

PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Association of water, sanitation, hygiene and food practices with enteric fever in a pediatric cohort in North India
AUTHORS	Rongsen-Chandola, Temsunaro Dudeja, Nonita Sinha, Bireswar Goyal, Nidhi Arya, Alok Revi, Anitha Dutta, Ankita More, Deepak Chakravarty, Aparna Kumar, Chandra Mohan

VERSION 1 – REVIEW

REVIEWER	Reviewer name: Dr. Sarah Nevitt Institution and Country: University of Liverpool, United Kingdom of Great Britain and Northern Ireland Competing interests: None
REVIEW RETURNED	31-Jan-2022

GENERAL COMMENTS	<p>I have conducted a statistical review of the manuscript "Association of water, sanitation, hygiene and food practices with enteric fever in a pediatric cohort in North India"</p> <p>The authors describe a cohort study recruiting around 6000 children recruited from households in South Delhi, India examining the association between WASH practices and enteric fever.</p> <p>My main query after reading this manuscript is what this study contributes to the existing literature. The authors refer to a recent systematic review examining the association between WASH practices and enteric fever in both the Introduction and the Discussion section. In general, one would assume that a systematic review which pools multiple studies examining a topic would provide more information than a single study on the topic, so why is this new study required in addition to the existing systematic review? Are the studies in the systematic review out of date / low quality? Have practices now changed compared to the studies included in the systematic review? Does the present cohort study examine additional questions or factors which the systematic review does not?</p> <p>Please add some emphasis of what this study adds to the literature in the context of the existing systematic review to both the Introduction and to the Discussion.</p> <p>Other specific comments: 1) The authors refer to 'WASH factors' throughout the manuscript but I don't think that a clear definition or description of what 'WASH factors' actually are is provided. Some definitions related to water safety are provided in the 'Data Variables and Analysis' section but I suggest that 'WASH factors' should be clearly defined the first time</p>
-------------------------	---

	<p>this term is used (i.e. the Introduction).</p> <p>2) I have a few comments on the 'Data Collection' section:</p> <ol style="list-style-type: none"> a. "In the event of fever, our team made daily contacts till the end of the fever episode." Presumably fever cases are recorded in only the child / children recruited into the study rather than any fever present within the household? b. Related to the above sentence, if I understand correctly, eligibility screening was performed via door-to-door surveys (i.e. in person) and information on fever cases was collected via mobile phones (Discussion, second paragraph). How was data collected on WASH and food hygiene practices, in person survey or via phone? Please clearly describe all modes of data collection. c. "Culture confirmed enteric fever cases were followed up, till the 90th day from fever onset, for outcome assessment." I am not sure I understand this sentence. My understanding is that the 'outcome' of interest to this study was presence of enteric fever (yes or no) so what is assessed within the follow-up after enteric fever is confirmed? Please clarify d. "Information related to food practices were obtained for the eldest enrolled child in the household" So where multiple children are recruited from a household, it is assumed that food practices are the same for all children (in other words, food practices are recorded at the household-level rather than the child-level – also see comment 3b below about methods to adjust for clustering). e. The sample size selected for the analysis of microbiologically safe drinking water may well be reasonable but I don't understand the assumption "50% households in the community would have access to microbially safe drinking water with an allowable error of 20%." Does this mean that it is assumed that between 30% and 70% of households may have access to microbiologically safe drinking water? This seems like quite a big range and given the statement within the Introduction section (third paragraph) that less than 50% of households have access to safe drinking water, is the upper bound of 70% too high? <p>3) I have a few comments on the 'Data variables and analysis' section</p> <ol style="list-style-type: none"> a. Please note that DAG models are a useful means of visually presenting the relationships between variables (e.g. exposure, outcome and confounders, structural pathways, assumptions) but they are not a method of performing univariable analyses as implied in this section. Please rephrase and describe how the covariates were identified as potential confounders. b. Please describe in more detail how the univariable and multivariable analyses were conducted 'while accounting for household clustering.' Particularly, which variables were treated as household-level variables and which were treated as child-level variables? How were variables measured at different levels included in the same analysis (e.g. multi-level or mixed effects regression analyses)? c. Please either add references for 'Moran's Index' and 'Buffer analysis' or please add further explanation of what these methods do. <p>4) Results: The numbers within Table 1 and summary statistics quoted within the text reflect 5916 children from 3123 households (i.e those who completed the survey on demographic characteristics and WASH and food practices, rather than 6000 children from 3172 households. In other words, I think that the sentence at the start of the second paragraph needs to come right at the start of the section for clarity.</p> <p>5) Results: The results of the buffer analysis are stated at the end of the results section but don't really seem to be interpreted. Is this the interpretation of the results of buffer analysis (Discussion, first paragraph)?</p>
--	---

	<p>"Geospatial analysis revealed probability of higher risk of a case of enteric fever in surrounding zones of households with contaminated drinking water as compared to those without contaminated drinking water."</p> <p>If so, please note that the output of the buffer analysis are odds ratios rather than risk ratios so these results should be interpreted in terms of odds rather than risk.</p> <p>6) Discussion, second paragraph: "We believe that unequal distribution of the undetected cases amongst the exposed and unexposed groups is unlikely, but the smaller number of cases may be a reason for lesser power and the observed wide confidence intervals." I'm not sure I understand the use of the word 'smaller' here, smaller than what? Do the authors mean 'small number of cases' rather than 'smaller'?</p> <p>Also, if the authors are concerned about undetected cases, they could calculate the incidence of enteric fever during the follow-up time and compare to population-based estimates of incidence (i.e. those quoted within the Introduction) to get an idea of the incidence within this cohort is lower than expected (i.e. if some 'expected' cases may have been undetected).</p> <p>7) Discussion, third paragraph: When quoting the results of the previous systematic review, what does n correspond to? The number of studies? I suggest that it is also important to state the number of participants and/or households the estimates correspond to for context when comparing the results of this study to the results of the systematic review.</p>
--	--

