

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	How sustainable household environment and knowledge of healthy practices relate to childhood morbidity in South Asia: Analysis of survey data from Bangladesh, Nepal and Pakistan
<b>AUTHORS</b>	Hasan, Masud; Richardson, Alice

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Dr Jennifer Saxton Not presently linked to any institution, UK.
<b>REVIEW RETURNED</b>	21-Nov-2016

<b>GENERAL COMMENTS</b>	<p><b>Notes on DHS analysis of predictors of diarrhoea and ARI in three South Asian countries.</b></p> <p><i>Terminology:</i></p> <ol style="list-style-type: none"><li>1. Where you have written 'impact' it should say 'association' given that you are using cross-sectional data; the word impact implies causality.</li><li>2. Where you refer to 'incidence', I think you really mean 'prevalence'; you also mention prevalence in one of your results tables.</li><li>3. Nowhere in the paper or tables have you defined your outcome or predictor measures: ARI for instance is defined differently by different people, as is diarrhoea. You should include definitions of all variables that you are using.</li><li>4. You have not provided references to support the construction of your predictor or outcome variables. These will need to be added to increase the credibility of your analysis and findings.</li></ol> <p><i>Quality of writing and layout of paper:</i></p> <ol style="list-style-type: none"><li>5. This is an interesting paper, and although parts of it are well written, there are a lot of typos (I've pointed out some, but not all), inappropriate changes of tense, inappropriate abbreviations (e.g. in the tables), and missing words. The manuscript needs a thorough proof-read and edit before it is resubmitted.</li><li>6. P.8, lines 18-20 – slightly awkward phrasing. You could write something like: 'Place of residence was significantly associated with child ARI in Bangladesh and Pakistan' (give p-values).</li><li>7. I suggest you split your results into appropriate subheadings so it's easier to read. At least it could be split by outcome,</li></ol>
-------------------------	--

but could also be split by predictor.

*Data cleaning and preparation:*

8. There was a lack of information about data cleaning, and any exclusions that were made to the dataset. I would like to know which cases were excluded, and why, as this could have biased your findings.
  - For example, what happened to the households that didn't have access to a toilet and instead use open defecation?
  - What happened to households who did not do their cooking inside the house?

*General reporting in the manuscript:*

9. Page 6, line 44: shouldn't 100% of the sample be accounted for here, rather than 80%?
10. Page 6, lines 52+53: was open defecation included somewhere? I would be surprised if everyone in the sample had access to a toilet, given the high proportion of rural respondents.
11. Page 6, lines 46-60: you should add references to support your use of these categorisations, as well as adding missing definitions (e.g. what constitutes 'hygienic' water).
12. Page 6, line 59 – were there any cases of people cooking outside of the home? And if so, what happened to these cases in your analyses?
13. Page 7, lines 17-24 – add references
14. Page 7, line 43 – by 'co-morbidity' I think you mean ARI and Diarrhoea. This should be clarified, as co-morbidity is a term that can be used to refer to generic other illnesses.
15. Page 8, line 3 households compared to what?
16. Page 8, line 8: an extra 'are' that doesn't belong there.
17. P8, line 11: is it actually number of ARI episodes, or just ARI in the last 2 weeks? Again, your variables need to be defined.
18. P8, lines 41-42: presentation of results is a bit confusing and contains errors.
  - '0.61...more likely' could be changed to 61% more likely.
  - Nepal figures show an OR of both 2.11 and 3.11. Judging by your CIs it should be 3.11
  - Bangladesh figures show ORs of both 1.67 and 2.67 – I think it should be 2.67?
19. P8, line 49: check your estimate and CI as the lower bound of the CI and the estimate are the same?
20. P8, line 57: what are you classing as statistically significant? State the p-value.
21. P9, line 5: ORs and AORs?
22. P9, line 55: cross-sectional
23. P9, line 10: I'm not convinced you are reporting an incidence rate here. See my earlier points about incidence vs prevalence, as well as the need for clear definitions of how your outcome variables are measured.
24. P9, line 45: It might just be my preference, but it's usual to

have the data limitations after the discussion of the results. Data limitations should also include mention that you have excluded important confounders from your analysis (measured and unmeasured), which could account for your findings. I would in fact recommend including all potential confounders in a re-run of your analysis. You should also mention any possible bias in your analyses (e.g. if you excluded any cases at the beginning, before you ran the analyses which could affect the representativeness of the sample).

25. P10, 31-33: as mentioned above, I don't understand your explanation about not including confounding variables in your model. These should be included so you can get a 'truer' understanding of the association between your predictor and outcome, over and above the influence of the confounder.
26. Page 10, line 47: immune **system**

*Important omissions in your statistical analysis:*

27. Maternal age. Given that this is an epidemiological paper, and that maternal age is such an important driver of child health I suggest re-running your analyses with maternal age included. It is likely to be an important effect modifier.
28. I'm confused about your interpretation of confounders, and your statement that they should not be included in your models. It is my understanding that you **should** include all potential confounders that are available (sample size permitting), to observe your main effects after adjusting for confounders. Your models could be overly simplistic as a result of this, and your findings may actually be accounted for a variable that you have not included in your model.
29. It is not clear to me whether you conducted a series of univariate regressions (i.e. a single predictor and outcome) in Tables 3 and 4, or a single multivariable regression model (i.e. all predictors in the model together regressed against a single outcome) presented in Table 3, and another single multivariable model presented in Table 4. This needs to be clarified in the text, and in the tables. It is my view that **if** you have conducted only a series of univariate regression models then this paper doesn't go far enough in its analysis, and you should run the multivariable models as well; there would also be little point to the chi squared analyses in Table 2 and it would be preferable to have univariate and then multivariate regression models, and to ignore the chi-square analysis as it doesn't add value to the paper.

*Your discussion and conclusion:*

30. I would like to see a statement about what this paper adds to our current understanding of the issues you explore in your paper
31. The discussion is rather thin at the country level, and in your country comparisons. It needs greater depth, including references to other studies.
32. I would also have liked to have seen greater reference to the

SDGs, as was indicated in the introduction section.

33. Some of your conclusion section actually belongs in the discussion. My understanding is that a conclusion should include a small number of succinct statements about your findings and their implications.
34. Your conclusion is worded to imply that only maternal education is important to reduce the burden of ARI and diarrhoea. I think this sends the wrong message. To reduce the burden of these infections there needs to be simultaneous improvements in all these areas.
35. Page 11, line 30: Secondary education for women is not 'highly ambitious', it's a human right. Your conclusion seems to suggest that secondary education will not be possible for women, and to focus on primary instead. I think this needs rewording! And in line 36 – yes, there may be high enrolment, but that says nothing about attendance or quality of education, and I don't think you should be using those statistics in a face value manner.
36. The fact that you have women in your sample who are mothers at 12 years of age says something about the need to protect girls/women from early marriage. It's a shame you haven't included maternal age in your analyses, as I feel it could also be an important risk factor for child ARI and diarrhoea – it's certainly a risk factor for undernutrition. I would advise including maternal age in the re-run of your analysis.

*Your figures and tables:*

37. Page 16, Figure 1:

- This figure should show the actual % and 95% CIs.
- Clarify that 'co-morbidity' means both ARI and diarrhoea
- State 'last 2 weeks' on the x-axis; same applies to the tables.

38. Table 1:

- Line 14: Age of child labels – the oldest group should have an upper age limit, rather than 36 months+; this also applies to the other tables.
- Lines 8 and 9: you don't need to show ARI and Diarrhoea prevalence as you've already included it in Figure 1.
- Line 22 – what is 'elec' media?
- Line 27 – why has 'unhygienic' been combined with 'not on premises'? It is possible to get hygienic water from public hand pumps. Again, this links to the need for definitions and references for your variables and how you've constructed them.
- Line 31 – why have unhygienic unshared, and hygienic shared been grouped together?
- Line 34: aren't children also included in the definition of household overcrowding? You only mention adults here.

39. Table 2:

- Title would be better at the top, in accordance with your other tables.

	<ul style="list-style-type: none"> <li>• The columns are not labelled. I presume these are percentages that show prevalence by sub-category?</li> <li>• The data in the table are starred (presumably to indicate statistical significance), but there is no key to the stars to help understand the significance level.</li> <li>• Is this table showing chi-squared results? If so, this should be mentioned in the title. Also, the <math>X^2</math> and df should be shown.</li> <li>• Lines 4 and 5 – should mention the last 2 weeks</li> <li>• Line 23: 'rich' implies a value judgement. Perhaps better to say lowest, middle, and highest.</li> <li>• Some of the abbreviations in the table look a bit odd (e.g. 'Hy'). It would be better to spell them out.</li> </ul> <p>40. Table 3 and 4:</p> <ul style="list-style-type: none"> <li>• The titles have confused me. Are these results of adjusted regression models? If so, they should say AORs, not ORs. This needs clarifying in the tables and text.</li> <li>• See my earlier point – multivariable analyses would add credibility to your findings (if these tables don't already show AORs from multivariable analyses).</li> <li>• There is a problem with the title placement of table 4 (it's split over the top and bottom of the table).</li> </ul>
--	--

<b>REVIEWER</b>	GRACE K.M.MUHOOZI UNIVERSITY OF OSLO, NORWAY; KYAMBOGO UNIVERSITY, UGANDA
<b>REVIEW RETURNED</b>	03-Dec-2016

<b>GENERAL COMMENTS</b>	<p>Q4 Sample size calculation not shown and the sampling strategy not elaborated.</p> <p>Q9 Results answer prevalence of disease (from cross sectional surveys) not Impact which the authors state in their title</p> <p>Q10 Clarifications to be made in the text and in the tables. Details are in the comments attached</p> <p>Q13 I did not see any of those materials</p> <p>Q15 English need some editing by someone whose English is first language.</p> <p><b>Impact of sustainable household environment and knowledge of healthy practices on childhood morbidity in South Asia: Analysis of survey data from Bangladesh, Nepal and Pakistan</b></p> <p><b>The title:</b></p> <ol style="list-style-type: none"> <li>1. The word "impact" seems to be very strong for this study which is looking at data that was collected from cross-sectional DHSs.</li> <li>2. I hope the "sustainable household environment" is defined along</li> </ol>
-------------------------	--

the way in introduction.

**Abstract**

**Objective:** Could be “background and objective”

The authors now give “The aim” instead of the objective. However, the aim given and title do not rhyme very well.

**Participants:** This information could be put better to indicate that the women were the mothers of the children, i.e Information was obtained from mothers (12-49 years) of 23,940 children under age of five years.....Is it important to mention the ever-married?

**Conclusion:** The last sentence in the conclusion is very so long that by the end of it, one has lost meaning and intension. Consider making 2-3 short meaningful sentences.

**Introduction:**

The definition/ explanation of “sustainable household sanitation” is required maybe in the second paragraph. What aspects are we looking at? Maybe the authors could have taken time to define what it is.

Line 27, page 4, the sentence begins with “Educated mothers.....their “child”..... Since it is plural then it should follow that word becomes “children”

The two sentences that follow in line 32 and 33 begin with “In addition...” the words become monotonous to the readers. Authors could try to find another way of starting the next sentence.

Line 6 on page 5 “Objective” should be replaced with “Aim”.

Line 9, the “first aim” should then be the “first objective”...

