proportional odds model with a categorical variable is an advantage to this study. Did you do a sensitivity analysis using a binary outcome for satisfaction? **Results:** • Pg. 9 "This is an indication that most of the providers had worked in their organization long enough to understand and give credible information to pregnant women" – was there not inclusion criteria for HW participation in the survey? The criteria should have been related to education and professional qualification not number of years of professional experience. • Pg. 11 – Careful with use of very satisfied/dissatisfied with scores of 3.7 and 2.5 – these are really more mid-range on a 5-point scale. • Table 4: It would be helpful to have % of ANC clients here instead of the N • Table 5 and 6 are presented out of order in the paper on pg. 12. Consider reversing table numbers and appearance in the paper. • Pg. 13 the Cronbach's alpha for the indices appears a bit out of place. I'd move that to coincide with the presentation of results for each index earlier in the results section. Also, was there a single score created for structure and process? If so, is this score used in the regression analysis? If not, why not? • Pg. 13 "After controlling for…" – Were you controlling for some covariates to assess the association of others? Relates back to the overall research question. May want to clarify the language here. **Discussion:** • Findings on satisfaction amongst rural ANC clients vs lower educated clients are contradictory. Perhaps some explanation would be useful here to explain that. • Paragraphs on reproductive-related factors should be re-evaluated in light of a decision on whether or not these are variables you are controlling for or assessing for a relationship with client satisfaction. **Abstract and strengths/limitations bullets:**

	• Update and clarify based on the changes to the main text.

REVIEWER	Md Mizanur Rahman Universiti Malaysia Sarawak
REVIEW RETURNED	16-Mar-2020

GENERAL COMMENTS	1) The title of the study does not match with research objectives. Rewrite the title or objectives 2) No need to add STATA basic command in the test, e.g. at beginning of the discussion, the author stated gologit2, delete the sentence 3) In the collection procedure, the ethical issues concerning this study to be spelt out.

REVIEWER	Vanessa Brizuela World Health Organization, Switzerland
REVIEW RETURNED	12-Apr-2020

GENERAL COMMENTS	Thanks for providing me the opportunity to review this manuscript. The authors attempt to assess quality of antenatal care (ANC) services in Northwest Ethiopia through a series of measurements, specifically women's satisfaction with the services. While this is an important study, there are some important gaps that I find critical to address. I understand the authors want to measure women's satisfaction as an outcome of quality care, but I believe there should be a mention to the ultimate goal/outcome of quality of maternal and newborn care, which is improved health outcomes (decreased mortality and morbidity, for example) and therefore more explicitly make this linkage The methods section needs to be further developed as it is difficult to understand the decisions and assumptions made given the results the authors present. It would help to understand how the authors constructed the indices as it is hard to assess without knowing how they were constructed, especially since there is quite a bit of literature on this topic. It would also help determine why specific measures were used and not others, etc. Relatedly, there are several publications that deal with measurement of quality of maternal and newborn care, which includes antenatal care, which have been omitted from this manuscript. I would suggest that the authors look at these and assess whether they would be better suited to include/replace some of the existing cited literature: i. Trends in maternal mortality 2000 to 2017: Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO. ii. Moxon SG, et al. Service readiness for inpatient care of small and sick newborns: what do we need and what can we measure now? J Glob Health 2018; 8: 010702. iii. Brizuela V, et al. Measuring quality of care for all women and newborns: how do we know if we are doing it right? A review of facility assessment tools. The Lancet Global Health. 2019 May 1;7(5):e624-32. iv. Sheffel A, et al. Use of Service Provision Assessments and Service Availability and Readiness Assessments for monitoring

