## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | The effects of short birth interval on neonatal, infant and under-five child mortality in Ethiopia: a nationally representative observational study using inverse probability of treatment weighting |
| **AUTHORS** | Shifti, Desalegn Markos; Chojenta, Catherine; Holliday, Elizabeth; Loxton, Deborah |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Saha, Unatti<br>Tilburg University |
| **REVIEW RETURNED** | 14-Mar-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | The effects of short birth interval on neonatal, infant and under-five child mortality in Ethiopia<br><br>This study addressed important issue child mortality and statistical methods propensity score matching to balance covariates across two groups “a group when children were born after short birth intervals-less than 33 months -treatment” and “a group when children were not born after short birth intervals- less than 33 months –control).<br><br>Authors argued that their study is innovative over the existing studies of examining short birth intervals’ effect on mortality outcome. They referred that they have used the short birth interval cutoff 33 months according to WHO guidelines, however, other studies did not do so. Authors also argued that existing studies included all mothers in their analysis; however, mothers with only one child are not eligible to study birth intervals effect on the mortality of one child. Also, existing studies did not include appropriately all covariates in the analysis of determining the effects of short birth intervals on mortality outcome. In this study, authors tried to incorporate accordingly as they have argued here.<br><br>First, they have used the short interval cutoff less than 33 months (following WHO cutoffs)<br>Second, they have used a range of covariates in their analysis<br>Third, they have performed confounding analysis by using direct acyclic graph (DAG).<br>Fourth, they have computed propensity matching score and balanced the characteristics of individuals in their analysis across two groups and child mortality.<br><br>My comments are given below. |

| | |
|---|---|
| | For the analysis, authors used the DHS 2016 dataset from Ethiopia which in nature is observational data. Literature suggests that in observational data treatment assignment is not at random. This may leads to selection bias where measured and unmeasured characteristics of individuals are associated with likelihood of receiving treatment and with the outcome. Propensity scores provide a way to balance measured covariates across treatment and control groups and better approximate the counterfactual for treatment individuals.<br><br>By performing propensity score matching authors balanced the measured covariates across two groups treatment vs control, however, it is not clear to me how the authors accounted for the causal pathways (see DIA in Figure 1) in their analysis. Authors perhaps can explain this in detail and also how covariates are confounded to birth intervals and child mortality.<br><br>I also concern how authors addressed the correlated mortality outcomes because existing literature of child mortality demonstrates that sibling's deaths are correlated. If correlated deaths are not addressed it may bias the birth intervals effect.<br>I am also concern about the birth intervals' effect on mortality in different age groups of children neonatal (AOR=1.53), infant (AOR=1.94), under-five (AOR=2.02). These are in fact cumulative effects and authors probably missed to interpret them accordingly. I mean that the short birth intervals more likely (equally) to affect child deaths during neonatal period and infancy.<br><br>It could be more interesting if authors could have adjusted the denominators for different age categories in their analysis. Authors matched covariates at birth by whether children born after short birth intervals (treatment group) or not (control group). It may change the values of covariates over time so authors need to mention this as limitation of their study.<br><br>Page 8, line 136, the sentence is not clear to me<br><br>Page 8, lines 152-156, it perhaps would be explicit if authors refer the Table number in which they have reported the computed values based on the formulas for both continuous and dichotomous variables.<br><br>Page 11, lines 203-204, the interpretation is not clear to me.<br><br>Supplemental material III, Table 2, values before and after weighting for overall sample, but perhaps it could be better if they would have given values for two groups (as Figure 3) before and after weighting.<br><br>Supplemental material II, Table 1, Maternal educational level, it is not clear 3=Secondary+ (or Educated and Uneducated)<br><br>Some variables are coded with more than two categories, e.g., birth weight, delivery care etc. however, code was not assigned for all variables.<br><br>Figure 1 (DAG), treatment (SBI) affects green lines (arrow) ANC (blue circle), this is not clear to me how? Also, this is not clear how authors constructed short birth intervals preceding or succeeding |

| | |
|---|---|
| | short birth intervals? So far, I see that it is not mentioned all through the manuscript.<br><br>Page 14, lines 50-55, not matched with supplemental material III and table 2.<br><br>Page 12, Table 1, do authors reported the column percentage? So, how they would interpret the distribution and p-value? I think authors should report row percentage to conclude on the differences of each category and mortality outcome yes vs no.<br><br>Page 11, lines 198-204, interpretation is based on column percentages in Table 1. For example, neonatal mortality was higher (88.0) in rural than in urban areas (p=0.004), is it correct interpretation? I can see that the percentage was higher (91.2) in rural residents among children who did not die at neonatal period too.<br><br>What % of sample children born after SBI (less than 33 months), it is not clear from this manuscript.<br><br>Page 6, lines 106-11, sentences need to revise e.g., “requesting children’s dates of birth” – reuesting or interviewed?<br><br>Using the same dataset as the authors used for their analysis, it is observed that (see the publication which is referred in number 29 in this manuscript) child mortality was significantly higher for births who born in less than 18 months preceding birth intervals following a reductions in child mortality in other categories 18-23, 24-29, 30-35 etc. About 12% of all sample children born in the category of short birth intervals <18 months.<br><br>So, authors categorized the birth intervals between <33 months or not, it may suppress the policy goal where the policy should target precisely.<br><br>Overall, the manuscript is not written in organized way, several problems in statistical method use and interpretations are not clear. To accept for the publication it needs rigorous editing all through the manuscript including statistical analysis. As I am not expert in direct acyclic analysis see graph (DAG), the manuscript may benefit with expert review on direct acyclic DAG analysis. |

