

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

TITLE (PROVISIONAL)	Environmental and behavioral exposure pathways associated with diarrhea and enteric pathogen detection in five month old, peri-urban Kenyan infants: a cross-sectional study
AUTHORS	Baker, Kelly; Mumma, Jane; Simiyu, Sheillah; Sewell, Daniel; Tsai, Kevin; Anderson, John; MacDougall, Amy; Dreibelbis, Robert; Cumming, Oliver

VERSION 1 – REVIEW

REVIEWER	Page, Nicola National Institute for Communicable Diseases, Centre for Enteric Diseases
REVIEW RETURNED	05-May-2022

GENERAL COMMENTS	<p>The paper describes an assessment of various environmental and behavioral factors in the exposure pathways (sanitation and animal management) that may influence diarrhea and enteric pathogen detection in young children in Kisumu, Kenya. The study analyzed baseline data from six-month old infants enrolled onto the Safe Start cluster-randomized controlled trial of infant food hygiene behavior change intervention conducted in Kisumu, Kenya in March 2018 – November 2019.</p> <p>Assistance with review of the statistical methods provided by Siobhan Johnstone.</p> <p>The investigations probing exposure pathways in low-income settings are valuable to design appropriate interventions. The following concerns were noted:</p> <p>The title is a little confusing – is it 20 six week old or 26 week old? Perhaps just revise for clarity.</p> <p>Line 18: Reconsider the word “improved” as the aim is to assess then exposure pathways without a predetermined outcome.</p> <p>Line 70: “...will consistently use it for ...”</p> <p>Line 80: “At a community scale, ...”</p> <p>Line 82, 83: Repeat of reference – delete 19.</p> <p>Line 84: Reconsider the word “modifying” as the aim is to assess then exposure pathways without a predetermined outcome. Not all of the factors had a modifying role in the pathway.</p> <p>Line 87: “...animal fecal sources.”</p> <p>Line 93: “...between fecal sources...”</p> <p>Line 113: “...and CHVs provided input...”</p> <p>Line 132: “...within five hours of collection from the household.”</p> <p>Line 141: Why were only half of the specimens spiked with MS2? Were more intervention/control arm participant specimens assessed in this manner or was it evenly split between the two groups? How were the results for the unspiked specimens dealt with/analysed? This needs to be listed as a study limitation.</p> <p>Line 150: Clostridioides difficile</p>
-------------------------	--

	<p>Line 165: Please provide reference</p> <p>Line 167: The manner in which the authors have structure the data analysis section is confusing, making it hard to follow the analysis strategy employed for the study. The authors explain independent variables for point sources of feces and environmental conditions but are not clear on what analysis was done for these.</p> <p>Line 173: Some of the variables described as “intermediate pathogen exposure pathways” don’t seem to be intermediate at all e.g. water sources, source of infant feeding etc.</p> <p>Line 186: Not sure why rotavirus vaccination status is a confounder?</p> <p>Lines 191-198: The process for designing the model is a little unclear; as written now it appears as if the authors added all explanatory variables and confounders into one. Please clarify</p> <p>Line 200: “... resulting in the exclusion of some redundant exposure variables.” Please specify the variables that were excluded to clarify for the reader.</p> <p>Line 201: “In addition, some confounder variables had had insufficient variability to have estimable effects and hence were removed.” This sentence is a little unclear. Were these variables found to be non-significant in the multivariable model? If so, this should be explained in the modelling step (which says they included all variables).</p> <p>Line 202: The authors excluded breastfeeding from the model, as there were very few that did not breastfeed. This is still an important risk factor for childhood diarrhoea and I’m not sure I agree with completely excluding it. Also if this variable is excluded it should not be listed as one of the included variables in the model (as it is currently).</p> <p>Lines 210 – 252: I’m struggling to understand the hypothesis testing using interaction terms. This seems unnecessary – if authors build a robust model to see which factors are associated with diarrhoea, most of these things can be determined from the model without the need for an additional step.</p> <p>Line 216: The term human sanitation is a little unclear. Do the authors mean latrine access and type of latrine? Were these two exposure variable combined for a sanitation exposure variable?</p> <p>Line 222: “Association between ...”</p> <p>Lines 263-279: There seems to be a difference in the numbers reported in the text and the numbers shown in Table 1. Please recheck and correct.</p> <p>Line 265: Preterm birth rate percentage in text 17% vs 16.0% in table 1.</p> <p>Line 268: “...a majority of ...”</p> <p>Line 270: Household who owned animals 13.9% in text vs 12.8% in table 1.</p> <p>Line 270: Animals slept inside house 11.7% in text vs 10.5% in table 1.</p> <p>Line 276: Hand wash station present in 6.8% of homes in text vs 6.6% in table 1.</p> <p>Line 270: Hand washing after touching animals in 12.*% in text vs 12.0% in table 1.</p> <p>Table 1 - please label column 2 i.e. percentage (n)</p> <p>Line 300: This sentence is unclear as washing hands after self-defecation and washing before infant feeding have been reported above (lines 298-300) and remain significant in the adjusted model. Using improved latrines is no longer significant after adjusting for other WASH factors. Please revise.</p> <p>Line 322 – Please indicate Table 3.</p> <p>Line 326 – Spacing in table 3 is incorrect. Please adjust.</p>
--	---