REVIEWER	Reviewer name: Dr. Sina Aziz Institution and Country: Department of Paediatrics, Pakistan Competing interests: None
REVIEW RETURNED	14-Feb-2022

GENERAL COMMENTS	<p>I have conducted a statistical review of the manuscript "Association of water, sanitation, hygiene and food practices with enteric fever in a pediatric cohort in North India"</p> <p>The authors describe a cohort study recruiting around 6000 children recruited from households in South Delhi, India examining the association between WASH practices and enteric fever.</p> <p>My main query after reading this manuscript is what this study contributes to the existing literature. The authors refer to a recent systematic review examining the association between WASH practices and enteric fever in both the Introduction and the Discussion section. In general, one would assume that a systematic review which pools multiple studies examining a topic would provide more information than a single study on the topic, so why is this new study required in addition to the existing systematic review? Are the studies in the systematic review out of date / low quality? Have practices now changed compared to the studies included in the systematic review? Does the present cohort study examine additional questions or factors which the systematic review does not?</p> <p>Please add some emphasis of what this study adds to the literature in the context of the existing systematic review to both the Introduction and to the Discussion.</p> <p>Other specific comments: 1) The authors refer to 'WASH factors' throughout the manuscript but I don't think that a clear definition or description of what 'WASH</p>
-------------------------	--

	<p>factors' actually are is provided. Some definitions related to water safety are provided in the 'Data Variables and Analysis' section but I suggest that 'WASH factors' should be clearly defined the first time this term is used (i.e. the Introduction).</p> <p>2) I have a few comments on the 'Data Collection' section:</p> <ol style="list-style-type: none"> "In the event of fever, our team made daily contacts till the end of the fever episode." Presumably fever cases are recorded in only the child / children recruited into the study rather than any fever present within the household? Related to the above sentence, if I understand correctly, eligibility screening was performed via door-to-door surveys (i.e. in person) and information on fever cases was collected via mobile phones (Discussion, second paragraph). How was data collected on WASH and food hygiene practices, in person survey or via phone? Please clearly describe all modes of data collection. "Culture confirmed enteric fever cases were followed up, till the 90th day from fever onset, for outcome assessment." I am not sure I understand this sentence. My understanding is that the 'outcome' of interest to this study was presence of enteric fever (yes or no) so what is assessed within the follow-up after enteric fever is confirmed? Please clarify "Information related to food practices were obtained for the eldest enrolled child in the household" So where multiple children are recruited from a household, it is assumed that food practices are the same for all children (in other words, food practices are recorded at the household-level rather than the child-level – also see comment 3b below about methods to adjust for clustering). The sample size selected for the analysis of microbiologically safe drinking water may well be reasonable but I don't understand the assumption "50% households in the community would have access to microbially safe drinking water with an allowable error of 20%." Does this mean that it is assumed that between 30% and 70% of households may have access to microbiologically safe drinking water? This seems like quite a big range and given the statement within the Introduction section (third paragraph) that less than 50% of households have access to safe drinking water, is the upper bound of 70% too high? <p>3) I have a few comments on the 'Data variables and analysis' section</p> <ol style="list-style-type: none"> Please note that DAG models are a useful means of visually presenting the relationships between variables (e.g. exposure, outcome and confounders, structural pathways, assumptions) but they are not a method of performing univariable analyses as implied in this section. Please rephrase and describe how the covariates were identified as potential confounders. Please describe in more detail how the univariable and multivariable analyses were conducted 'while accounting for household clustering.' Particularly, which variables were treated as household-level variables and which were treated as child-level variables? How were variables measured at different levels included in the same analysis (e.g. multi-level or mixed effects regression analyses)? Please either add references for 'Moran's Index' and 'Buffer analysis' or please add further explanation of what these methods do. <p>4) Results: The numbers within Table 1 and summary statistics quoted within the text reflect 5916 children from 3123 households (i.e those who completed the survey on demographic characteristics and WASH and food practices, rather than 6000 children from 3172 households. In other words, I think that the sentence at the start of the second paragraph needs to come right at the start of the section for clarity.</p> <p>5) Results: The results of the buffer analysis are stated at the end of</p>
--	---

	<p>the results section but don't really seem to be interpreted. Is this the interpretation of the results of buffer analysis (Discussion, first paragraph)?</p> <p>"Geospatial analysis revealed probability of higher risk of a case of enteric fever in surrounding zones of households with contaminated drinking water as compared to those without contaminated drinking water."</p> <p>If so, please note that the output of the buffer analysis are odds ratios rather than risk ratios so these results should be interpreted in terms of odds rather than risk.</p> <p>6) Discussion, second paragraph: "We believe that unequal distribution of the undetected cases amongst the exposed and unexposed groups is unlikely, but the smaller number of cases may be a reason for lesser power and the observed wide confidence intervals." I'm not sure I understand the use of the word 'smaller' here, smaller than what? Do the authors mean 'small number of cases' rather than 'smaller'?</p> <p>Also, if the authors are concerned about undetected cases, they could calculate the incidence of enteric fever during the follow-up time and compare to population-based estimates of incidence (i.e. those quoted within the Introduction) to get an idea of the incidence within this cohort is lower than expected (i.e. if some 'expected' cases may have been undetected).</p> <p>7) Discussion, third paragraph: When quoting the results of the previous systematic review, what does n correspond to? The number of studies? I suggest that it is also important to state the number of participants and/or households the estimates correspond to for context when comparing the results of this study to the results of the systematic review.</p>
--	--

