Reviewer's report

Title: Water sources as reservoirs of Vibrio cholerae O1 and non-O1 strains in Bepanda, Douala (Cameroon): relationship of isolation and physico-chemical factors.

Version: 4 Date: 17 February 2014

Reviewer: Andrew Azman

Reviewer's report:

This is a re-review of “Water sources as reservoirs of Vibrio cholerae O1 and non-O1 strains in Bepanda, Douala (Cameroon): relationship of isolation and physico-chemical factors.” I have used the old comments as a template and indicate new comments with ** and ***'s.

Major Compulsory Revisions:

1. Study Design/Methods

- The authors say little about how the sampling sites were selected. Did the authors enumerate all wells and taps in the community and then randomly sample from those or was this a convenience sample?

** The authors should be a little more specific about how they were chosen. They now say random but it would be nice to understand how since in many settings like this, the sampling frame is extremely hard to define. If they figured out a clever way to do this it would be useful for readers.

** “Sampling sites were identified using a Global Positioning System (Guangzhou Making Electronic Technology Co., Ltd, China).” This is not a useful addition to the manuscript unless there is a bit more explanation. Were the coordinates for all the wells in the city available and then they used these to locate the randomly selected ones?

- The authors state that samples were collected each month from February to July 2009. Does this mean that one sample from each source was collected once per month (seems like this is not true based on the number of samples)? Were they all collected on the same day each month?

** The reader should be able to figure out which sources were not sampled and when.

** The authors still did not indicate anything about when the samples were collected.

- The authors give insufficient details on the wells, taps, and stream(s) that were sampled. In particular, were the wells protected? What was the origin of the tap water? Was it part of a larger drinking water treatment system that uses chemical disinfectant?
** Please reference the guidelines you are referring to in this statement: “Wells investigated were not protected and did not meet the guidelines for well construction”

** In the letter you say that the wells have been previously described in Akoachere 2013. Please add this citation when describing the wells.

- The authors seem to use the number of isolates from a sample in many of their analyses. Considering that each sample was (1) enriched in APW, (2) then streaked on the TCBS, (3) then suspected colonies were inoculated on HAI slants, I don’t think that using the actual counts of isolates obtained in this manner serves as a reliable (nor validated) quantitative metric. I believe with the methods used, the authors should only count each sample as either cholera positive or cholera negative.

** The authors clarify that the isolates are counted on the TCBS but still don’t explain why the number of isolates is a reliable/meaningful quantitative measure. I still believe that analyses should only be done on the binary outcomes of positive or negative.

** Statements made about the proportion of isolates (e.g. “Among the 96 isolates characterized as V. cholerae, 32 (33.3%) were Vibrio cholerae O1 while the rest 64 (66.7%) belonged to the non-O1/non-O139 serogroup (Table 2).”) really aren’t meaningful and should be reconsidered.

- Since many of these environmental factors are correlated, looking at raw correlations (Kendall tau) alone may not be sufficient. The authors, do perform analyses stratified by season to control for this factor, but should consider using regression models that allows for the inclusion of multiple covariates.

** The authors now use logistic regression models in their manuscript.

** The authors added: “Binary logistic regression model was used to strengthen the associations and identify confounding, bias, and modification.”

*** I am not sure how logistic regression can be “used to strengthen the associations.” Please explain more or revise statement.

*** Did you actually look for effect modification?

*** Did you see evidence of confounding from your models?

*** What biases are you talking about in this statement?

- Previous studies have shown significant lagged associations between environmental variables and clinical cases (e.g. Huq 2005). The same may be true in with environmental isolates. The authors should explore the lagged relationships.

** The authors did not address this comment. Please explain why you chose to explore unlagged relationships only.

2. Results/Discussion

-pg. 7, p2: "Well water samples had the highest occurrence of O1 (64.3%) while
the highest rate of isolation of non-O1/non-O139 strains (100%) was from tap samples." It isn't clear where 64.3% comes from. (also minor spelling error)

** OK but see isolates comment above

- pg 7 p3: "There was no significant difference (#2=2.565, df=2, P=0.277) in isolation of Vibrio cholerae non-O1/non-O139." No significant difference between water sources?

** Sentenced revised but was this done on number of isolates or binary +/- measure? See isolates comment above.

- pg 8 p4: "The highest rate of isolation of 20.8% and 60.4% obtained respectively in the dry and rainy seasons were from well samples (Table 2)." This sentence is unclear.

** OK

pg 8 p1: "Eighty-seven (27.4%) of the 318 samples analyzed were contaminated with V. cholerae." then the next paragraph says "Among the 96 isolates characterized as V. cholerae,...".

** OK but see isolates comment above

pg 10 p4: "The highest rate of isolation of the organism was from wells where 81.3 % of isolates were obtained." This an artifact of the sampling design, right? Also, I don't think you mean rate here.

** Thanks for clarifying but this sentence is still not very meaningful. You have a lot more samples from wells so it isn't surprising that you have more isolates.

pg 11 p1: "The middle and lower parts of the stream which are accessible to the inhabitants were found to be more contaminated than the upper part." How did you quantify/test this?

- It is really hard to understand when and where the cholera was found. The authors should consider revising the supplement by adding the number of cholera positive samples isolated at each source in each month.

** The authors edited the statement but it still isn't sufficiently justified.

Minor Essential Revisions:
- The authors misuse the word "rate" throughout. It is usually used in reference to a proportion.

** OK

- pg 5 p3. You should clarify that while you sampled from a single stream, you sampled at three locations.

** OK

- pg 5. p1 "The main consequence of the use of contaminated water has been the occurrence of several outbreaks of cholera." --> Please reference this statement.
** A single news report isn’t great evidence but this is an understudied area.

- pg 5, p2 "Most cholera outbreaks have started in Bepanda." Please cite this statement.

** This source only refers to the 2004 outbreak starting in Bepanda not any others. Please revise statement or find another source.

- "Samples from the stream were more contaminated (11/21, 52.4%)," more contaminated than what? (plus punctuation typo)

** OK

- pg 12, p3: "Isolation of V. cholerae O1 strains in our study area, likely support the consideration of Bepanda as the starting point of most cholera outbreaks in Douala." - The authors only sampled from Bepanda therefore their findings really can’t support the hypothesis that it serves as a starting point for outbreaks in the city, rather it fails to exclude this as a potential hypothesis.

** OK

- pg 16, p1: "Prompt repairs of broken pipes is highly solicited." This seems out of context. This study mostly found cholera in streams and wells, not piped water systems.

** Thanks for revising this sentence. The new version says that the water “undergoes adequate treatment prior to distribution.” If possible, please cite this.

Discretionary Revisions:

- pg 9, p2: Why drop the p-values from the correlations in this sentence? "pH positively correlated with isolation in the dry season (+0.184) and rainy season (+0.09) but was not significant."

** OK

- The discussion section is unnecessarily long (7 pages), repetitive, and difficult to read. Consider editing down this section.

** OK

- The importance of non-O1/non-O139 in cholera disease either direct through their pathogenicity or indirect through gene transfer is still widely debated. - pg 3, p2: The authors should acknowledge that this is controversial before making statements like: "Thus for an effective diarrheal disease prevention programme in cholera endemic localities, it is important to include the surveillance for non-O1/non-O139 strains in environmental samples."

** OK

- Were there any cholera cases reported during this time period? If so, it would add a lot to this manuscript to connect the environmental isolations to clinical cases.

** The authors note in the letter that no clinical cases occurred during this period. This should be noted in the discussion section.
Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests