Author’s response to reviews

Title: Risk Factors of Morbidity among Children under Age Five in Ethiopia

Authors:

Kasahun Geneti (kastake10@gmail.com)
Temesgen Zewotir (zewotir@ukzn.ac.za)
Denis Ndanguza (dndanguza@gmail.com)

Version: 2 Date: 12 Apr 2019

Author’s response to reviews:

Response to reviewers’ comments on ‘Risk Factors of Morbidity among Children under Age Five in Ethiopia.’

We are very grateful for the reviews provided by the editors and each of the external reviewers of this manuscript. The comments are encouraging and the reviewers appear to share our judgement that this study and its results are important to support strategies and interventions to address diarrhea and fever in children aged 5 or younger. Please see below, in blue, our detailed response to comments. All changes to the manuscript are indicated in the text by highlighting yellow color. We look forward to hearing from you regarding our submission and are available to respond to any further queries that may arise.
Reviewer #2

Introduction:

Reviewer comment: Page 2, lines 53-57: results from your literature review are unclear. You explain that "most of the investigations used logistic regression", followed by a sentence explaining that "most articles... using simple statistics". If you could clarify and explain previously research on this topic.

Authors’ response: Thank very much for this comments. References added

Reviewer comment: Page 3, line 14-29: This paragraph should focus on what you are doing. The additional details included such as "an extension of generalized linear model", "among marginal model family" are not necessary. Your research focus can become a lot clearer with a succinct statement about the intent of this manuscript. Additionally "subjects for the same cluster" should be explained as it directly pertains to your research. In this case children from the same household.

Authors’ response: Thank you so much. We agree and removed the sentence ‘Consequently, in this paper we use an extension of generalized linear model, generalized estimating equation and alternating logistic regression among marginal model family, which account for the correlation between responses of interest in subjects from the same household to identify the risk factors of childhood morbidity.’
Reviewer comment: Page 3, lines 46-59: there needs to be a clearer presentation of the variables you used in your analysis. What are the survey questions that produced these variables? For example "mothers education level" Is this a dichotomous variable or continuous? What were the survey options for education level? What levels of education did you look at in your analysis. Similarly "Marital status" can include: single, co-habiting, married, divorced, widow, etc. Which 'marital status' categories did you include. This should also be done for diarrhea and fever. How were these defined. Did the respondent self-report diarrhea or fever in the 2 weeks? Did they attend a medical appointment for diarrhea or fever? If only self-reported what questions were used to ascertain if the respondent correctly labelled diarrhea or fever? Was a description of diarrhea and fever provided to the respondent?

Reviewer comment: The majority of the methods section is not study specific and should not be included. Although interesting, the Marginal Models, Generalized Estimating Equations, Alternating Logistic Regressions and model selection sections are more of a literature review than a description of what was done. This information should be summarized in a few sentences in a Strengths and Limitations section, not the focus of the methods section. Can you please review this section and provide a synthesized description of the (1) data source, (2) what model you used, (3) how you defined the variable, (4) description of your specific analysis. When discussing the methods they should be in related to study specific reasons, data, and proposed outcomes. I can review references provided for further details on the statistical models used. Which variables were controlled for in GEE, what p-value was used to denote significance, did you do any descriptive statistics to obtain prevalence or incidence rates and understand relationships between variables, ect. What about the cross-tabulation analysis you discuss in the results section?

Authors’ response: Thank very much for this important comments. We agree to the reviewer comments and modified:
Materials and Methods

Data source

Ethiopia is a landlocked country located in East Africa, which categorized as sub-Saharan Africa. In this paper, we use the data set available from the 2016 Ethiopian DHS. This survey is the fourth cross-sectional investigation administered at household level. A stratified two-stage cluster sampling procedure was used. Enumeration areas (EA) were the sampling units for the first stage. The sample included 645 enumeration areas among 202 in urban areas and 443 in rural areas. The survey included questions designed to explore sociodemographic, socioeconomic, child health, maternal and environmental conditions at household level. The data was collected from women interviewees aged 15–49 years. In the survey, the health status of each interviewee’s children aged less than 60 months in the 2016 survey was assessed by asking the interviewee ‘Has your child had diarrhea, and/or fever in the last two weeks?’ The occurrence of diarrhea and/or fever categorized as ‘Yes’ or ‘No.’ Overall, data on 8,742 ‘young’ children was collected in the survey [11]. Table 1 provides information on categorical socioeconomic and demographic covariates, their categories, frequencies and association with diarrhea and fever. Childhood diarrhea characterized as the frequent (three or more times per day) loss of liquid stools within two weeks period preceding to survey. Fever is an abnormally high body temperature, accompanied by shivering, headache, and restlessness [19].

