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: 3 Date: 9 October 2013

Reviewer: Christopher Troeger

Reviewer’s report:

*General Comments

1. This is an interesting study that investigates an important missing component of cholera in Douala, Cameroon, and Africa in general- how and where does cholera persist in Africa outside of epidemics? The data presented herein provide a compelling piece of evidence that V. cholerae O1 may survive between outbreaks in the urban water supplies. Additionally, there is widespread multi-drug resistance in the isolates. The findings are very important for public health, cholera control and prevention, and guiding policy and action to reduce future cholera outbreaks in Douala. Very nicely done. Following are mostly comments that will help to focus the manuscript and make the findings more easily accessible. There are several comments that must be addressed in the review process but most of the others are minor and/or stylistic advice.

2. A major consideration to start with is that the manuscript is too long, particularly the Discussion. Please refer to the very similar to “Multi-drug resistant toxigenic Vibrio cholerae O1 is persistent in water sources in New Bell-Douala, Cameroon” (Akoachere, 2013). This study and manuscript needs to stand independently of that work but it could provide a guide for how the present paper could be shortened and condensed.

3. The Methods section could be shortened by citing “Standard methods” or “methods described elsewhere” to avoid going into detail about culturing and antibiotic resistance testing. When investigating the association between water temperature, pH, and salinity, consider a simple binomial or logistic regression model. This could be useful in identifying confounding or effect modification (described more below).

4. In the Results section, it is not crucial to report every statistic calculated. If they are included in the Tables, trust the reader to look them up if interested.

5. The Discussion section often feels unstructured. It is not necessary to speculate on every finding from the study. Identify the most pertinent or interesting findings and go into detail about these. For me, the most important findings are that V. cholerae O1 was isolated in the stream (unimproved) and wells samples (an improved water source!) and that a large proportion of isolates were resistant or multi-drug resistant. What does this result mean for people who
rely on well water for their drinking and cooking? For severe cholera treatment? This could have major impacts on cholera outbreak dynamics and control.

6. I’m not so interested in the statistical significance in the association between salinity and Vibrio isolation. The association between salinity and isolation is also mentioned in the Abstract. A large body of literature has made this connection and although a range of salinity (might) be necessary for vibrio survival, it is certainly not sufficient for its existence in the environment. I would also suggest that salinity might be a confounder- the tap water was much less saline, on average, than the stream and well water. How might this affect the association between salinity and vibrio isolation? Either investigate this further or acknowledge this in the discussion of the water samples.

6b. This suggestion is revisited later in the review but consider performing a simple logistic or binomial regression model for the seasonal variation and physico-chemical characteristics of water samples and the isolation of V. cholerae. This would strengthen the associations and help identify confounding, bias, and modification.

7. Proof read the manuscript for typos and grammatical mistakes, there are a modest number of these.

8. Pay careful attention to the use of “Prevalence”. For example, in the 7th paragraph of the Discussion section, “Prevalence was higher in the rainy season (75%) than the dry season (25%).” This statement means that 75% of samples in the rainy season were positive. Be clear that 75% of positive samples came from the rainy season, instead.

9. Be specific with objective and factually stated information. For example, Bepanda is cited several times in the manuscript as the source of cholera outbreaks in Douala. This may be true but provide evidence to back up this statement. Out of how many years? Are there epidemiological investigations? Surveillance data?

*Major Compulsory Revisions

1a. Sample Collection- Paragraph 1- Please include additional information about how the wells and taps were selected. Was there a randomization protocol used? How were the sites established or identified?

1b. Sample Collection Paragraph 1- When sampling from the stream, it is stated that “various locations” were sampled. Were these locations consistent during each sampling period? The sampling times were monthly between February and July, indicating that there should be five separate sample times. Therefore it is unclear why the number of samples presented in Table 1 are not divisible by 5. Were a different number of samples taken each month? Additionally, were the sample sites consistent over time? If different sources were sampled each month, this undermines associations between the season and cholera isolation because it may add a potential confounder that is not accounted for in the analysis. Please clarify on the sampling methods.
2a. Table 1- I am not able to replicate the Chi2 values that are reported in this table. Please explain how they were calculated. In order to investigate if the proportion of samples that were positive for V. cholerae varied by source, my contingency table looks like this:

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well</td>
<td>72</td>
<td>169</td>
<td>241</td>
</tr>
<tr>
<td>Tap</td>
<td>4</td>
<td>52</td>
<td>56</td>
</tr>
<tr>
<td>Stream</td>
<td>11</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>87</td>
<td>231</td>
<td>318</td>
</tr>
</tbody>
</table>

The Chi2 with 2 degrees of freedom for this table is 18.9. That value doesn’t match the Chi2 reported in the manuscript.