| | |
|---|---|
| **REVIEWER** | Ahammed, Benojir<br>Khulna University, Statistics Discipline |
| **REVIEW RETURNED** | 16-Apr-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | Comments to the Author<br><br>Thanks for the chance to review the paper. The authors tried to measure the effects of short birth interval on neonatal, infant and under-five child mortality in Ethiopia. Their effort is good enough, but the analysis and evidence are not sufficient to explain the causes of neonatal, infant and under-five child mortality. My specific comments are bellow.<br><br>Abstract<br>The objective of this paper is clear but not it does not play the important role in a broad sense. Even the results of this paper may |

| | |
|---|---|
| | not explain the causes of neonatal, infant and under-five child mortality in Ethiopia.<br>Introduction<br>The description of introduction is in well formatted. Authors firstly display the scenarios of short birth interval and then neonatal, infant and under-five child mortality. Then they describe the different factors associated with neonatal, infant and under-five child mortality. However, the motive of this research is well described but in a broad sense, the objectives of this paper is not clear.<br>Methods<br>1. Authors consider only on exposure variable. That mean when they apply model, they get crude results. In my knowledge, I think it is not the actual findings of any research. Some times this type of results misleads the current situation and intervention of public health program.<br>2. Is it important to use propensity score? Why?<br>3. What is the study area? Which sampling technique they use? Why they choose same number of exposure and non-exposure?<br>4. Authors collect primary data but why they considered leadership, management and governance intervention?<br>Results<br>1. Authors write the results as too much limited format. It needs to write about more important variables which are associated with neonatal, infant and under-five child mortality.<br>Discussion<br><br>Similarly, discussion also needed for the more important variables those are associated with neonatal, infant and under-five child mortality. |

**VERSION 1 – AUTHOR RESPONSE**

| **Editorial/Reviewers' comments and suggested change** | **Change made and authors' reply** |
|---|---|
| **Reviewer #1** | |
| This study addressed important issue child mortality and statistical methods propensity score matching to balance covariates across two groups "a group when | Thank you. We agree with your summary. |

children were born after short birth intervals-less than 33 months - treatment" and "a group when children were not born after short birth intervals- less than 33 months –control).

Authors argued that their study is innovative over the existing studies of examining short birth intervals' effect on mortality outcome. They referred that they have used the short birth interval cutoff 33 months according to WHO guidelines, however, other studies did not do so. Authors also argued that existing studies included all mothers in their analysis; however, mothers with only one child are not eligible to study birth intervals effect on the mortality of one child. Also, existing studies

did not include appropriately all covariates in the analysis of determining the effects of short birth intervals on mortality outcome. In this study, authors tried to incorporate accordingly as they have argued here.

First, they have used the short interval cutoff less than 33 months (following WHO cutoffs)

Second, they have used a range of covariates in their analysis

Third, they have performed confounding analysis by using direct acyclic graph (DAG).

Fourth, they have computed propensity matching score and balanced the characteristics of individuals in their analysis across two groups and child mortality.

For the analysis, authors used the DHS 2016 dataset from Ethiopia which in nature is observational data. Literature suggests that in observational data treatment assignment is not at random. This may leads to selection bias where measured and unmeasured characteristics of individuals are associated with likelihood of receiving treatment and with the outcome. Propensity scores provide a way to balance measured covariates across treatment and control groups and better approximate the counterfactual for treatment individuals. By performing propensity score matching authors balanced the measured covariates across two groups treatment vs

Please find the below reply to your question on the application of Direct Acyclic Graph in our study.

A confounder is defined as a variable that has a direct causal effect on both the main exposure variable and the outcome of interest (Zhang Z, 2019). The traditional approach to confounding is multivariable adjustment, meaning all potential confounders are included as covariates in a multiple regression model. In particular, to be defined as a confounder, a covariate should cause, and not be affected by the exposure. As illustrated in Figure 1 in the manuscript, variables with blue circles such as antenatal care (ANC), place of delivery, postnatal care (PNC), birth weight, and TT vaccine were affected by the treatment variable (short birth interval) in the presumed causal model, and they, then, affected the outcomes. For example, when the women conceived in a short birth-to-pregnancy interval (which would be a short birth-to-birth interval when the women deliver), they may not be able to start their antenatal care on time or/and may not attend the whole ANC follow up due to the burden of care she has for her closely spaced previous child. This means short birth-to-pregnancy interval (which would be a short birth-to-birth interval when the women deliver) affect ANC utilization. This may in turn affect neonatal, infant, and under-five survival. As a result, ANC and the above-mentioned other variables cannot be considered confounders. Besides, although some of the variables, such as child sex, latrine type, water source, child respiratory infection, child fever status, child diarrhoeal disease, may affect the outcome variables (neonatal, infant, and under-five mortality), they cannot affect the treatment variable (short birth interval). Therefore, to objectively identify minimum adjustment sets, which defined the set of explanatory variables for the propensity score model select, an epidemiological tool, Direct Acyclic Graph (DAG), was used in this

