Prioritizing interventions for cholera control in Kenya, 2015-2020

Waqo Boru\textsuperscript{1,2*}, Shaoming Xiao\textsuperscript{3*}, Patrick Amoth\textsuperscript{1}, David Kareko\textsuperscript{1}, Daniel Langat\textsuperscript{1}, Ian Were\textsuperscript{1}, Mohammad Ali\textsuperscript{2}, David A. Sack\textsuperscript{2}, Elizabeth C. Lee\textsuperscript{4}, Amanda K. Debes\textsuperscript{5**}

\textsuperscript{1}Ministry of Health, Nairobi, Kenya
\textsuperscript{2}Field Epidemiology and Laboratory Training Program, Nairobi, Kenya
\textsuperscript{3}Johns Hopkins University School of Medicine, 200 N. Wolfe Street, Baltimore, Maryland 21205, United States of America
\textsuperscript{4}Johns Hopkins School of Public Health, Department of Epidemiology, 615 N. Wolfe Street, Baltimore, Maryland 21205, United States of America
\textsuperscript{5}Johns Hopkins School of Public Health, Department of International Health, 615 N. Wolfe Street, Baltimore, Maryland 21205, United States of America

*authors contributed equally

** Correspondance:
Amanda K. Debes
debes1@jhu.edu

Keywords: cholera, control, Kenya, interventions, prioritization, hotspots

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Kenya has experienced cholera outbreaks since 1971, with the most recent wave beginning in late 2014. Between 2015-2020, 32 of 47 counties reported 30,431 suspected cholera cases. The Global Task Force for Cholera Control (GTFCC) developed a Global Roadmap for Ending Cholera by 2030, which emphasizes the need to target multi-sectoral interventions to priority areas known as “cholera burden hotspots.” This study utilizes the GTFCC’s hotspot method to identify hotspots in Kenya at the county and sub-county administrative levels from 2015 through 2020. 32 of 47 (68.1%) counties reported cholera cases during this time while only 149 of 301 (49.5%) sub-counties reported cholera cases. The analysis identifies hotspots based on the mean annual incidence (MAI) over the past five-year period and cholera’s persistence in the area. Applying a MAI threshold of 90th percentile and the median persistence at both the county and sub-county levels, we identified 13 high risk sub-counties from 8 counties, including the 3 high risk counties of Garissa, Tana River and Wajir. This demonstrates that several sub-counties are high level hotspots while their counties are not. In addition, when cases reported by county versus sub-county hotspot risk are compared, 1.4 million people overlapped in the areas identified as both high-risk county and high-risk sub-county. However, assuming that finer scale data is more accurate, 1.6 million high risk sub-county people would have been misclassified as medium risk with a county-level analysis. Furthermore, an additional 1.6 million people would have been classified as living in high-risk in a county-level analysis when at the sub-county level, they were medium, low or no-risk sub-counties. This results in 3.2 million people being misclassified when county level analysis is utilized rather than a more-focused sub-county level analysis. This analysis highlights the need for more localized risk analyses to target cholera intervention and prevention efforts towards the populations most vulnerable.
Introduction

Cholera remains a public health threat to over one billion people globally,¹ ² ³ and the majority of cases reported to the World Health Organization (WHO) are from sub-Saharan Africa.⁴ Cholera was first reported in Kenya in 1971 and Kenya has since experienced outbreaks regularly every few years, with the most recent wave of outbreaks beginning in December 2014.⁵ ⁶ ⁷ Kenya reported the “index case” of the epidemic wave onset in Eastern and Southern Africa in December 2014, which spread to other African nations including Tanzania, Uganda and South Sudan.⁸ As such, cholera control efforts in Kenya not only protect vulnerable people in Kenya, but could also stop spread to neighboring countries.

To combat cholera in Kenya, the Ministry of Health (MoH) developed a national cholera control plan based on the guidance from the WHO-led Global Task Force on Cholera Control (GTFCC). The GTFCC’s Global Roadmap for Ending Cholera by 2030⁹ stresses the importance of targeting control activities to priority intervention areas, also known as “cholera burden hotspots.” These are administrative units with high incidence rates, persistent cholera transmission, or high risk of cholera introduction relative to other locations in the country.⁹-¹¹ Identification of these high priority areas can help with efficient targeting of oral cholera vaccination (OCV) campaigns, surveillance system investments, and water and sanitation infrastructure, all of which could be used to coordinate a comprehensive cholera control strategy within the country.
In this analysis, we identified priority intervention areas for cholera using data from the Kenya MoH from 2015-2020. We performed this analysis at the county and sub-county administrative units following the guidance issued by GTFCC.\textsuperscript{12}

**Methods**

**Cholera surveillance data**

Notifiable diseases are reported via national line list data from county and sub-county health facilities to the Kenya MoH as part of the Integrated Disease Surveillance and Response (IDSR) program.\textsuperscript{13} We examined all suspected cholera cases across Kenya from the national line list data from January 1, 2015 through July 29\textsuperscript{th}, 2020; no cholera cases were reported after July 2020 so we assumed this to be a complete year of 2020 data. The study team cleaned location names and performed name disambiguation to match sub-county names to GADM sub-county locations.\textsuperscript{14} For sub-county names with no corresponding label on the map, the names were cross-referenced with sub-counties listed in the *Kenya 2019 Population and Housing Census Report Volume 2* (2019 Census)\textsuperscript{15} and subsequently aligned with the map. Further data cleaning details are described in Supplemental Figure 1.

IDSR guidelines define a suspected cholera case as the sudden onset of three or more episodes of watery diarrhea within 24 hours in a person $\geq$ 2 years of age or a younger patient in whom a clinician suspected cholera. A confirmed cholera case was defined as a suspected case that was confirmed by culture or PCR with *Vibrio cholerae* O1 or O139. Cases were assigned to the date they sought care at a health facility, or if missing, the date of diarrhea onset.
Population data

The annual population counts from 2015-2020 were aggregated from 100-meter-resolution population raster of Kenya to sub-county and county administrative units\textsuperscript{16,17} using the R package “exactextractr” (version 0.6.1).\textsuperscript{18} Sub-county and county-level administrative maps of Kenya were obtained from the GADM database.\textsuperscript{14}

Prioritizing cholera burden hotspots

The prioritization exercise was performed separately at the county and sub-county units of analysis. For each geographic scale, we identified cholera burden hotspots, also known as priority intervention areas, according to the GTFCC-recommended method,\textsuperscript{12} which uses two quantitative indicators: 1) mean annual incidence (MAI), the mean of the annual cholera incidence rate (cases divided by population) over the analysis period; 2) persistence, the percentage of weeks in the analysis period where at least one cholera case was reported.\textsuperscript{12}

We identified thresholds for each indicator, and locations were categorized as high priority if both indicators exceeded their thresholds and medium priority if only one indicator exceeded its threshold. Locations that remained below both MAI and persistence thresholds were low priority.

Based on discussions with the MoH pertaining to the logistics for targeting interventions, at the sub-county scale, we selected the 90th percentile value as the MAI threshold (33 per 100,000) and the median value as the persistence threshold (2.88%). We also performed sensitivity analyses for all combinations of MAI and persistence thresholds in decile increments; a subset of these is displayed in the Supplementary Material.
Software Used

The software programs used include Microsoft Excel for data management, ArcGIS 10.6 for the management of geographic information files and mapping, and R (version 4.1.0) for statistical computing and graphics.

Ethics

The study used de-identified public health disease surveillance data. Therefore, no ethical approval was required for conducting this study.

Results

From 2015 to 2020, the total annual suspected cases ranged from a maximum of 9464 in 2015 to a minimum of 711 in 2020, with 32 of 47 counties reporting 30,431 cases overall. After excluding records with missing date or missing or unidentifiable geographic information, we successfully geolocated 27,267 cases to counties and 22,825 to sub-counties, and these were the final datasets used for the county and sub-county analyses, respectively (See additional information on data cleaning in Supp Figure 1).

Cholera cases varied by month and year, and while there was not a readily discernable seasonal or cyclical pattern, there were extended periods with relatively high cholera transmission (May 2015-August 2016, June 2017-July 2018, and February 2019-November 2019) interspersed with months of relatively sparse or low activity (Figure 1). Out of the 32 counties affected during this 6-year period, 1 county (Garissa) reported cases in all 6 years, 3 counties reported cases in 5 of 6 years, 5 counties reported cases in 4 of the 6 years, 6 counties reported cases in 3 of the 6 years, 11 counties reported cases in 2 of the 6 years, and 6 reported cases in only one year. The sub-counties with the highest case count and incidence varied by year, suggesting that outbreaks
in different regions of the country contributed to the periods with relatively high transmission (Supp Figure 2; Figure 2).

Mean annual incidence (MAI) and persistence were calculated for each sub-county in order to perform the prioritization activity according to the 90th percentile and median thresholds, respectively (Figure 3/Supp Figure 3). The sub-counties with high mean annual incidence tended to have high persistence, and few locations exceeded the 90th percentile MAI while remaining below the median persistence value. Thirteen sub-counties (6% of the population) were categorized as high priority, 61 sub-counties (24% of the population) as medium priority, and 76 sub-counties (25% of the population) as low priority (Table 1). There were 151 sub-counties (45% of the population) that had no reported cases in the analysis period.

When performing the same prioritization analysis at the county scale instead of the sub-county scale, classification becomes much less efficient (Figure 4). While the county-level analysis identified three high priority counties in Northeast Kenya (Garissa, Tana River and Wajir), the sub-county analysis identified thirteen high priority sub-counties from eight counties, including those in Nairobi County and five additional counties that border neighboring countries. When finer resolution data are not available, significant misclassification in both directions may occur (Table 2). For example, roughly 660,000 individuals living in sub-counties with no reported cases were classified as living in a high priority county, and 1.6 million living in high priority sub-counties were classified as living in a medium priority county. While no populations living in high priority sub-counties would be classified as living in low priority counties, there would be significant “over-prioritization” of populations living in sub-counties with no or few cases reported.

Due to challenges with geolocation, we assessed the sensitivity of the results when making different location name cleaning assumptions. There were many records that could only be
geolocated to the county scale, as the sub-county field was missing or otherwise unidentifiable. To assess the sensitivity of our hotspots to these missing data, we performed a separate analysis on a sub-county dataset, where in one arm we excluded records that had insufficient data to geolocate the case to a specific sub-county. In the comparison arm, we included those same non-geolocatable records by assigning them proportionally according to the distribution of geolocatable sub-county records in the associated county and reporting year. The sensitivity analysis showed the same high-risk hotspots in both arms of the analysis, identifying over 3 million people at high-risk regardless of which data set is used for the analysis (Supp Figure 4/Supp Table 3).

We also explored the impact of alternate thresholds for MAI (Supp Figure 5/Supp Table 4). For example, when reducing the primary analysis’s 90\textsuperscript{th} percentile threshold for MAI to the 80\textsuperscript{th} percentile threshold, 11\% or over 5 million people lived in priority areas (compared to the 6\% or roughly 3 million people in the primary sub-county analysis).

Discussion

In this analysis we collated cholera IDSR line list data from 2015 to 2020 in Kenya to identify priority intervention areas at both county and sub-county scales following the proposed GTFCC method. At the sub-county scale, there were 13 high priority areas from a total of 8 counties that represented 6\% of the Kenyan population. When comparing the same analyses at the county and sub-county scales, we found that sub-counties were more likely to be over-prioritized (given a higher priority at the county scale than what they would have received at the sub-county scale) than under-prioritized when data were analyzed in the same way at the county scale. However, high priority sub-counties were relatively stable under different cleaning methods and threshold settings.
Previous work has reported that Kenya has early-year (e.g. January) and relatively low amplitude cholera seasonality, and that there were several cholera outbreaks in the country during the 2015-2019 period. The four seasons in Kenya include a dry season from January to March, a rainy season from March to May, a dry season from May to October and a rainy season from October to December. According to the 2009 census, while ~90% of urban households have access to improved water sources, only approximately 50% of rural households have such access. In addition, the type of improved water source greatly differs, with urban households predominantly having piped water into the dwelling while rural households primarily have access to dug wells as an improved water source. Access to adequate sanitation is also a significant issue in Kenya, with the 2009 census reporting that only 30% of urban and 20% of rural households have access to an improved toilet facility. Per UNICEF’s 2012 Update on the Progress of Drinking Water and Sanitation, Kenya is among ten countries with the largest population without access to an improved drinking water source. Systematic collection of risk factor data by sub-county would improve understanding of the role of risk factors in high-risk areas and facilitate intervention decisions.

The current GTFCC method for priority intervention areas recommends that cholera data are collated at an administrative level 2 or smaller level to facilitate operational planning. At the request of the MoH, this analysis was performed at the sub-county level. The counties in Kenya are comprised of populations with a median size greater than 1 million people, thus, it would be challenging to target interventions at the county scale. By focusing on the sub-county, targeting becomes more efficient, and the population sizes are reduced from populations in the millions to those in the low hundreds of thousands, increasing the feasibility and likely impact of interventions. Further, when Kenya introduces OCV, county level vaccination targets are likely both too large...
for the vaccine stockpile to consider and too challenging logistically. A recent study focusing on hotspots in Nigeria concluded that targeting locations at the ward level (finer geographic scale) was recommended over hotspots at the local government level. Local knowledge (opinion) was used to focus OCV in specific sub-counties of the districts identified as high risk. This study in Kenya attempts to identify these high risk sub-counties more objectively. These results highlight the value of surveillance, as the sub-county level data changes the interpretation at the spatial scale and optimizes the vaccine target size for the best cost-benefit scenario.

There were several limitations to this study. While the IDSR program reports cholera line lists at the county level, sub-county level reporting is not as consistent. We used sub-county level data reported to the MoH for this analysis, however only 22,825 records out of the total 30,431 line list records were able to be cleaned and their sub-counties identified. Further, the county and sub-county datasets were different due to location name data cleaning challenges, which means that these analyses are not exactly comparable. Our analysis may have dropped records from the same, repeated location names that could not be identified successfully, which would bias the spatial distribution of cases and the subsequent results. Another limitation is due to reporting of cases on the basis of clinical presentation. The outbreaks were initially confirmed as cholera by culture, but subsequent cases were reported based on clinical criteria. Some of the cases reported as cholera may have had illness due to another pathogen, and conversely, cholera cases may be under-reported, especially in remote areas. Surveillance may be improved in the future if rapid diagnostics tests are used more widely to confirm cases.
This analysis may provide information useful to the Kenya MoH to support their National Cholera Control Plan and intervention efforts, including the use of OCV, in targeting cholera control in the high-risk sub-counties identified through this analysis. The cholera hotspots mapping helps the MOH Kenya to efficiently utilize available resources through an evidence-based approach to prioritize high risk sub-counties in the implementation of the available interventions. In addition, we hope that documentation of the methods of this exercise could aid other countries engaged in national cholera control planning.

12. GTFCC. Guidance and tool for countries to identify priority areas for intervention.
   Geneva: GTFCC. https://www.gtfcc.org/resources/guidance-and-tool-for-countries-to-

13. Fall IS, Rajatonirina S, Yahaya AA, et al. Integrated Disease Surveillance and Response
   (IDSR) strategy: current status, challenges and perspectives for the future in Africa. BMJ


   2019.

16. Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. Population distribution, settlement


18. exactextractr [computer program]. 2021.


20. Macro' KNBoSKaI. Kenya Demographic and Health Survey 2008-2009. Calverton,
   Maryland2010.


   Nigeria, 2010-2019—analysis of patient characteristics, Spatio-temporal patterns and

23. Debes AK, Murt KN, Waswa E, et al. Laboratory and Field Evaluation of the Crystal VC-O1
   Cholera Rapid Diagnostic Test. The American journal of tropical medicine and hygiene.

Figure 1: Weekly suspected cholera cases in Kenya from 2015-2020. No more cases were
reported in 2020 after July so the figure is truncated after that point.

Figure 2: Total annual cholera incidence rate by sub-counties, Kenya, 2015-2020.

Figure 3. A) Mean annual incidence and B) persistence in Kenya by sub-county from 2015-2020.
C) Scatterplot for mean annual incidence (on the log scale) versus persistence, the percentage of
total weeks reporting at least one suspected cholera case. The horizontal and vertical lines
represent the mean annual incidence and persistence thresholds, respectively

Figure 4a. Priority intervention areas at the county scale in Kenya 2015-2020 (MAI threshold:
90 percentile – 33 per 100,000, Persistence threshold: median – 9.78%)

Figure 4b. Priority intervention areas at the sub-county scale in Kenya from 2015-2020 (MAI
threshold: 90 percentile – 39 per 100,000, Persistence threshold: median – 2.88%)

Table 1. Population and number of sub-counties under each hotspot (priority) level
<table>
<thead>
<tr>
<th>Hotspot (Priority) Level</th>
<th>Population</th>
<th>Percentage of population</th>
<th>Number of sub-counties</th>
<th>Population Weighted Average MAI (cases per 100,000): Mean (Min, Max)</th>
<th>Population Weighted Average Persistence: Mean (Min, Max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>3021238</td>
<td>6</td>
<td>13</td>
<td>74.02 (39.39, 221.55)</td>
<td>36.60 (25.00, 47.12)</td>
</tr>
<tr>
<td>Medium</td>
<td>12124262</td>
<td>24</td>
<td>61</td>
<td>12.25 (1.35, 67.26)</td>
<td>20.78 (7.69, 41.67)</td>
</tr>
<tr>
<td>Low</td>
<td>12559717</td>
<td>25</td>
<td>76</td>
<td>2.35 (0.07, 32.46)</td>
<td>5.06 (0.64, 9.29)</td>
</tr>
<tr>
<td>No case reported</td>
<td>23041691</td>
<td>45</td>
<td>151</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2. Comparison of population living in different priority areas in the county and sub-county analyses. When data are not available at the finer resolution scale, classification of priority areas becomes less efficient. For example, roughly 660,000 people living in sub-counties with no reported cases would be categorized as living in high priority counties.
<table>
<thead>
<tr>
<th>Sub-county Hotspot</th>
<th>No case reported</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>13859714</td>
<td>5768177</td>
<td>2755295</td>
<td>658505</td>
</tr>
<tr>
<td>Low</td>
<td>0</td>
<td>5112447</td>
<td>7015429</td>
<td>431841</td>
</tr>
<tr>
<td>Medium</td>
<td>0</td>
<td>2638988</td>
<td>8933536</td>
<td>551738</td>
</tr>
<tr>
<td>High</td>
<td>0</td>
<td>0</td>
<td>1626394</td>
<td>1394844</td>
</tr>
</tbody>
</table>
Figure 3a
Figure 4