PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The changing temporal association between caesarean birth and neonatal death in Ethiopia: secondary analysis of nationally representative surveys
AUTHORS	Yisma, Engida; Mol, Ben; Lynch, John; Smithers, Lisa

VERSION 1 - REVIEW

REVIEWER	A/Prof Kristen Gibbons Mater Research, Brisbane, Australia
REVIEW RETURNED	26-Nov-2018

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript. My primary concern with the manuscript is the title indicates investigation of the changing association (i.e. trend over time) is the primary objective of the manuscript, however the figures and output relating to this objective is not fully described in the manuscript. Other comments are below. “What this study adds” line 16: Add ‘be’ to “which may be aggravated by” “What this study adds” second bullet point: Consider changing to two sentences; it is currently a long sentence which is difficult to read. Abstract: Results line 21: Missing closing) Abstract: Results line 21: It is worth noting that the association with the lowest quintile of household wealth is not statistically significant, as the previous result (comparison for rural women) is Strength and limitations line 17: Insert ‘the’ before ‘Three Delays Model’ Introduction line 4: United Nation should be United Nations Introduction page 5 lines 21-28: Reporting on the previous studies in this area is highly important to this research; the Introduction would benefit from the authors expanding on this information and providing more detailed accounts of the results of these studies so the reader can understand how diverse the findings of the previous studies are. Introduction page 6 line 5: The statement referring to previous studies and concluding that null effects may be suggested is too broad and is not supported by evidence or references. While previous finds may be inconsistent, even if there are methodological/design limitations, this does not suggest null effects. The authors are requested to revise this statement by either providing evidence of studies where null effects could be

	suggested or altering the statement to discuss more broadly the issues with previous studies without drawing the significant conclusions that are currently included in the text. Either option must be referenced with the studies that the authors are referring to. Introduction page 6 line 8: Indication for caesarean intervention has not previously been identified as a contributor to understanding the association between CS and NND. If it is a key component in this relationship, it needs to be explored in the Introduction further. Introduction page 6 line 17: Change 'describe' to 'describes' Introduction page 6 line 17: The description of the 'Three Delays Model' could be moved to the Methods section, or if crucial for the reader to understand, be explored further in the Introduction, prior to the statement of aims. Methods page 7 line 13: Do women self-report births in Ethiopia? Or is it done by the hospital? For women answering the survey about a birth five years ago, is there any estimate of recall bias that may be introduced? Similarly, for women who may have had more than one birth in the five year period, has work been undertaken to validate their responses? Based on the Stanton et al article, it would be beneficial to expand on the statement of "The self-reported data collected for caesarean birth are thought to be reliable" to reassure the reader that the data are valid. Methods page 7 line 26: What standard/reference is used to define the size of the baby? Is it the mother's self-assessment? Methods page 7 line 28: For birth order, why are 2-3 grouped together? If possible they should be used separately. Methods page 8 line 8: It would be helpful to specify which analyses are on an individual patient level and which are on an aggregate region/city level. Methods page 8 line 11: Spell out CI (confidence interval) as it is the first use. How were variables entered into the multivariable model determined? Methods page 8 line 12: Were the subgroup analyses pre-defined? It is stated that the association is more pronounced in the 2016 DHS – were the analyses conducted on the other surveys and results not reported? Also, insert a 'the' before '2016 DHS'. Methods page 8 line 14: The subgroup analysis restricted to regions with the highest CS rates is being used to examine if the increased access to CS affects NND rate. Given the number of analyses in this section, it would be worth stating that these analyses are exploratory in nature. Methods page 8 line 23: It is recommended to combine the final paragraph under 'Statistical analysis' with the first paragraph under 'Statistical analysis' to keep the details on the modelling techniques together. Methods page 8: How is missing data handled? Is it highly prevelant in the DHS surveys? What type I error was used? Methods page 8 line 25: Change 'using unit of analysis's' to 'using the unit of analysis'. Methods page 8: There is no discussion of the Three Delays Model in the Methods, and how it will be applied to the data. Please include this in the manuscription. Results page 9 line 6: The reader expects that the Results section will begin with a description of the participants from all four surveys, rather than just 2016, as this would follow the description in the Statistical analysis section. Additionally, the title of the manuscript includes "Interpreting the changing association", indicating that the trend over time is central to this paper. Please

	reorder the Results section, and include the information from the 2000, 2005, 2011 surveys in the primary Tables. Results page 9 line 20: The title of the article specifies examining the changing association between CS rate and NND rate, however the primary figure/analysis that examines this is in a Supplementary Figure and only warrants one line in the article. Include the results from the regression analysis in text. Figure 1: Ensure the number of decimal places for each bar is consistent. Change “Survey years” to “Survey year”. Figures 2, 3: Label the y-axes in a similar format (ie. have the % at the end for both label). Change “Survey years” to “Survey year”. These figures are very busy and could be included as supplementary figures, rather than primary figures. Table 3: What variables were used in the multivariable analysis for the analyses restricted to quintiles of wealth? Supplementary Figures 1, 2: These figures address the primary aim of the study, and should be primary figures, rather than supplementary figures. Include a confidence interval around the line in both. It’s also recommended to start the y-axis at 0, rather than its lower limit which appears to be approximately 1.2. Label the two axes in a similar format (ie. have the % at the end for both label).

REVIEWER	Ties Boerma University of Manitoba, Winnipeg, Canada
REVIEW RETURNED	06-Jan-2019

GENERAL COMMENTS	Review Caesarean section and neonatal mortality in Ethiopia Main comments 1. Because Ethiopia has such a low C-section rate the numbers of deaths following a C-section is very small. This is a major limitation of the study. The confidence intervals are wide. I estimated from the tables that there were 4, 7, 10 and 17 neonatal deaths following C Sections in the four DHS surveys in Ethiopia respectively. On page 10 the analysis is done separately for rural and urban women. This is the right decision given the very large rural urban differences in C-section rate (and neonatal mortality). The numbers however will become even smaller. I find the confidence interval surprisingly small (probably owing to decent denominators). This needs to be made explicit in both the methods and the limitations in the discussion. The small numbers make the efforts to explain pathways and differentials rather futile. 2. Ethiopia is a special case. Even within Africa its C-section and institutional delivery rates are extremely low. The major increases in the past decade are notable, but Ethiopia is still one of the lowest coverage countries in the world. To study the impact of an advanced life saving intervention – C-section – on a fairly common outcome (neonatal mortality) is worthwhile but it is also special in the sense that the findings cannot easily be generalized to other countries. The explanations given in the discussions are likely to be correct: at very low levels of C-section rate, most cases are emergency referrals with major problems and the likelihood of an adverse outcome of pregnancy is high. The paper’s discussion is weak however in explaining the specific context in Ethiopia, and how this affects generalizability of the results. The low coverage rates for especially rural Ethiopia need to be explained to the reader in the introduction and also in the discussion.

	3. Page 16: The most important limitation of our study is that we lack data on intrapartum indications for caesarean delivery. This is not entirely correct. The DHS 2016 has questions on the timing of the decision to have a C-section, which are considered a proxy for the indications. While this is likely to be a noisy association, it would be good to add this variable to the 2016 data. Minor comments: It would be useful for the paper to include references to the recently published series in the Lancet on lack of access and overuse of C-section. Page 5: WHO suggests a C-section rate of 10-15%. This is not correct. WHO considers a population C-section rate of 10-15% indicative of an optimal C-section rate. The population rate however does not indicate if all those who receive a C-section were indeed those women who needed a C-section. Page 9: "The overall proportion of neonatal death was 2.9% in 2016." Needs to be phrased correctly. Page 12: "Like other developing settings, these problems are especially applicable in Ethiopia, where distance to a health facilities limits access to health care for 50% of women." Would be good to point out that the access to C-section situation in Ethiopia is worse than in most other settings. Figure 3: there must be a better way to present these results

REVIEWER	Adeline Boatin MGH, USA
REVIEW RETURNED	08-Jan-2019

GENERAL COMMENTS	This is a study based on a secondary analysis of the Ethiopia DHS to interpret the changing association between caesarean birth and neonatal death. To add to the existing literature examining the relationship between CS rates and neonatal mortality their main aim is to examine country level data in Ethiopia over time and then apply the three-delay model to interpret the changing association between CS and neonatal death over time. They also make note of taking account of contextual factors such as unequal access, infrastructure and workforce constraints. Although this is stated in the introduction and again reviewed in the discussion, there is no explanation of how the three-delay model is applied to facilitate this interpretation in the methods. Similarly, the explanation of how their subgroup analysis correlate to the contextual factors noted – unequal access, infrastructural and workforce constraints needs to be improved and expanded. For example, it is not exactly clear how exploring the association between CS and neonatal birth in high vs low CS regions, or by wealth is directly related to infrastructure and workforce constraints – it appears a few assumptions have been made to make that leap this should be further explained. This then relates to the interpretation where again, the link between the differences found in the subgroups and contextual factors seems based more on assumptions than directly supported by the data presented in the study. Similarly, the interpretation based on the three-delay model is not necessarily supported by data presented in the results but rather by assumptions made by the authors. These may be indeed relevant

	factors and explanations, but they are not directly supported by data demonstrated and that should be made clear. For e.g. the authors discuss attendance by a skilled provider as a factor and related to the 2nd and 3rd delay but this information or factor was not included in the analysis even though that data is captured in the DHS. Thus, though it may be true, it is unclear how the findings presented e.g. “increasing timely access to CS and timely decision for CS” are directly supported by the results of this study. Some additional specific comments - Reference for WHO recommendation on a specific CS rate is outdated and has been revised by WHO – authors should use the more up to date recommendation which avoids citing a specific rate - How were a priori confounders chosen? Based on expert opinion? prior studies? Something else? - Can you include some indication of statistical significance in the differences between groups in Table 1 – e.g. p-values.