| | |
|---|---|
| control, however, it is not clear to me how the authors accounted for the causal pathways (see DIA in Figure 1) in their analysis. Authors perhaps can explain this in detail and also how covariates are confounded to birth intervals and child mortality. | study. DAG is a formal system of mapping variables and the direction of causal relationships among them, thus distinguishing confounders from mediators. Hence, the DAG depicted in Figure 1, illustrates the relationship between the treatment variable (short birth interval in this case), wide ranges of covariates, and outcome variables (neonatal, infant, and under-five mortality). The below further description of DAG is added in this revised version of the manuscript (lines 143-150).<br><br>*' DAG is a formal system of mapping variables and the direction of causal relationships among them.[48 49] This graphical representation of causal effects among variables helps understand whether bias is potentially reduced or increased when conditioning on covariates. Moreover, it illustrates covariates that lie in the causal pathway between the treatment and outcomes, which should not be included in the analysis as a confounder. These variables are indicated by green lines in Figure 1. This is because a propensity score that includes covariates affected by the treatment (i.e., variables on the causal pathway between treatment and outcome) obscures part of the treatment effect that one is trying to estimate.50'*<br><br>**References**<br><br>1. Attia JR, Oldmeadow C, Holliday EG, Jones MP. Deconfounding confounding part 2: using directed acyclic graphs (DAGs). Medical Journal of Australia. 2017;206(11):480-3.<br><br>2. Zhang Z. Distinguishing between mediators and confounders is important for the causal inference in observational studies. AME Med J 2019;4(35).<br><br>3. Rothman KJ, Greenland S: Precision and validity in epidemiologic studies. In Modern Epidemiology Volume 2. Edited by: Rothman KJ, |

| | |
|---|---|
| | Greenland S. Philadelphia: Lippencott-Raven Publishers; 1998:115-134. |
| I also concern how authors addressed the correlated mortality outcomes because existing literature of child mortality demonstrates that sibling's deaths are correlated. If correlated deaths are not addressed it may bias the birth intervals effect. | The potential confounding effect of the survival status of the preceding child, which was not considered in the previously submitted manuscript, has now been considered in the analysis. Accordingly, the DAG and results including the standardized differences are updated. |
| I am also concern about the birth intervals' effect on mortality in different age groups of children neonatal (AOR=1.53), infant (AOR=1.94), under-five (AOR=2.02). These are in fact cumulative effects and authors probably missed to interpret ccordingly. I mean that the short birth intervals more likely (equally) to affect child deaths during neonatal period and infancy. | Our reply to this comment is based on the updated results of the revised manuscript.<br><br>When only the adjusted odds ratio (AOR) without their confidence interval (CI) is considered, the odds of neonatal, infant, and under-five mortality looks different (increasing), which are 1.85 (AOR=1.85), 2.16 (AOR=2.16), and 2.26 (AOR=2.26) times higher among women with short birth interval than those without. However, technically (statistically), no differences were observed in AORs (and their associated effect of short birth interval in different age groups). This can be observed when the adjusted odds ratio with their corresponding confidence interval (CI) are presented together. This is because CI gives a range of plausible values for the effect size and AOR alone cannot be interpreted to give meaning to the observed results. Specifically, the confidence interval of each outcome's effect size overlap with each other; (AOR=1.85, 95% CI= 1.19, 2.89) for neonatal mortality, (AOR=2.16, 95% CI= 1.49, 3.11) for infant mortality, and |

| | |
|---|---|
| | (AOR=2.26, 95% CI= 1.60, 3.17) for under-five mortality. Let us illustrate this using the AOR and CI of neonatal mortality (AOR=1.85, 95% CI= 1.19, 2.89) as an example. The AOR and CI (AOR=1.85, 95% CI= 1.19, 2.89) show that the odds of neonatal mortality were 85% higher among women with short birth interval than those without. Besides, we are 95% confident that the odds ratio for neonatal mortality could fall anywhere between 1.19, the lower bound 95% CI, to 2.89, the upper bound 95% CI. This means that the 95% CI of neonatal mortality (95% CI= 1.19, 2.89) could include an AOR of 2.16, which is the effect size for infant mortality, and AOR=2.26, which is the effect size for under-five mortality. This indicates a lack of significant differences in the effect of short birth interval among neonatal, infant, and under-five mortality, despite small differences in the point estimates. This means that the authors did not misinterpret the results accordingly but the findings are due to the above-mentioned statistical reality. The above-mentioned justification is also applicable to the findings of the previously submitted version of the manuscript. |
| It could be more interesting if authors could have adjusted the denominators for different age categories in their analysis. Authors matched covariates at birth by whether children born after short birth intervals (treatment group) or not (control group). It may change the values of | Yes, the effect of treatment/exposure (short birth interval in this case) on each outcome variable (neonatal, infant, and under-five mortality) was assessed separately (one at a time). The data was obtained from a cross-sectional survey, where outcomes, treatment/exposure, and covariates were measured at a point in time. Unlike longitudinal studies, where the study subjects were followed over time, the changes in the value of covariates over time are not a concern for our study. |