REVIEWER	Reviewer name: Dr. Mila Shakya Institution and Country: United Kingdom of Great Britain and Northern Ireland Competing interests: None
REVIEW RETURNED	19-Feb-2022

GENERAL COMMENTS	<p>General: "Enteric fever" and "typhoid fever" have been capitalized throughout the paper, not a proper noun, please correct.</p> <p>Methods: 1. All enrollees are children under the age of 16 years. The authors mention consent for participation form children. Please clarify. 2. The authors mention weekly follow-up. Although the authors refer to a previously published protocol, can the authors mention how and why the weekly follow-ups were done? 3. Were the fever cases identified in the weekly follow-ups? If yes, does that potentially mean that febrile cases may have gone unidentified? 4. It is unclear who the respondents for the WASH / food practices questionnaires are. Can this lead to misclassification?</p> <p>Results 1. Please include the p-values for the Ors and RRs.</p> <p>Discussion 1. Can the authors elaborate on what the findings of the the geospatial analysis mentioned in page 8 lines 14 - 17 could imply? 2. Page 8 lines 40 - 45, reference for systematic review missing. Also not clear what "this review" in line 47 is referring to. 3. Page 8 lines 58 - 60 - how does the study show the importance of</p>
-------------------------	--

	<p>active case finding for typhoid fever? Is active case finding done programmatically anywhere that has helped reduce the burden?</p> <p>4. The authors claim association but a lack of precision in the first paragraph of the discussion. The authors mention low sample size as a limitation in the second paragraph. However, it is not just lack of precision or wide CI. Of note, all (except one) 95% CI of RRs /ORs presented in the result section contain the value 1, which would suggest there is no association. Please rephrase your statement in the discussion.</p> <p>5. Page 9 Line 19, please specify what other factors.</p>
--	---

REVIEWER	<p>Reviewer name: Dr. Abhilasha Karkey Institution and Country: University of Oxford, United Kingdom of Great Britain and Northern Ireland Competing interests: None</p>
REVIEW RETURNED	10-Feb-2022

GENERAL COMMENTS	<p>Page3: Line 23-26: RR 1.71 does not mean, "The risk of enteric fever was 71% higher in children belonging to households eating food from outside once a week or more". It means, "The risk of enteric fever was 1.71 times more in children belonging to households eating food from outside once a week or more."</p> <p>Page 4; Line 10: South east Asia needs to be converted to South Asia. Most recent global analysis on the burden of disease shows that South Asia has the highest burden of disease(https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(18)30685-6/fulltext).</p> <p>Page 5; Line 37: How often were water samples taken for analysis? One time analysis might not reflect any MPN variance that could be happening.</p> <p>Page 5; Line 39: I am not sure what the TUV SUD stands for. Are you referring to performance testing? All abbreviations need to be expanded on in the manuscript.</p> <p>Page 6; Line 6,7,8: I am sure good food practices are not defined as "consumption of uncooked foods, consumption of food from street vendors and unnamed ice creams less than once a week." Please change definition. Good food practices are the opposite to what is mentioned in the text.</p> <p>Page 6; Line 38: It should be 17% and not 71%</p> <p>Page 7; Line 8: RR 1.71 does not mean, "The risk of enteric fever was 71% higher in children belonging to households eating food from outside once a week or more". It means, "The risk of enteric fever was 1.71 times more in children belonging to households eating food from outside once a week or more."</p> <p>Page 7; Line 13-14: RR 1.32 means, "The risk of enteric fever was 1.32 times higher in children living in households sharing toilets."</p> <p>The authors need to reference more recent data on global burden of enteric fever, as they have the region with the highest burden of disease incorrect. Certain interpretations on datasets cannot be made as I have mentioned above.</p> <p>The paper does not add anything new with what we know of enteric fever and how they are interrelated with food and water. However, I do accept that information needs to be updated from time to time, but how often is too often and how important is this publication? That is subjective.</p>
-------------------------	--

VERSION 1 – AUTHOR RESPONSE

AUTHOR RESPONSE LETTER

Dear Prof. Imti Choonara

Editor-in-chief

BMJ Paediatrics Open

We thank you for your response and invitation to resubmit the paper in the BMJ Paediatrics Open.

We have taken note of the comments and suggestions made by the reviewers and have revised the manuscript accordingly.

As per the editorial requirements, our responses to the comments are presented using a tabular structure with the original reviewer's comments and authors responses. Where revision to the manuscript text has been made, the text is copied into this letter with yellow highlight along with page numbers and section in italics (as they appear in the new/clean version).

Both this clean version of the revised manuscript as well as a version with the changes tracked (highlighted) are also submitted.

Best regards,

Temsunaro Rongsen Chandola

COMMENTS

RESPONSES

Editor in Chief Comments to Author :

What this study adds 1st statement delete the 2nd sentence (not statistically significant and the statement is too long)

Thank you. We have removed the statement.

What this study adds delete the 2nd statement (Confidence intervals are large and we try not to present numerical data in what this study adds)

Thank you. We have removed the statement.