	quality of maternal and newborn health services in low-income and middle-income countries. BMJ Glob Health 2018; 3: e001011. v. Tunçalp Ö, et al. Quality of care for pregnant women and newborns—the WHO vision. BJOG 2015;122:1045–1049. vi. Sharma J, et al. Poor Quality for Poor Women? Inequities in the Quality of Antenatal and Delivery Care in Kenya. PLoS One. 2017;12(1):e0171236. Published 2017 Jan 31. doi:10.1371/journal.pone.0171236 vii. Larson E, et al. When the patient is the expert: measuring patient experience and satisfaction with care. Bull World Health Organ. 2019;97(8):563–569. doi:10.2471/BLT.18.225201 viii. World Health Organization. Standards for Improving Quality of Maternal and Newborn Care in Health Facilities. Geneva, Switzerland 2016. Available at: http://apps.who.int/iris/bitstream/10665/249155/1/9789241511216-eng.pdf?ua=1 I have made comments directly onto the pdf, but I include them below for easy reference. Abstract: 1- In the setting you should add that it was healthcare facilities providing antenatal care services in Northwest Ethiopia. 2- In the outcome measures you state you used principal component analysis (PCA) but this is not later explained as such in the manuscript. Please be consistent. 3- In results, please use the same language as in the manuscript. It is not evident that low satisfaction is the same as unsatisfactory. Strengths and limitations: 4- There are barely any limitations (as well as in the body of the manuscript). I suspect if the methods were better explained, there would be more limitations. Introduction: 5- You state that “99% of those deaths occur in developing nations including Ethiopia” and you reference a paper by Alkema. I’m not sure this paper states that %. I would also suggest using the more updated numbers provided in the Trends in maternal mortality document suggested above. 6- Please refrain from using double negatives –suggest replacing “not inevitable” with “avoidable” or “preventable” (line 32) 7- Suggest replacing reference 4 (line 22, page 5) with more updated references. See suggestions above. I would also include the framework developed by Tunçalp and others at WHO to describe quality of care. It’s more updated than Donabedian’s, is based on the same principles, and is specific to maternal and newborn care. 8- In the following paragraph you state about quality of antenatal care in Ethiopia but I think you should also reference other work on QoC in ANC in other settings (see references by Sharma and Larson as examples). Methods: 9- Under study population and sampling, you include part of the setting (health facilities). 10- Under data collection you state using information from WHO guidelines. Why did you not use the WHO Standards for quality of care that include specific measures for QoC? I included the complete reference above. 11- When you state “Experienced midwives who were not affiliated

	with the surveyed healthcare facilities collected the data." Does this mean that the data collectors went around the facility observing for presence or absence of a specific structural measure or you interviewed hospital management? In that same paragraph, how was the healthcare provider interview handled? This is not described. 12- You state that you evaluated whether HCW conformed in conducting "key ANC measures." How were these key measures identified? 13- Under structure and process attributes, it would be good to either attach the data collection tools as an appendix or explain what was measured for each of the variables. I don't feel I can assess the methodology well if I don't know whether your data collection tools captured the essentials --hence, not knowing whether your study limitations are accurate. 14- Also need to understand how you constructed the indices better. How many variables per construct? How many constructs? Did you weigh each item equally? 15- You state you assess availability of physical structure and essential equipment: How were these determined? From the below you assess for magnesium sulphate (which I assume is for the treatment of pre-eclampsia/eclampsia (PE/E) during childbirth and not to prevent PE/E during pregnancy). 16- Reference 29 is not referencing WHO criteria but to a paper written by non-WHO staff. Which criteria were used? 17- Under outcome variable, how were the cut-off points determined/decided upon? Results: 18- When you state "more than half of them booked between 3 and 6 months after conception" I suggest you use the same wording as in the table, which is clearer: "more than half of the women interviewed had their first ANC visit during their second trimester of pregnancy" 19- Table 3: under health facility survey I cannot tell what is being measured through these indices. The items you should be measuring with regards to ANC should refer to these alone and not childbirth or newborn care. Also, remove the year from the table heading (I am assuming this is what 2018 refers to). 20- Also remove 2018 from table 4 header. 21- Suggest moving table 5 as an appendix as this is detailing part of the methodology that went behind defining variables for the models. What does "service year of the provider" mean? 22- On page 14 it is unclear when you present OR1=0.29 and state this relates to increased odds. 23- Table 6, remove the year from table title. Also remove the type of modelling used in the header, you should describe the output not the methodology. Also in table 6, are these crude odds ratios or adjusted odds ratios? I thought it was the latter in which case please replace OR with aOR. What does address mean in the variables? Are you referring to whether the woman lived in a rural or urban environment? If so, then write that and specify which is the reference. Why is gestational age measured in months and not weeks (which is the norm)? What does service year of the provider (year) mean and what is the reference in this variable? What does health screening mean? It sounds extremely broad and non-specific, which might explain why the OR are so high. Discussion: 24- The description of the methods does not belong in the