The use of the word “incidence” in line 7 on page 5 and everywhere in the manuscript should be checked. It could be replaced by “prevalence”.

Line 14, this having been data from a cross-sectional surveys, it is not possible to measure the impact rather an association/relationship or likelihood. The word “impact” has also been used in many other places.

Line 16 “sustainable household environment” comes up, this time as a variable/predictor? Without being defined earlier in introduction, it is not easy to connect with the other variables.

Line 38 Sentence begins with “On the other hand.....” there is need to edit that sentence. It is unfair to those countries, to depict them as “representing”.....

Line 53, 54.....here is need to show how sample size was obtained from each country at the various stages to arrive at the

figures seen in the results.

Finally, “household sustainable environment” is described in 46-49 on page 6; This is good but should have been done earlier in the introduction for readers to know what the authors are talking about.

### Analysis

- Authors should give the rationale for excluding some variables in the various models.

### Results

Line 38, the use of the word “incidence” is not correct. Rather use prevalence.

Figure 1

Check the bars of Bangladesh in Diarrhoea and ARI, they don't seem to say the same as the prevalence of the two conditions in the test and in table 1.

Table one, presents results in percentages. It would be good to include numbers beside the percentages and “n (%)” in the footnotes.

Title of table 1 refers to predictors. Why then include the prevalence of diarrhea and ARI in the first line? Prevalence is already shown in figure 1 or these could be placed below the figure as foot notes.

Table 2

Be consistent: you had titles for table1, 3 and 4 at the top and title for tale 2 at the bottom of the table.

What are the figures in table 2?

I will assume they are percentages from bivariate chi-square test and proceed as follows:

- Indicate numbers against percentages
- Also indicate this in the foot notes below the table.
- Some editorial correction in line 8, page 8 “are were”

Page 8 line 8, “In the bivariate analyses, children suffering from ARI “are were” more likely to suffer from diarrhea and vice-versa (Table 2).” Does the statement above refer to the figures below extracted from table 2?

<i>Childhood ARI/ Diarrhoea</i>						
	<i>Diarr</i>	<i>ARI</i>	<i>Dairr</i>	<i>ARI</i>	<i>DiarrARI</i>	
<i>Suffered</i>	30.2***	23.7***	9.8***	9.3***	30.1***	10.1***
<i>Not suffered</i>			5.4	5.1	13.1	3.8

If yes, the authors need to think a little more about this. What I get from the figures is the prevalence of diarrhea and ARI at the time of the survey and not the likelihood of a child suffering from ARI also to suffer from diarrhea or vice –versa as they seem to indicate;(i.e the proportion of children who suffered a particular disease was significantly higher than those who did not suffer in each country at the time of survey).

The paragraph on page 8, the word “impact” has been used repeatedly, but then Chi-square test measures association rather than impact. Therefore replace this word in the text appropriately.

Again the word “incidence” appears in this paragraph. It should also appropriately be replaced.

In the text describing results presented in Table 3 and 4, there are values presented as proportions (could they be the ***B=Logistic Co-Efficients?***), alongside the OR and CI.

- I suggest space be created in the tables so that all variable have these ***B=Logistic Co-Efficient*** values presented for each.
- Also there should be consistency in the way of presentation of these values (e.g either as 0.70 or 70%), Check line 41 and line 49 page 8; line 15 and 25 on page 9.

Table 3 and 4, foot notes would be good to use in order to reduce the crowding in the tables and to make it easy for readers.

### **Discussion**

I think the authors should acknowledge using that data from cross sectional survey provides information regarding the disease episodes for a single point of time (line 57 page9), such study cannot yield “impact” of a given variable on disease episodes.

Page 10, line 1 and may be others to follow: this study should not be talking of incidence of disease/ incidence rates.

Line 21 and 28, the word “effect” is still inappropriate in this study because the design of this study cannot bring out the “cause- effect” relationship.

It would have been interesting to show the levels of exclusive breastfeeding in the background characteristics for the different countries.

### **Conclusion**

Line 26 -28, While it could be true, this study has not confirmed it, therefore state with caution that “maternal education could be effective.....”



## VERSION 1 – AUTHOR RESPONSE

### **Response to Reviewers**

**Original Paper Title:** Impact of sustainable household environment and knowledge of healthy practices on childhood morbidity in South Asia: Analysis of survey data from Bangladesh, Nepal and Pakistan.

**Manuscript ID:** bmjopen-2016-015019

### **Response to the comments of first reviewer:**

#### **Responses to the comments on Terminology:**

*Page and line numbers in this response refer to the clean version of the paper*

**Comment 1:** Where you have written 'impact' it should say 'association' given that you are using cross-sectional data; the word impact implies causality.

**Response 1:** Changes are now made to the manuscript considering the nature of the study (page 2, line 3; page 5, line 7; page 7, line 3; page 8, lines 31, 32, 34; page 9, lines 3, 4, 5 and elsewhere).

**Comment 2:** Where you refer to 'incidence', I think you really mean 'prevalence'; you also mention prevalence in one of your results tables.

**Response 2:** 'incidence' is now replaced with 'prevalence' (Page 2, lines 2, 20, 22; Page 3, line 6; Page 4, line 7, Page 5, lines 4, 6, 8 and elsewhere).

**Comment 3:** Nowhere in the paper or tables have you defined your outcome or predictor measures: ARI for instance is defined differently by different people, as is diarrhoea. You should include definitions of all variables that you are using.

**Response 3:** A clear definition of the dependent variables is now provided in the first paragraph of the *Dependent and independent variables* section. The independent variables are defined in the subsequent paragraphs.

**Comment 4:** You have not provided references to support the construction of your predictor or outcome variables. These will need to be added to increase the credibility of your analysis and findings.