Reviewer: 1

Dr. Sarah Nevitt, University of Liverpool

I have conducted a statistical review of the manuscript "Association of water, sanitation, hygiene and food practices with enteric fever in a pediatric cohort in North India" The authors describe a cohort study recruiting around 6000 children recruited from households in South Delhi, India examining the association between WASH practices and enteric fever.

-

My main query after reading this manuscript is what this study contributes to the existing literature. The authors refer to a recent systematic review examining the association between WASH practices and enteric fever in both the Introduction and the Discussion section. In general, one would assume that a systematic review which pools multiple studies examining a topic would provide more information than a single study on the topic, so why is this new study required in addition to the existing systematic review? Are the studies in the systematic review out of date / low quality? Have practices now changed compared to the studies included in the systematic review? Does the present cohort study examine additional questions or factors which the systematic review does not?

We thank the reviewer for their comment. There are certain points we would like to highlight:

Three Indian studies were included in the systematic review mentioned, two studies were from West Bengal and one from Delhi. All these studies were done in the early 2000s and showed high heterogeneity in the burden, ranging from 215 per 100 000 PY in Kolkata to 980 per 100 000 PY in New Delhi. There have been a lot of change in the population dynamics, practices related to water sanitation and hygiene and immunization coverage in the past decade. It was expected that the burden of enteric fever may have decreased over the last two decades. However, we found that currently the incidence of Enteric Fever in the present population was high at 703.7 (95% CI 560.5–

874.7) per 100 000 PY. No recent study has been done in the Indian setting that studies to revisit the strength of association between WASH factors and incidence of enteric fever in the current scenario. Several initiatives came into force during these two decades. The Swachh Bharat mission was launched in the country in 2014 with a primary focus on improving sanitation, especially, reducing open free defecation. Availability of water and sanitation services have improved across states. Additionally, the typhoid vaccine was rolled out in the state of Delhi at 2 yrs of age for all children under the immunization program. The vaccine however was not available in all the states leading to irregular coverage.

Given the significant changes in the population dynamics, government policies and immunization practices in the country over the years, it becomes important to understand if Water Sanitation and Hygiene practices still have a role in the incidence of Enteric fever, specifically in an urban slum population.

Please add some emphasis of what this study adds to the literature in the context of the existing systematic review to both the Introduction and to the Discussion.

Thank you for your comment we have added the rationale in both Introduction and Discussion.

Introduction, Page 4, lines 26-38,

“The studies on association of WASH and enteric fever from the Indian settings included in the review by Brocket et al, were published in early 2000s. 8-10 Also, there was a lot of heterogeneity in the burden of enteric fever across different states ranging from 215 per 100 000 PY in Kolkata to 980 per 100 000 PY in New Delhi. 11-14 Multiple initiatives during these years have affected the WASH and food hygiene practices as well as immunization schedules across different states in the country. Contrary to the speculations of a lower burden, our recent study reported a high burden of enteric fever i.e., 703.7 (95% CI 560.5–874.7) per 100 000 PY. 15

In the present scenario, with improved access to water and sanitation, the role of WASH practices towards disease burden of enteric fever remains to be reasserted. Using the data from our paediatric longitudinal cohort in Delhi, 16 we aim to estimate the association between WASH and food practices with culture-confirmed enteric fever.”

Other specific comments:

1) The authors refer to ‘WASH factors’ throughout the manuscript but I don’t think that a clear definition or description of what ‘WASH factors’ actually are is provided. Some definitions related

to water safety are provided in the 'Data Variables and Analysis' section but I suggest that 'WASH factors' should be clearly defined the first time this term is used (i.e. the Introduction).

We thank the reviewer for their comment. We have added the description of the term WASH in the introduction. The line reads as below:

Introduction, Page 4 lines 16-17

"A recent systematic review on association between WASH (acronym for water, sanitation and hygiene) factors and Typhoid fever revealed that improved water...."

2) I have a few comments on the 'Data Collection' section:

-

"In the event of fever, our team made daily contacts till the end of the fever episode." Presumably fever cases are recorded in only the child / children recruited into the study rather than any fever present within the household

Yes, that is correct.

Related to the above sentence, if I understand correctly, eligibility screening was performed via door-to-door surveys (i.e. in person) and information on fever cases was collected via mobile phones (Discussion, second paragraph). How was data collected on WASH and food hygiene practices, in person survey or via phone? Please clearly describe all modes of data collection.

We thank the reviewer for pointing this out.

Potentially eligible children were identified by the study personnel by home visits. In case an eligible child contracted fever, daily follow up in the form of home visits was done till the fever episode resolved. Water sanitation hygiene and food practices were assessed using a structured interview schedule and direct observation during home visit by study personnel. We have added this in the data collection section.

Methods, Data Collection, Page 5, lines 15-16

"WASH and food hygiene practices were assessed using a structured interview schedule that was administered during home visit in the study households."

"Culture confirmed enteric fever cases were followed up, till the 90th day from fever onset, for outcome assessment." I am not sure I understand this sentence. My understanding is that the

'outcome' of interest to this study was presence of enteric fever (yes or no) so what is assessed within the follow-up after enteric fever is confirmed? Please clarify

Thank you for the comment. We have removed this statement as it is not relevant in this article's context.

For your information we would like to state that this sub-cohort of culture confirmed enteric fever cases was formed to determine the complications (if any) and outcomes at three months after an episode of enteric fever.

"Information related to food practices were obtained for the eldest enrolled child in the household" So where multiple children are recruited from a household, it is assumed that food practices are the same for all children (in other words, food practices are recorded at the household-level rather than the child-level – also see comment 3b below about methods to adjust for clustering).

Thank you.

Yes, that is correct.

In the present study we have assumed that the household practices related to food are the similar for all the children in that household. Hence, the food practices were recorded at the household level rather than the child level.

The sample size selected for the analysis of microbiologically safe drinking water may well be reasonable but I don't understand the assumption "50% households in the community would have access to microbially safe drinking water with an allowable error of 20%." Does this mean that it is assumed that between 30% and 70% of households may have access to microbiologically safe drinking water? This seems like quite a big range and given the statement within the Introduction section (third paragraph) that less than 50% of households have access to safe drinking water, is the upper bound of 70% too high?

We thank the reviewer for their comment. Our assumption of '50% households having access to microbiologically safe drinking water is based on the following reference:

Published literature by Bandopadhyay et al in 1992 showed that ~43% of the drinking water samples collected were contaminated with coliforms in Delhi.

(Bandyopadhyay S, Banerjee K, Khanna KK, Sharma RS, Verghese T. Drinking water quality and diarrhoea in Delhi. *J Commun Dis.* 1992 Sep;24(3):156-8. PMID: 1344946.)

We have added this reference.

We have considered relative allowable error not absolute.

The sample size for this exercise was calculated assuming that 50% households in the community would have access to microbially safe drinking water with a relative allowable error of 20%¹⁹.

3) I have a few comments on the 'Data variables and analysis' section

-

Please note that DAG models are a useful means of visually presenting the relationships between variables (e.g. exposure, outcome and confounders, structural pathways, assumptions) but they are not a method of performing univariable analyses as implied in this section. Please rephrase and describe how the covariates were identified as potential confounders.

We agree with the reviewer. We have modified this section to improve clarity.

We used the DAG model as a conceptual framework to identify the covariates to be adjusted in the multivariable analysis. This as a technique that has also been recommended previously. We have added reference for this.

However, we did not solely base our adjustments on the DAG model. We included the covariates in our final multivariable model which were either suggested by conceptual DAG or significant at $p < 0.25$ in the univariable analysis. We have mentioned this in the Data variables and analysis section as below.

Methods, Data variables and analysis, Page 6, Lines 9-12

"Based on the Directed Acyclic Graph (DAG) conceptual model (daggity.net), age, education, family size and receiving typhoid vaccine were identified as the potential confounders. We included the covariates in our final multivariable model which were either suggested by the conceptual DAG model or significant at $p < 0.25$ in univariable analysis (Figure 1)."

Please describe in more detail how the univariable and multivariable analyses were conducted 'while accounting for household clustering.' Particularly, which variables were treated as household-level variables and which were treated as child-level variables? How were variables measured at different levels included in the same analysis (e.g. multi-level or mixed effects regression analyses)?

Thank you.

We have done a child based analysis in the study. For all the WASH and food related practices, household level information was taken. However, in case of multiple children belonging to the same household, we used the variance-covariance matrix of the estimators (VCE) command in STATA to account for household clustering.

Methods, Data variables and analysis, Page 5, Lines 42-46

“We conducted a child-based analysis to study the association between WASH practices and enteric fever. To account for the design effects of household clustering, when more than one child were included in the cohort from a household, we used Stata’s robust variance estimator (cluster) option.”

Please either add references for ‘Moran’s Index’ and ‘Buffer analysis’ or please add further explanation of what these methods do.

Thank you for the comment. We have added the reference.

Methods, Data variables and analysis, Page 6, Lines 14-18

We estimated Moran’s index to assess clustering of enteric fever cases in the area.²¹ We conducted a buffer analysis to estimate the odds of a case of enteric fever within a coverage area of 5m, 10m, and 25m radius around the study households with and without contaminated drinking water.²²

4) Results: The numbers within Table 1 and summary statistics quoted within the text reflect 5916 children from 3123 households (i.e those who completed the survey on demographic characteristics and WASH and food practices, rather than 6000 children from 3172 households. In other words, I think that the sentence at the start of the second paragraph needs to come right at the start of the section for clarity.

We agree with the reviewer. The sentence has been moved to the start of the paragraph.

Results, Page 6, Lines 38-39

“The survey on demographic characteristics and WASH and food practices was completed in 3123 households with 5916 children.”

5) Results: The results of the buffer analysis are stated at the end of the results section but don’t really seem to be interpreted. Is this the interpretation of the results of buffer analysis (Discussion, first paragraph)?

“Geospatial analysis revealed probability of higher risk of a case of enteric fever in surrounding zones of households with contaminated drinking water as compared to those without contaminated drinking water.”

If so, please note that the output of the buffer analysis are odds ratios rather than risk ratios so these results should be interpreted in terms of odds rather than risk.

We thank the reviewer for their comments. We have reframed the sentence as follow:

Discussion, Page 8, Lines 9-11

“Geospatial analysis revealed higher odds of enteric fever in surrounding zones of households with contaminated drinking water as compared to those without contaminated drinking water.”

6) Discussion, second paragraph: “We believe that unequal distribution of the undetected cases amongst the exposed and unexposed groups is unlikely, but the smaller number of cases may be a reason for lesser power and the observed wide confidence intervals.”

I'm not sure I understand the use of the word 'smaller' here, smaller than what? Do the authors mean 'small number of cases' rather than 'smaller'?

Also, if the authors are concerned about undetected cases, they could calculate the incidence of enteric fever during the follow-up time and compare to population-based estimates of incidence (i.e., those quoted within the Introduction) to get an idea of the incidence within this cohort is lower than expected (i.e. if some 'expected' cases may have been undetected).