	discussion paper (it would, in a methods paper) (first paragraph). I would use this first paragraph to describe what you set out to do and what your main/most important findings are. 25- Last paragraph on page 16 contradicts your findings and what you present in the prior paragraph. How do you explain this? 26- The limitations section is quite weak. I anticipate that a more detailed description of the methods will render some additional limitations that are impossible to asses at the moment. Ethics approval: 27- Was verbal consent obtained from the women and the healthcare providers observed during client consultations? What provisions were put in place to secure the confidentiality of the data obtained? References: 28- Please check the references as some seem incomplete or need links to be able to find them.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1
RESEARCH OBJECTIVE:
The research objective is not clearly defined and differs between the abstract and introduction.
Response: Thank you for your suggestion. We fully agree with your suggestion that it would be better if we could present two research objectives. We, accordingly, have rewritten our objective in line with your recommendation. The change has been highlighted in the revised manuscript (Page2, line 30-32)
METHODS;
2.1 There are some weaknesses in describing the sample size calculation and sampling procedures.
Responses: We thank the reviewer for highlighting this important point. We have clarified our intent and used the method proposed by Anthony G. Turner et al 2001 as a reference for the sample size calculation (Pages 5-6, line 138-162).
In the Ethiopian context, health providers are accredited health professionals (such as doctors, nurses, midwives, or public health officers) who have been educated and trained to proficiency in the skills needed to manage normal (uncomplicated) pregnancies, childbirth, and the immediate postnatal period, and in the identification, management and referral of complications in women and newborns.
2.2 More detail is needed on the items included in the structure and process indices.
Responses: We agree that clarity is needed. Hence, we further clarified this section in the revised version (Pages 8,9 line 209-237). In addition, the individual items that were used in constructing each dimension are annexed. (Page 24,annex, Tables 4and 5).
2.3 The criteria for categorizing facilities and providers as poor, fair, and good needs further explanation
Responses: Thank you for pointing this out. The Heredia and colleagues in the WHO bulletin used the highest quintile of the procedures received as cut-offs for the upper category. (At this juncture, we apologies for considering this as the WHO's work). Similarly, we applied their criteria (80% cut-offs) just in determining the upper category of the classification. However, instead of dichotomizing the response we generated a three category. This is because while dichotomizing is simple, the simplicity would be gained at some cost of losing power. Beeckman et al, 2013, which has been cited by Heredia and colleagues, also grouped the content of care in to four groups.
2.4. Methods for creation of the outcome variable are unclear
Responses: For the satisfaction items, since the first component explained only 58 % of the variances, we did not use the first component. We rather created a summed index of the retained items from the two components that explained 68% of the variances. For structural and process quality dimensions, we did not apply the PCA because some subcomponents components were constructed from few variables. This has been highlighted in the revised version(Pages7-8, 195-204)
2.5 Data analysis lacks information on model building

o What are the other co-variates? How were they selected? Was bivariate analysis conducted prior to the final multi-variate model?

Responses: The steps undertaken in model building were missed in the previous version, we now considered in the revised version. Bivariate analysis was used to select candidate variables for the multivariate analysis using p-value= 0.2 as a cutoff point (Page 9, line 257-59).
o Did you check for co-linearity of predictor variables?

Responses: Yes, we have already mentioned that multi-collinearity test between independent variables was done using the Variance Inflation Factor (VIF), and no serious problem was detected (the highest value was 2.7).(Page 9, 259-61)
o Why aren't all the structural indices included in the model?
Responses: Thank you for this important observation. Structural attribute and process indicators were the main independent variables considered in this study. However, during the bivariate analysis none of the structural attributes showed any statistically significant association with the outcome variable.
o How did you account for the nested structure of the data (i.e. women within providers and providers within facilities)? How was the survey design accounted for in the analysis?
Responses: Owing to the hierarchical structure of our data, our first intention was to analyze the data using the multilevel analysis modeling. Accordingly, we checked the assumptions for multilevel logistic regression. Unfortunately, the small and insignificant variance of the assumption tested in null model did not allow us to run random effects at level 2 (i.e. at facility level). Hence, we are forced to apply the flat model.
o Was the complex survey design accounted for in the analysis (i.e. weights, stratification, clustering)?
Responses: Yes, stratification was considered during sampling. First, five districts of the zone (4 rural districts and 1 town administration) were selected after first stratifying the study area in to 13 rural districts and 2 town administrations. We have added this statement in the revised manuscript

RESULTS
3.1 The availability of individual structural and process items should be presented including disaggregation by facility type.

Response: Thanks for this suggestion. Tables showing the availability each individual items have been annexed as per your suggestion, (Page 24, annex 1, Tables 4 and 5 Page)

3.2 Table 6 needs information on the reference categories for each variable that is categorical.
Responses: We are grateful for the comments. The reference category for each categorical variable has been indicated as per your comment (Pages 16-17, line 400).
DISCUSSION AND CONCLUSION
The discussion could use some revisions to clearly state the main findings and secondary findings.
Responses: Thank you for these insightful comments. We tried to indicate the area of focus in the first paragraph of the discussion(Page 17 line 406-10). We boldly gave due emphasis in redrafting the limitation and conclusion parts as per your suggestion(Page 19-29, line 458-87).
MINOR COMMENTS
Introduction:
1. Pg. 4 says "According to this model, the ultimate goal for quality is to produce client satisfaction…". I think the goal of the model is to improve health outcomes, which includes but is not limited to client satisfaction. Many people perceive changes in health status as the ultimate goal, so I think it's important not to lose sight of this.

Response: Thanks. Accepted and amended (Page 4, line 102-104)

2. The justification for the study only relates to the first objective and not much to the second.
Response: Thank you for the suggestion. We have revised lines 119-124 (page 5) to clearly outline the main difference between our study and the previous studies w.r.t the second objective.

Concerning SARA studies in Ethiopia, ANC quality was measured mainly from INPUT indicators using ONLY ten tracer items categorized in four domains: Trained Staff & guidelines, Equipment, Diagnostics, and Medicines & commodities. Importantly, previous studies did not consider the association between client satisfaction and measures of structural and process quality.