Marginal Models

Generalized Estimating Equation (GEE) and Alternating Logistic Regression (ALR) were used to model childhood morbidity, and compare the odds of a child being had diarrhea and/or fever given the various risk factors considered. Within these GEE and ALR models framework the households were considered to be clustered and not independent within each household. The response variables were the occurrence of diarrhea and/or fever (Yi: binominal variable, i.e., yes or no), and for this study we considered that the event occurred if the child had diarrhea and/or fever in the last two weeks of the survey. Consequently, the response variable was coded with 1 for child had diarrhea and/or fever and with 0 for child had no diarrhea and/or fever. The link function between the mean value Yi and the model covariates considered for GEE and working correlation structure is defined by:
\[ g(\omega_i) = \text{logit}(\omega_j) = \left[ x_i \right]^\top \beta \]  \hspace{1cm} (1)

Where \( g(\omega_i) \) is logit link function, \( x_i \) is \( n_i \times x_i \) dimensional vector of known covariates, \( \beta = (1 \times p) \) dimensional vector of unknown parameters, \( E(Y_i) = \omega_i \) is expected value of the response \( Y_i \) in the cluster \( i \) which is binomially distributed as \( y_i \sim \text{bin}(n_i, \omega_i) \). In addition, GEE is non-likelihood method that captures the association within households in terms of marginal correlations [18]. With this GEE model, the correlation structure of the data within each household was assumed to be of the independence, unstructured, exchangeable, and first order auto-regressive [20]. The parameter \( \beta \) are estimated by quasi-likelihood.

However, when the cluster sizes become larger, the simultaneous estimation of marginal mean and dependence structure can become computationally prohibitive using GEE. As a result, alternating logistic regression which measures pairwise association of two observations in the same household and follow the precision estimates for both the regression (\( \beta \)) and the association (\( \alpha \)) parameters considered [18, 21]. Furthermore, unlike GEE, no working assumptions about the third order and fourth-order odds ratios are required. ALR measures the association using the odds ratio, which is interpretable and more applicable for binary data [18, 21]. Let \( \gamma_{ijk} \) be the log odds ratio between outcomes \( y_{ij} \) and \( y_{ik} \), let \( \mu_{ij} = p(y_{(ij)=1}) \) and \( v_{ijk} = p(y_{ij}=1,y_{ik}=1) \), then the association of the two responses is defined as [27]:

\[
\logit p(Y_{ij}=1/Y_{ik}=y_{ik}) = \gamma_{ijk} Y_{ik} \log((\mu_{ij} - v_{ijk})/(1 - \mu_{ij} - \mu_{ik} + \mu_{ijk}))
\]  \hspace{1cm} (2)

Similar to GEE, the parameter \( \beta \) are estimated by quasi-likelihood. Descriptive statistics and chi-squared tests were carried out to identify associations between risk factors and outcome variables (diarrhea and fever) using version 21 of SPSS. P-values of less than 0.05 were considered statistically significant. GEE and ALR model fit for this study was carried out with the SAS 9.4 version.

Model Selection: The purpose of building a statistical model is to find an optimal model characterized by principles of generalizability, goodness-of-fit and parsimony based on model selection criteria. In this study, three working correlation assumptions exchangeable, AR (1) and independence were used to select the best model with minimum QIC and to select a working correlation structure [23]. In order to select important covariates related to diarrhea and fever backward selection method was used. With this method, we started with a full model containing all main effects and interactions to a more parsimonious model.
Results:

Reviewer comment: Somali does not appear higher than a couple other regions such as Afar? Can you provide the proportion of cases by population in the region? Are there more cases because of the higher population. Some additional context can help the reader understand these results. For example, why are you comparing to Addis Ababa? This does not appear to the lowest incidence of fever? And given there are lower responses to "No" fever, is the population in Addis Ababa smaller? This might suggest a higher proportional incidence rate.