Thank you. Yes, we mean to say that the number of cases was small. We have reframed the sentence as follows:

Discussion, Page 8, Lines 21-23

We believe that unequal distribution of the undetected cases amongst the exposed and unexposed groups is unlikely, but the small number of cases may be a reason for less power and the observed wide confidence intervals.

When we talk of undetected cases, we mean primarily the intrinsic limitation of the testing method, that is the poor sensitivity of blood culture (60%). Other reason for non-detection could be the surveillance strategy that excludes children with possible diagnosis of enteric fever with only two days of fever who were not tested as per the protocol.

We beg to differ with the reviewers statement that “ incidence of enteric fever during the follow-up time and compare to population-based estimates of incidence (i.e. those quoted within the Introduction) to get an idea of the incidence within this cohort is lower than expected” as

Our cohort incidence estimates are already higher than the pooled estimates in the country. The previously done studies in the same settings were done at different time periods and it has been almost 20 years since the last incidence estimates were obtained for a slum area of New Delhi. Current estimates for our population are unavailable for comparison, so it is difficult to comment on the expected number of cases at present.

1) Discussion, third paragraph: When quoting the results of the previous systematic review, what does n correspond to? The number of studies? I suggest that it is also important to state the number of participants and/or households the estimates correspond to for context when comparing the results of this study to the results of the systematic review.

Thank you , we have written as number of studies. However, the number of participants are not mentioned in the manuscript.

Discussion, Page 8, Lines 32-28

“A systematic review by Brockett S et al reported that an improved water source (16 studies, OR = 0.73, 95% CI = 0.56–0.95) and treatment of water before consumption (9 studies OR = 0.59, 95% CI = 0.45–0.75) are associated with significantly lower odds of typhoid fever. Similar to our findings, it was reported that food and drink consumed outside home (39 studies) was significantly associated with higher odds of typhoid (OR = 1.6, 95% CI = 1.4–1.8).⁵ The study also reported that that ice cream consumption was significantly associated with higher odds of Typhoid (10 studies, OR = 1.5, 95% CI = 1.2–1.9, I² = 31%).⁵”

Reviewer: 2

Dr. Abhilasha Karkey, University of Oxford, Patan Academy of Health Sciences

Page3:

Line 23-26: RR 1.71 does not mean, " The risk of enteric fever was 71% higher in children belonging to households eating food from outside once a week or more". It means, "The risk of enteric fever was 1.71 times more in children belonging to households eating food from outside once a week or more.

We thank the reviewer for their comment, we have rephrased the interpretation.

Results, Page 7, Lines 9-10

The risk of enteric fever was 1.71 times higher in children belonging to households eating food from outside once a week or more (adjusted RR 1.71, 95% CI 1.00 to 2.94).

Page 4; Line 10:

South east Asia needs to be converted to South Asia. Most recent global analysis on the burden of disease shows that South Asia has the highest burden of disease([https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(18\)30685-6/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(18)30685-6/fulltext)).

We thank the reviewer for their comment. We have changed South-east Asia to South Asia and added the reference as well.

Introduction, Page 4, lines 7-8

“Low and middle-income countries have high burden of enteric fever with South Asia and Sub-Saharan African countries accounting for majority of the cases.¹”

Page 5; Line 37: How often were water samples taken for analysis? One time analysis might not reflect any MPN variance that could be happening.

This was a one-time assessment done for checking the water quality.

Page 5; Line 39: I am not sure what the TUV SUD stands for. Are you referring to performance testing? All abbreviations need to be expanded on in the manuscript.

Thank you for the comment.

TÜVs (German pronunciation: [ˈtʏf]; short for German: Technischer Überwachungsverein, English: Technical Inspection Association) are internationally active, independent service companies

which provide certification and testing services for a variety of industrial needs. We have expanded this abbreviation in the manuscript.

Methods, Data Collection, Page 5, Lines 27-28

“The laboratory tests were conducted in TÜVs (Technischer Überwachungsverein) (TUV SUD), Gurgaon, Haryana, India.”

Page 6; Line 6,7,8: I am sure good food practices are not defined as "consumption of uncooked foods, consumption of food from street vendors and unnamed ice creams less than once a week." Please change definition. Good food practices are the opposite to what is mentioned in the text.

Thank you. We have reframed the sentence.

Methods, Data variables and analysis, Page 6, Lines 5-7

"Consumption of cooked foods along with rare consumption of outside food (consumption of food from street vendors and/or unnamed ice creams less than once a week) was labelled as good food practices.

Page 6; Line 38: It should be 17% and not 71%

Thank you. It may be noted that as given in table 1, 70.6 % of the children belonged to overcrowded households, we had rounded off to 71%. We have replaced the value with 70.6%.

Results, Page 6, Lines 34-35

"Three-fourth (73%) of the children belonged to nuclear families and 70.6% belonged to overcrowded households."

Page 7; Line 8: RR 1.71 does not mean, " The risk of enteric fever was 71% higher in children belonging to households eating food from outside once a week or more". It means, "The risk of enteric fever was 1.71 times more in children belonging to households eating food from outside once a week or more.

We have rephrased the statement.

Results, Page 7, Lines 9-10

"The risk of enteric fever was 1.71 times more in children belonging to households eating food from outside once a week or more (adjusted RR 1.71, 95% CI 1.00 to 2.94)."

Page 7; Line 13-14: RR 1.32 means, "The risk of enteric fever was 1.32 times higher in children living in households sharing toilets.

We have rephrased the statement.