Methods:

3. The tools used for data collection are referred to inconsistently in the paper
Response: Thank you. This has been corrected in the revised manuscript
4. No mention of IRB approval.
Response: Thank you for your observation. However, we have already mentioned the ethical issues under 'ethical approval' section, which is located below the conclusion part as per the journals format (Page 21, lines 501-508).

5. Pg. 7 – "52.7% of ANC clients booked between 3 and 6 months after conception"; this value doesn't appear in Table 1. Suggest aligning result and table.
Response: We are grateful for this comment. We have revised this sentence as per your comment (Page 10, line 276-77).

6. You mention that the use of the proportional odds model with a categorical variable is an advantage to this study. Did you do a sensitivity analysis using a binary outcome for satisfaction?

Response: To be honest we did not do any sensitivity analysis. Yet, a major strength of gologit2 is that it can fit three special cases of the generalized model: the proportional odds model, the partial proportional odds model, and the logistic regression model as well.(Williams R. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. Stata Journal 2006; 6: 58–82. It is also true that the ordinal regression does not alter an ordinal outcome as a dichotomous measure for logistic regression analysis, which may lead to the loss of information inherent.
Results:
7. Pg. 9 "This is an indication that most of the providers had worked in their organization long enough to understand and give credible information to pregnant women" – was there not inclusion criteria for HW participation in the survey? The criteria should have been related to education and professional qualification not number of years of professional experience.

Response: Whilst we agree that the educational and professional qualification has a paramount importance for proper management of clients, we also believe that the number of years of professional experience would have its own role.

8. Pg. 11 – Careful with use of very satisfied/dissatisfied with scores of 3.7 and 2.5 – these are really more mid-range on a 5-point scale.

Response: Thank you for this pointing out that. We have rephrased the sentence to show that the difference is minimal (Page 14, line 339)

9. Table 4: It would be helpful to have % of ANC clients here instead of the N

Response: This point is accepted, and the changes have been reflected in the revised manuscript (Page 14, line 349).

10. Table 5 and 6 are presented out of order in the paper on pg. 12. Consider reversing table numbers and appearance in the paper.

Response: We thank you. Slight modification has been made in numbering. Yet, still we feel that the assumption test needs to be presented ahead of the final analysis. Moreover, table 5 (table for the assumption test) has been annexed as per the third reviewers suggestion.

11. Pg. 13 the Cronbach's alpha for the indices appears a bit out of place. I'd move that to coincide with the presentation of results for each index earlier in the results section. Also, was there a single score created for structure and process? If so, is this score used in the regression analysis? If not, why not?

Response: Thank you for your suggestion. We moved the Cronbach's alpha to the "result" section(Page12 and 14; line 310-11; 333-34; and 344-45). Cronbach's alpha value was calculated separately for the structural and process attributes. Fortunately, their calculated values are similar. As stated in the methods section these are composite variables computed from its own individual

variables.

12. Pg. 13 “After controlling for…” – Were you controlling for some covariates to assess the association of others? Relates back to the overall research question. May want to clarify the language here
Response: Thank you for your comment. It has been corrected as ‘After adjusting for all the covariates ….’(Page15, line 364)

Discussion:
13. Findings on satisfaction amongst rural ANC clients vs lower educated clients are contradictory. Perhaps some explanation would be useful here to explain that.
Response: Note that we measured the perceived quality of care that the women felt according to their own level of understanding, which might not necessarily reflect the actual care received. Taking this in mind, the direct relation of educational status of the women with satisfaction might be related to differentiating real from the non-real procedures during consultation. On the other hand giving high credit for every service they received might be one possible explanation for the direct relationship between rural residence and high satisfaction

14. Paragraphs on reproductive-related factors should be re-evaluated in light of a decision on whether or not these are variables you are controlling for or assessing for a relationship with client satisfaction.
Response: Reproductive-related factors are among the variables considered in assessing the relationship with client satisfaction.

ABSTRACT AND STRENGTHS/LIMITATIONS BULLETS:
15. Update and clarify based on the changes to the main text.
Response: We thank you for your comment. We tried to make changes in these sections (Page 3, line 60-61 and 65-67).

Reviewer #2
1) The title of the study does not match with research objectives. Rewrite the title or objectives.

Response: Thank you for this insightful comment. We have rewritten our objective in line with your and the first reviewer’s recommendations (Page2, line 30-32)

2) No need to add STATA basic command in the test, e.g. at beginning of the discussion, the author stated gologit2, delete the sentence

Response: We accept your comment. This has been deleted as per your suggestion.

3) In the collection procedure, the ethical issues concerning this study to be spelt out.