Authors’ response: Thank so much. We have corrected Somali region related questions. Why are you comparing to Addis Ababa? Because Addis Ababa is the capital of the country and relatively developed region.

Reviewer comment: page 8, Line 16-17: male children do not have a significantly higher rate of fever in Table 1, as suggested in the text. P-value is 0.65?

Authors’ response: We thank the reviewer for this comment. We agree to the reviewer comment and corrected. Moreover, from Table 1, the male children are more likely to have diarrhea compared to female counterpart (12.2%).

Reviewer comment: page 8, line 27-28: "The prevalence of diarrhea and fever are among non breastfeeding children than breastfeeding children" - This relationship needs to be clarified. This sentence suggests a different finding from your table 1. Table 1 shows that non-breastfed children have a lower rate of diarrhea. But breastfed children have a similar rate of fever compared to non-breastfed children (14.4 vs. 14.3%)

Authors’ response: Table 1 shows that non-breastfed children have a lower rate of diarrhea. However, breastfed children have a similar rate of fever compared to non-breastfed children (14.4 vs. 14.3%)

Reviewer comment: Table 1: The title needs to explain the analysis in the title, and what the p value is referring to (i.e. significance between yes vs. no or between variable categories)? Additionally, there needs to be consistency in the number of decimal places used both in the table and the text.

Authors’ response: We thank the reviewer for this question. We have changed the title to “Distribution of childhood morbidity and its associated selected risk factors”. The inconsistency in the number of decimal places used both in the table and the text is corrected. The p-value is referring to significance between the children morbidity (diarrhea and fever) and risk factors.
Reviewer comment: Results from Marginal Models: section is a literature review, not results from your study. It should not be presented as your results. The explanation of methods used should be presented in the methods, strengths and limitation section.

Authors’ response: We have removed the explanation of methods from the section. The important predictors related to morbidity were selected based on backward selection method. As stated in [29], with backward selection predictors which its removal has least damaging effect on the model are successively removed based on score statistics (p-value > 0.05). In other words, the predictors with a least contribution to the model based on the highest p-value were eliminated successively and each time a reduced model with the remaining predictors were fitted. Then, predictors fit well the data for both diarrhea and fever outcomes were remained. On the other hand, none of the interaction effects were found to be significant for both diarrhea and fever outcomes.

Reviewer comment: page 10 line 12-13: Children 12-23 months have the highest odds ratio for diarrhea and fever, but your text is focused on children 0 - 11 months.

Authors’ response: We have included “Furthermore, child diarrhea increased by 132% for children age group between 12-23 months as compared to children with had no diarrhea.”

Reviewer comment: There is interpretation within your results section. Interpretation of the results should be presented in the discussion section.

Authors’ response: we have corrected.

Reviewer comment: page 10, line 25-26: There is a need to be careful with the language used in the paper. For example, "with respect to mothers work effect, it was observed that mothers work is significantly related to child diarrheal disease" This sentence is not necessary. Instead of "mothers work" I think you intent to focus on 'mothers work outside of the home' OR 'mother paid employment'. Mothers who are at home all day are still working, running a household is still work. If you are presenting this manuscript to an international audience, the language you use is important.

Authors’ response: We thank the reviewer for this important comment and we agree that, as it needs care. We have adjusted for the variable as “mother paid employment.”
Reviewer comment: Table 2: define all abbreviations.
Authors’ response: we have incorporated abbreviation section to the manuscript.

Reviewer comment: page 11 line 49-58: this is not appropriate for the results section of your manuscript.
Authors’ response: We share some part of the comments and removed: The QIC for GEE and ALR are almost the same. However, empirically corrected standard errors under ALR are somewhat less than those of GEE and the small differences in parameter estimates attributed to the fact that ALR considers the association unlike the GEE treat the association as a nuisance parameter. Removed from the results section.

Reviewer comment: page 12 line 18-19: "this indicated that child from separated and never in union mother is more exposed to fever disease". This is not correct. You are assessing association not exposure. You are using self-reported survey data which does not control for all factors/variables and can present a self-report bias.
Authors’ response: We thank the reviewer for this important comment. This indicates that child from separated and never in union mother are associated with fever disease.