VERSION 1 – AUTHOR RESPONSE

Response to Reviewer #1: A/Prof Kristen Gibbons

Comment: Thank you for the opportunity to review this manuscript. My primary concern with the manuscript is the title indicates investigation of the changing association (i.e. trend over time) is the primary objective of the manuscript, however the figures and output relating to this objective is not fully described in the manuscript.

Authors’ response: Our work does describe the changing temporal association between caesarean section and neonatal death. This requires estimating an association for each DHS survey year, so the ‘trend’ is in describing changing adjusted Prevalence Ratios. This is precisely what we have described in the first sentence of the Results in the Abstract, and in Table 2. At the request of the Editor, the title has been amended to be consistent with the Journal style.

Comment: “What this study adds” line 16: Add ‘be’ to “which may be aggravated by”

Authors’ response: Based on the editor’s comment this section is not required and has been removed (see page 2).

Comment: “What this study adds” second bullet point: Consider changing to two sentences; it is currently a long sentence which is difficult to read.

Authors’ response: Based on the editor’s comment this section is not required and has been removed (see page 2).

Comment: Abstract: Results line 21: Missing closing)

Authors’ response: We have included the missing closing round bracket (page 3, line 21).

Comment: Abstract: Results line 21: It is worth noting that the association with the lowest quintile of household wealth is not statistically significant, as the previous result (comparison for rural women) is

Authors' response: In accordance with modern epidemiological and statistical practices, we give more emphasis to effect estimate and confidence interval rather than the significant versus not significant dichotomy of p-values. This is consistent with recommendations by the American Statistical Association who state "Practices that reduce data analysis or scientific inference to mechanical "bright-line" rules (such as "$p<0.05$") for justifying scientific claims or conclusions can lead to erroneous beliefs and poor decision making."1

Reference

1. Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. The American Statistician 2016; 70(2): 129-33.

Comment: Strength and limitations line 17: Insert 'the' before 'Three Delays Model'.

Authors' response: We have revised it accordingly, page 4, line 16.

Comment: Introduction line 4: United Nation should be United Nations

Authors' response: We have revised it accordingly (page 5, line 4).

Comment: Introduction page 5 lines 21-28: Reporting on the previous studies in this area is highly important to this research; the Introduction would benefit from the authors expanding on this information and providing more detailed accounts of the results of these studies so the reader can understand how diverse the findings of the previous studies are.

Authors' response: We have revised and expanded what previous studies reported about the association between caesarean birth and neonatal mortality. This revision can be found in track changes, page 5, lines 32-33 & page 6, line 1-6.

Comment: Introduction page 6 line 5: The statement referring to previous studies and concluding that null effects may be suggested is too broad and is not supported by evidence or references. While previous findings may be inconsistent, even if there are methodological/design limitations, this does not suggest null effects. The authors are requested to revise this statement by either providing evidence of studies where null effects could be suggested or altering the statement to discuss more broadly the issues with previous studies without drawing the significant conclusions that are currently included in the text. Either option must be referenced with the studies that the authors are referring to.

Authors' response: We have removed the statement about null effects. We also revised the statement and provided a supporting reference. Our revision can accessed in track changes on page 6, line 14.

Comment: Introduction page 6 line 8: Indication for caesarean intervention has not previously been identified as a contributor to understanding the association between CS and NND. If it is a key component in this relationship, it needs to be explored in the Introduction further.

Authors' response: We have elaborated the description regarding medical indications for caesarean intervention, page 6, lines 16-18.

Comment: Introduction page 6 line 17: Change 'describe' to 'describes'

Authors' response: As the description of the 'Three Delays Model' is moved to the Methods section, this section has been removed (see page 6, line 29).

Comment: Introduction page 6 line 17: The description of the 'Three Delays Model' could be moved to the Methods section, or if crucial for the reader to understand, be explored further in the Introduction, prior to the statement of aims.

Authors' response: We have moved the description of the 'Three Delays Model' from the "Introduction" to "Methods" as per the suggestion. We have also provided detail description of the 'Three Delays Model' in "Methods" under sub-title The 'Three Delays Model' on page 10, line 17.

Comment: Methods page 7 line 13: Do women self-report births in Ethiopia? Or is it done by the hospital? For women answering the survey about a birth five years ago, is there any estimate of recall bias that may be introduced? Similarly, for women who may have had more than one birth in the five year period, has work been undertaken to validate their responses? Based on the Stanton et al article, it would be beneficial to expand on the statement of "The self-reported data collected for caesarean birth are thought to be reliable" to reassure the reader that the data are valid.