	<p>Line 438: "...6 months of age."</p> <p>Figure 1 – please indicate what different shaded lines in figure represent.</p> <p>Figure 2 – Please label figure properly, especially boxes to the left of pathogen count which presumably contain ORs with 95% CI. Alternatively, indicate in figure heading what numbers in diagram are.</p> <p>Supplementary tables 1 and 2 – please indicate what bold numbers in the tables mean either below table (as done for Tables 1-3) or in table heading.</p>
--	---

REVIEWER	Havelaar, AH University of Florida, Emerging Pathogens Institute
REVIEW RETURNED	16-May-2022

GENERAL COMMENTS	<p>The authors present an analysis of baseline data during enrollment for a Randomized Controlled Trial on the impact of improved sanitation and animal management in Kenya. The outcome variable is caretaker-reported diarrhea and the author associate this with major risk factors and then study the interaction between significant risk factors and 7 putative effect modifiers. This is an interesting design that takes the complex nature of the transmission of enteric pathogens in the household environment into account. The study is clearly described, and the conclusions are of interest. I recommend accepting the manuscript for publication, considering the following comments.</p> <p>Line 60: should "other" be inserted before "early childhood infections"? Please provide some examples of infections for which this has been established and references.</p> <p>Line 80: scale, not scales.</p> <p>Line 86 and elsewhere: there are confusing statements about the age of the enrolled children. Here 6 months is mentioned; in line 107 it is under six months of age, in line 109 22 +/- 1 week. Please clarify what was the target age at enrollment, and what was achieved upon completion of enrollment.</p> <p>Line 104: please write out abbreviations.</p> <p>Line 165: please provide a reference</p> <p>Line 169: were animals kept in the home all day or only at night?</p> <p>Line 235: how is food-related handwashing differentiated from overall handwashing?</p> <p>Line 276: there appears to be a major difference between observed presence of hand-washing stations and self-reported handwashing. What has been done to understand whether self-reported behavior was biased by social desirability? Have the authors observed handwashing behavior or otherwise probed deeper? See also lines 462 and 463 where it is stated that enumerators verified conditions. These verifications do not seem to have been taken into account.</p> <p>Lines 283-284: the sentence on confounding is confusing.</p> <p>Line 285: Nearly perfect is a strong statement in the face of a 14% difference.</p> <p>Table 1: please provide more details on the animal species owned. Was sleeping inside household related to species, in the text not only rodent droppings but also visible presence of rodents was mentioned.</p> <p>Line 303: not clear what the statement on other WASH variables is based on.</p> <p>Table 2: lower confidence bound for fresh packed milk is rounded to three significant digits, all other results to 2. Rounding to 1.00</p>
-------------------------	--

