

1 **Prioritizing interventions for cholera control in Kenya, 2015-2020**

2 **Waqo Boru^{1,2*}, Shaoming Xiao^{3*}, Patrick Amoth¹, David Kareko¹, Daniel Langat¹, Ian**

3 **Were¹, Mohammad Ali², David A. Sack², Elizabeth C. Lee⁴, Amanda K. Debes^{5**}**

4 ¹Ministry of Health, Nairobi, Kenya

5 ²Field Epidemiology and Laboratory Training Program, Nairobi, Kenya

6 ³Johns Hopkins University School of Medicine, 200 N. Wolfe Street, Baltimore, Maryland 21205,

7 United States of America

8 ⁴Johns Hopkins School of Public Health, Department of Epidemiology, 615 N. Wolfe Street,

9 Baltimore, Maryland 21205, United States of America

10 ⁵Johns Hopkins School of Public Health, Department of International Health, 615 N. Wolfe

11 Street, Baltimore, Maryland 21205, United States of America

12 *authors contributed equally

13

14

15 **** Correspondance:**

16 Amanda K. Debes

17 adebes1@jhu.edu

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

19

20 **Abstract**

21 Kenya has experienced cholera outbreaks since 1971, with the most recent wave beginning in late
22 2014. Between 2015-2020, 32 of 47 counties reported 30,431 suspected cholera cases. The Global
23 Task Force for Cholera Control (GTFCC) developed a *Global Roadmap for Ending Cholera by*
24 *2030*, which emphasizes the need to target multi-sectoral interventions to priority areas known as
25 “cholera burden hotspots.” This study utilizes the GTFCC’s hotspot method to identify hotspots
26 in Kenya at the county and sub-county administrative levels from 2015 through 2020. 32 of 47
27 (68.1%) counties reported cholera cases during this time while only 149 of 301 (49.5%) sub-
28 counties reported cholera cases. The analysis identifies hotspots based on the mean annual
29 incidence (MAI) over the past five-year period and cholera’s persistence in the area. Applying a
30 MAI threshold of 90th percentile and the median persistence at both the county and sub-county
31 levels, we identified 13 high risk sub-counties from 8 counties, including the 3 high risk counties
32 of Garissa, Tana River and Wajir. This demonstrates that several sub-counties are high level
33 hotspots while their counties are not. In addition, when cases reported by county versus sub-county
34 hotspot risk are compared, 1.4 million people overlapped in the areas identified as both high-risk
35 county and high-risk sub-county. However, assuming that finer scale data is more accurate, 1.6
36 million high risk sub-county people would have been misclassified as medium risk with a county-
37 level analysis. Furthermore, an additional 1.6 million people would have been classified as living
38 in high-risk in a county-level analysis when at the sub-county level, they were medium, low or no-
39 risk sub-counties. This results in 3.2 million people being misclassified when county level analysis
40 is utilized rather than a more-focused sub-county level analysis. This analysis highlights the need
41 for more localized risk analyses to target cholera intervention and prevention efforts towards the
42 populations most vulnerable.

43

44 **Introduction**

45 Cholera remains a public health threat to over one billion people globally,^{1 2 3} and the majority of
46 cases reported to the World Health Organization (WHO) are from sub-Saharan Africa.⁴ Cholera
47 was first reported in Kenya in 1971 and Kenya has since experienced outbreaks regularly every
48 few years, with the most recent wave of outbreaks beginning in December 2014.^{5 6,7} Kenya
49 reported the “index case” of the epidemic wave onset in Eastern and Southern Africa in December
50 2014, which spread to other African nations including Tanzania, Uganda and South Sudan.⁸ As
51 such, cholera control efforts in Kenya not only protect vulnerable people in Kenya, but could also
52 stop spread to neighboring countries.

53

54 To combat cholera in Kenya, the Ministry of Health (MoH) developed a national cholera control
55 plan based on the guidance from the WHO-led Global Task Force on Cholera Control (GTFCC).
56 The GTFCC’s *Global Roadmap for Ending Cholera by 2030*⁹ stresses the importance of targeting
57 control activities to priority intervention areas, also known as “cholera burden hotspots.” These
58 are administrative units with high incidence rates, persistent cholera transmission, or high risk of
59 cholera introduction relative to other locations in the country.⁹⁻¹¹ Identification of these high
60 priority areas can help with efficient targeting of oral cholera vaccination (OCV) campaigns,
61 surveillance system investments, and water and sanitation infrastructure, all of which could be
62 used to coordinate a comprehensive cholera control strategy within the country.

63

64 In this analysis, we identified priority intervention areas for cholera using data from the Kenya
65 MoH from 2015-2020. We performed this analysis at the county and sub-county administrative
66 units following the guidance issued by GTFCC.¹²

67

68 **Methods**

69 **Cholera surveillance data**

70 Notifiable diseases are reported via national line list data from county and sub-county health
71 facilities to the Kenya MoH as part of the Integrated Disease Surveillance and Response (IDSR)
72 program.¹³ We examined all suspected cholera cases across Kenya from the national line list data
73 from January 1, 2015 through July 29th, 2020; no cholera cases were reported after July 2020 so
74 we assumed this to be a complete year of 2020 data. The study team cleaned location names and
75 performed name disambiguation to match sub-county names to GADM sub-county locations.¹⁴
76 For sub-county names with no corresponding label on the map, the names were cross-referenced
77 with sub-counties listed in the *Kenya 2019 Population and Housing Census Report Volume 2*
78 (2019 Census)¹⁵ and subsequently aligned with the map. Further data cleaning details are described
79 in Supplemental Figure 1.

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

85

86 **Population data**

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

91

92 **Prioritizing cholera burden hotspots**

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

99 We identified thresholds for each indicator, and locations were categorized as high priority
100 if both indicators exceeded their thresholds and medium priority if only one indicator exceeded its
101 threshold. Locations that remained below both MAI and persistence thresholds were low priority.
102 Based on discussions with the MoH pertaining to the logistics for targeting interventions, at the
103 sub-county scale, we selected the 90th percentile value as the MAI threshold (33 per 100,000) and
104 the median value as the persistence threshold (2.88%). We also performed sensitivity analyses for
105 all combinations of MAI and persistence thresholds in decile increments; a subset of these is
106 displayed in the Supplementary Material.