Authors' response: In Ethiopia, women self-report births in surveys. There is no established electronic system in health facilities to report births which can be used for research purposes, particularly in rural health facilities. Health facilities record all births in their own delivery record books. However, DHS collects quality data on fertility, family planning, maternal and child health, nutrition, mortality, environmental health, HIV/AIDS, malaria, and provision of health services in over 90 low-and middle-income countries. The DHS uses a standardized questionnaire to collect data. The DHS questionnaire is designed to the extent that the responses of each participant can be validated. For example, first the name, sex and month/year of all births in each household, whether alive or not at the time of the survey, are recorded, starting with the first birth that women had. Then, by cross-checking the name and year of each children from the recoded section of the questionnaire, the interviewer asks women about all of their births within the five years before the day of the survey separately, starting with the last birth. An example of question asked include: When you got pregnant with (NAME), did you want to get pregnant at that time?

We have expanded the statement stated as: "The self-reported data collected for caesarean birth are thought to be reliable" using Stanton et al article as per the comment. Our revision can be found in track changes, page 7, line 15-17.

Comment: Methods page 7 line 26: What standard/reference is used to define the size of the baby? Is it the mother's self-assessment?

Authors' response: All DHS in low-and middle-income countries collect information on baby's size at birth based on mother's perception (self-assessment) by asking women as "was the newborn very large, larger than average, average, smaller than average, very small and don't know?" This is because the majority of births in these countries occur at home and birth weight is not recorded. We have indicated this as limitation in the discussion, page 21, line 23.

Comment: Methods page 7 line 28: For birth order, why are 2-3 grouped together? If possible they should be used separately.

Authors' response: DHS uses this grouping (2-3) to report the birth order in the DHS final reports. We, however, have separated birth order categories as 1, 2, 3, 4, 5 and 6+ in Table 1. For multivariable regression analyses model, we used the existing category. Our revision can be found in Table 1 in our revised manuscript, page 31.

Comment: Methods page 8 line 8: It would be helpful to specify which analyses are on an individual patient level and which are on an aggregate region/city level.

Authors' response: In response to this comment we introduced new subheadings to the 'Statistical Analysis' section of the manuscript. We now have two sections that specify which analyses are at the individual level (from Page 8 Line 20) and at the aggregate level (page 10, line 8).

Comment: Methods page 8 line 11: Spell out CI (confidence interval) as it is the first use.

Authors' response: We have revised it accordingly, page 8, line 23.

Comment: How were variables entered into the multivariable model determined?

Authors' response: We have identified all variable (confounders) entered in the models based on a priori subject-matter and expert knowledge. This is the preferred method of identifying and selecting the potential confounders in most epidemiologic literature. We have included information on how the confounders were identified in the revised manuscript, page 7, line 27-28.

Comment: Methods page 8 line 12: Were the subgroup analyses pre-defined? It is stated that the association is more pronounced in the 2016 DHS – were the analyses conducted on the other surveys and results not reported? Also, insert a 'the' before '2016 DHS'.

Authors' response: All sub-group analyses were decided a priori. We first conducted analyses for each of the surveys in 2000, 2005, 2011, and 2016 separately. We then compared the strength of the association between caesarean birth and neonatal death across all surveys analysed. After noting that the association between caesarean section and neonatal death increased over time and become pronounced in 2016 DHS, we then explored the 2016 year further to better understand what was happening within the context of Ethiopia. This allowed us understand the role of the contextual factors in the association between caesarean birth and neonatal death. We have revised and indicated this in the revised manuscript, page 8. Lines 24-29.

Comment: Methods page 8 line 14: The subgroup analysis restricted to regions with the highest CS rates is being used to examine if the increased access to CS affects NND rate. Given the number of analyses in this section, it would be worth stating that these analyses are exploratory in nature.

Authors' response: We have included this information in the revised manuscript as per the suggestion, page 9, line 17-18.

Comment: Methods page 8 line 23: It is recommended to combine the final paragraph under 'Statistical analysis' with the first paragraph under 'Statistical analysis' to keep the details on the modelling techniques together.

Authors' response: We have combined the final paragraph with the first one as per the recommendation. Our revision can be found in track changes in the revised manuscript, page 8, lines 10-18.

Comment: Methods page 8: How is missing data handled? Is it highly prevalent in the DHS surveys? What type I error was used?

Authors' response: Missing information is uncommon in DHS because the data is collected by a trained researcher at a face-to-face interview. For the variables used in our analyses, missing data ranged from 0% (for maternal age at birth) to 0.08% (for place of delivery).

Comment: Methods page 8 line 25: Change 'using unit of analysis's' to 'using the unit of analysis'.

Authors' response: We have revised it accordingly (see page 8, line 13).

Comment: Methods page 8: There is no discussion of the Three Delays Model in the Methods, and how it will be applied to the data. Please include this in the manuscript.

Authors' response: We have provided the detail discussion of the 'Three Delays Model' under 'Methods' section. We have also identified a number of contributing factors to the "three delays" from Ethiopian DHS and conducted additional analysis presented in Table 4 in the revised manuscript. This

allowed us, in addition to the theoretical application of the model, provide empirical evidence to the three delays to help facilitate the interpretation of the changing association between caesarean birth and neonatal death within the context of Ethiopia. This revision can be found in track changes, on page 10, line 17-28 & page 11, line 1-18.