	<p>would make the result significant, just as UHT milk. The point estimate for fresh packed milk is higher than for UHT milk, so the decision to declare this not significant is debatable.</p> <p>Line 317: a threshold of Ct = 35 is quite high and may lead to false positives. Suggest doing a sensitivity analysis with Ct = 30, or Ct = 35 AND < mean + 1 standard deviation of negative controls.</p> <p>Line 325: keeping or owning?</p> <p>Table 3: define higher vs. lower number of enteric pathogens. It would also be interesting to compare the results of the ordinal logistic regression with standard logistic regression in which the number of pathogens is an integer. Compressing high numbers to 6 would not be highly influential.</p> <p>Line 340: here the term pathogen count suggests an integer number.</p> <p>Line 343: I would not consider p-values between 0.01 and 0.05 as strong evidence.</p> <p>Line 393: add processed before cow milk?</p> <p>Line 446-447: this statement cannot be deduced from the cited reference.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 Comments to the Author:

The title is a little confusing – is it 20 six week old or 26 week old? Perhaps just revise for clarity.

Response: The title and throughout has been revised to clarify “5 month old”, with the exception of a few locations where we say < 6 months and discuss our population as still being within the window of age where exclusive breastfeeding is promoted.

Line 18: Reconsider the word “improved” as the aim is to assess then exposure pathways without a predetermined outcome.

Response: The word “improved” has been deleted.

Line 70: “...will consistently use it for ...”

Response: The word “it” has been added to line 79.

Line 80: “At a community scale, ...”

Response: “Scales” has been revised to “scale”.

Line 82, 83: Repeat of reference – delete 19.

Response: The repetitive Harris et al. citation has been removed.

Line 84: Reconsider the word “modifying” as the aim is to assess then exposure pathways without a predetermined outcome. Not all of the factors had a modifying role in the pathway.

Response: The word “modifying” is a common way of discussing statistical analysis of effect modification. In our case it refers to our pre specified stated objective of examining whether intermediation conditions are effect modifiers of pathogen source-outcome relationships. You are

correct that not all, most in fact, of our hypothesized effect modifiers did not meet statistical criteria of significance for modifying source-outcome relationships. However, the results of analytical tests should not influence how we describe the stated research purpose of the study in the introduction. We have modified this sentence slightly to use “effect modification” to address the reviewer’s point.

Line 87: “...animal fecal sources.”

Response: Line 87 has been modified to state “animal fecal contamination sources”.

Line 93: “...between fecal sources...”

Response: Line 93 has been modified to use “fecal sources”.

Line 113: “...and CHVs provided input...”

Response: “Community health workers” has been converted into “CHVs”.

Line 132: “...within five hours of collection from the household.”

Response: We have added “from the household” to line 132.

Line 141: Why were only half of the specimens spiked with MS2? Were more intervention/control arm participant specimens assessed in this manner or was it evenly split between the two groups? How were the results for the unspiked specimens dealt with/analysed? This needs to be listed as a study limitation.

Response: Only half of the specimens were spiked with MS2 for several reasons. First, the Taqman Array Cards that we use have a built in PCR process control that allows us to screen for broad-sample polymerase problems and sample-level evidence of inhibition. The purpose of the MS2 spike is purely for surveillance of issues posed by failures in storage and transport conditions among a group of samples as they make their way from freezers in Kenya to Iowa, and evaluation of extraction efficiency. A small group of samples with false negative MS2 is adequate to indicate cold chain problems, and minimizing how much unnecessary work required of our technicians reduces the chances that they will make mistakes. As far as evaluating extraction efficiency, our prior work with specimen preservation and extraction protocol indicated high extraction efficiency with low molecular inhibition issues using these Zymo products. Therefore, we determined a random selection of samples each week for spiking was adequate for quality control surveillance. There was no difference in number of intervention and control specimens spiked. Had we found false negative MS2 results, we would have consulted the PCR process control and if evidence of inhibition was detected, then repeated PCR with a 1:5 DNA/RNA dilution. If no inhibition was detected, then we would have examined whether other false negative MS2 issues were occurring among other specimens in that same transport group and either performed confirmation testing or asked the lab in Kenya to revert to the frozen stool specimen and resample for a second extraction. We had no false negative MS2 detection so are confident that there are no sample transport failures to acknowledge. We have added a sentence to limitations to acknowledge that we cannot assess variability in extraction efficiency for all individual samples.

Line 150: Clostridioides difficile

Response: This correction has been made.

Line 165: Please provide reference

Reference: The reference to Liu et al. 2013 has been added.

Line 167: The manner in which the authors have structure the data analysis section is confusing, making it hard to follow the analysis strategy employed for the study. The authors explain independent variables for point sources of feces and environmental conditions but are not clear on what analysis was done for these.

Response: Our analysis section was structured to first describe variables and our hypotheses about where they fit in a source to infant transmission system, followed by a paragraph describing our software, the type of model used to model bivariate vs categorical outcomes, and clarifying that we ran bivariate and then fully adjusted models for all point source, non-point source, and intermediate variables first. Our analysis section then transitions to describe adding terms that tested for pre-specified effect modification relationships in these fully adjusted models. We cannot describe models before variable definition and cannot describe effect modification before stating what variables are in the models that modification terms are added to. So, we are uncertain how to address this comment without a more detailed recommendation from the reviewer.

Line 173: Some of the variables described as “intermediate pathogen exposure pathways” don’t seem to be intermediate at all e.g. water sources, source of infant feeding etc.

Response: In our analysis section where we explain each of our theoretical hypotheses about pathogen sources and intermediate effect modifiers, Hypothesis 6 and 7 agree with this point that food and water sources external to the household could be contaminated by pathogens from non-household sanitary issues. These sections also explain what household food and water conditions could alter pathogen presence or absence conditions prior to infant feeding or drinking. We also display this in our conceptual Figure places boxes for cow milk sources – a leading supplemental food type in Kenya – on the far left side in alignment with human and animal sanitation conditions, to represent farm or food production sources of contamination, with arrows to pathogen outcomes in infants flowing through “Food Preparation” steps alongside food-related hand hygiene. Line 173 and the associated paragraph was inconsistent with these other discussions on food and water sources and pathogen sources, but has been edited to improve clarity.

Line 186: Not sure why rotavirus vaccination status is a confounder?

Response: Rotavirus vaccination status was included as a potential confounder as it is plausibly associated with independent and dependant variables of interest and not on the causal pathway. More specifically, we included RV vaccination because of concerns that non-vaccination would be correlated with other poor hygiene conditions (caregivers who do not take infants to get vaccine more likely to be poor or have poor health risk perception behaviors) and with the pathogen detection outcome in infants.

Lines 191-198: The process for designing the model is a little unclear; as written now it appears as if the authors added all explanatory variables and confounders into one. Please clarify Line 200: “...

resulting in the exclusion of some redundant exposure variables.” Please specify the variables that were excluded to clarify for the reader.

Response: This is correct, we built bivariate models for each variable, then adjusted as one with all explanatory and confounder variables that could support parameterization of such a model. Fitting the model and the information learned about variable distributions and relationships is a result of the pre-specified methodological process. Therefore, variables excluded for collinearity are reported in the second paragraph of results.

Line 201: “In addition, some confounder variables had had insufficient variability to have estimable effects and hence were removed.” This sentence is a little unclear. Were these variables found to be non-significant in the multivariable model? If so, this should be explained in the modelling step (which says they included all variables).

Response: Excluding a variable with insufficient variability means that the data distribution for a variable of interest is not analyzable. For example, with the variable “still breastfeeding”, 99.7% said yes while 0.3% said no. In a bivariate model with no other adjustments, you can technically estimate an effect for this variable with the outcome because you are really only looking at data in a 2 by 2 fashion, in the case of still breastfeeding our “no” group only had 14 participants, two of which had an infant with diarrhea and 12 which did. Adding a third variable to this bivariate model, let’s take flooring (3 categories) for example, now requires that our groups must be split by a third level to accommodate 3 x 3 structure. If we only have 2 caregivers in the group with no breastfeeding and diarrhea, we cannot break this group out according to type of flooring without winding up with a zero in one of the sub-groups. This overparametrization of the model results in quasi or complete separation and maximum likelihood estimates simply do not exist. The data distributions does not support the complexity of that specified multivariable model and must be excluded from the model or - if it is the primary variable of interest (still breastfeeding was only a confounder for us) - then you report bivariate analysis as the best the data permits. As with the prior response, variables excluded for insufficient data after vetting distribution variability and collinearity are reported in the second paragraph of results.

Line 202: The authors excluded breastfeeding from the model, as there were very few that did not breastfeed. This is still an important risk factor for childhood diarrhoea and I’m not sure I agree with completely excluding it. Also if this variable is excluded it should not be listed as one of the included variables in the model (as it is currently).

Response: We believe strongly that transparency in scientific reporting requires reporting the original research design used to implement a study, and then clarifying why the research was not implemented as intended, rather than omitting information about steps that were intended but not implemented. In the case of breastfeeding, excluding reference to this as a factor we considered and tried to model may alarm readers in this field as to why we did not measure and adjust for breastfeeding, and call into question our understanding of conditions that could influence enteric infections in children. The breastfeeding variables cannot be included because the lack of variability in responses results in non-convergence of the model.

Lines 210 – 252: I’m struggling to understand the hypothesis testing using interaction terms. This seems unnecessary – if authors build a robust model to see which factors are associated with

diarrhoea, most of these things can be determined from the model without the need for an additional step.

Response: You cannot detect and measure effect modification, or differences in exposure-outcome effects by sub-group, by simply including variables in a model. You can derive averaged exposure-outcome effects between groups, but not individual group-level effects. Effect modification must be tested by examining whether interaction between two variables is associated with the outcome. If detected and disparities between groups is important to address the research question, then you report sub-group level effects.

Line 216: The term human sanitation is a little unclear. Do the authors mean latrine access and type of latrine? Were these two exposure variable combined for a sanitation exposure variable?

Response: This term was unclear. We have revised this term here, as well as at the beginning of Hypothesis 1 to state “Lack of access to a latrine with a barrier between users and excreta, and/or sharing unclean latrines with others”, which reflects the two types of sanitation data we collected and reported: any access to a latrine of improved design and whether that latrine was shared by multiple households.

Line 222: “Association between ...”

Response: This now reads “An association between...”.

Lines 263-279: There seems to be a difference in the numbers reported in the text and the numbers shown in Table 1. Please recheck and correct.

Line 265: Preterm birth rate percentage in text 17% vs 16.0% in table 1.

Line 270: Household who owned animals 13.9% in text vs 12.8% in table 1.

Line 270: Animals slept inside house 11.7% in text vs 10.5% in table 1.

Line 276: Hand wash station present in 6.8% of homes in text vs 6.6% in table 1.

Line 270: Hand washing after touching animals in 12.*% in text vs 12.0% in table 1.

Response: Thank you for catching these mistakes. The text has been modified to match the accurate data in Table 1.

Line 268: “...a majority of ...”

Response: Line 268 in the prior version read “used secondary water sources and nearly all households shared their latrine with multiple other” and after carefully re-reading this paragraph, we cannot infer what needs editing.

Table 1 - please label column 2 i.e. percentage (n)

Response: Table 1 column 2 is now labeled with percent (n).

Line 300: This sentence is unclear as washing hands after self-defecation and washing before infant feeding have been reported above (lines 298-300) and remain significant in the adjusted model. Using improved latrines is no longer significant after adjusting for other WASH factors. Please revise.

Response: The second line about latrines was reporting a bivariate association between latrines and diarrhea that was explained in some part (both effect size and consistency of the association) by accounting for other variables in the fully adjusted model. This is a useful observation because it supports the interaction effects we find between latrines and handwashing in the effect modification tests.

Line 322 – Please indicate Table 3.