| | |
|---|---|
| covariates over time so authors need to mention this as limitation of their study | |
| Page 8, line 136, the sentence is not clear to me | The sentence is modified as follows (line 170-178):<br>*'Given that the outcomes (i.e., neonatal, infant, and under-five mortality) were relatively infrequent, the unbiased effect of short birth interval on each outcome was estimated using propensity scores (PS) with stabilized inverse probability of treatment weighting (IPTW).'*<br><br>As it is described in lines 189-190 of this manuscript, the weighted prevalence of neonatal, infant, and under-five mortality was 2.9% (95% CI: 2.39, 3.61), 4.8% (95% CI: 4.11, 5.58), and 5.5% (95% CI: 4.73, 6.44), respectively. That is why we said the outcomes are infrequent.<br><br>Regarding the phrase 'stabilized inverse probability of treatment weighting (IPTW)':<br><br>The three common propensity score matching methods are 1) IPTW using propensity score with normalized weight, 2) IPTW with stabilized weight, and (3) greedy algorithms with 1:1 matching of the propensity score.<br><br>The first one, normalized weights for the IPTW matching method can be applied to avoid extreme values of weight by dividing each individual propensity score by the mean of all propensity scores. This allows a comparison to samples representative of specific populations. This method does not result in a loss of observations. However, IPTW with normalized method often produces large variance estimates that result in high type I error rates. Therefore, we used stabilized IPTW method to account for high propensity scores in treatment and control groups |

and this method better validates point estimation compared to normalized weight for the IPTW method. The third method, the greedy algorithm with 1:1 matching of propensity score is not recommended for the rare outcomes such as the one seen in our study.

An additional statement regarding the reason for selecting stabilized IPTW is now added in this revised version of our manuscript (lines 174-178).

'*A previous study[50] has shown that IPTW with stabilized weights preserves the sample size of the original data, provides an appropriate estimation of the variance of the main effect, and maintains an appropriate type I error rate. The other methods, such as IPTW with normalized weight and greedy algorithm with 1:1 matching methods, are discussed elsewhere.'*

References

1. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.

2. Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. Stat Med 2006;25:2084-106.

3. Coca-Perraillon M. Local and global optimal propensity score matching. Proceedings of the SAS Global Forum 2007; 2007 Apr 16-19; Orlando, FL. p. 1-9.

4. Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. Value Health 2010;13:273-7.

| | |
|---|---|
| Page 8, lines 152-156, it perhaps would be explicit if authors refer the Table number in which they have reported the computed values based on the formulas for both continuous and dichotomous variables. | The text now includes a reference to the Supplementary Table number in which standardized differences are reported (line 196 and line 253). |
| Page 11, lines 203-204, the interpretation is not clear to me. | The interpretation has been rewritten to improve clarity (lines 241-242). |
| Supplemental material III, Table 2, values before and after weighting for overall sample, but perhaps it could be better if they would have given values for two groups (as Figure 3) before and after weighting. | The purpose of kernel densities (Figure 2) is to graphically demonstrate the propensity score balance in the treatment group (i.e., women with short birth interval) and control groups (women with non-short birth interval). Balance in propensity scores was considered to be achieved when the kernel density line for the treatment group and control group lay closer together (line 205-209 and lines 247-249). It is a subjective way of assessing the balance in propensity score in the treatment group before and after weighting.<br><br>The second and relatively objective method of assessing the balance is computing standardized differences of the covariates as presented in supplemental material II, Table 2. The value of standardized differences illustrates the balance of measured covariates/confounders assessed across treatment groups before and after weighting. A standard difference less than 0.1 has been suggested as indicating a negligible difference in the mean or prevalence of a covariate between treatment and control groups. |

| | |
|---|---|
| | Technically, it is not possible to compute a single standard difference for the overall sample. This is because the standard difference for continuous and categorical variables can be computed using different mathematical equations as shown below (lines 196-202):<br><br>For a continuous covariate, the standardized difference is defined as:<br><br>$$d = \frac{(\bar{x}_{treatment} - \bar{x}_{control})}{\sqrt{\frac{s^2_{treatment} + s^2_{control}}{2}}}$$<br><br>where $\bar{x}_{treatment}$ and $\bar{x}_{control}$ denote the sample mean of the covariate in treated and untreated subjects, respectively and $s^2_{treatment}$ and $s^2_{control}$ denote the corresponding sample variances of the covariate. The standardized difference for a dichotomous variable is given as:<br><br>$$d = \frac{(\hat{p}_{treatment} - \hat{p}_{control})}{\sqrt{\frac{\hat{p}_{treatment}(1-\hat{p}_{treatment}) + \hat{p}_{control}(1-\hat{p}_{control})}{2}}}$$<br><br>where $\hat{p}_{treatment}$ and $\hat{p}_{control}$ denote the prevalence of the dichotomous variable in treated and untreated subjects, respectively.<br><br>Even if the variables were either continuous or categorical, it does not make sense to compute a single value of standardized difference. This is because a single standard difference would not show from which specific covariate or covariates does the imbalance occurred if any. The one presented in our study, however, explicitly shows the balance status of each covariate considered in the study. |
| Supplemental material II, Table 1, Maternal educational level, it is not clear 3=Secondary+ (or | Our apologies, this was a typo that has now been corrected. The corrected text for maternal education reads as 1= Uneducated, 2=Primary and 3=Secondary+ |