Comment: Results page 9 line 6: The reader expects that the Results section will begin with a description of the participants from all four surveys, rather than just 2016, as this would follow the description in the Statistical analysis section. Additionally, the title of the manuscript includes "Interpreting the changing association", indicating that the trend over time is central to this paper. Please reorder the Results section, and include the information from the 2000, 2005, 2011 surveys in the primary Tables.

Authors' response: We have provided the description of participants using each of the data for 2000, 2005, 2011 and 2016 in Table 1 and modified the text to comment on each of the surveys on page 12 lines 4-6. We have also re-ordered the information in Table 2 in the logical order.

Comment: Results page 9 line 20: The title of the article specifies examining the changing association between CS rate and NND rate, however the primary figure/analysis that examines this is in a Supplementary Figure and only warrants one line in the article. Include the results from the regression analysis in text.

Authors' response: The key information on the changing individual-level association is reported in Table 2 and presented in the results, and in the discussion. We have also revised and included analysis based on aggregate data. We have also further disaggregated the data on caesarean section rate and proportion of neonatal deaths by urban-rural areas for each of the nine regions and two city administrations in Ethiopia for all four DHS surveys. However, the urban-rural classification for Addis Ababa is only available for the 2005 DHS data. These results in a total of 85 data points. Based on this data, we re-analysed and provided the results in the main text of the manuscript. Our revision can be found in the revised manuscript in track changes, page 13, lines 12-16.

Comment: Figure 1: Ensure the number of decimal places for each bar is consistent. Change "Survey years" to "Survey year".

Authors' response: We have rounded up the figures to one decimal point, so all bar is consisted now. We have also changed "Survey years" to "Survey year".

Comment: Figures 2, 3: Label the y-axes in a similar format (ie. have the % at the end for both label). Change "Survey years" to "Survey year". These figures are very busy and could be included as supplementary figures, rather than primary figures.

Authors' response: We have presented Figure 3 in the form of table by including additional information and have then moved it to the 'Supplementary material' as supplementary Table A2. We have made all revisions as per the comment for Figure 2 and we feel that the message being conveyed is clearer if figure 1 is presented to the main text of the manuscript than in the Supplementary material.

Comment: Table 3: What variables were used in the multivariable analysis for the analyses restricted to quintiles of wealth?

Authors' response: We apologize here that we did not include a legend as an indicator of what variables were adjusted for in error for the analysis restricted to quintiles of wealth. We have now revised and provided a legend that explain this in track changes in the revised manuscript on page 32—Table 3.

Comment: Supplementary Figures 1, 2: These figures address the primary aim of the study, and should be primary figures, rather than supplementary figures. Include a confidence interval around the line in both. It's also recommended to start the y-axis at 0, rather than its lower limit which appears to be approximately 1.2. Label the two axes in a similar format (ie. have the % at the end for both label).

Authors' response: We have included the Supplementary figures 1 & 2 in the main text of in the revised manuscript. We have also added confidence intervals to the figures and labeled the axes in a similar format.

Response to Reviewer #2: Ties Boerma

Major comments

Comment: Because Ethiopia has such a low C-section rate the numbers of deaths following a C-section is very small. This is a major limitation of the study. The confidence intervals are wide. I estimated from the tables that there were 4, 7, 10 and 17 neonatal deaths following C Sections in the four DHS surveys in Ethiopia, respectively. On page 10 the analysis is done separately for rural and urban women. This is the right decision given the very large rural urban differences in C-section rate (and neonatal mortality). The numbers however will become even smaller. I find the confidence interval surprisingly small (probably owing to decent denominators). This needs to be made explicit in both the methods and the limitations in the discussion. The small numbers make the efforts to explain pathways and differentials rather futile.

Authors' response: We agree that the national caesarean section rates in Ethiopia are very low and also there are substantial disparities among regions. Since the sample is randomly selected and our analyses are weighted, the weighting improves the representativeness of the data in terms of size, distribution and characteristics of the Ethiopian population. The confidence interval for some subgroup analyses are wide and we have acknowledged this as limitation of our study. We have revised and added comments about this to the revised manuscript both in methods and limitations section of the discussion. Our revision can be found in track changes on page 9 (line 21-22) & page 21 (lines 10-18).

Comment: Ethiopia is a special case. Even within Africa its C-section and institutional delivery rates are extremely low. The major increases in the past decade are notable, but Ethiopia is still one of the lowest coverage countries in the world. To study the impact of an advanced life saving intervention –

C-section – on a fairly common outcome (neonatal mortality) is worthwhile but it is also special in the sense that the findings cannot easily be generalized to other countries. The explanations given in the discussions are likely to be correct: at very low levels of C-section rate, most cases are emergency referrals with major problems and the likelihood of an adverse outcome of pregnancy is high. The paper's discussion is weak however in explaining the specific context in Ethiopia, and how this affects generalizability of the results. The low coverage rates for especially rural Ethiopia need to be explained to the reader in the introduction and also in the discussion.

Authors' response: We have revised and added new sentences to the discussion section to acknowledge that this is a limitation of our study. We agree that the interpretation of our study is only confined to the context of Ethiopia and may not necessarily generalizable to any other poor resource country settings in Africa or elsewhere. In addition to the proportion of institutional deliveries and caesarean section rates, the contextual factors may be different in other settings. We have revised our 'Interpretation and discussion' sections to include more detail on the specific context in Ethiopia. We have also indicated that additional new analysis regarding the "Three Delays Model" is done in methods section (Page 11, line 7-18) to describe factors contributing to the delays in accessing caesarean section in the context of Ethiopia in the revised manuscript (see the 'Interpretation' section on page 14-18). We have included description about the lower coverage rates, especially for rural Ethiopia in introduction and discussion sections. We have also stated that the results of our study may not be generalizable to settings outside of Ethiopia under 'discussion' section as limitations in the revised manuscript. Our revision can be found in track changes under introduction (page 6, lines 23-26), 'Interpretation' (page 14-18) and 'discussion' (page 21, line 10-18) sections.

Comment: Page 16: The most important limitation of our study is that we lack data on intrapartum indications for caesarean delivery. This is not entirely correct. The DHS 2016 has questions on the timing of the decision to have a C-section, which are considered a proxy for the indications. While this is likely to be a noisy association, it would be good to add this variable to the 2016 data.

Authors' response: We understand that the DHS 2016 collected information regarding 'timing of decision to conduct caesarean section (i.e., whether it was before or after the onset of the labour pains).' This may provide some information about why the caesarean section was performed for women because emergency caesarean section is conducted when there is labour and medical complications such as 'fetal distress', 'cord prolapse', 'prolonged and obstructed labour', 'fetal mal-presentation', 'major antepartum haemorrhage', and 'placenta praevia'. To test if this can be used as a proxy to obstetric complications (underlying medical indications for caesarean intervention), we have included this variable in the regression models for 'timing of decision to conduct caesarean section' and 'delivery by caesarean section' together. However, due to collinearity the variable could not be retained in the model 'timing of decision to conduct caesarean section'. We note that these two variables are highly correlated (correlation coefficient=0.9512). We then run the model again assuming that 'timing of decision to conduct caesarean section' as a proxy for type of caesarean birth (indicative of intrapartum or elective caesarean section). We found almost comparable estimate as our original analysis, but with larger standard error. Thus, 'timing of decision to conduct caesarean section' may be used as a proxy to the type of caesarean section (indicative of intrapartum or elective caesarean section).

We have indicated this in the revised manuscript and have provided the results of our analysis in supplementary Table A2. We have also provided description about this under 'methods' section in the revised manuscript, page 9 (lines 28-31) & page 10 (lines 1-2).

Minor comments:

Comment: It would be useful for the paper to include references to the recently published series in the Lancet on lack of access and overuse of C-section.

Authors' response: We have added the reference suggested and also revised the manuscript based on the article. Our revision can be found in track changes under 'interpretation' section, page 9, line 19-20.

Comment: Page 5: WHO suggests a C-section rate of 10-15%. This is not correct. WHO considers a population C-section rate of 10-15% indicative of an optimal C-section rate. The population rate however does not indicate if all those who receive a C-section were indeed those women who needed a C-section.

Authors' response: We apologize for not reporting the latest WHO recommendation on caesarean section rates. We have revised the sentence accordingly. Our revision can be found in track changes under 'introduction' section, page 5, line 20-21.

Comment: Page 9: "The overall proportion of neonatal death was 2.9% in 2016." Needs to be phrased correctly.

Authors' response: We have removed this sentence and provide clear description of the proportion of neonatal deaths over time in the revised manuscript. Our revision can be found on page 12, line 16-18.

Comment: Page 12: "Like other developing settings, these problems are especially applicable in Ethiopia, where distance to a health facilities limits access to health care for 50% of women." Would be good to point out that the access to C-section situation in Ethiopia is worse than in most other settings.

Authors' response: We have used this information in the revised manuscript. Our revision can be found in track changes on page 17, line 11-12.

Comment: Figure 3: there must be a better way to present these results

Authors' response: We have presented Figure 3 in the form of table by including additional data from DHS and have then moved it to the 'Supplementary material' as supplementary Table A2. As the main figures from this table are already presented in figure 1, we feel that the more detail information about the proportion of neonatal death by urban-rural residence and region can be accessible in Supplementary material on page 3.

Response to Reviewer #3: Adeline Boatin

Comment: This is a study based on a secondary analysis of the Ethiopia DHS to interpret the changing association between caesarean birth and neonatal death. To add to the existing literature examining the relationship between CS rates and neonatal mortality, their main aim is to examine country level data in Ethiopia over time and then apply the three-delay model to interpret the changing association between CS and neonatal death over time. They also make note of taking account of contextual factors such as unequal access, infrastructure and workforce constraints. Although this is stated in the introduction and again reviewed in the discussion, there is no explanation of how the three-delay model is applied to facilitate this interpretation in the methods.