Response: The reference to Table 3 has been added.

Line 326 – Spacing in table 3 is incorrect. Please adjust.

Response: We are not completely clear on this comment but think perhaps it relates to indentation for handwashing at critical times, where the reported odds appeared to be on a different level of indentation than the variable name. This should remain appropriately indented if the journal formats this table such that single indent lines are shown as such, but for the sake of presentation here, we have added an extra line of indentation to the lines under the second and third column. Please clarify if we have missed the point completely.

Line 438: "...6 months of age."

Response: This has been edited to 5 months age, which is defined as 22 weeks +/- 1 week in methods.

Figure 1 – please indicate what different shaded lines in figure represent.

Response: We have edited the figure legend to indicate that the circle indicates mean odds ratio estimate and the shaded lines reflect the 95% confidence interval in that estimate.

Figure 2 – Please label figure properly, especially boxes to the left of pathogen count which presumably contain ORs with 95% CI. Alternatively, indicate in figure heading what numbers in diagram are.

Response: We have expanded the Figure legend to include much more detail about shading, bolding, and dashing patterns, as well as the mean of the numbers.

Supplementary tables 1 and 2 – please indicate what bold numbers in the tables mean either below table (as done for Tables 1-3) or in table heading.

Response: We have added text clarifying that bolding refers to statistically significant associations, as for Tables 2 and 3.

Reviewer: 2 Comments to the Author:

The authors present an analysis of baseline data during enrollment for a Randomized Controlled Trial on the impact of improved sanitation and animal management in Kenya. The outcome variable is caretaker-reported diarrhea and the author associate this with major risk factors and then study the interaction between significant risk factors and 7 putative effect modifiers. This is an interesting design that takes the complex nature of the transmission of enteric pathogens in the household environment into account. The study is clearly described, and the conclusions are of interest. I recommend accepting the manuscript for publication, considering the following comments.

Line 60: should “other” be inserted before “early childhood infections”? Please provide some examples of infections for which this has been established and references.

Response: The word “other” has been inserted at the specified location and the last sentence has been modified to clarify how enteric infections increase vulnerability to re-infection/incidence of infection through enteric dysfunction and malnutrition, with citation.

Line 80: scale, not scales.

Response: The “s” has been removed from this word.

Line 86 and elsewhere: there are confusing statements about the age of the enrolled children. Here 6 months is mentioned; in line 107 it is under six months of age, in line 109 22 +/- 1 week. Please clarify what was the target age at enrollment, and what was achieved upon completion of enrollment.

Response: thank you for pointing out this inconsistency. We have edited to clarify in abstract and methods that enrollment age was 5 months (22 weeks +/- 1 week), with the exception of a few locations in introduction and conclusion where we retain “less than six months” to discuss the contributions of this study as far as risk factors among very young infants. In the case of cow milk, we reference these infants being younger than the six month benchmark considered to be a common transition from breastfeeding.

Line 104: please write out abbreviations.

Response: The abbreviations have been written in completion.

Line 165: please provide a reference

Response: A reference to Liu et. Al. 2013 has been added.

Line 169: were animals kept in the home all day or only at night?

Response: Our survey question asked about general animal keeping practices and unfortunately, we lack study data about variation over the course of the day. We added the word “typically” to characterize the generality of responses.

Line 235: how is food-related handwashing differentiated from overall handwashing?

Response: This sentence stated “We hypothesize that food-related handwashing also moderates the association between feces sources (latrine access and the presence of animals or rodents in the household) and pathogens in infants such that caregivers washing hands before preparing food and eating or feeding the infant decreases the association between latrines, domestic animals, and rodents with pathogens in infants, compared to caregivers who do not wash hands at these times.” We have expanded the three underlined handwashing behaviors in the text and numbered them 1 through 3 to better highlight what food-related handwashing means and that these are three different behaviors.

Line 276: there appears to be a major difference between observed presence of hand-washing stations and self-reported handwashing. What has been done to understand whether self-reported behavior was biased by social desirability? Have the authors observed handwashing behavior or otherwise probed deeper? See also lines 462 and 463 where it is stated that enumerators verified conditions. These verifications do not seem to have been taken into account.