| | |
|---|---|
| Educated and Uneducated) | (NB. Supplemental material II has now become Supplemental material I ) |
| Some variables are coded with more than two categories, e.g., birth weight, delivery care, etc. however, code was not assigned for all variables. | Our apologies, this was an oversight; codes have now been assigned for all variables. |
| Figure 1 (DAG), treatment (SBI) affects green lines (arrow) ANC (blue circle), this is not clear to me how? Also, this is not clear how authors constructed short birth intervals preceding or succeeding short birth intervals? So far, I see that it is not mentioned all through the manuscript. | A short birth interval, a birth-to birth-interval of less than 33 months, is a sum of <u>less than 24 months of birth-to-conception</u> and 9 months of pregnancy period. When the conception occurred in less than 24 months, it may affect women's antenatal care utilization. One of the reasons could be the women may be busy taking care of their young preceding child or may not have someone to look after their preceding young child when she visits the health facility for maternal health services such as antenatal care (ANC). Therefore, a short birth interval (implicitly referring to a short birth-to-conception interval when the variable is ANC) affects the ANC utilization and other maternal and child health service utilization. That is why the arrow runs from short birth interval (short birth-to-conception section period) to the ANC.<br><br>The current study considered the preceding birth interval and was mentioned in line 99. Additionally, we have now added additional statement as follow (line 127-136):<br><br>*"A preceding birth interval, the amount of time between the birth of the child under study (index child) and the immediately preceding birth, was considered in this study."* |
| Page 14, lines 50-55, not matched with | There were no lines 50-55 on page 14. We assume you might mean the 'covariate balance', which was mentioned on page 14 and annexed |

| | |
|---|---|
| supplemental material III and table 2. | supplementary material III (it is now supplementary material II). It is true that after weighting adjustment, standardized differences of covariates were all less than 0.1 (10%) (lines 251-253). |
| Page 12, Table 1, do authors reported the column percentage? So, how they would interpret the distribution and p-value? I think authors should report row percentage to conclude on the differences of each category and mortality outcome yes vs no. | The selection of either the column or row percentages can be of interest to the researcher, and on the question to be answered. The below column percent (weighted) illustrates that <u>from neonatal mortality</u>, 88.0% of them occurred in rural areas. |

| Variable | Neonatal mortality | |
|---|---|---|
| Residence | No | Yes |
| Urban | 1264 (8.8) | 22 (12.0) |
| Rural | 6933 (91.2) | 229 (88.0) |

On the other hand, the below row percent (weighted) demonstrates that <u>from rural residents</u>, there was 2.8% of neonatal mortality.

| Variable | Neonatal mortality | |
|---|---|---|
| Residence | No | Yes |
| Urban | 1264 (96.0) | 22 (4.0) |
| Rural | 6933 (97.2) | 229 (2.8) |

Generally, as long as it is interpreted correctly, either of the above methods is acceptable for use. Column percentage, however, best suited the context of our study, and Table 1 presents the weighted proportion (column percentage) of outcome variables (neonatal, infant,

| | |
|---|---|
| | and under-five mortality) by the background characteristics of the respondents.<br><br>However, the p-value was calculated using Pearson's chi-squared test, which is the same regardless of whether column or row percentages are used. This is now specified in the analysis section (lines 170-171).<br><br>*"Participants' characteristics were described using frequency with percent. P-values were calculated using Pearson's chi-squared test."* |
| Page 11, lines 198-204, interpretation is based on column percentages in Table 1. For example, neonatal mortality was higher (88.0) in rural than in urban areas (p=0.004), is it correct interpretation? I can see that the percentage was higher (91.2) in rural residents among children who did not die at neonatal period too. | Yes, the interpretation is correct. As you may see from the table, neonatal mortality was also higher in rural (88.0%) than in urban areas (12.0%) (p=0.004).<br><br>The one you mentioned (i.e., 91.2%) is about those who survived the neonatal period (it is not about those who died). That means that majority of the neonate who survived reside in the rural areas (91.2%) than urban areas (8.8%). Generally, out of 8,488 study participants included in this analysis, 1,286(15.2%) women were urban residents and 7,162(84.8) women were rural residents. Please also see our response given for the previous comment on the column and row percentages. |
| What % of sample children born after SBI (less than 33 months), it is not clear from this manuscript. | The below information has now been included in the revised version of the manuscript (line 255).<br><br>*'The prevalence of short birth interval in Ethiopia was 45.8% (95% CI: 42.91–48.62).'* |

| | |
|---|---|
| Page 6, lines 106-11, sentences need to revise e.g., “requesting children’s dates of birth” – requesting or interviewed? | Although ‘requesting’ refers to ‘politely or formally asking for information (children’s date of birth in this case)’ and the sentence is revised as follows with detailed information (lines 129-136).<br><br>*‘Women’s birth interval data were collected through extracting the date of birth of their biological children data from children’s birth /immunization certificate, and/or asking information regarding their children’s date of birth from the women. Mothers were asked to confirm the accuracy of the information before documenting children’s date of birth from children’s birth/immunization certificates. This crosschecking was performed to avoid errors, since in some cases the documented birth date may represent the date when the birth was recorded, rather than the actual birth date. In the absence of children’s birth certificates, information regarding children’s date of birth was obtained from their mothers.’* |
| Using the same dataset as the authors used for their analysis, it is observed that (see the publication which is referred in number 29 in this manuscript) child mortality was significantly higher for births who born in less than 18 months preceding birth intervals following a reductions in child mortality in other categories 18-23, 24-29, 30-35 etc. About 12% of all sample children born in the | The article (Laelago T. 2019) mentioned has several key limitations, as outlined below:<br><br>The base for classifying the birth interval into less than 18 months, 18-23 months, 24-29 months, 30-35 months, 36-47 months, 48-59 months, and 60 and above in the study you mentioned (Laelago T. 2019) is not clear and looks haphazard classifications. Our study, on the other hand, used the most recent recommendation by the World Health Organization (the guideline cited in the manuscript). Understanding the impact of short birth interval on neonatal, infant, and under-five mortality, using the WHO definition, is necessary for the formulation of valid, consistent policies and health planning strategies and interventions to improve child health outcomes (lines 85-88).<br><br>Secondly, the study conditioned only on limited covariates such as ‘maternal age at birth, educational status of women, wealth index of HH, |