**Response 4:** The construction of the predictor variables is now explained with references in the *Dependent and independent variables* section (see page 6, line 23 - page 7, line 26).

#### **Quality of writing and layout of paper:**

**Comment 5:** This is an interesting paper, and although parts of it are well written, there are a lot of typos (I've pointed out some, but not all), inappropriate changes of tense, inappropriate abbreviations (e.g. in the tables), and missing words. The manuscript needs a thorough proof-read and edit before it is resubmitted.

**Response 5:** Typos and grammar have been addressed in the resubmitted manuscript with many improvements made throughout.

**Comment 6:** P.8, lines 18-20 – slightly awkward phrasing. You could write something like: 'Place of residence was significantly associated with child ARI in Bangladesh and Pakistan' (give p-values).

**Response 6:** The sentence is revised. (page 8, line 34 - page 9, line 2).

**Comment 7:** I suggest you split your results into appropriate subheadings so it's easier to read. At least it could be split by outcome, but could also be split by predictor.

**Response 7:** The results section has been split by outcome (diarrhea then ARI) guided by Tables 3 and 4.

**Data cleaning and preparation:**

**Comment 8:** There was a lack of information about data cleaning, and any exclusions that were made to the dataset. I would like to know which cases were excluded, and why, as this could have biased your findings. For example, what happened to the households that didn't have access to a toilet and instead use open defecation? What happened to households who did not do their cooking inside the house?

**Response 8:** The data cleaning and exclusions are now explained in the data section. For example, the present study excluded children from women who were not de-jure resident at the time of interview (see page 6, lines 1–2). The households that didn't have access to a toilet and households who did not do their cooking inside the house are also included in the analysis. The procedure is explained in last paragraph of the *Dependent and independent variables* section (Page 7, lines 16–26).

**General reporting in the manuscript:**

**Comment 9:** Page 6, line 44: shouldn't 100% of the sample be accounted for here, rather than 80%?

**Response 9:** The typo is fixed (page 7, line 10)

**Comment 10:** Page 6, lines 52+53: was open defecation included somewhere? I would be surprised if everyone in the sample had access to a toilet, given the high proportion of rural respondents.

**Response 10:** Open defecation was also considered in one of the categories, and it is now mentioned now in the definition of the variable (page 7, lines 18–19).

**Comment 11:** Page 6, lines 46-60: you should add references to support your use of these categorisations, as well as adding missing definitions (e.g. what constitutes 'hygienic' water).

**Response 11:** The definition of categorisation for safe drinking water is mentioned with reference (page 7, lines 12–15).

**Comment 12:** Page 6, line 59 – were there any cases of people cooking outside of the home? And if so, what happened to these cases in your analyses?

**Response 12:** People cooking outside of the home also included in one of the categories and now a better definition is provided in the manuscript (page 7, lines 22–24).

**Comment 13:** Page 7, lines 17-24 – add references

**Response 13:** Reference is added as suggested (page 8, line 4).

**Comment 14:** Page 7, line 43 – by 'co-morbidity' I think you mean ARI and Diarrhoea. This should be clarified; as co-morbidity is a term that can be used to refer to generic other illnesses.

**Response 14:** The co-morbidity in the context is now defined (page 8, lines 13–14).

**Comment 15:** Page 8, line 3 households compared to what?

**Response 15:** The sentence is revised (page 8, lines 25–26)

**Comment 16:** Page 8, line 8: an extra 'are' that doesn't belong there.

**Response 16:** The typo is fixed (page 8, lines 27–28).

**Comment 17:** P8, line 11: is it actually number of ARI episodes, or just ARI in the last 2 weeks? Again, your variables need to be defined.

**Response 17:** The sentence is revised as advised, the variables are now defined (page 8, lines 27–28, and see page 6, lines 15–22 for the definition of the variable).

**Comment 18:** P8, lines 41-42: presentation of results is a bit confusing and contains errors. '0.61...more likely' could be changed to 61% more likely. Nepal figures show an OR of both 2.11 and 3.11. Judging by your CIs it should be 3.11, Bangladesh figures show ORs of both 1.67 and 2.67 – I think it should be 2.67?

**Response 18:** All the results from logistic regression are now presented as percentages (page 9, lines 9–26). For Nepal, the odds ratio for reference category (aged above three years) is 1.00, and for younger children the odds ratio is 3.19. The difference is 2.19, and hence, the younger children were 219% more likely to suffer from ARI than those belong to reference group.

**Comment 19:** P8, line 49: check your estimate and CI as the lower bound of the CI and the estimate are the same?

**Response 19:** The typo is fixed (page 9, lines 18).

**Comment 20:** P8, line 57: what are you classing as statistically significant? State the p-value.

**Response 20:** The p-value is mentioned (Page 9, lines 22–24).

**Comment 21:** P9, line 5: ORs and AORs?

**Response 21:** The Adjusted odds ratios are used and now mentioned (Page 9, lines 9, 27).

**Comment 22:** P9, line 55: cross-sectional

**Response 22:** Typo is fixed (Page 11, line 23)

**Comment 23:** P9, line 10: I'm not convinced you are reporting an incidence rate here. See my earlier points about incidence vs prevalence, as well as the need for clear definitions of how your outcome variables are measured.

**Response 23:** As a response of the earlier comments, the term "Incidence" has been replaced with "prevalence". The definitions of how the outcome variables are measured is provided in first paragraph of *Dependent and independent variables* section (Page 6, lines 15–22).

**Comment 24:** P9, line 45: It might just be my preference, but it's usual to have the data limitations after the discussion of the results. Data limitations should also include mention that you have excluded important confounders from your analysis (measured and unmeasured), which could account for your findings. I would in fact recommend including all potential confounders in a re-run of your analysis. You should also mention any possible bias in your analyses (e.g. if you excluded any cases at the beginning, before you ran the analyses which could affect the representativeness of the sample).