Response: We did not conduct research on potential response bias but agree that based upon our discrepancies and literature on bias in self-reported handwashing, exposure misclassification is likely quite high. For this reason we also used observed validation of handwashing stations (lines 462 and 463). We have edited the text to clarify that many more households indicated they had handwashing stations in the household, but the presence of soap and water required for that handwashing was rare. This may not be due purely to courtesy bias. Qualitative evidence from our study communities has indicates that visual inspections of handwashing stations for soap and water could bias conclusions on handwashing behaviors the opposite direction because households will often store soap in a secure location where it must be requested for use. This occurs for a variety of reasons, including fear of theft, child misuse or loss of soap, and animals eating the soap. Structured observation is another approach, but also one prone to over practice of behaviors due to observation. We do not have sufficient information to conclude from this study, although other ongoing studies by this team are collecting data on physical presence of handwashing materials, self-reported behaviors, and structured observation of caregiver behaviors over time that will let us triangulate the consistency of these different indicators. We have embellished upon our original sentences in the limitations section of the discussion to add more nuance to interpreting data discrepancies.

Lines 283-284: the sentence on confounding is confusing.

Response: We have edited this sentence to better distinguish between the three different ways we asked about breastfeeding and the capacity to use any of those variables to control for breastfeeding influencing exposure-outcome effects.

Line 285: Nearly perfect is a strong statement in the face of a 14% difference.

Response: we have altered this wording to “strongly predicts”, although to clarify from Table 1, 12.8% of households owned animals of which 10.5% kept animals inside the household and 2.3% kept them outside.

Table 1: please provide more details on the animal species owned. Was sleeping inside household related to species, in the text not only rodent droppings but also visible presence of rodents was mentioned.

Response: We have added a sentence to the text and a Supplemental Table reporting housing location by type of animal. We have added detail on the type of evidence of rodents to Table 1.

Line 303: not clear what the statement on other WASH variables is based on.

Response: We have changed “WASH variables” to “exposure pathways”.

Table 2: lower confidence bound for fresh packed milk is rounded to three significant digits, all other results to 2. Rounding to 1.00 would make the result significant, just as UHT milk. The point estimate for fresh packed milk is higher than for UHT milk, so the decision to declare this not significant is debatable.

Response: We have corrected the lower confidence bound in table 2 to 1.00 and bolded it as significant.

Line 317: a threshold of Ct = 35 is quite high and may lead to false positives. Suggest doing a sensitivity analysis with Ct = 30, or Ct = 35 AND < mean + 1 standard deviation of negative controls.

Response: The threshold of Ct=35 is standard practice for presence/absence determination in clinical diagnostics studies, and was used only after screening negative controls for every single pathogen in our analysis to confirm assays did not produce false positives rates below 35. Our lower limit of detection varies slightly by pathogen, but generally a Ct of 35 is equivalent to roughly 10¹ (most of our bacteria ad protozoans) to 10² (viruses) per well on the TAC card, which is well within the range of sensitivity for PCR assays. We are also concerned that setting lower thresholds would result in outcome misclassification where positives would be classified as false negatives. In a study of symptomatic status, loss of low concentration detections might not influence conclusions much – in another paper we are examining the relationship between Ct and self-reported diarrhea. However, here we treat detection as an indicator of infant exposure and detection of low concentrations is very important. Therefore, given these reasons and our methodological validation of false positives between 30 and 35 Ct, we ask the reviewer to not request inclusion of this information. Discrepancies between high versus low concentration shedding and exposures could occur simply due to statistical probability, but could very much complicate the interpretation of study results.

Line 325: keeping or owning?

Response: This has been corrected to “keeping”.

Table 3: define higher vs. lower number of enteric pathogens.

Response: This modeling approach assesses the odds that the pathogen count in one exposure group is higher or lower than the other, so there is no specific cutoff value. We have modified the title of Table 3 to clarify: “Odds ratios (OR) and confidence intervals (CI) from ordinal regression models of detecting one additional enteric pathogen type in feces of five”.