| | |
|---|---|
| category of short birth intervals <18 months. So, authors categorized the birth intervals between <33 months or not, it may suppress the policy goal where the policy should target precisely. | sex of the child, place of residence, the child wanted or not.' In contrast, our study conditioned on a wide range of covariates after reviewing relevant literature. These were maternal age at the birth of the index child, maternal education, maternal occupation, husband's education, husband's occupation, household wealth status, the total number of the preceding child, survival status of the preceding child, place of residence (urban/rural), administrative regions, access to media, and decision making autonomy. We have now indicated this limitation of the previous study (Laelago T. 2019) in this revised version of our manuscript (lines 95-97). The findings of the study performed by Laelago T, 2019 would have been different if a wide variety of covariates had been considered. |
| Overall, the manuscript is not written in organized way, several problems in statistical method use and interpretations are not clear. To accept for the publication it needs rigorous editing all through the manuscript including statistical analysis. As I am not expert in direct acyclic analysis see graph (DAG), the manuscript may benefit with expert review on direct acyclic DAG analysis. | First of all, thank you very much for taking your time to review our paper and providing us your comments.<br><br>We have endeavoured to increase the organisation and clarity of statistical descriptions and interpretation in the revised manuscript. Please also note the authors include a senior statistician with expertise in the use of directed acyclic graphs (DAGs) for causal inference (E Holliday). Other co-authors have also previously used DAGs in their research, e.g.,:<br><br>Please see the below articles previously published by the co-authors which used, or have provided advice on the use of, these methods:<br><br>1. Harris ML, Hure AJ, **Holliday E**, **Chojenta C**, et al. Association between preconception maternal stress and offspring birth weight: findings from an Australian longitudinal data linkage study. BMJ open 2021;11(3):e041502. |

| | |
|---|---|
| | 2. Attia JR, Oldmeadow C, **Holliday EG**, et al. Deconfounding confounding part 2: using directed acyclic graphs (DAGs). *Medical Journal of Australia* 2017;206(11):480-83.<br><br>3. Harris ML, Oldmeadow C, Hure A, Luu J, **Loxton D**, et al. Stress increases the risk of type 2 diabetes onset in women: A 12-year longitudinal study using causal modelling. PloS one 2017;12(2):e0172126.<br><br>Regarding your statistical concerns, we have provided further information regarding how the p-values, in Table 1, were calculated (lines 170-171). We trust we have addressed your comments and concerns in our revised manuscript. Thank you. |
| **Reviewer #2** | |
| Thanks for the chance to review the paper. The authors tried to measure the effects of short birth interval on neonatal, infant and under-five child mortality in Ethiopia. Their effort is good enough, but the analysis and evidence are not sufficient to explain the causes of neonatal, infant and under-five child mortality. | Thank you. We will respond to your comments accordingly. |
| Abstract | As mentioned in the introduction section (lines 55-56 and line 255), short birth interval is one of the public health concerns in Ethiopia with a prevalence of 45.8%. Previous studies have shown that short birth |

| | |
|---|---|
| The objective of this paper is clear but not it does not play the important role in a broad sense. Even the results of this paper may not explain the causes of neonatal, infant and under-five child mortality in Ethiopia. | interval associated with adverse child and maternal health outcomes (line 59-62). However, no conclusive evidence has been found regarding the effect of short birth interval (as per the WHO definition) on neonatal, infant, and under-five mortality in Ethiopia. Given that our study identified an effect of short birth interval on neonatal, infant, and under-five mortality in Ethiopia, we believe this paper provides important additional information to inform policy and public health interventions in this country. Please note this study was not designed to comprehensively assess all possible causes of neonatal, infant, and under-five mortality in Ethiopia, and does not report on these. Rather we have assessed, and reported on, a single, important cause of this mortality, which is amenable to public health intervention. |
| Introduction<br><br>The description of introduction is in well formatted. Authors firstly display the scenarios of short birth interval and then neonatal, infant and under-five child mortality. Then they describe the different factors associated with neonatal, infant and under-five child mortality. However, the motive of this research is well described but in a broad sense, the | Thank you.<br><br>The objective of this paper is described in the first sentence of the abstract: “To assess the effect of short birth interval on neonatal, infant and under-five mortality in Ethiopia”. The objective is also stated in the last sentence of the Introduction: “This paper aimed to assess the effect of short birth interval on neonatal, infant, and under-five mortality using the most recent WHO definition and adjusting for a comprehensive set of potential confounders”. We hope this provides reassurance that the paper’s objective is clear. |