Authors' response: We have revised and included an explanation about the 'Three Delays Model' in the 'methods' section and how it was applied to interpret the findings in our study. Our revision can be found in track changes under sub-title 'the Three Delays Model' in methods, page 10 & 11.

Comment: Similarly, the explanation of how their subgroup analysis correlate to the contextual factors noted – unequal access, infrastructural and workforce constraints needs to be improved and expanded. For example, it is not exactly clear how exploring the association between CS and neonatal birth in high vs low CS regions, or by wealth is directly related to infrastructure and workforce constraints – it appears a few assumptions have been made to make that leap this should be further explained. This then relates to the interpretation where again, the link between the differences found in the subgroups and contextual factors seems based more on assumptions than directly supported by the data presented in the study.

Authors' response: In the revision we have added further detail about how the different subgroup analyses give contextual insight into the situation in Ethiopia. For example, in Ethiopia, there is substantial disparities in caesarean section rates among administrative regions, with very low rates in rural areas and among the poorest women, suggesting unequal access. The stronger association found between caesarean section among regions with low caesarean section rate and among women with lowest quintile of household wealth are indicative of unequal or lack of access in caesarean which is likely related to poor health infrastructure (e.g., shortage of medical care institutions, deficiencies in surgical facilities, surgical and anesthesia personnel and equipment, and blood transfusion capacity) and due to lack of skilled birth attendants. This is because insufficient number of skilled birth attendants at any health facility and poor infrastructure (shortage of medical care institutions, poor road network) will lead to delay in accessing emergency caesarean section, which is usually accessible at specialized health facilities.

We further located DHS and other government documents in Ethiopia in order to find data/information that help us explain the specific context of Ethiopia such as proportion of births delivery by skill births attendants and problems in accessing healthcare by women according to sociodemographic characteristics as well as number of health facilities. Our revision can be found in track changes in the revised manuscript on pages 14-18.

Comment: Similarly, the interpretation based on the three-delay model is not necessarily supported by data presented in the results but rather by assumptions made by the authors. These may be indeed relevant factors and explanations, but they are not directly supported by data demonstrated and that should be made clear. For e.g. the authors discuss attendance by a skilled provider as a factor and related to the 2nd and 3rd delay but this information or factor was not included in the analysis even though that data is captured in the DHS. Thus, though it may be true, it is unclear how the findings

presented e.g. “increasing timely access to CS and timely decision for CS” are directly supported by the results of this study.

Authors’ response: We have added data from DHS to help make our interpretation supported by Data. We have identified factors such as delivery by skilled birth attendants, problems in accessing healthcare by women of reproductive age (distance to health facility; getting money for treatment; getting permission to go for treatment; and not wanting to go alone), and women’s sociodemographic characteristics (see ‘the Three Delays Model’ section, page 11, lines 7-18). These factors are important to understand and address the barriers that women face in seeking care during pregnancy and delivery. Thus, we have now introduced new table (Table 4) that summarizes these factors in the revised manuscript. We have also revised the text of the manuscript to reflect that our theoretical interpretation based on the three-delay is now also supported by data in the ‘Interpretation’ section (page 16 & 17).

Comment: Reference for WHO recommendation on a specific CS rate is outdated and has been revised by WHO – authors should use the more up to date recommendation which avoids citing a specific rate

Authors’ response: We have introduced this latest reference to the revised manuscript. Our revision can be found in track changes under introduction section, page 5, line 20-21.

Comment: - How were a priori confounders chosen? Based on expert opinion? prior studies? Something else?

Authors’ response: We have identified the confounders based on a priori subject-matter and expert knowledge. We now indicate this in the revised manuscript, page 7, line 27-28.

Comment: Can you include some indication of statistical significance in the differences between groups in Table 1 – e.g. p-values.

Authors’ response: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement explicitly discourages the use of p values to compare groups in Table 1 specifically stating “significance tests should be avoided in descriptive tables”. 1 STROBE Statement is a reporting guideline that consist of a checklist of 22 items to improve the reporting of observational studies. It was developed by renowned academics (a group of epidemiologists, methodologists, statisticians, researchers and journal editors) and endorsed by over 100 journals, including this journal. This is the most important reason why we have not made this change. Although it is common to report p-values as an indicator of statistical differences between groups in Table 1, this practice does not provide information about the differences between groups. Readers are able to compare the percentage of the characteristics of the study participants across all the surveys waves. While we believe that statistical comparison in Table 1 are counter to current best practice, we have provided sufficient information for readers to make statistical comparisons themselves if they desire.

Reference

1. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. International Journal of Surgery 2014; 12(12): 1500-24.

VERSION 2 – REVIEW

REVIEWER	Kristen Gibbons Mater Research, Australia
REVIEW RETURNED	09-Apr-2019

GENERAL COMMENTS	Thank you for your responses to the comments provided. They are generally satisfactory. * The manuscript would still benefit from further explanation on how the data is collected, i.e. self-report, as well as further explanation of the Stanton et al article, to assure the reader that the data is reliable. * While I appreciate the discussion around assessing baby size in the Discussion, it should also be noted in the methods that it is mother's assessment of baby's size. * Note the low levels of missing data in the Methods section of the manuscript.