It would also be interesting to compare the results of the ordinal logistic regression with standard logistic regression in which the number of pathogens is an integer. Compressing high numbers to 6 would not be highly influential.

Response: While we appreciate the reviewer’s suggestion, we have decided to refrain from additional analyses for the following reasons. First, the count of unique pathogen types is a sum of non-independent non-identically distributed Bernoulli random variables, and hence it cannot be thought of as a binomial random variable as is assumed in logistic regression. Second, an alternative approach to applying logistic regression would be to dichotomize the data; this, however, requires an arbitrarily selected cutoff (e.g., ≤ 2 pathogen types vs. > 2 pathogen types) which would lose a large amount of information contained in the data. Third, while the data are counts, for the same reasons we cannot consider this count to be a binomial random variable, we cannot consider this to be Poisson or negative binomial random variable, ruling out these common generalized linear models. Treating this as continuous is inappropriate due to low numbers and small number of unique possible values, but also because it is not correct to assume the difference between, say, 0 and 1 pathogen types is the same as between 5 and 6 pathogen types. Ordinal logistic regression is, in our view, the best approach to overcome all the shortcomings of the above mentioned methods.

Line 340: here the term pathogen count suggests an integer number.

Response: That is correct, please see above response.

Line 343: I would not consider p-values between 0.01 and 0.05 as strong evidence.

Response: The word “strongly” has been deleted.

Line 393: add processed before cow milk?

Response: The word “processed” has been added before cow milk.

Line 446-447: this statement cannot be deduced from the cited reference.

Response: This sentence has been revised to emphasize zoonotic transmission potential, without appropriate treatment: “Animal-based foods, like cow milk, can pose a risk for zoonotic enteric pathogen transmission⁴¹ although pasteurization is an effective means for making milk safer.”

VERSION 2 – REVIEW

REVIEWER	Page, Nicola National Institute for Communicable Diseases, Centre for Enteric Diseases
REVIEW RETURNED	08-Sep-2022
GENERAL COMMENTS	The paper describes an assessment of various environmental and behavioral factors in the exposure pathways (sanitation and animal management) that may influence diarrhea and enteric pathogen detection in young children in Kisumu, Kenya. The study analyzed baseline data from six-month old infants enrolled onto the Safe Start cluster-randomized controlled trial of infant food

	<p>hygiene behavior change intervention conducted in Kisumu, Kenya in March 2018 – November 2019.</p> <p>The investigations probing exposure pathways in low-income settings are valuable to design appropriate interventions. The following corrections still need to be implemented</p> <p>Figure 1: Please indicate the difference between the black and grey lines on the figure.</p> <p>Figure 2: Please indicate what the abbreviations Yhwd and Yhwcd mean in the diagram</p>
REVIEWER	Havelaar, AH University of Florida, Emerging Pathogens Institute
REVIEW RETURNED	18-Sep-2022
GENERAL COMMENTS	Thank you for your revision and clarifying statements. Please adjust the titles in your supplementary information to reflect the clarified statements about age of the infants at enrollment.

VERSION 2 – AUTHOR RESPONSE

This is great news, thank you.

Regarding the revision request of Reviewer 1, I have revised the main manuscript figure legend for Figure 1 to clarify "The ORs shown in grey are from the main model with no interaction terms. The ORs in black are from the interaction model." I also revised the legend for Figure 2 to define YHMD and YHMCD: "The term YHWD refers to a sub-group analysis of shared sanitation and pathogen count conditional on whether the caregiver reported washing his/her hands after self-defecation. Similarly, YHWCD refers to a subgroup analysis of shared sanitation and pathogen count conditional on whether the caregiver reported washing his/her hands after cleaning a child that defecated. "

Regarding the revision request of Reviewer 2, I have also revised Supplemental Figures Tables 2 and 3 to replace "26 weeks age" with "five months old" to match the description of age in the main manuscript.

I look forward to reviewing a proof.