Results, Page 7, Lines 13-14

"The risk of enteric fever was 1.32 times higher in children living in households sharing toilets (adjusted RR 1.32, 95% CI 0.76 to 2.30)...."

The authors need to reference more recent data on global burden of enteric fever, as they have the region with the highest burden of disease incorrect. Certain interpretations on datasets cannot be made as I have mentioned above.

The paper does not add anything new with what we know of enteric fever and how they are interrelated with food and water. However, I do accept that information needs to be updated from time to time, but how often is too often and how important is this publication? That is subjective.

Thank you for your comment. We have added the available recent references on enteric fever burden as suggested.

We have rephrased the language of interpretation of RR.

As mentioned in our response to the previous reviewer, we have modified the rationale of the study in the introduction section to clarify why it is important to study the association of WASH factors with enteric fever in the current scenario.

Reviewer: 3
Dr. Sina Aziz

Not clear how it was decided that 50% of the population of India does not have access microbial safe water

This assumption was based on data from the following source-

Published literature by Bandopadhyay et al in 1992 showed that ~43% of the drinking water samples collected were contaminated with coliforms in Delhi. (Bandyopadhyay S, Banerjee K, Khanna KK, Sharma RS, Verghese T. Drinking water quality and diarrhoea in Delhi. J Commun Dis. 1992 Sep;24(3):156-8. PMID: 1344946.)

We have added the reference.

Methods, Data Collection, Page 5, lines 23-25

“The sample size for this exercise was calculated assuming that 50% households in the community would have access to microbially safe drinking water with a relative allowable error of 20%.19”

A reference for the GIS used should be added

Thank you. We have added the reference.

Methods, Data variables and analysis, Page 6, Lines 12-14

“The cases of Typhoid and Paratyphoid along with the households with contaminated drinking water were plotted using ArcGIS 10.8 on a Google earth base map of July 2021 for the study area.20”

About, 13.7% of the kids were vaccinated by the government program, would be interesting to know why the remaining had not been vaccinated, as it is a free vaccination program and available to all?

Thank you for the comment. Though this is beyond the scope of this article, we agree with the reviewer that it is interesting to know why the coverage was low.

The typhoid vaccine was only rolled out in the state of Delhi in India. Beyond issues in supply, it is possible that children may have missed the dose at 2 years of age, given the fact that this is a migratory population and maybe because that there are no other vaccines to be administered at that particular age (if the booster for DPT, measles, OPV are already administered earlier).

It should be added as to how many children had both E. coli and enteric fever co-infection

The E coli concentration was determined in the household drinking water samples to see the contamination. Among the households with and without contaminated drinking water 6.25 % (1/16) and 1.1 % (1/92) had cases of enteric fever, respectively.

Title was " Association of water, sanitation, hygiene and food practices with enteric fever in a pediatric cohort in North India" why has E.coli been added, please add a few lines in the rationale

One of our primary objectives was to study association of water, sanitary and hygiene (WASH) factors with enteric fever.

One of the many WASH parameters studied was microbiological quality of water, which was evaluated by measuring the most probable number of thermotolerant coliform in the water (as per definition of WHO). We have mentioned in the methods that "We collected drinking water samples (1.5 L) from the drinking water storage point in these households to test for the most probable number (MPN) of coliform per 100 mL, pH, and total dissolved solids."

Add a definition of "pucca house" for all the readers

Thank you, we have added the definition as a footnote of the table.

Page 11, Lines 7-8

"Pucca house is a term that refers to housing in South Asia built of substantial material such as stone, brick, cement, concrete or timber."

What was the filtration method used by the family? What is its sensitivity and specificity? Add with a reference

19.7% (1166) of the households used a candle filter while 10.9% (643) households used an electric water purifier. We have added this as a footnote to Table 2.

What is meant by adequate water treatment, please add with a definition

Adequate water treatment was defined as treatment of water by boiling, bleach/chlorine or filtration. The definition is mentioned in the data variables and analysis section.

Methods, Data variables and analysis, Page 6, Lines 1-2

Adequate water treatment was defined as treatment of water by boiling, bleach/chlorine or filtration.

In fig add what DAG abbreviation stands for.

Thank you we have added the full form of DAG in the figure.

Figure 1

Reviewer: 4
Dr. Mila Shakya

"Enteric fever" and "typhoid fever" have been capitalized throughout the paper, not a proper noun, please correct.

Thank you for the comment, we have modified this.

Methods:

All enrollees are children under the age of 16 years. The authors mention consent for participation form children. Please clarify.

We thank the reviewer for their comment. Kindly note we have mentioned that consent was taken from the parents of the eligible children while assent was taken from children > 7 years of age. This has been mentioned in the ethical clearance section.

Methods, Ethical Clearance, Page 6, Line 22-24

"Written informed consent was obtained from the parents and assent was obtained from children > 7 years of age, before the interview."

The authors mention weekly follow-up. Although the authors refer to a previously published protocol, can the authors mention how and why the weekly follow-ups were done?

We thank the reviewer for their comment.

We conducted active weekly contacts either by phone or by home visits, with ≥ 1 face-to-face contact every 4 weeks to collect information on fever, illness, or hospitalizations. In addition, a monthly mobile recharge of 50 Indian rupees was provided to all enrolled families to encourage early reporting of fever to the study team, by telephone. Quality checks were done by the study supervisor if contact with primary caregiver was not made for 2 consecutive weeks. If a child was identified to have a

febrile illness in the period since the last contact a Fever Case Report Form (CRF) was filled. The interval for the follow-up visits was kept at 1 week to reduce the chance of missing out a fever episode and more importantly to improve sensitivity of blood culture for timely diagnosis of enteric fever.