Reviewer comment: page 12 line 19-23: this is an interpretation. The word "observed" is used inappropriately throughout the results section.
Authors’ response: We agree with the reviewer and we have replaced by appropriate words throughout the result section.

Reviewer comment: page 12 line 23-24: this needs to be rephrased. Was there a question about literacy on the survey? How did the results move from education level to literacy? Higher education level of the husband was associated with a 0.747 (exp -0.2904) less odds of fever in a child compared to husbands with a lower education. (OR 25% lower odds).
Authors’ response: We thank the reviewer for this question. We have changed to “Higher education level of the husband was associated with a 0.747 less odds of fever in a child compared to husbands with a lower education.”
Reviewer comment: If abbreviating Demographic and Health Survey, DHS should be used consistently.

Authors’ response: we have changed accordingly throughout the manuscript implies “Demographic and Health Survey” used.

Discussion:

Reviewer comment: page 13, line 60/page 14 lines 4-6: covariates should be explained. i.e. male children, etc. were significantly associated with higher odds of diarrhea

Authors’ response: We thank the reviewer for this important comment. male children, 0-11 months aged children, 12-23 months aged children, anemic children, husband with a lower education, mothers paid employment, Afar, Amhara, Dire Dawa, Gambela, Oromia, SNNPR and Tigray regions were significantly associated with higher odds of diarrhea. For fever as an outcome variable, 0-11 months aged children, 12-23 months aged children, children of never in union mother, children of separated mother, non-breastfed children, husband with a lower education, mothers paid employment, Benishangul, Somali and Tigray were significantly associated with higher odds of fever.

Reviewer comment: page 14 line 21: gender discrimination? can you elaborate?

Authors’ response: In Ethiopia there is a gender preference culturally.

Reviewer comment: page 14, line 38: "This may be due to the mothers are the most frequent primary caregivers for their children". This needs to be explained. How does a mother being the most frequent primary caregiver for their children" affect associations with fever and diarrhea? This is following a sentence about mothers with paid employment outside of the home. Are you suggesting that mothers are bringing home illnesses, or that if children are cared for by someone else they are more likely to have a fever or diarrhea? You should be careful how you word and explain this statement as it could suggest that the mother should not work. You have not considered illness duration, severity or sequela, which may be different among different population subgroups. You have also used self-reported data, with perception bias.

Authors’ response: We have removed the sentence. This may be due to the mothers are the most frequent primary caregivers for their children.
Reviewer comment: page 14, line 27: education is one of the socioeconomic variables. By this statement do you intend to suggest that education is more important then income and employment? I do not this it is necessary to include this statement "husband education level is an important factors..."

Authors’ response: we have changed to “Furthermore, the study found that the husband education level is an important factor that affects the health status of children in Ethiopia.”

Reviewer comment: page 14, lines 40 - 46: are these based on multivariate or cross-tab analysis. The two analysis that were reported suggested differences.

Authors’ response: It is based on the separate ALR model results.

Reviewer comment: page 14, lines 50-51: "Some communities living in Somali and Afar regions are nomadic so that they lack access to based health-case facilities and sanitation services" This statement is not supported by your research. These two regions did not experience higher odds of both fever and diarrhea, if this was related to access to health care and sanitation I would expect to see higher incidence and OR for these regions. But this was not found.

Authors’ response: we have changed to “Likewise, children from Tigray region are at higher risk of fever morbidity. The finding is in line with that of [6, 26]. This may be due to the difference in environmental and socio-demographic characteristics of households.” However, we have removed the sentence, “Some of communities living in Somali and Afar regions are nomadic so that they lacking access to basic health-care facilities and sanitation services.”

Conclusions:

Reviewer comment: page 14, line 56: "children and mothers... determinants". How do you define children and mothers determinants; you have listed socioeconomic determinants separate. Is socioeconomic status not a determinant of the mother?

Authors’ response: we have changed to “In this paper, we studied socioeconomic, demographic, household and spatial determinants of children morbidity measured through prevalence of diarrheal and fever using Ethiopian Demographic and Health Survey data.”
Reviewer comment: page 15, line 10-12: define the covariates. As in the discussion section.

Authors’ response: The covariates defined as in the discussion section and the details indicated in the manuscript.

Reviewer #1

Response: We thank the reviewer for those comments. We have managed all the comments via the response to reviewer #2.