2b. Table 2- The same problem exists in replicating the Chi2 values but I think that this is because the total number of samples in each season and source is not included.

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry Season</td>
<td>24</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Rainy Season</td>
<td>72</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Total</td>
<td>96</td>
<td>222</td>
<td>318</td>
</tr>
</tbody>
</table>

*Minor Essential Revisions*

1. Background, Third paragraph- Include a citation for the survival of V. cholerae in association with zooplankton (consider Rita Colwell’s body of work [Colwell RR, Science 1996]).

2. Methods, Study Area, Paragraph 2- “Most cholera outbreaks have started in Bepanda.” Please include a citation for this or at least some quantitative support for this statement. How many is most? 4/5? 2/3? During what time period? The statement is important for the meaningfulness of the study but not specific enough.

3. Results, Contamination of Water, Paragraph 2- Pay careful attention to the specificity of the reported results. “There was a significant difference in isolation of Vibrio cholerae O1 ( ) from samples” This is reporting a significant difference in the source of O1 isolates, right? Rephrase the sentence to reflect this. There is a similar phrasing problem in “Well water samples had the highest occurrence of O1 (64.3%)”. This sentence implies that 64.3% of well water samples were positive for O1 when really 64.3% of O1 isolates were from well water samples. This is an important distinction.

4. Results, Contamination of Water, Paragraph 3- Don’t include the well number, this is distracting UNLESS this helps guide specific policy or public health action. As a reader, the included map provides some context for location but it doesn’t
otherwise factor into the interpretation of the results.

5. Discussion- In general, this section is way too long. Take some time to consider what the main results of the study are and focus on those. It is preferable to be detailed about the most pertinent findings rather than repeating and speculating on every finding from the results. Currently the Discussion is 7 pages and 12 (long) paragraphs. This could reasonably be shortened by being concise, more focused, and removing redundant information.

6. Discussion, Paragraph 1- “Several outbreaks,” how many is several? During what time period?

7. Discussion, Paragraph 1- The results presented do not really “drive the survival” of cholera, more accurately they are associated with the presence of cholera. Is it proposed the vibrios are able to survive for long periods of time in these water bodies, there is constant low-level infection and contamination of the water supply, or a different explanation altogether?

8. Discussion, Paragraph 2- “The highest rate of isolation.” This phrase suggests that the percentage reported will be the percent of wells that were positive for cholera but that is not what is meant. Try “Most of the positive isolates were from wells (81.3%).”

9. Discussion, Paragraph 2- The sentence starting with “Although the rate of contamination…” compares the results in this study with those obtained in rural areas of Islamabad. This isn’t a very good comparison. Presumably, Islamabad and Douala are very different settings. It this comparison is included, more context needs to be provided such as a statement or two about why the comparison is being made. Convince the reader of the validity of the comparison.

10. Discussion, Paragraph 3- There is a statement about “a better adaptability” of non-O1/O139 strains compared to O1/O139 strains. Antibiotic resistance testing does show more resistance in non-O1 strains but the ubiquity of strains, presumably referring to the rate of isolation of O1/non-O1 strains, doesn’t really support the statement. It could be worthwhile to reconsider this statement.

11. Discussion, Paragraph 4- This paragraph explores the absence of O139 isolates. Good. Why, though, does one sentence say that O139 is no public health importance in Africa and the next sentence says that Perez et al. found the strain in Mozambique and that O139 should be included in surveillance even in areas where it hasn’t been reported. Please reconcile this contradiction.

12. Discussion, Paragraph 6- This paragraph describes O1 isolates exhibiting auto-agglutination. This doesn’t really add to the importance or interpretation of the study and could be left out completely. However, if it is decided to keep this discussion point, these results need to be presented and described in the Results section of the manuscript.