Were the fever cases identified in the weekly follow-ups? If yes, does that potentially mean that febrile cases may have gone unidentified?

We thank the reviewer for their comment.

The fevers were identified through weekly follow-up as mentioned above. We used a combination of active weekly contacts along with incentivization in form of monthly phone recharge for promoting early passive reporting of fever. However, given that the median (IQR) duration of a fever episode is 2(1–4) days, a biweekly contact would have allowed better capture of fever episodes at the cost of operational feasibility. With our approach we may have missed a few fever episodes because many of these households have a common mobile phone, not always available for the primary caregiver of the child to report fever. We expect that the number of missed fever episodes would be very less.

These details are published earlier.

Bireshwar Sinha, Temsunaro Rongsen-Chandola, Nidhi Goyal, Alok Arya, Chandra Mohan Kumar, Aparna Chakravarty, Mohammed Aslam, Deepak More, SEFI tier 1 collaborators, Incidence of Enteric Fever in a Pediatric Cohort in North India: Comparison with Estimates from 20 Years Earlier, *The Journal of Infectious Diseases*, Volume 224, Issue Supplement_5, 15 November 2021, Pages S558–S567, <https://doi.org/10.1093/infdis/jiab046>

It is unclear who the respondents for the WASH / food practices questionnaires are. Can this lead to misclassification?

Thank you.

The WASH practices were recorded by the study personnel during home visits by direct observation and interview schedules. The interview schedules were administered to an adult member of the household. The personnel were well trained in the assessment as per the WHO guidelines for assessment of WASH practices. Multiple quality checks were a part of the process which included cross checking of the findings by field supervisors. Therefore, the likelihood of misclassification very less for our study.

Result

Please include the p-values for the ORs and RRs.

Thank you, we have added the p values in the table.

Page 13, Table 3

Discussion

Can the authors elaborate on what the findings of the geospatial analysis mentioned in page 8 lines 14 - 17 could imply?

We thank the reviewer for their comment.

We wanted to ask a question – Does a child living-in an area nearby to households with contaminated drinking water compared to those living-in areas without contaminated drinking water have higher risk/odds of enteric fever?

The findings imply that the odds of having a case of enteric fever in a child is higher if they are living-in a house closer to households with contaminated drinking water compared to households without contaminated drinking water.

Page 8 lines 40 - 45, reference for systematic review missing. Also not clear what "this review" in line 47 is referring to.

Thank you. We have added the reference and reframed the statement.

Discussion, Page 8, Lines 32-38

“A systematic review by Brockett S et al reported that an improved water source (16 studies, OR = 0.73, 95% CI = 0.56–0.95) and treatment of water before consumption (9 studies OR = 0.59, 95% CI = 0.45–0.75) are associated with significantly lower odds of typhoid fever. Similar to our findings, it was reported that food and drink consumed outside home (39 studies) was significantly associated with higher odds of typhoid (OR = 1.6, 95% CI = 1.4–1.8).⁵ The study also reported that that ice cream consumption was significantly associated with higher odds of Typhoid (10 studies, OR = 1.5, 95% CI = 1.2–1.9, I² = 31%).⁵”

Page 8 lines 58 - 60 - how does the study show the importance of active case finding for typhoid fever? Is active case finding done programmatically anywhere that has helped reduce the burden?

Thank you. We have removed the statement.

The authors claim association but a lack of precision in the first paragraph of the discussion. The authors mention low sample size as a limitation in the second paragraph. However, it is not just lack of precision or wide CI. Of note, all (except one) 95% CI of RRs /ORs presented in the result section contain the value 1, which would suggest there is no association. Please rephrase your statement in the discussion.

Thank you, we have rephrased the statement.

Discussion, Page 8, Lines 4-8

“Our findings suggest that consumption of safe drinking water and safe sanitation practices were associated with a lower risk of enteric fever while sharing of toilets and eating food from outside were associated with a higher risk of enteric fever. However, the precision for some of these associations were wide and the association is non significant”

Page 9 Line 19, please specify what other factors.

Thank you

We have detailed the statement.

Discussion, Page 9, Lines 9-11

“The findings also suggest that apart from WASH factors there may be several other factors that affect the transmission dynamics for eg demographic profile, socioeconomic status and Typhoid vaccine coverage.”

VERSION 2 – REVIEW

REVIEWER	Reviewer name: Dr. Sarah Nevitt Institution and Country: University of Liverpool, United Kingdom of Great Britain and Northern Ireland Competing interests: None
REVIEW RETURNED	29-Mar-2022

GENERAL COMMENTS	Thank you to the authors for their responses and for edits to their manuscript. All of my comments have been suitably addressed.
-------------------------	---

REVIEWER	Reviewer name: Institution and Country: Competing interests:
REVIEW RETURNED	

GENERAL COMMENTS	
-------------------------	--

VERSION 2 – AUTHOR RESPONSE

VERSION 3 – REVIEW

REVIEWER	Reviewer name: Institution and Country: Competing interests:
REVIEW RETURNED	

GENERAL COMMENTS	
-------------------------	--

REVIEWER	Reviewer name: Institution and Country: Competing interests:
REVIEW RETURNED	

GENERAL COMMENTS	
-------------------------	--

REVIEWER	Reviewer name: Institution and Country: Competing interests:
REVIEW RETURNED	

GENERAL COMMENTS	
-------------------------	--

VERSION 3 – AUTHOR RESPONSE