| | |
|---|---|
| objectives of this paper is not clear. | |
| Methods<br><br>1. Authors consider only on exposure variable. That mean when they apply model, they get crude results. In my knowledge, I think it is not the actual findings of any research. Some times this type of results misleads the current situation and intervention of public health program. | Please note that while we defined a single exposure variable (short birth interval), the effects of this exposure were estimated while adjusting for twelve covariates considered as control variables/potential confounders. These covariates were maternal age at the birth of the index child, maternal education, maternal occupation, husband's education, husband's occupation, household wealth status, the total number of the preceding child, survival status of the preceding child, place of residence (urban/rural), administrative regions, access to media, and decision making autonomy. The study participants were kept similar in all the above-mentioned control variables/confounders except the treatment/exposure variable. This was done to control for the effect of the above-mentioned potential confounders and estimate the unbiased effect of short birth interval on neonatal, infant, and under-five mortality. This was ascertained by checking the covariates balance between the control and treatment groups by computing the standardized difference of covariates and constructing the density plot (line 193-209). Please see the variable section (i.e., in the methods) and balance diagnostics (in the results section) of this manuscript for further information. Unlike the standard regression analysis where one may report the unadjusted and adjusted effect size of all covariates included in the analysis, this analysis employed a more robust statistical method, propensity scores with stabilized inverse probability of treatment weighting (lines 171-225). As it is presented in Table 2, the effect of short birth interval on neonatal, infant, and under-five mortality was estimated after adjusting for the control variables/potential confounders. |

| | |
|---|---|
| 2. Is it important to use propensity score? Why? | Yes, in this study there were advantages to using propensity scores, as outlined below (please see lines 171-190, lines 209-225).<br><br>A randomized control trial would have been a gold standard method to investigate the effect of treatment (short birth interval in this case) on outcomes (neonatal, infant, and under-five mortality in this case). However, for ethical reasons mainly and logistical, economic, or other reasons, it would not be feasible to investigate the effect of short birth interval on the above-mentioned child health outcomes using randomized control trials. Therefore, using propensity score analysis, data obtained from the observational study can be used to estimate the effect of treatment (short birth interval in this case) on the outcomes (neonatal, infant, and under-five mortality). Propensity scores are used to account for imbalance in covariates between treated (women with short birth interval) and control group (women with non-short birth interval), which make the identification of the unbiased effect of treatment (i.e., short birth interval) on the outcomes (neonatal, infant, and under-five mortality) easier.<br><br>Additionally, when the outcome of interest is a rare event, it results in data sparsity, and the application of multivariable models would not be sound. For example, if researchers were concerned that educational level might affect both treatment selection and outcome, one strategy would be to compare women with similar educational levels in both treatment and comparison groups. As variables, such as maternal age, employment status, place of residence, wealth status, region, etc., are added to the matching process, however, it becomes more and more difficult to find exact matches for women (i.e., it is unlikely to find individuals in both the treatment and comparison groups with identical educational level, maternal age, employment status, place of residence, wealth status, region, etc.). Propensity scores solve this dimensionality |

| | |
|---|---|
| | problem by compressing the relevant factors into a single score. Individuals with similar propensity scores are then compared across treatment (women with short birth interval) and comparison groups (women with non-short birth interval). Propensity scores with inverse probability of treatment weights (IPTW) in particular is one of the statistical methods used for this purpose, especially when the outcome of interest is rare.<br><br>Reference<br><br>1. Deb S, Austin PC, Tu JV, Ko DT, Mazer CD, Kiss A, et al. A review of propensity-score methods and their use in cardiovascular research. Canadian Journal of Cardiology. 2016;32(2):259-65.<br><br>2. Requena CC, Muriel A, Peñuelas O. Analysis of causality from observational studies and its application in clinical research in Intensive Care Medicine. Medicina Intensiva (English Edition). 2018;42(5):292-300.<br><br>3. Listl S, Jürges H, Watt RG. Causal inference from observational data. Community dentistry and oral epidemiology. 2016;44(5):409-15.<br><br>4. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41-55. |
| 3. What is the study area? Which sampling technique they use? Why they choose same number of exposure and non-exposure? | Source of data, study area, sampling procedure, and sample size are described under the ‘study design and study area’ section of methods (lines 107-118). |

| | |
|---|---|
| 4. Authors collect primary data but why they considered leadership, management and governance intervention? | First, the current study used secondary data (not primary data) from the 2016 Ethiopia Demographic and Health Survey (EDHS). The EDHS is a nationally representative cross-sectional study conducted in nine geographical regions (Tigray, Afar, Amhara, Oromia, Somali, Benishangul-Gumuz, Southern Nations Nationalities and Peoples (SNNP), Gambela, and Harari) and two administrative cities (Addis Ababa and Dire Dawa) of Ethiopia. It is designed to provide estimates of key indicators for the country as a whole, for urban and rural areas separately, and for each of the nine regions and the two administrative cities so that informed decision making can be made by the policy makers and program planners. The Demographic and Health Surveys (DHS) Program, in general, has collected and disseminated nationally representative data on fertility, maternal and child health, domestic violence, HIV/AIDS, nutrition, and other health and health-related issues through more than 400 surveys in over 90 developing countries. Therefore, we believe that the findings of our study obtained from nationally representative data will inform decision making by policy makers and health program planners. |
| Results<br><br>1. Authors write the results as too much limited format. It needs to write about more important variables which are associated with neonatal, infant and under-five child mortality. | We would like to explain this question from the objective of our study perspective. The objective of the current study is to assess the effect of short birth interval on neonatal, infant, and under-five mortality in Ethiopia. Now, the point should be whether the results meet the objective of the study or not. Table 2 presents the treatment (short birth interval in this case) effect estimation. Accordingly, short birth interval has a significant effect on neonatal (AOR=1.85, 95% CI= 1.19, 2.89), infant, (AOR=2.16, 95% CI= 1.49, 3.11), and under-five mortality (AOR=2.26, 95% CI= 1.60, 3.17) in Ethiopia. We believe that the results answer the objective of the study. The main treatment/exposure variable of our study is short birth interval and it is beyond the scope of |

| | |
|---|---|
| | the current study to illustrate the effect of other variables on the outcomes in this causal analysis. |
| Discussion<br><br>Similarly, discussion also needed for the more important variables those are associated with neonatal, infant and under-five child mortality. | This comment is similar to the previous comments regarding the objective and results of the current study, as responded to above. The discussion section has five paragraphs. The first one summarizes the objective and overall findings of the study. The second, third, and fourth paragraphs discuss the effect of short birth interval on neonatal, infant, and under-five mortality as per the findings of the study, respectively. The final paragraph acknowledges the limitations and strengths of our study. We believe the discussion is written as per the objective and results of the study. |

**VERSION 2 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Saha, Unatti<br>Tilburg University |
| **REVIEW RETURNED** | 31-May-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | Thank you for replying all of my comments in a structured way and take into account for where it was necessary.<br><br>The estimator IPTW should be perhaps IPW.<br><br>Also, the authors need to mention the name of software that they used for their statistical analysis/model estimation, e.g. STATA or R or SPSS. |

| | |
|---|---|
| **REVIEWER** | Ahammed, Benojir<br>Khulna University, Statistics Discipline |
| **REVIEW RETURNED** | 18-May-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | After updating this paper similar problems are arises. If authors focus only one explanatory variables. Short birth interval is significantly associated with neonatal, infant and under-five child mortality. A lot of study found similar findings. It has no significant merits in future plaining to reduce the problem of short birth interval. Even this study has no new dimension of research. |

**VERSION 2 – AUTHOR RESPONSE**

Reviewer #1
Comment: Thank you for replying all of my comments in a structured way and take into account for where it was necessary.
Response: Thank you.
Comments: The estimator IPTW should be perhaps IPW.
Response: Literature uses 'inverse probability of treatment weighting (IPTW)' and 'inverse probability weighting (IPW)' interchangeably. Although, both terms describe the same statistical method, the first one, IPTW, is the more explanatory term in describing the analysis method than the second one, IPW. Therefore, we prefer to use IPTW consistently throughout this paper.
Comment: Also, the authors need to mention the name of software that they used for their statistical analysis/model estimation, e.g. STATA or R or SPSS.
Response: The statistical software used to perform the analysis in this study was already mentioned under the 'Data analysis' section (line 220-222):
"Statistical analysis was performed using Stata version 14 statistical software (StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP. 2015). Figure 2 presents a schematic summary of the overall analysis procedure."

Reviewer #2
Comment: After updating this paper similar problems are arises. If authors focus only one explanatory variables. Short birth interval is significantly associated with neonatal, infant and under-five child mortality. A lot of study found similar findings. It has no significant merits in future plaining to reduce the problem of short birth interval. Even this study has no new dimension of research.
Response: Although previous studies have attempted to assess the association between short birth interval and neonatal, infant, and under-five mortality, they have several key limitations that suggest the need for further studies. We have already discussed those important limitations of the previous studies and their implications on policy and program. Please see the below statements that are also presented in the introduction section of our manuscript (line 82-99):
"Although previous studies18-20 24 25 28-32 have suggested birth interval as one factor influencing neonatal, infant, under-five mortality, these studies have several limitations. One of the key limitations is that these studies18-20 24 25 28-32 did not use the World Health Organization (WHO) recommended1 definition of short birth interval. Understanding the impact of short birth interval on neonatal, infant, and under-five mortality, using the WHO definition1 is necessary for the formulation of valid, consistent policies and health planning strategies and interventions to improve child health outcomes. Second, women who were not eligible to provide birth interval information (i.e., those who had given birth only once) were included in the analysis of some studies.20 25 29 This may result in underestimation or obscuration of the true effect of birth interval on child mortality. Third, even among studies using the same definition of short birth interval, findings have been inconsistent.20 25 One of the studies using national data20 did not control for a range of potential confounders including maternal education, wealth status, number of children, and region of residence, even though these data were available in the datasets used for analysis. Similarly, another previous study30 that used national data did not condition on maternal occupation, husband education, husband occupation, the total number of preceding children, regions, access to mass media, and women's decision making autonomy. In addition, various studies did not consider short birth interval as a potential predictor of neonatal,22 26 27 33-36 infant,19 37 38 and under-five mortality39-42 in their studies."
Therefore, since our study filled the above-mentioned key limitations and the existing information gap, its findings provide new insight regarding the association between short birth interval as per the WHO definition and neonatal, infant, and under-five mortality. This, in turn, will have significant contributions to the formulation of policy and development of intervention programs.