Response: We clearly stated all the ethical issues under the heading ‘Ethics approval’, that has been located below the conclusion section as per the journal’s format (page 21, line 500-507).
Reviewer # 3
Abstract:
Comment #1- In the setting you should add that it was healthcare facilities providing antenatal care services in Northwest Ethiopia.
Response: Thank you for the comment. We have accepted the comment and included in the revised version (page 2, line 34).
Comment #2- In the outcome measures you state you used principal component analysis (PCA) but this is not later explained as such in the manuscript. Please be consistent.
Response: We acknowledge your comment. Though we are unable to specify all of the details included in the methods section due to the word count limit of the abstract, we have briefly explained how PCA was computed in the methods section under the heading’ measurement and variables’ (page 7-8, line 195-204)
.Comment #3- In results, please use the same language as in the manuscript. It is not evident that low satisfaction is the same as unsatisfactory.
Response: We have accepted the comment and modified it accordingly. We have toned down the language by substituting “unsatisfactory” to “low satisfaction” in the first sentence of the result section

of the abstract(page 2, line 44).
Strengths and limitations:
Comment #4- There are barely any limitations (as well as in the body of the manuscript). I suspect if the methods were better explained, there would be more limitations.
Response: Thank you for pointing out this. The authors already stated the following as a limitation in the 'strength and limitation' section and in the body of the manuscript as well: 'the presence of an observer during the client-provider interaction may bias the results obtained in a positive direction '. However, we accept your suggestion and we feel that this is not the only limitation that this paper suffered. Accordingly, we have included some other limitations in both the strength and limitation section(Page 3, line 65) and the body of the paper (pages 19-20, line 457-74).
Introduction:
Comment #5- You state that "99% of those deaths occur in developing nations including Ethiopia" and you reference a paper by Alkema. I am not sure this paper states that %. I would also suggest using the more updated numbers provided in the Trends in maternal mortality document suggested above.
Response: We agree with the reviewer; and we acknowledged that there is more work in the literature. We, accordingly, have updated our reference based on your suggestion (Page3-4). With regard to our previous reference, we were considering the worst scenario from the figure reported by Alkema and colleague, that is by taking the lowest boundary of the uncertainty interval of mortality in developed countries (11 per 100,000 LB) and the upper boundary of the interval in LMC (652 deaths per 100000 LB)
Comment #6- Please refrain from using double negatives–suggest replacing "not inevitable" with "avoidable" or "preventable" (line 32).
Response: We are grateful for this observation; we have made the suggested change (Page4, line 74).
Comment #7- Suggest replacing reference 4 (line 22, page 5) with more updated references. See suggestions above. I would also include the framework developed by Tunçalp and others at WHO to describe quality of care. It's more updated than Donabedian's, is based on the same principles, and is specific to maternal and newborn care.
Response: Thank you for suggesting these important references we missed in the previous version. We have extended this section to provide a better coverage of the literature in line with that suggested by the reviewer(page 4).
Comment #8- In the following paragraph you state about quality of antenatal care in Ethiopia but I think you should also reference other work on QoC in ANC in other settings (see references by Sharma and Larson as examples).
Response: Thank you for this important comment. Despite the references you suggested are not referring to Ethiopian context, we found your suggestions as an opportunity to improve the quality of our paper. Hence, we tried to incorporate the suggested references in the revised manuscript to show the global situations and conditions in developing countries (Page 4).
Methods:
Comment #9- Under study population and sampling, you include part of the setting (health facilities).
Response: Thanks you for your suggestion. However, apart from interviewing pregnant women and the health workers assessing the capacity of health facilities also was our objective and the results have been presented as structural attributes at facility level. Hence, we thought that health facilities could deserve being a population, as population includes all subjects to be studied and does not necessarily mean people.
Comment #10- Under data collection you state using information from WHO guidelines. Why did you not use the WHO Standards for quality of care that include specific measures for QoC? I included the complete reference above.
Response: We thank the reviewer for this comment. Whilst we appreciate the difference (i.e. both in structure and contents of quality) between the Donabedian model and other frameworks, we feel that they appeared to be similarly beneficial in women regardless of this differences.

Comment #11- When you state "Experienced midwives who were not affiliated with the surveyed healthcare facilities collected the data." Does this mean that the data collectors went around the facility observing for presence or absence of a specific structural measure or you interviewed hospital management? In that same paragraph, how was the healthcare provider interview handled? This is not described.
Response: the midwives who were involved in observing the client-provider interaction and to interviewing pregnant women at exit were recruited from other health facilities. The rationale of using midwives who were not affiliated with the surveyed healthcare facilities was to minimize the courtesy