107 **Software Used**

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

111

112 **Ethics**

113
114 The study used de-identified public health disease surveillance data. Therefore, no ethical approval
115 was required for conducting this study.

116

117 **Results**

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

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

132 in different regions of the country contributed to the periods with relatively high transmission
133 (Supp Figure 2; Figure 2).

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

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

153 Due to challenges with geolocation, we assessed the sensitivity of the results when making
154 different location name cleaning assumptions. There were many records that could only be

155 geolocated to the county scale, as the sub-county field was missing or otherwise unidentifiable. To
156 assess the sensitivity of our hotspots to these missing data, we performed a separate analysis on a
157 sub-county dataset, where in one arm we excluded records that had insufficient data to geolocate
158 the case to a specific sub-county. In the comparison arm, we included those same non-geolocatable
159 records by assigning them proportionally according to the distribution of geolocatable sub-county
160 records in the associated county and reporting year. The sensitivity analysis showed the same high-
161 risk hotspots in both arms of the analysis, identifying over 3 million people at high-risk regardless
162 of which data set is used for the analysis (Supp Figure 4/Supp Table 3).

163 We also explored the impact of alternate thresholds for MAI (Supp Figure 5/Supp Table
164 4). For example, when reducing the primary analysis's 90th percentile threshold for MAI to the
165 80th percentile threshold, 11% or over 5 million people lived in priority areas (compared to the 6%
166 or roughly 3 million people in the primary sub-county analysis).

167 **Discussion**

168 In this analysis we collated cholera IDSR line list data from 2015 to 2020 in Kenya to identify
169 priority intervention areas at both county and sub-county scales following the proposed GTFCC
170 method. At the sub-county scale, there were 13 high priority areas from a total of 8 counties that
171 represented 6% of the Kenyan population. When comparing the same analyses at the county and
172 sub-county scales, we found that sub-counties were more likely to be over-prioritized (given a
173 higher priority at the county scale than what they would have received at the sub-county scale)
174 than under-prioritized when data were analyzed in the same way at the county scale. However,
175 high priority sub-counties were relatively stable under different cleaning methods and threshold
176 settings.

177 Previous work has reported that Kenya has early-year (e.g. January) and relatively low
178 amplitude cholera seasonality, and that there were several cholera outbreaks in the country during
179 the 2015-2019 period.¹⁹ The four seasons in Kenya include a dry season from January to March, a
180 rainy season from March to May, a dry season from May to October and a rainy season from
181 October to December.²⁰ According to the 2009 census, while ~90% of urban households have
182 access to improved water sources, only approximately 50% of rural households have such access.²⁰
183 In addition, the type of improved water source greatly differs, with urban households
184 predominantly having piped water into the dwelling while rural households primarily have access
185 to dug wells as an improved water source. Access to adequate sanitation is also a significant issue
186 in Kenya, with the 2009 census reporting that only 30% of urban and 20% of rural households
187 have access to an improved toilet facility.²⁰ Per UNICEF's 2012 Update on the Progress of
188 Drinking Water and Sanitation, Kenya is among ten countries with the largest population without
189 access to an improved drinking water source.²¹ Systematic collection of risk factor data by sub-
190 county would improve understanding of the role of risk factors in high-risk areas and facilitate
191 intervention decisions.

192 The current GTFCC method for priority intervention areas recommends that cholera data
193 are collated at an administrative level 2 or smaller level to facilitate operational planning. At the
194 request of the MoH, this analysis was performed at the sub-county level. The counties in Kenya
195 are comprised of populations with a median size greater than 1 million people, thus, it would be
196 challenging to target interventions at the county scale. By focusing on the sub-county, targeting
197 becomes more efficient, and the population sizes are reduced from populations in the millions to
198 those in the low hundreds of thousands, increasing the feasibility and likely impact of interventions.
199 Further, when Kenya introduces OCV, county level vaccination targets are likely both too large

200 for the vaccine stockpile to consider and too challenging logistically. A recent study focusing on
201 hotspots in Nigeria concluded that targeting locations at the ward level (finer geographic scale)
202 was recommended over hotspots at the local government level.²² Local knowledge (opinion) was
203 used to focus OCV in specific sub-counties of the districts identified as high risk. This study in
204 Kenya attempts to identify these high risk sub-counties more objectively. These results highlight
205 the value of surveillance, as the sub-county level data changes the interpretation at the spatial scale
206 and optimizes the vaccine target size for the best cost-benefit scenario.

207

208 There were several limitations to this study. While the IDSR program reports cholera line lists at
209 the county level, sub-county level reporting is not as consistent. We used sub-county level data
210 reported to the MoH for this analysis, however only 22,825 records out of the total 30,431 line list
211 records were able to be cleaned and their sub-counties identified. Further, the county and sub-
212 county datasets were different due to location name data cleaning challenges, which means that
213 these analyses are not exactly comparable. Our analysis may have dropped records from the same,
214 repeated location names that could not be identified successfully, which would bias the spatial
215 distribution of cases and the subsequent results. Another limitation is due to reporting of cases on
216 the basis of clinical presentation. The outbreaks were initially confirmed as cholera by culture,
217 but subsequent cases were reported based on clinical criteria. Some of the cases reported as cholera
218 may have had illness due to another pathogen, and conversely, cholera cases may be under-
219 reported, especially in remote areas. Surveillance may be improved in the future if rapid
220 diagnostics tests are used more widely to confirm cases.²³

221

222 This analysis may provide information useful to the Kenya MoH to support their National Cholera
223 Control Plan and intervention efforts, including the use of OCV, in targeting cholera control in the
224 high-risk sub-counties identified through this analysis. The cholera hotspots mapping helps the
225 MOH Kenya to efficiently utilize available resources through an evidence-based approach to
226 prioritize high risk sub-counties in the implementation of the available interventions. In addition,
227 we hope that documentation of the methods of this exercise could aid other countries engaged in
228 national cholera control planning.

229

- 230 1. World Health O. Cholera vaccine: WHO position paper, August 2017 -
231 Recommendations. *Vaccine*. 2018;36(24):3418-3420.
- 232 2. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of
233 diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study
234 2016. *The Lancet Infectious diseases*. 2018;18(11):1211-1228.
- 235 3. Ali M, Nelson AR, Lopez AL, Sack DA. Updated global burden of cholera in endemic
236 countries. *PLoS neglected tropical diseases*. 2015;9(6):e0003832.
- 237 4. Lessler J, Moore SM, Luquero FJ, et al. Mapping the burden of cholera in sub-Saharan
238 Africa and implications for control: an analysis of data across geographical scales.
239 *Lancet*. 2018;391(10133):1908-1915.
- 240 5. Mutonga D, Langat D, Mwangi D, et al. National surveillance data on the epidemiology
241 of cholera in Kenya, 1997-2010. *The Journal of infectious diseases*. 2013;208 Suppl
242 1:S55-61.
- 243 6. George G, Rotich J, Kigen H, et al. Notes from the Field: Ongoing Cholera Outbreak -
244 Kenya, 2014-2016. *MMWR Morbidity and mortality weekly report*. 2016;65(3):68-69.
- 245 7. Zheng Q, Luquero FJ, Ciglenecki I, et al. Cholera outbreaks in sub-Saharan Africa during
246 2010-2019: A Descriptive Analysis. *medRxiv*. 2021:2021.2010.2025.21265347.
- 247 8. Cholera, 2015. *Wkly Epidemiol Rec*. 2016;91(38):433-440.
- 248 9. Global Task Force on Cholera Control G. Overview of Ending Cholera - A Global Roadmap
249 to 2030. The World Health Organization, WHO.
250 <https://www.who.int/cholera/publications/global-roadmap-summary.pdf>. Published
251 2017. Accessed July 10, 2019.
- 252 10. Lessler J, Azman AS, McKay HS, Moore SM. What is a Hotspot Anyway? *The American
253 journal of tropical medicine and hygiene*. 2017;96(6):1270-1273.
- 254 11. Global Task Force on Cholera Control G. *FRAMEWORK FOR THE DEVELOPMENT AND
255 MONITORING OF A MULTISECTORAL NATIONAL CHOLERA PLAN*. Geneva, Switzerland
256 June 2019 2019.

- 257 12. GTFCC. Guidance and tool for countries to identify priority areas for intervention
258 Geneva: GTFCC. <https://www.gtfcc.org/resources/guidance-and-tool-for-countries-to-identify-priority-areas-for-intervention/>. Published 2021. Accessed.
259
- 260 13. Fall IS, Rajatonirina S, Yahaya AA, et al. Integrated Disease Surveillance and Response
261 (IDSR) strategy: current status, challenges and perspectives for the future in Africa. *BMJ*
262 *Glob Health*. 2019;4(4):e001427.
- 263 14. GADM maps and data. <https://gadm.org/>. Accessed September 15, 2021.
- 264 15. Statistics KNBo. *2019 Kenya Population and Housing Census Results*. November, 4, 2019
265 2019.
- 266 16. Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. Population distribution, settlement
267 patterns and accessibility across Africa in 2010. *PLoS one*. 2012;7(2):e31743.
- 268 17. SDI W. <http://www.worldpop.org>. Accessed March 9, 2022.
- 269 18. *exactextractr* [computer program]. 2021.
- 270 19. Perez-Saez J, Lessler J, Lee EC, et al. The seasonality of cholera in sub-Saharan Africa: a
271 statistical modelling study. *The Lancet Global health*. 2022;10(6):e831-e839.
- 272 20. Macro' KNBoSKal. *Kenya Demographic and Health Survey 2008-2009*. Calverton,
273 Maryland2010.
- 274 21. UNICEF. *Progress on Drinking Water and Sanitation*. New York, NY2012.
- 275 22. Ngwa MC, Ihekweazu C, Okwor TJ, et al. The micro-hotspots of cholera in Kano State,
276 Nigeria, 2010-2019—analysis of patient characteristics, Spatio-temporal patterns and
277 contextual determinants at the ward level. *medRxiv*. 2021:2021.2008.2020.21262313.
- 278 23. Debes AK, Murt KN, Waswa E, et al. Laboratory and Field Evaluation of the Crystal VC-O1
279 Cholera Rapid Diagnostic Test. *The American journal of tropical medicine and hygiene*.
280 2021;104(6):2017-2023.

281

282

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

285

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

287

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

292

293 Figure 4a. Priority intervention areas at the county scale in Kenya 2015-2020 (MAI threshold:
294 90 percentile – 33 per 100,000, Persistence threshold: median – 9.78%)

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

297 Table 1. Population and number of sub-counties under each hotspot (priority) level

298

299

Hotspot (Priority) Level	Population	Percentage of population	Number of sub-counties	Population Weighted Average MAI (cases per 100,000): Mean (Min, Max)	Population Weighted Average Persistence Mean (Min, Max)
High	3021238	6	13	74.02 (39.39, 221.55)	36.60 (25.00, 47.12)
Medium	12124262	24	61	12.25 (1.35, 67.26)	20.78 (7.69, 41.67)
Low	12559717	25	76	2.35 (0.07, 32.46)	5.06 (0.64, 9.29)
No case reported	23041691	45	151	0	0

300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321

322 Table 2. Comparison of population living in different priority areas in the county and sub-county
323 analyses. When data are not available at the finer resolution scale, classification of priority areas
324 becomes less efficient. For example, roughly 660,000 people living in sub-counties with no
325 reported cases would be categorized as living in high priority counties.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

326

		County Hotspot			
		No case reported	Low	Medium	High
Sub-county Hotspot	No case reported	13859714	5768177	2755295	658505
	Low	0	5112447	7015429	431841
	Medium	0	2638988	8933536	551738
	High	0	0	1626394	1394844

327

328

329

330

331

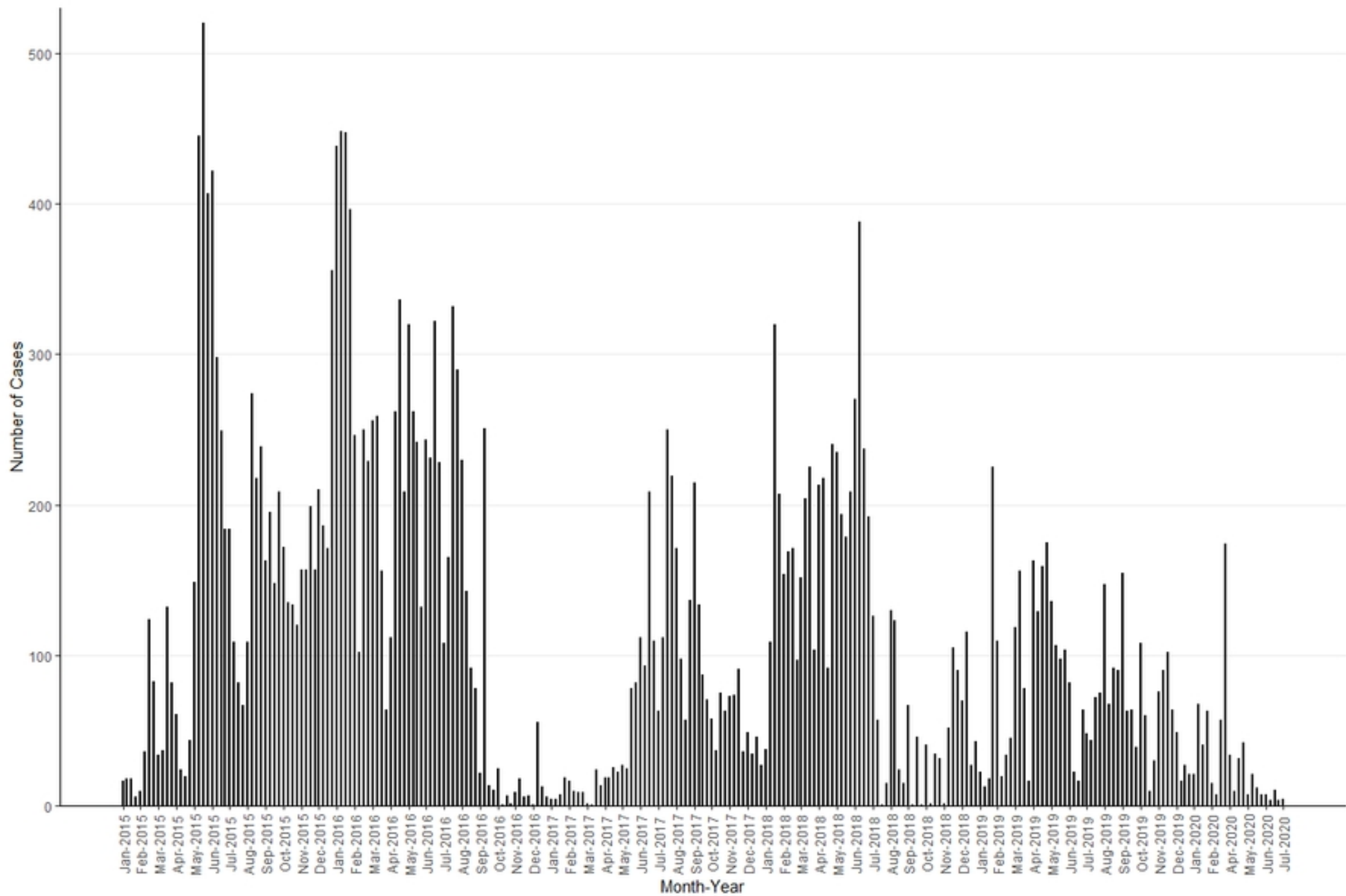
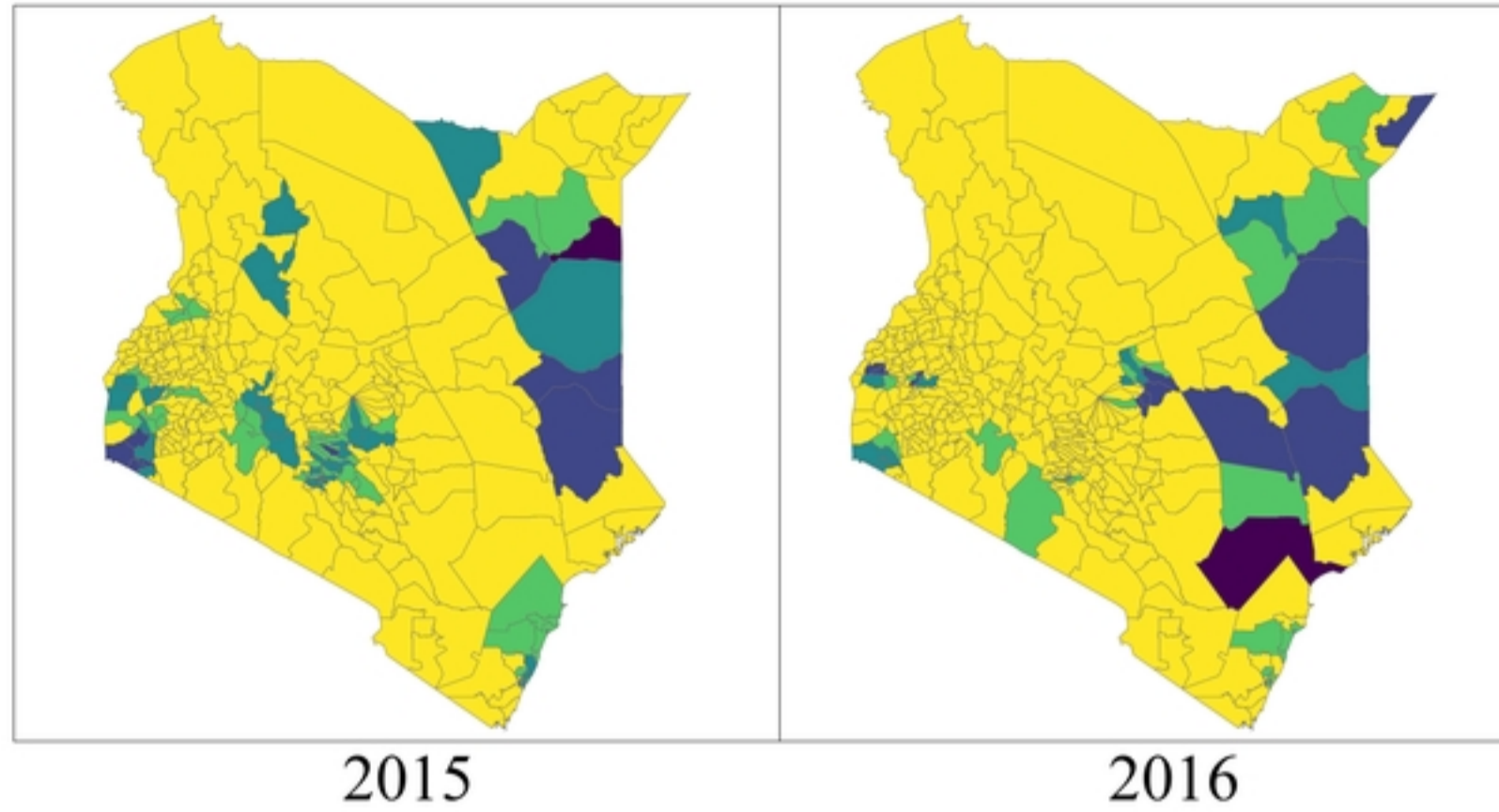
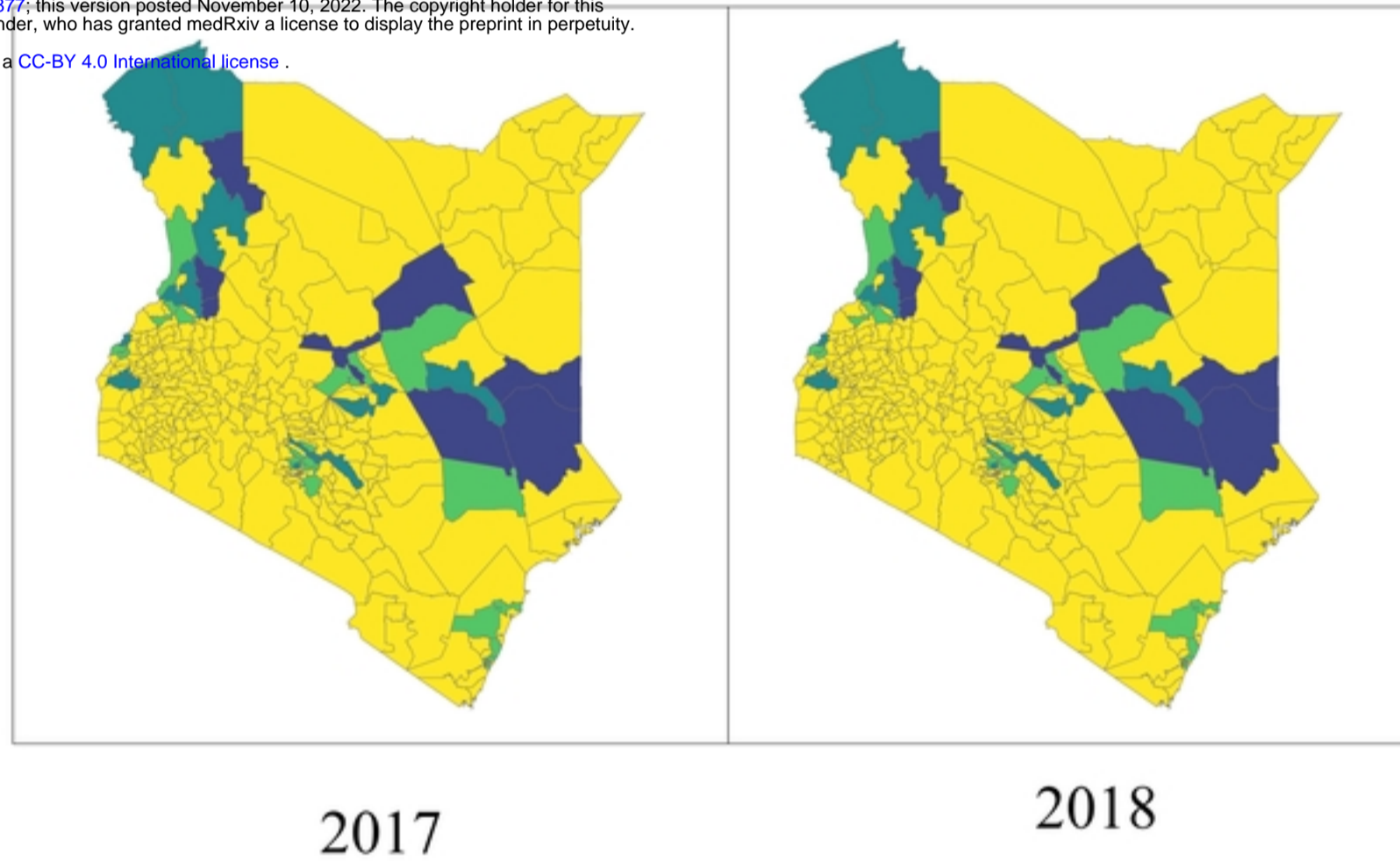


Figure 1



medRxiv preprint doi: <https://doi.org/10.1101/2022.11.07.22281877>; this version posted November 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



Total Annual Incidence Rate

- No case reported
- 0.00 - 10.00
- 10.01 - 100.00
- 100.01 - 500.00
- >500.00

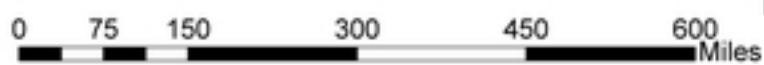
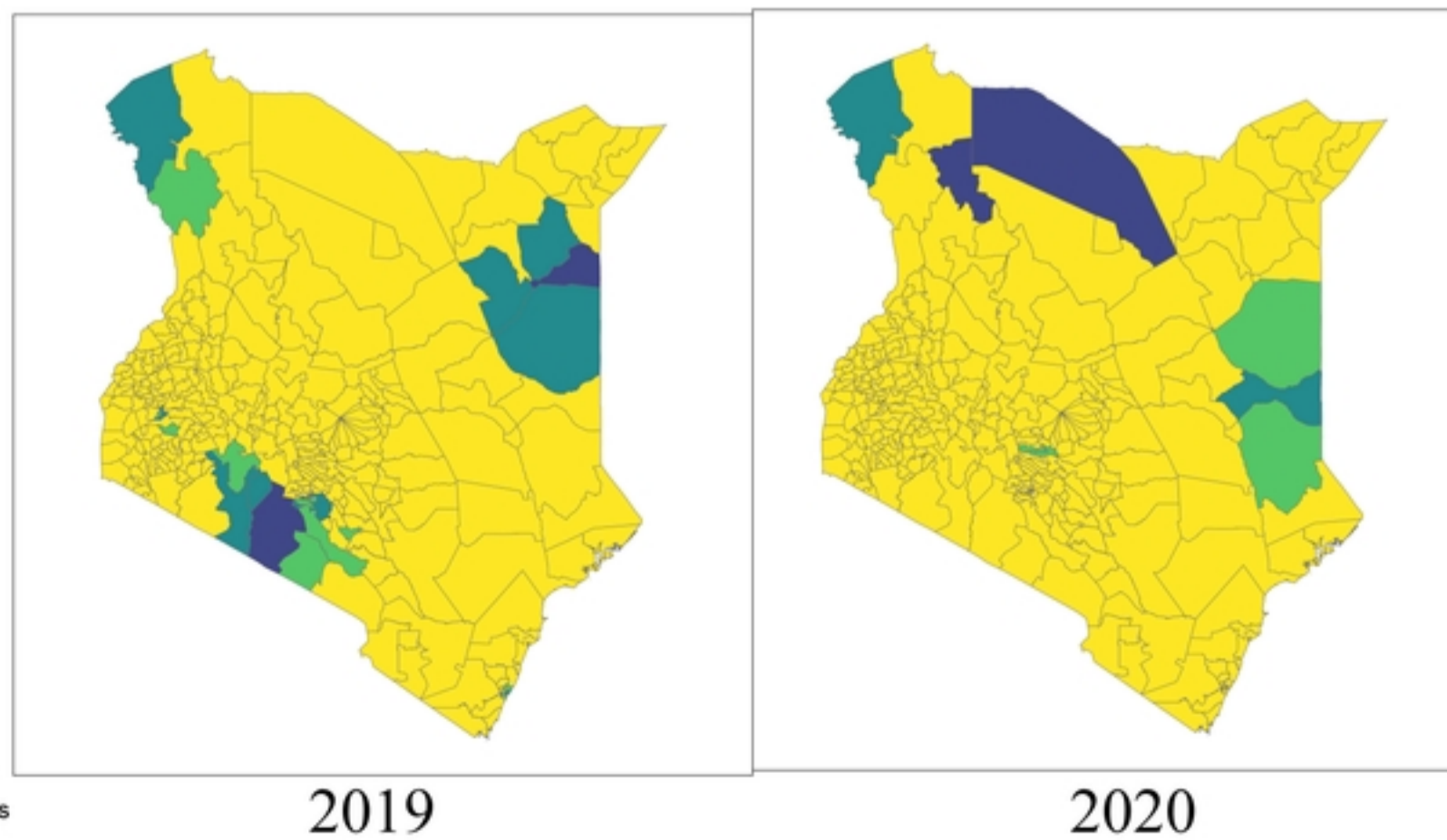


Figure 2

A

medRxiv preprint doi: <https://doi.org/10.1101/2022.11.07.22281677>; this version posted November 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

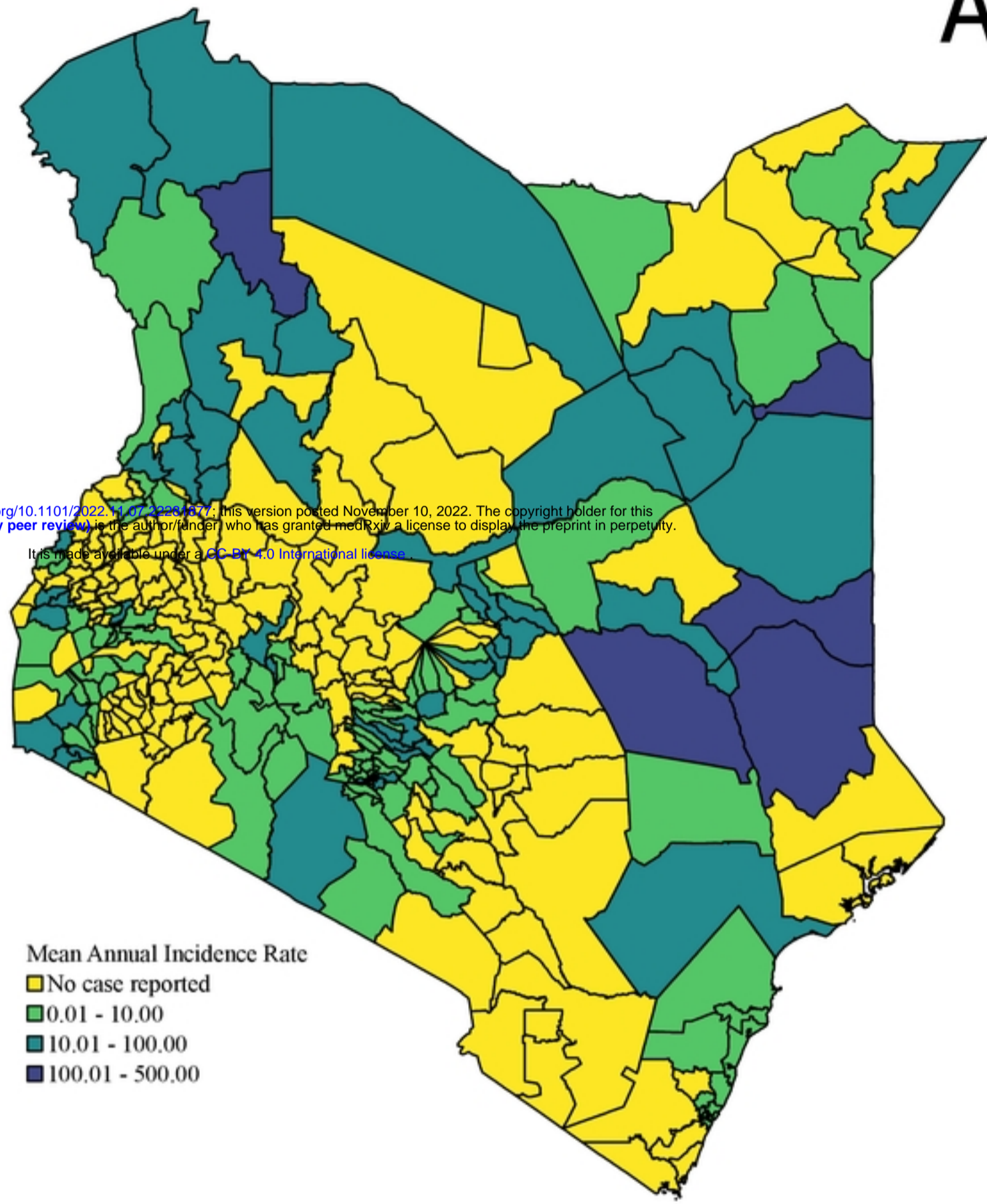


Figure 3a

B

medRxiv preprint doi: <https://doi.org/10.1101/2022.11.07.22281877>; this version posted November 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

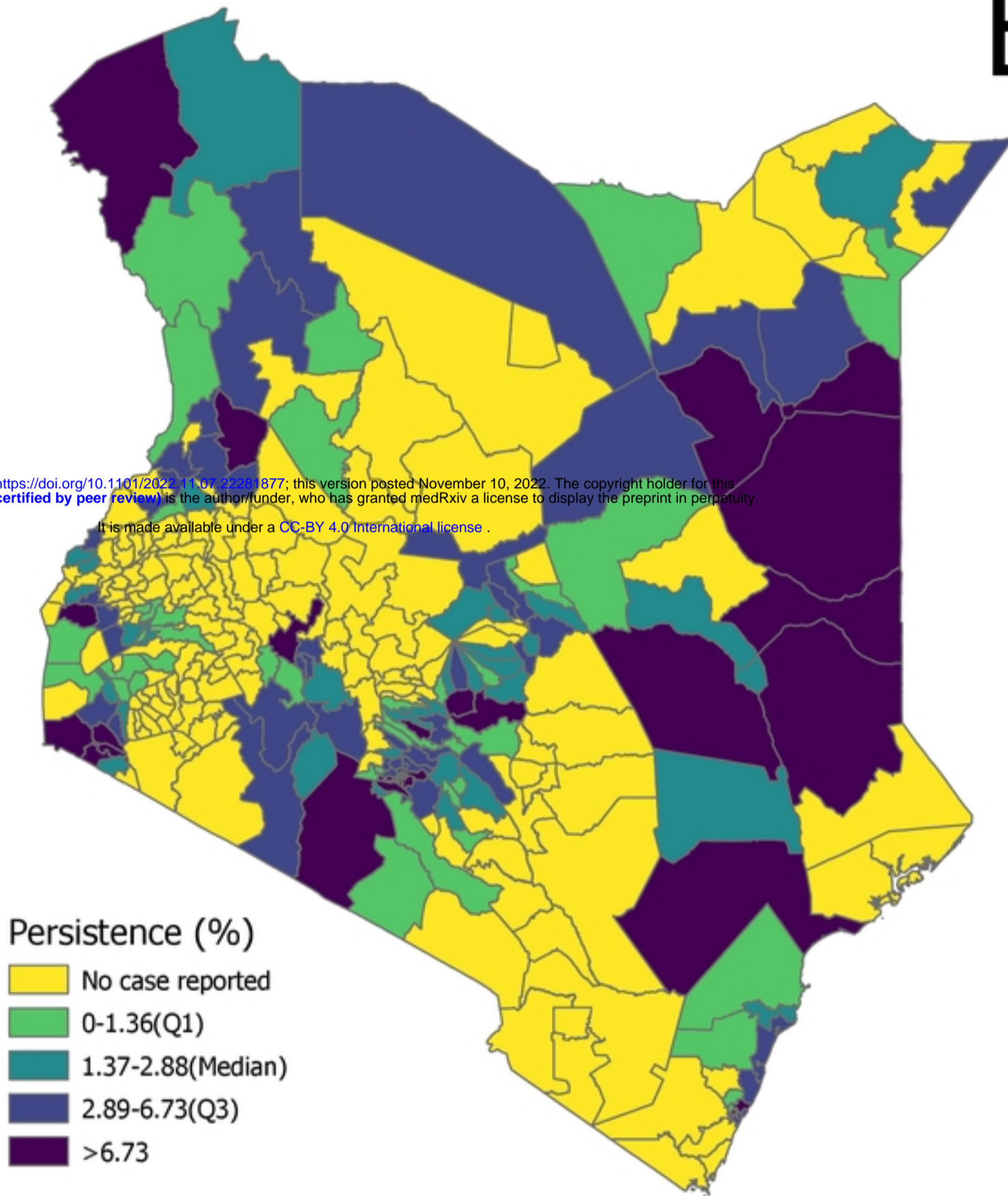


Figure 3b

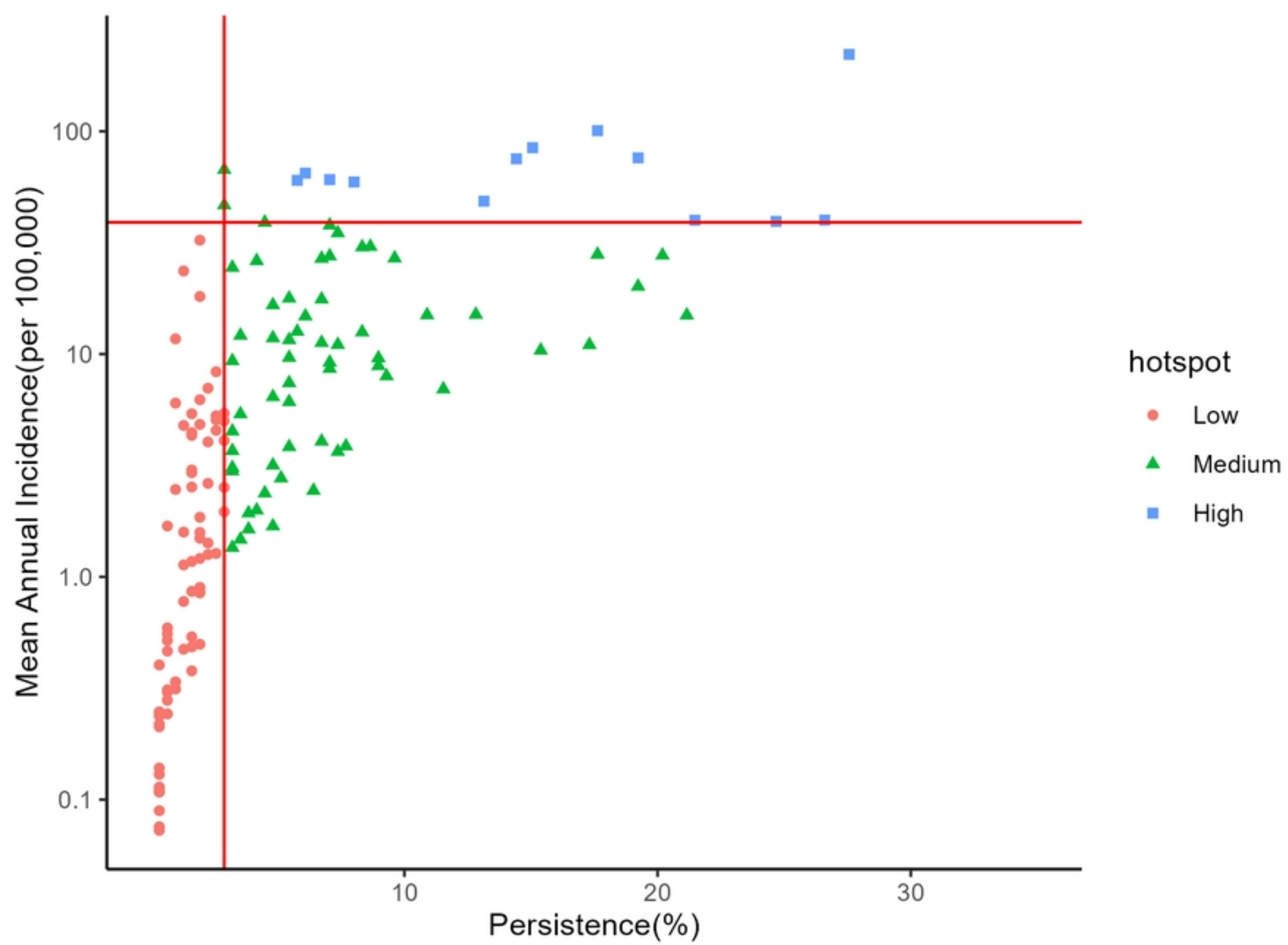