**Response 24:** Data limitation is now moved after the discussion section (page 11, lines 23 –32). All relevant confounders e.g. crowding and poor housing are included in the models. A clearer explanation of why diarrhea is not in the ARI model and vice versa is given, with the word "confounding" removed as it was a poor choice (see response 25 below). Possible biases are also included in the Limitations section (page 11, lines 27 – 31).

**Comment 25:** P10, 31-33: as mentioned above, I don't understand your explanation about not including confounding variables in your model. These should be included so you can get a 'truer' understanding of the association between your predictor and outcome, over and above the influence of the confounder.

**Response 25:** The term "confounding variables" was a poor choice and has been changed. All predictors (risk factors) of interest have been included in the models for ARI and diarrhea separately, with an explanation about the overlapping risk factors for the diseases leading to their exclusion from the relevant models. The rationale for the modelling decisions regarding ARI in the diarrhea mode and vice versa is now on page 10 line 32 – page 11 line 6.

**Comment 26:** Page 10, line 47: immune system

**Response 26:** The suggestion is incorporated (Page 11, lines 16–17).

**Important omissions in your statistical analysis:**

**Comment 27:** Maternal age. Given that this is an epidemiological paper, and that maternal age is such an important driver of child health I suggest re-running your analyses with maternal age included. It is likely to be an important effect modifier.

**Response 27:** New model is fitted considering maternal age as a predictor.

**Comment 28:** I'm confused about your interpretation of confounders, and your statement that they should not be included in your models. It is my understanding that you should include all potential confounders that are available (sample size permitting), to observe your main effects after adjusting for confounders. Your models could be overly simplistic as a result of this, and your findings may actually be accounted for a variable that you have not included in your model.

**Response 28:** See response 24 and 25, "confounder" was a poor choice of word in the epidemiological context and has been removed. Our models include all relevant predictors (risk factors) for each disease separately.

**Comment 29:** It is not clear to me whether you conducted a series of univariate regressions (i.e. a single predictor and outcome) in Tables 3 and 4, or a single multivariable regression model (i.e. all predictors in the model together regressed against a single outcome) presented in Table 3, and another single multivariable model presented in Table 4. This needs to be clarified in the text, and in the tables. It is my view that if you have conducted only a series of univariate regression models then this paper doesn't go far enough in its analysis, and you should run the multivariable models as well; there would also be little point to the chi squared analyses in Table 2 and it would be preferable to have univariate and then multivariate regression mode

**Response 29:** Tables 3 and 4 represent results from single multivariable regression model (i.e. all predictors in the model together regressed against a single outcome).

#### ***Your discussion and conclusion:***

**Comment 30:** I would like to see a statement about what this paper adds to our current understanding of the issues you explore in your paper

**Response 30:** The beginning of the *Conclusion* has been reworded to capture the important variables that have not, and then have, been found to be significantly associated with diarrhea and ARI in the countries studied.

**Comment 31:** The discussion is rather thin at the country level, and in your country comparisons. It needs greater depth, including references to other studies.

**Response 31:** Country comparisons have been strengthened and extra references included in the *Discussion* and *Conclusion*.

**Comment 32:** I would also have liked to have seen greater reference to the SDGs, as was indicated in the introduction section.

**Response 32:** The SDG are now referred to in the *Conclusion*.

**Comment 33:** Some of your conclusion section actually belongs in the discussion. My understanding is that a conclusion should include a small number of succinct statements about your findings and their implications.

**Response 33:** The BMJ Open website states that the Conclusion section of the Abstract is to contain "primary conclusions and their implications, suggest areas for further research if appropriate". We include that material in the *Conclusion*, and seek the Editor's further advice as to whether a Conclusion section is consistent with journal policy.

**Comment 34:** Your conclusion is worded to imply that only maternal education is important to reduce the burden of ARI and diarrhoea. I think this sends the wrong message. To reduce the burden of these infections there needs to be simultaneous improvements in all these areas.

**Response 34:** The phrase "along with" is used in the *Conclusion* to emphasise the importance of simultaneous improvements.

**Comment 35:** Page 11, line 30: Secondary education for women is not 'highly ambitious', it's a human right. Your conclusion seems to suggest that secondary education will not be possible for

women, and to focus on primary instead. I think this needs rewording! And in line 36 – yes, there may be high enrolment, but that says nothing about attendance or quality of education, and I don't think you should be using those statistics in a face value manner.

**Response 35:** According to the suggestions of the reviewers, the manuscript is now revised (page 12, lines 10 –22).

**Comment 36:** The fact that you have women in your sample who are mothers at 12 years of age says something about the need to protect girls/women from early marriage. It's a shame you haven't included maternal age in your analyses, as I feel it could also be an important risk factor for child ARI and diarrhoea – it's certainly a risk factor for undernutrition. I would advise including maternal age in the re-run of your analysis.

**Response 36:** Maternal age is now incorporated into the model and the tables and figures are updated accordingly.

### **Your figures and tables:**

Comment 37: Page 16, Figure 1:

**Comment:** This figure should show the actual % and 95% CIs.

**Response:** The actual percentages and CIs are added in the figure.

**Comment:** Clarify that 'co-morbidity' means both ARI and diarrhoea

**Response:** The term co-morbidity is now defined in the context (Page 8, lines 13–14).

**Comment:** State 'last 2 weeks' on the x-axis; same applies to the tables.

**Response:** The suggestion is accepted and the x-axis label is changed accordingly.

Comment 38: Table 1:

**Comment:** Line 14: Age of child labels – the oldest group should have an upper age limit, rather than 36 months+; this also applies to the other tables.

**Response:** Upper age limit is set for the mentioned age group of the children.

**Comment:** Lines 8 and 9: you don't need to show ARI and Diarrhoea prevalence as you've already included it in Figure 1.

**Response:** The mentioned lines have been removed from table 1.

**Comment:** Line 22 – what is 'elec' media?

**Response:** The variables are now phrased in a better way.

**Comment:** Line 27 – why has 'unhygienic' been combined with 'not on premises'? It is possible to get hygienic water from public hand pumps. Again, this links to the need for definitions and references for your variables and how you've constructed them.

**Response:** The variables are now defined. Ideal situation in this context is considered as those households using water from improved source located on the premises.

**Comment:** Line 31 – why have unhygienic unshared, and hygienic shared been grouped together?

**Response:** The variable and classification is now reworded and presented in full.

**Comment:** Line 34: aren't children also included in the definition of household overcrowding? You only mention adults here.

**Response:** The study considered only the number of adults sharing a room to define crowding.

Comment 39: Table 2:

**Comment:** Title would be better at the top, in accordance with your other tables.

**Response:** The title moved while creating the pdf file. Now the authors will take special care while creating the .pdf file from .docx one.

**Comment:** The columns are not labelled. I presume these are percentages that show prevalence by sub-category?

**Response:** The column labels were also moved while creating the pdf file. Now the authors will take special care while creating the .pdf file from .docx one.

**Comment:** The data in the table are starred (presumably to indicate statistical significance), but there is no key to the stars to help understand the significance level.

**Response:** The significance level is now come as footnote of the table.

**Comment:** Is this table showing chi-squared results? If so, this should be mentioned in the title. Also, the  $X^2$  and df should be shown.

**Response:** The title is revised; the chi-square values are also mentioned. The df is now mentioned to save space and it can be easily obtained from the number of levels of the variables considered.

**Comment:** Lines 4 and 5 – should mention the last 2 weeks.

**Response:** It is now mentioned.

**Comment:** Line 23: 'rich' implies a value judgement. Perhaps better to say lowest, middle, and highest.

**Response:** The suggestion is incorporated in the tables and elsewhere in the manuscript.

**Comment:** Some of the abbreviations in the table look a bit odd (e.g. 'Hy'). It would be better to spell them out.

**Response:** The suggestion is incorporated into the manuscript.

Comment 40: Table 3 and 4:

**Comment:** The titles have confused me. Are these results of adjusted regression models? If so, they should say AORs, not ORs. This needs clarifying in the tables and text.

**Response:** These results in tables 3 and 4 are from multivariable logistic regression models, and hence, the adjusted odds ratios are used. It is now mentioned in the table and respective sections of the manuscript.

**Comment:** See my earlier point – multivariable analyses would add credibility to your findings (if these tables don't already show AORs from multivariable analyses).

**Response:** The multivariable analyses were used to produce the tables.

**Comment:** There is a problem with the title placement of table 4 (it's split over the top and bottom of the table).

**Response:** Parts of the title was moved down while creating the pdf file. Now the authors will take special care while creating the .pdf file from .docx one.

## Response to reviewer 2:

### The title:

**Comment 1:** The word “impact” seems to be very strong for this study which is looking at data that was collected from cross-sectional DHSs.

**Response 1:** The title of the manuscript is now revised as: *How sustainable household environment and knowledge of healthy practices relate to childhood morbidity in South Asia: Analysis of survey data from Bangladesh, Nepal and Pakistan* (Page 1, lines 1–3).

**Comment 2:** I hope the “sustainable household environment” is defined along the way in introduction.

**Response 2:** The “sustainable household environment” is now defined in the introduction (Page 4, lines 22–27).

### Abstract

**Objective:** Could be “background and objective” The authors now give “The aim” instead of the objective. However, the aim given and title do not rhyme very well.

**Response:** The objective is now revised (Page 2, lines 2–5).

**Participants:** This information could be put better to indicate that the women were the mothers of the children, i.e Information was obtained from mothers (12-49 years) of 23,940 children under age of five years.....Is it important to mention the ever-married?

**Response:** The information is now clarified as suggested (Page 2, lines 9 –10).

**Conclusion:** The last sentence in the conclusion is very so long that by the end of it, one has lost meaning and intension. Consider making 2-3 short meaningful sentences.

**Response:** The last part of conclusion is rewritten (Page 2, lines 18 –24).

### Introduction:

**Comment:** The definition/ explanation of “sustainable household sanitation” is required maybe in the second paragraph. What aspects are we looking at? Maybe the authors could have taken time to define what it is.

**Response:** The “sustainable household environment” is now defined in third paragraph of introduction section (Page 4, lines 22–27).

**Comment:** Line 27, page 4, the sentence begins with “Educated mothers.....their “child”..... Since it is plural then it should follow that word becomes “children”

**Response:** The sentence is now rewritten (Page 4, lines 15–17).

**Comment:** The two sentences that follow in line 32 and 33 begin with “In addition...” the words become monotonous to the readers. Authors could try to find another way of starting the next sentence.

**Response:** The sentence is rewritten (Page 4, lines 18–22).

**Comment:** Line 6 on page 5 “Objective” should be replaced with “Aim”. Line 9, the “first aim” should then be the “first objective”.

**Response:** The suggestions are incorporated into the manuscript (Page 5, lines 3–14).

**Comment:** The use of the word “incidence” in line 7 on page 5 and everywhere in the manuscript should be checked. It could be replaced by “prevalence”.

**Response:** The term ‘incidence’ is now replaced with ‘prevalence’ (Page 2, lines 2, 20, 22; Page 3, line 6; Page 4, line 7, Page 5, lines 4, 6, 8 and elsewhere).