13. Discussion, Paragraph 6- “Isolation of V. cholerae… likely support the consideration of Bepanda as the starting point of most cholera outbreaks in
Douala.” This is not very well supported. Be careful establishing a causal relationship. This study does not describe individual outbreaks, does not prove that the outbreaks started in Bepanda, and doesn’t report any cholera incidence or surveillance. This study does show that O1 strains are present in the water in Bepanda and could be a source of infection.

14. Discussion, Paragraph 7- Prevalence of cholera isolates is reported but this is not accurate. Try something more like “75% of isolates came from the rainy season” instead of “Prevalence was higher in the rainy season (75%).” The way that it is currently written suggests that 75% of samples during the rainy season were positive for cholera.

15. Discussion, Paragraph 10- There is a pretty good body of research that shows that V. cholerae prefers relatively high salinity for growth and this is mentioned in this paragraph. However, “Our results therefore indicate that salinity could be one of the major factors influencing the occurrence of V. cholerae,” is not the most accurate representation of the findings. Temperature was significantly correlated with isolation and pH was correlated, even if not strongly, with isolation. For what reason beyond statistical significance is salinity emphasized in the discussion? This is not a new finding- it might be preferable to discuss the salinity correlation in the context of the existing body of research. Is it proposed that salinity is more important than fecal contamination in the water bodies?

16. Discussion, Paragraph 10- Seroconversion is introduced as a concept in this paragraph but is not explored elsewhere, consider not including it in the discussion.

17. Discussion- Dedicate a small paragraph to discussion the limitations of the study, including the lack of data on cholera isolate virulence.

18. Discussion- What are the next steps? Is this part of a larger study of cholera in Douala? How should this information be applied?

19. Data/Tables- In the Table, the Statistics (Chi2, df, p-value) are reported under O1 Isolates, Non-O1 Isolates, and Total Isolates. This implies that they are compared to a common total which doesn’t appear to be the case. It would improve the readability of the table if the statistics corresponded clearly with the columns that they are in. Similarly, the (Percent) reported in the Total Row is 27.4% of the total number of samples for the No. of Samples Positive Column but then the (Percent) reported in the next three columns are of 96. This inconsistency is confusing. I can see why the percent of isolates and samples are different but try to find a way to make that change in fraction reported more obvious.

*Discretionary Revisions

1. Background, First paragraph- sentence starting with “This enteric disease is characterized” implies that hypovolemic shock and renal failure is the clinical course followed by all infections. Consider revising to indicate that this is the
lethal outcome in 30-40% of untreated cases.

2. Background, Third paragraph- First sentence: V. cholerae is autochthonous in aquatic environments but that doesn't explain the occurrence of sporadic outbreaks in non-endemic settings (outbreaks often associated with floods, imported cases [Haiti], infected water supply [Angola], etc). Identifying and explaining V. cholerae isolation in Cameroon, where an autochthonous reservoir is not obvious, is a major goal of the present study.

3. Methods, Study Area, Paragraph 1- Including information about the average annual temperature and humidity in Douala doesn’t add anything to manuscript or results. Consider removing this and other information that is not directly related to the research questions, results, or interpretation of the study.

4. Results, Seasonal Distribution, Paragraph 1- “However, non-O1/non-O139…” This sentence is redundant, it was stated in the previous section.

5. Results, Physico-chemical Characteristics, Paragraph 1, 2, & 3- A small comment but as F-statistics are reported, it would be more consistent to report the range in mean temperatures, pH, and salinities instead of low-high range.

6. Results, Physico-chemical Characteristics- This section would read more easily if it were condensed from 3 paragraphs into 1. In general, refer to Table 3 and report only the most pertinent F-statistics and correlations, specifically those that relate directly to the research questions. There are too many data points reported that are not qualitatively important.

7. Discussion- Many of the paragraphs are too long. Consider breaking them into smaller, more manageable blocks.

8. Discussion- There is a mention, several times, of the broken pipes and inadequate sterilization of wells. This could very well be a hugely important factor in the existence of vibrios in the samples and water supply but it would strengthen this assertion if numbers were used.

8. Table 3- The switch between columnar to row data presentation between the Sample Source and One-Way ANOVA sections is jarring. I don’t have a better way to display it, necessarily, but suggest putting some thought into reformatting the table. As an additional aside, please make every corresponding row and column the same width and chose whether to align the text on the top, middle, or bottom of each cell but keep this decision consistent.

**Overall, great work. This is really important to understanding endemic cholera in urban Africa, especially in an epidemic prone location.

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published
**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.