Figure 3c

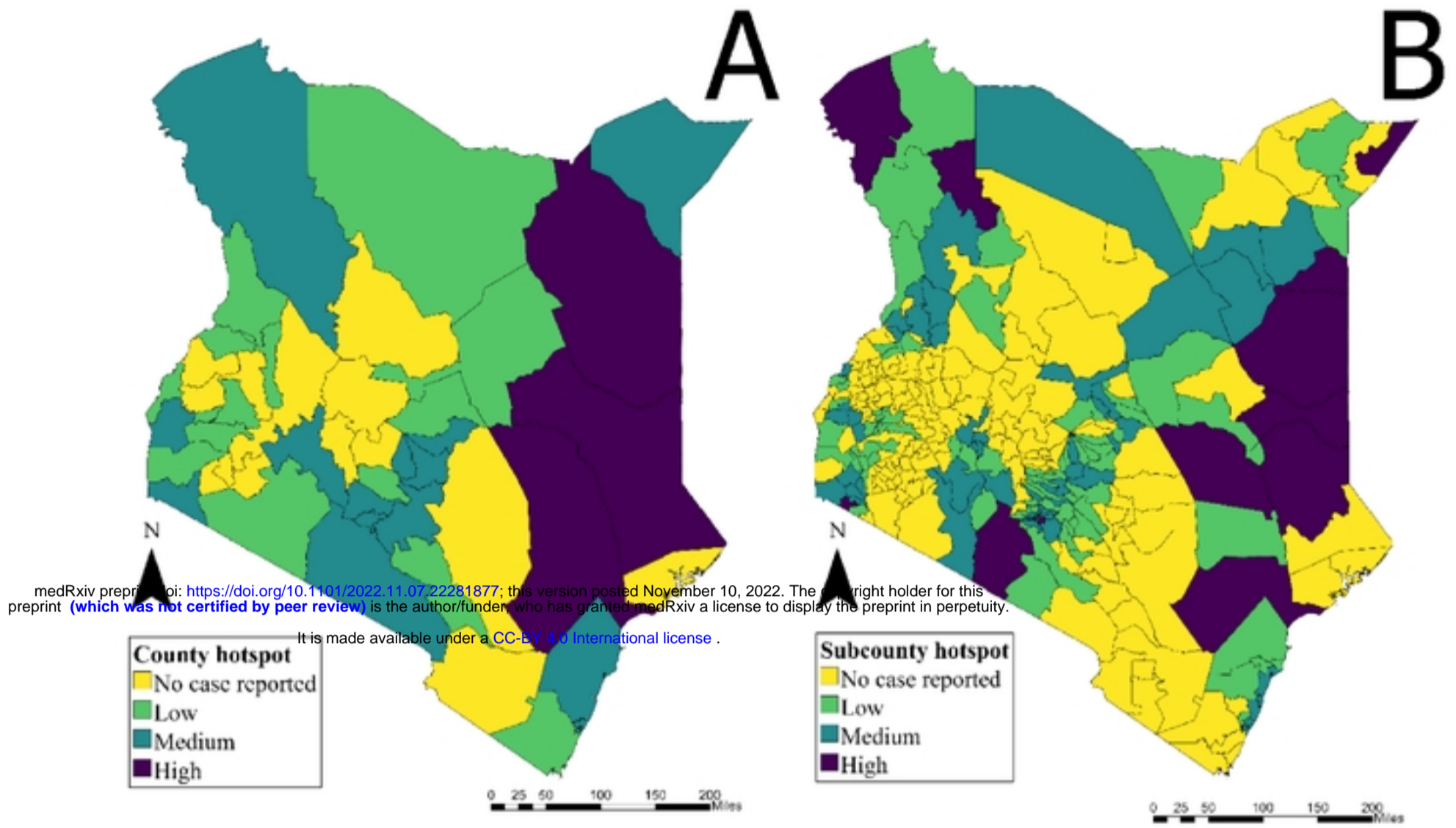


Figure 4