**Comment:** Line 14, this having been data from a cross-sectional surveys, it is not possible to measure the impact rather an association/relationship or likelihood. The word “impact” has also been used in many other places.

**Response:** Changes are now made to the manuscript considering the nature of the study (page 2, line 3; page 5, line 7; page 7, line 3; page 8, lines 31, 32, 34; page 9, lines 3, 4, 5 and elsewhere).

**Comment:** Line 16 “sustainable household environment” comes up, this time as a variable/predictor? Without being defined earlier in introduction, it is not easy to connect with the other variables.

**Response:** The “sustainable household environment” is now defined in the introduction (Page 4, lines 22–27).

**Comment:** Line 38 Sentence begins with “On the other hand.....” there is need to edit that sentence. It is unfair to those countries, to depict them as “representing”.....

**Response:** The sentence is rewritten (Page 5, lines 21–23)

**Comment:** Line 53, 54.....here is need to show how sample size was obtained from each country at the various stages to arrive at the figures seen in the results.

**Response:** The sample size for each country and at various stages are now explained in the Data and Methods section (Page 5, lines 24–31).

**Comment:** Finally, “household sustainable environment” is described in 46-49 on page 6; This is good but should have been done earlier in the introduction for readers to know what the authors are talking about.

**Response:** The “sustainable household environment” is now defined in the introduction (Page 4, lines 22–27).

## Analysis

**Comment:** Authors should give the rationale for excluding some variables in the various models.

**Response:** See page 8 line 5 for the rationale (to focus the models) for excluding cooking fuel from the diarrhea model and source of drinking water from the diarrhea model. The rationale for the modelling decisions regarding ARI in the diarrhea mode and vice versa is now on page 10 line 32 – page 11 line 6.

## Results:

**Comment:** Line 38, the use of the word “incidence” is not correct. Rather use prevalence.

**Response:** The suggestion is incorporated into the manuscript.

**Figure 1:** Check the bars of Bangladesh in Diarrhoea and ARI, they don't seem to say the same as the prevalence of the two conditions in the test and in table 1.

**Response:** The bars are checked and revised plot is provided.

**Table one:** presents results in percentages. It would be good to include numbers beside the percentages and “n (%)” in the footnotes.

**Response:** Providing both information (number and percent) is a good idea. However, when the total number of respondents (N) and percentages of respondents for various subgroups are given, the number of respondents for any subgroup can be calculated. Hence, the authors did not put the numbers in the table to avoid duplication of information.

**Comment:** Title of table 1 refers to predictors. Why then include the prevalence of diarrhea and ARI in the first line? Prevalence is already shown in figure 1 or these could be placed below the figure as foot notes.

**Response:** The prevalence of diarrhoea and ARI in removed from the table and now placed on the top of the bars of Figure 1.

**Table 2:** Be consistent: you had titles for table1, 3 and 4 at the top and title for tale 2 at the bottom of the table. What are the figures in table 2?

**Response:** The title was moved down while creating the pdf file. Now the authors will take special care while creating the .pdf file from .docx one.



I will assume they are percentages from bivariate chi-square test and proceed as follows:

**Comment:** Indicate numbers against percentages

**Response:** The numbers are not added due to limited space.

**Comment:** Also indicate this in the foot notes below the table.

**Response:** The foot notes are added below the tables.

### Some editorial corrections:

**Comment:** In line 8, page 8 “are were”

**Response:** The typo is fixed.

**Comment:** Page 8 line 8, “In the bivariate analyses, children suffering from ARI “are were” more likely to suffer from diarrhea and vice-versa (Table 2).” Does the statement above refer to the figures below extracted from table 2?

*Childhood ARI/ Diarrhoea*

*Dairr ARI Diarr ARI Diarr ARI*

Suffered 9.8\*\*\* 9.3\*\*\* 30.1\*\*\* 10.1\*\*\* 30.2\*\*\* 23.7\*\*\*

Not suffered 5.4 5.1 13.1 3.8 20.4 13.6

If yes, the authors need to think a little more about this. What I get from the figures is the prevalence of diarrhea and ARI at the time of the survey and not the likelihood of a child suffering from ARI also to suffer from diarrhea or vice –versa as they seem to indicate;(i.e the proportion of children who suffered a particular disease was significantly higher than those who did not suffer in each country at the time of survey).

**Response:** Now the sentence has been changed to better reflect the interpretation of the percentages in Table 2 (page 8 line 28 –29).

**Comment:** The paragraph on page 8, the word “impact” has been used repeatedly, but then Chi-square test measures association rather than impact. Therefore replace this word in the text appropriately.

**Response:** The word “impact” is replaced with “association”.

**Comment:** Again the word “incidence” appears in this paragraph. It should also appropriately be replaced. In the text describing results presented in Table 3 and 4, there are values presented as proportions (could they be the  $B=Logistic\ Co-Efficients?$ ), alongside the OR and CI. I suggest space be created in the tables so that all variable have these  $B=Logistic\ Co-Efficient$  values presented for each.

**Response:** The word ‘incidence’ is now replaced with prevalence. Due to limited space, only the ORs are presented as they are easy to explain and understand. The use of only ORs are also not uncommon in the literature in public health discipline.

**Comment:** Also there should be consistency in the way of presentation of these values (e.g either as 0.70 or 70%), Check line 41 and line 49 page 8; line 15 and 25 on page 9.

**Response:** In the revised manuscript, only the percentages are used to explain the odds ratios.

**Comment:** Table 3 and 4, foot notes would be good to use in order to reduce the crowding in the tables and to make it easy for readers.

**Response:** The foot notes are added below the tables.

### Discussion

**Comment:** I think the authors should acknowledging using that data from cross sectional survey provides information regarding the disease episodes for a single point of time (line 57 page 9), such study cannot yield “impact” of a given variable on disease episodes. Page 10, line 1 and may be others to follow: this study should not be talking of incidence of disease/ incidence rates. Line 21 and 28, the word “effect” is still inappropriate in this study because the design of this study cannot bring out the “cause- effect” relationship. It would have been interesting to show the levels of exclusive breastfeeding in the background characteristics for the different countries.

**Response:** The first parts of the comments are accepted and the manuscript is revised accordingly (as mentioned in the response of the reviewer's earlier comments). Exclusive breastfeeding is not considered in the analysis as the information regarding exclusive breastfeeding is only available for children below six months (see page 11 lines 18 – 19).

## Conclusion

**Comment:** Line 26 -28, While it could be true, this study has not confirmed it, therefore state with caution that “maternal education could be effective.....”

**Response:** The suggestion is incorporated into the manuscript (page 12, lines 8 –10).

## VERSION 2 – REVIEW

<b>REVIEWER</b>	GRACE K M MUHOOZI KYAMBOGO UNIVERSITY, UGANDA UNIVERSITY OF OSLO, NORWAY
<b>REVIEW RETURNED</b>	28-Jan-2017

<b>GENERAL COMMENTS</b>	<p>The title, abstract and other sections of the paper have improved and strengthened the paper. Most of the specific peer review comments raised have also been addressed.</p> <p>However, I have a few concerns which need to be further addressed</p> <ol style="list-style-type: none"> <li>1. Strengths and limitations: differentiate which is which. They were simply lumped together.</li> <li>2. Introduction: Line 3 of page 4, “In addition to deaths.....”, better to use “Apart from deaths.....”</li> <li>3. Line 18 of page 4. The last part of the sentence that starts with “In addition.....” Is not clear. Does access to mass media form knowledge of communicable diseases?</li> <li>4. Line 1 page 5, the sentence should be in a reported form with past tense of “was” and not “is”.</li> <li>5. Similarly, lines 5-9, the objectives were stated in future tense, it should be in the past tense as this is a report.</li> <li>6. Lines 6,7 and 8 of page 6; With the sentence that follows, this sentence is not necessary here. Could this be placed among the limitations?</li> <li>7. I still see the use of the word “impact” in the manuscript; Line 2 page 8, line 8 page 12</li> <li>8. Line 33 on page 10 and line 3 on page 11; subheadings are not necessary.</li> <li>9. Line 20 page 12, the word “incorporating” could be replaced by “using or utilizing”</li> <li>10. The footnote on table 2,3,and 4 “a: <math>p &lt; 0.01</math>; b: <math>0.01 \leq p &lt; 0.05</math>; c: <math>0.05 \leq p &lt; 0.1</math>” is not clear to me. If I understand it, p should come before the values i.e a: <math>p &lt; 0.01</math>; b: <math>p \geq 0.01, &lt; 0.05</math>; c: does not make sense to me. Can it be clarified?</li> </ol>
-------------------------	--

## VERSION 2 – AUTHOR RESPONSE

Thankyou for the opportunity to respond to reviewer comments on the revised submission. Our response to the comments is given after each comment below.

Response to the comments from reviewer 2:

Comment 1: Strengths and limitations: differentiate which is which. They were simply lumped together.

Response 1: Two of the limitations are noted specifically as such. The last bullet point is expressed as a negative already, so no further changes made.

Comment 2: Introduction: Line 3 of page 4, “In addition to deaths.....”, better to use “Apart from deaths.....”

Response 2: The text is changed as suggested.

Comment 3: Line 18 of page 4. The last part of the sentence that starts with “In addition.....” Is not clear. Does access to mass media form knowledge of communicable diseases?

Response 3: The phrase is deleted.

Comment 4: Line 1 page 5, the sentence should be in a reported form with past tense of “was” and not “is”. Comment 5: Similarly, lines 5-9, the objectives were stated in future tense, it should be in the past tense as this is a report.

Response 4 and 5: Entire paragraph revised to incorporate the comments.

Comment 6: Lines 6,7 and 8 of page 6; With the sentence that follows, this sentence is not necessary here. Could this be placed among the limitations?

Response 6: The sentence is now moved to “Data limitations” section on page 11.

Comment 7: I still see the use of the word “impact” in the manuscript; Line 2 page 8, line 8 page 12

Response 7: The term “impact” is changed at Line 2 page 8 and Line 8, page 12. The remaining usages do not refer to the modelling but to broader concepts and so have remained in the manuscript.

Comment 8: Line 33 on page 10 and line 3 on page 11; subheadings are not necessary.

Response 8: The subheadings are removed.

Comment 9: Line 20 page 12, the word “incorporating” could be replaced by “using or utilizing”

Response 9: We prefer to leave “incorporating” because it suggests a coming together of disparate organisations and people to work towards a common goal.

Comment 10: The footnote on table 2,3 and 4 “a:  $p < 0.01$ ; b:  $0.01 \leq p < 0.05$ ; c:  $0.05 \leq p < 0.1$ ” is not clear to me. If I understand it, p should come before the values i.e a:  $p < 0.01$ ; b:  $p \geq 0.01, < 0.05$ ; c: does not make sense to me. Can it be clarified?

Response 10: Footnotes of the tables are now changed to match an existing article in BMJ (Open).

Kamal et al., (2015): Determinants of childhood morbidity in Bangladesh: evidence from the Demographic and Health Survey 2011,

<http://bmjopen.bmj.com/content/bmjopen/5/10/e007538.full.pdf>.

Comment: May seek services of a language expert.

Response: The entire manuscript has been checked for grammatical and typographical errors.