bias. These data collectors were also responsible for interviewing the facility managers and the healthcare providers as well.
Comment #12- You state that you evaluated whether HCW conformed in conducting “key ANC measures.” How were these key measures identified?
Response: The key ANC measures are to mean the basic components that the WHO and the Federal ministry of health of Ethiopia determined essential for every pregnant woman. The components included measuring body weight, blood pressure, urine analysis, blood testing, tetanus toxoid injection, deworming, iron and folic acid tablet, information provided about birth preparedness, and about pregnancy related danger signs.
Comment #13- Under structure and process attributes, it would be good to either attach the data collection tools as an appendix or explain what was measured for each of the variables. I don't feel I can assess the methodology well if I don't know whether your data collection tools captured the essentials --hence, not knowing whether your study limitations are accurate.
Response: Thank you for the comment. We have accepted the comment and details are presented in revised version (pages8-9, line210-37). In addition, all the results items are annexed (see annex 1 tables 3 and 4, page 24-29, line648).
Comment #14- Also need to understand how you constructed the indices better. How many variables per construct? How many constructs? Did you weigh each item equally?
Response: We agree that this issue should be elaborated further. We have tried to do so by expanding our information to make the issue more clear (pages8-9, line210-37).
Comment # 15- You state you assess availability of physical structure and essential equipment: How were these determined? From the below you assess for magnesium sulphate (which I assume is for the treatment of pre-eclampsia/eclampsia (PE/E) during childbirth and not to prevent PE/E during pregnancy).
Response: The checklists have been developed based on the general service availability and readiness assessment (SARA) indicators and ANC specific availability and readiness indicators of the WHO and SARA OF Ethiopian public health institute, and MCH guidelines developed by the FMOH of Ethiopia.
Regarding the issue of magnesium sulfate, there is no doubt that the magnesium sulfate ($MgSO_4$) is the ideal drug to prevent and control seizures though there are still questions regarding scheme and the optimal time of administration. This drug is among the essential drugs lists to be available throughout pregnancy, delivery and postpartum periods in Ethiopian context to managing severe eclampsis. Moreover, the WHO guideline also recommends health facilities to avail supplies of antihypertensive agents and magnesium sulfate in sufficient quantities at all times, in antenatal, labor, childbirth and postnatal areas for the management of women with pre-eclampsia. (WHO. Standards for improving quality of maternal and newborn care in health facilities, 2016, page 68/84. Whilst a large proportion of PE/E usually occur at the time of labor and delivery potentially it can happen at any stage of the late pregnancy and during the early post-partum periods as well.
Comment #16- Reference 29 is not referencing WHO criteria but to a paper written by non-WHO staff. Which criteria were used?
Response: Thank you for pointing out this. The paper by Heredia-Pi and colleagues has been published in the WHO bulletin and we were considering as if this was the WHO’s work, and your apologies please.
The Heredia and colleagues used the highest quintile of the procedures received as cut-offs for the upper category. Similarly, we applied their criteria (80% cut-offs) in determining the upper category of the classification. However, instead of dichotomizing the response we generated a three category, as we fear the impact of dichotomizing the variable. While dichotomizing is simple, the simplicity is gained at some cost of losing power. Even Beeckman et al, 2013, which has been cited by Heredia and colleagues, grouped the content of care in to four groups while ours.
Comment # 17- Under outcome variable, how were the cut-off points determined/decided up on?
Response: For the satisfaction items, we apply the steps of PCA to reduce the data. Since the first component of the PCA explained only 58 % of the variances, we did not use the first component to generate the outcome variable. We rather created a summed index of the retained 12 items from the two components that explained 68% of the variances. This has been explained in the revised version (page 7-8, 195-204)
Results:
Comment #18- When you state “more than half of them booked between 3 and 6 months after conception” I suggest you use the same wording as in the table, which is clearer: “more than half of the women interviewed had their first ANC visit during their second trimester of pregnancy”
Response: We are grateful for this comment. We revised this sentence as per the comment (Page 10,