REVIEWER	Ties Boerma University of Manitoba
REVIEW RETURNED	30-Mar-2019

GENERAL COMMENTS	The authors have responded adequately to my comments, I have no further comments

REVIEWER	Adeline Boatin Massachusetts General Hospital
REVIEW RETURNED	08-Apr-2019

GENERAL COMMENTS	My major remaining concern for this manuscript is with the use of the three delay model as an explanation or interpretation for the changing temporal association between cesarean delivery and neonatal mortality. The authors demonstrate that over time the association between CD and neonatal mortality is stronger, particular among poorer and rural women i.e. a neonatal death is more likely after a CS in 2016 compared to in 2000. The authors have made considerable effort to explain the contextual factors of the three delay model in Ethiopia, supported by data from the 2016. While this gives a comprehensive sense of the static picture in 2016 and factors that might contribute to a low CS rate and associated neonatal mortality, it still remains unclear to me how this explains the change in association over time. What has changed over time as regards the three delay model over the same time period that may explain the change in association between CS and neonatal mortality? This seems to me the crux of the interpretation.

VERSION 2 – AUTHOR RESPONSE

Response to Reviewer #2: Ties Boerma

Comment: The authors have responded adequately to my comments, I have no further comments.

Authors' response: Thank you.

Response to Reviewer #3: Adeline Boatin

Comment: My major remaining concern for this manuscript is with the use of the three delay model as an explanation or interpretation for the changing temporal association between cesarean delivery and neonatal mortality. The authors demonstrate that over time the association between CD and neonatal mortality is stronger, particular among poorer and rural women i.e. a neonatal death is more likely after a CS in 2016 compared to in 2000. The authors have made considerable effort to explain the contextual factors of the three delay model in Ethiopia, supported by data from the 2016. While this gives a comprehensive sense of the static picture in 2016 and factors that might contribute to a low CS rate and associated neonatal mortality, it still remains unclear to me how this explains the change in association over time. What has changed over time as regards the three delay model over the same time period that may explain the change in association between CS and neonatal mortality? This seems to me the crux of the interpretation.

Authors' response: Thank you for this insightful comment. The 'Three Delays Model' was used to help interpret the association between caesarean section and neonatal death using the 2016 data only. This was because the association between caesarean birth and neonatal death was more pronounced in 2016 data. We also note that the 2000 Ethiopian DHS data does not have the information required to explain the 'Three Delays Model' (i.e., all questions that may explain factors contributing to the 'Three Delays Model' are consistent only for 2005, 2011, and 2016 surveys). Therefore, the description of factors contributing to the 'Three Delays Model' we have provided for the 2016 data cannot explain the changing temporal association between caesarean section and neonatal death over time.

In response to your comment, we have removed any text where the 'Three Delays Model' has been interpreted to reflect the changing temporal association between caesarean section and neonatal death. Instead, we have amended the application of the 'Three Delays Model' to more clearly reflect the 2016 survey results. Our revision can be found in track changes in the revised manuscript (see Page(P)2 Line(L)14; P3 L14; P5 L28; P7 L23; P10 L1; P14 L26). We have also added description about why we applied the 'Three Delays Model' using the 2016 data to interpret the association between caesarean section and neonatal death in 2016 in 'Methods' (see Page 10, lines 3-5).

Response to Reviewer #1: A/Prof Kristen Gibbons

Comment: Thank you for your responses to the comments provided. They are generally satisfactory.

Authors' response: Thank you.

Comment: The manuscript would still benefit from further explanation on how the data is collected, i.e. self-report, as well as further explanation of the Stanton et al article, to assure the reader that the data is reliable.

Authors' response: In the revision, we have indicated that all the data in DHS was collected based self-report. We have added some more explanations about how the data is collected and also provided further explanation of the Stanton et al. article to assure reader that the data is reliable. This revision can be accessed in track changes on page 6, lines 15-21.

Comment: While I appreciate the discussion around assessing baby size in the Discussion, it should also be noted in the methods that it is mother's assessment of baby's size.

Authors' response: We have revised and added a new sentence which states, the baby size at birth was reported based on mother's perception (self-assessment) about the size of the baby at birth. We have also indicated, based on previous evidence, that when data on birth weight was not fully collected, mother's self-assessment of size of baby at birth can be used as a proxy to birth weight in developing countries. This revision can found in 'Methods' on page 7, lines 4-8.

Comment: Note the low levels of missing data in the Methods section of the manuscript.

Authors' response: We have added sentence to explain the low level of missing data in DHS in 'Methods' as per the comment. Our revision can be found in track changes on page 7, lines 17-18.

VERSION 3 - REVIEW

REVIEWER	Adeline Boatin Massachusetts General Hospital Harvard University USA
REVIEW RETURNED	29-Jul-2019

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.