line 276-78).
Comment #19- Table 3: under health facility survey I cannot tell what is being measured through these indices. The items you should be measuring with regards to ANC should refer to these alone and not childbirth or newborn care. Also, remove the year from the table heading (I am assuming this is what 2018 refers to).
Response: Thank you for your comment. We deleted the year 2018 from heading. We also agree that the numbers written might not be informative as it stands. This is because the table shows the summary of structure and process quality dimensions and the subheadings without any details. Hence, the detail of individual variables from which the structural and process quality dimensions have been generated annexed(page 24, line 648).
As a response to your question regarding the confusion of whether the items included were for ANC or delivery or both, the items included are referring to the ANC service despite sharing some items.
Comment #20- Also remove 2018 from table 4 header.
Response: Deleted, thank you.
Comment #21- Suggest moving table 5 as an appendix as this is detailing part of the methodology that went behind defining variables for the models. What does “service year of the provider” mean?
Response: This suggestion is acceptable. We moved table 5 down to the appendix (Pape 29, annex 2).. We used the term service year to indicate the number of years health care providers earned in their educational status regardless of the place of work
Comment # 22- On page 14 it is unclear when you present OR1=0.29 and state this relates to increased odds.
Response: Thank you for pointing out this. Despite the result of the Odds ratio is acceptable, the error has occurred during interpretation, and sorry for the error we made. Now we have modified and rephrased the statement by splitting the sentence into two (Page 16, line 378-81).
Comment # 23- Table 6, remove the year from table title. Also remove the type of modelling used in the header, you should describe the output not the methodology. Also in table 6, are these crude odds ratios or adjusted odds ratios? I thought it was the latter in which case please replace OR with aOR. What does address mean in the variables? Are you referring to whether the woman lived in a rural or urban environment? If so, then write that and specify which is the reference. Why is gestational age measured in months and not weeks (which is the norm)? What does service year of the provider (year) mean and what is the reference in this variable? What does health screening mean? It sounds extremely broad and non-specific, which might explain why the OR are so high.
Response: We are grateful for the comments. We have modified the table to make it clearer to the readers. We also indicated the reference category for each categorical variable (Page 16-17, line 399). Furthermore, we agree with your suggestion that GA would be more meaningful if it had been described in weeks than in months. In Ethiopian context, ultrasound (which is the reliable tool to estimating GA) is not available in most facilities to estimate GA. Rather; the LMP (with its limitation) is still used to calculate GA in many facilities, the reliability of which is depending on the memory status of women. Therefore, during history taking it is common for the health workers working in ANC clinics to ask first about the number of months lapsed since the LMP and then to convert to weeks and record in the registry. However, doing so is liable to recall bias. Hence, considering the gaps and to minimize recall bias we preferred to report GA in months than in weeks.
Health screening in our study are referring to laboratory and preventive interventions such as tetanus toxoid vaccination and deworming. Now it has been have corrected in this way. The high AOR and wide confidence level could be relate to the small number of variables any of the cells (Which I could not clarify more) during the chi-square calculation of the model
Discussion:
Comment # 24- The description of the methods does not belong in the discussion paper (it would, in a methods paper) (first paragraph). I would use this first paragraph to describe what you set out to do and what your main/most important findings are.
Response: we have accepted the comment and revised this section accordingly.
Comment # 25- Last paragraph on page 16 contradicts your findings and what you present in the prior paragraph. How do you explain this?

Response: Thank you for pointing out these inconsistences. As stated earlier women who were living within 60 minutes walking distance to the health care facility have lower level of satisfaction. In this context, the last paragraph of this page number could not be a problem rather the error was made in the sentence just immediately before this paragraph. And. now we have corrected the error in the revised manuscript (Page).
Comment #26- The limitations section is quite weak. I anticipate that a more detailed description of

the methods will render some additional limitations that are impossible to asses at the moment.
Response: We are very grateful for this comment. We have made the suggested revision in the revised manuscript, and the changes have been highlighted. Page 18, line 437)
Ethics approval:
Comment #27- Was verbal consent obtained from the women and the healthcare providers observed during client consultations? What provisions were put in place to secure the confidentiality of the data obtained?
Response: Yes, we took verbal consent as our study did not have any invasive procedure (like blood drawing, taking samples, etc.). The ethical guideline of Ethiopia also allows us to take verbal consent in case of noninvasive procedure. For instance, the observer in the observational study only observed the routine care given to mothers during ANC consultation (non-participatory). In addition, for exit interview women were asked about the routine ANC services they received. Personal identifiers were excluded from the data collection form to maintain privacy and confidentiality.
References:
Comment #28- Please check the references as some seem incomplete or need links to be able to find them.
Response: We appreciate your comment. Corrections have been made accordingly.

VERSION 2 – REVIEW

REVIEWER	Ashley Sheffel Johns Hopkins Bloomberg School of Public Health, USA
REVIEW RETURNED	09-Jun-2020

GENERAL COMMENTS	Thank you for the revised manuscript and your attention to the range of reviewer comments. The paper has been substantially revised. I have just a few minor comments remaining. • The research objectives were updated in the abstract and are now very clear. It would be helpful if this same language was used in the end of the introduction. • The supplementary tables are a very helpful resource, but they need to be properly cited in the main body of the paper so that the reader knows what information is contained in each supplementary table and can reference those tables if interested in looking at some of the more in-depth material. • More detail is still needed on the items included in the structure indices in the supplementary table. o For example, for the laboratory items, was it ability to conduct the test onsite + availability of certain equipment? o For equipment, was it observed available and functional? o For HR, how was this information on the standard vs. available converted into a binary score? • Pg. 14 – Still need to be careful with use of very satisfied/dissatisfied with scores of 3.7 and 2.5 – these are really more mid-range on a 5-point scale. This hasn't quite been adequately addressed.

REVIEWER	Vanessa Brizuela World Health Organization, Switzerland
REVIEW RETURNED	15-Jun-2020

GENERAL COMMENTS	Women's Satisfaction with the Quality of Antenatal Care Services Rendered at Governmental Health Facilities in Northwest Ethiopia: the application of partial proportional odds model Thanks for sending a revised manuscript that addresses most of the issues raised while responding to other issues brought up by the

	authors. I believe the revisions made to the manuscript have much improved it and their methods section is now clearer to follow and understand, and the inclusion of the supplementary files helps in making this assessment. I am pleased with the current version and would recommend publication after a few remaining minor issues: 1- Please revise the entire manuscript for grammar, spelling, and sentence structure. There were still some issues remaining making readability at times a bit challenging. 2- Please refrain from using the word “interview” to mean survey or questionnaire as the former is usually preferred to describe the qualitative data collection method and not quantitative ones. 3- Please make sure the tables can be read as stand-alone exhibits –this applies to supplementary files as well. This means that the reader should be able to understand them without needing to go back to the text for explanations of acronyms (especially in supplementary files). Table 4 would benefit from clarifying what the scoring range was (0-5) in a caption so that the mean score is understandable without needing to go back to the text. Table 5 needs some reconfiguration so that the names of the variables and what was used as a reference in the table 5 are a bit less confusing and easier to read. Also in table 5 I would not report significance other than $p<0.05$ since this is what you established in your methods section as your reference for significance (p 10, line 264). 4- I would suggest using the first paragraph of the discussion section to describe what you set out to do and what your main/most important findings are. It is currently still missing the main findings.

VERSION 2 – AUTHOR RESPONSE

Comments from Reviewer 1

Comment #1 The research objectives were updated in the abstract and are now very clear. It would be helpful if this same language was used in the end of the introduction.

Response: Thank you for your suggestion. We have revised this part as per your suggestion (Page 5, line 122-124)

Comment #2 The supplementary tables are a very helpful resource, but they need to be properly cited in the main body of the paper so that the reader knows what information is contained in each supplementary table and can reference those tables if interested in looking at some of the more in-depth material.

Response: We are grateful for this comment. These have been cited in the methods (page 8-9 line number 215 and 228) and the result section of the paper as well (page 13 line 316 and page 14 line 337)

Comment #3 More detail is still needed on the items included in the structure indices in the supplementary table. For example, for the laboratory items, was it ability to conduct the test onsite + availability of certain equipment? For equipment, was it observed available and functional? For HR, how was this information on the standard vs. available converted into a binary score?

Response: For laboratory items, we intended to assess the ability of the surveyed facilities in conducting the test onsite. For equipment and infrastructures, each item was observed if it was available and functional at the date of the survey. As stated in the methods section (Page8 line 210-12), each item/commodity was scored ‘1’ if the item was available and in good working condition, and ‘0’ if this was not the case. For HR, the number of health workers in each health facility was obtained and compared against the national standard by facility type. For each health worker, a value of “1”

was granted if the number of health workers are greater or equal to the minimum required number of health workers at each health facility, and “0” otherwise.
Comment #4 Pg. 14 – Still need to be careful with use of very satisfied/dissatisfied with scores of 3.7 and 2.5 – these are really more mid-range on a 5-point scale. This hasn’t quite been adequately addressed.
Response: We are grateful for this comment. We tried to revise this point as per your comment. The revised section is highlighted on line 340-344 (page 14)

Comments from Reviewer 3
Comment #1 - Please revise the entire manuscript for grammar, spelling, and sentence structure. There were still some issues remaining making readability at times a bit challenging.
Response: Thanks for this comment. We have carefully read the manuscript and made some editions as needed.

Comment #2 - Please refrain from using the word “interview” to mean survey or questionnaire as the former is usually preferred to describe the qualitative data collection method and not quantitative ones.
Response: Thank you for pointing out this. We agree that the term ‘in-depth interview’ is used exclusively for qualitative studies. However, in this study, we conducted pregnant women’s satisfaction at their exit after service utilization (exit interview). Data were also obtained from the health providers about their socio-demographic characteristics, regarding the trainings they took, and the working environment through interviewing them using a questionnaire (Provider interview).
Comment #3 - Please make sure the tables can be read as stand-alone exhibits –this applies to supplementary files as well. This means that the reader should be able to understand them without needing to go back to the text for explanations of acronyms (especially in supplementary files). Table 4 would benefit from clarifying what the scoring range was (0-5) in a caption so that the mean score is understandable without needing to go back to the text. Table 5 needs some reconfiguration so that the names of the variables and what was used as a reference in the table 5 are a bit less confusing and easier to read. Also in table 5 I would not report significance other than $p<0.05$ since this is what you established in your methods section as your reference for significance (p 10, line 264).
Response: We are grateful for these suggestions and comments. We tried to revise all the tables as per your suggestion.
Comment #4 I would suggest using the first paragraph of the discussion section to describe what you set out to do and what your main/most important findings are. It is currently still missing the main findings.
Response: We thank you; we revised it as per your suggestion (page 18 lines 410-417) .