Flexible scan statistic with a restricted likelihood ratio for optimized COVID-19 surveillance

Ernest Akyereko,[1,2] Frank B. Osei,[1] Kofi M. Nyarko,[2] Alfred Stein[1]

[1]Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, The Netherlands; [2]University of Environment and Sustainable Development, PMB, Somanya, ER, Ghana

Abstract

Disease surveillance remains important for early detection of new COVID-19 variants. For this purpose, the World Health Organization (WHO) recommends integrating of COVID-19 surveillance with other respiratory diseases. This requires knowledge of areas with elevated risk, which in developing countries is lacking from the routine analyses. Focusing on Ghana, this study employed scan-statistic cluster analysis to uncover the spatial patterns of incidence and Case Fatality Rates (CFR) of COVID-19 based on reports covering the four pandemic waves in Ghana between 12 March 2020 and 28 February 2022. Applying flexible spatial scan statistic with restricted likelihood ratio, we examined the incidence and CFR clusters before and after adjustment for covariates. We used distance to the epicentre, proportion of the population aged ≥ 65, male proportion of the population and urban proportion of the population as the covariates. We identified 56 significant spatial clusters for incidence and 26 for CFR for all four waves of the pandemic. The Most Likely Clusters (MLCs) of incidence occurred in the districts in south-eastern Ghana, while the CFR ones occurred in districts in the central and the north-eastern parts of the country. These districts could serve as sites for sentinel or genomic surveillance. Spatial relationships were identified between COVID-19 incidence covariates and the CFR. We observed closeness to the epicentre and high proportions of urban populations increased COVID-19 incidence, whiles high proportions of those aged ≥ 65 years increased the CFR. Accounting for the covariates resulted in changes in the distribution of the clusters. Both incidence and CFR due to COVID-19 were spatially clustered, and these clusters were affected by high proportions of the urban population, high proportions of the male population, high proportions of the population aged ≥ 65 years and closeness to the epicentre. Surveillance should target districts with elevated risk. Long-term control measures for COVID-19 and other contagious diseases should consider improving quality healthcare access and measures to reduce growth rates of urban populations.

Correspondence: Ernest Akyereko, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, The Netherlands.
Tel.: +31685310183
E-mail: e.akyereko@utwente.nl

Key words: flexible scan statistics, COVID-19, general linear model, Ghana.

Conflict of interest: the authors declare no conflict of interest.

Contributions: EA, conceptualization, methods, analysis, writing the initial draft. FB, AS & KN, investigation, supervision, validation, writing, reviewing and editing.

Availability of data and materials: all data generated or analyzed during this study are included in this published article.

Received: 4 January 2024.
Accepted: 9 August 2024.

Geospatial Health 2024; 19:1265
doi:10.4081/gh.2024.1265

Introduction

Measures, such as ban on travel, partial lockdown and closing schools, public markets and private workplaces, were introduced in the wake of the coronavirus disease 2019 (COVID-19) being declared a pandemic by the World Health Organisation (WHO) in early 2020 (https://www.who.int/europe/emergencies/situations/covid-19). This crippled several global economies and left a lasting fear of re-emergence of the disease as the risk of new variants could cause new surges of the disease (Adebowale *et al.* 2021). Following the reduced number of new infections and deaths reported by May 2023, WHO took the disease off the list of public health emergencies of international concern (Jacqui, 2023). However, it recommends the integration of COVID-19 surveillance with other respiratory diseases and the institution of geographically representative genomic surveillance as strong and timely surveillance is critical for the early detection of any changes in incidence and death numbers (WHO 2022). Similar to other African countries, the Ghana Health Service (GHS) monitors such situations by sentinel surveillance, a strategy requiring identification of areas with a higher risk of the disease for targeted surveillance (GHS, 2020). Similar to the surveillance of influenza-like-illness in Ghana, facilities situated in high-risk areas can be designated to periodically obtain representative clinical samples for testing and genomic sequencing (GHS 2020). According to the European Centre for Disease Prevention and Control, this could find new variant in circulation and potential reservoirs

(ECDC, 2021). For the successful integration of COVID-19 in a general surveillance system, it is important to understand the underlying spatial and temporal patterns.

So far, Ghana has recorded over 130,000 COVID-19 cases and 1,400 deaths detected as counts across all districts utilizing routine indicator-based surveillance (GHS, 2022). This approach has limited spatial resolution, which results in considerable informative limitation. Spatial epidemiological methods such as cluster detection offer an effective way of uncovering the underlying patterns for targeted intervention and surveillance (Arab-Mazar *et al.*, 2020; Zu *et al.*, 2020; Fatima *et al.*, 2021; Huang *et al.*, 2021; Bermudi *et al.*, 2021a; Siljander *et al.*, 2022). In the context of surveillance, detecting districts with elevated risk of COVID-19 incidence and Case Fatality Rates (CFR) could identify changes in viral characteristics and virulence. Hence, it might provide information on sites that are the most useful for representative sampling for genomic sequencing.

This study aimed to identify clusters of small areas (districts) with elevated COVID-19 risk that could support precision surveillance. To do so, we used the spatial scan statistic method proposed by Kulldorff (Alves *et al.*, 2021; Paul *et al.*, 2021; Siljander *et al.*, 2022), which is powerful and relatively easy to implement, although restricted by only detecting circular and elliptical-shaped clusters (Toshiro & Kunihiko, 2005). We adapted the flexible scan statistic with restricted likelihood ratio (Toshiro & Kunihiko, 2012) that would allow us to detect clusters of varying shapes.

Spatial cluster detection without accounting for potential confounders can only tell part of the story of spatial trends. An important aspect of disease cluster analysis is to account for confounders affecting the clustering behaviour. Studies have associated the COVID-19 pandemic with overcrowding and emphasized its prevalence in big cities (Chen *et al.*, 2020). Therefore, an additional objective was to determine how variation of the proportion of certain variables affect the clustering, *i.e.* proximity to epidemic centres, urban concentration of people, the numbers of males and those aged ≥ 65 years. This should provide valuable information to guide health professionals and governments in developing and implementing interventions.

Materials and Methods

Study area

Ghana is located in West Africa, with an area of approximately 240,000 km^2 and an estimated population of about 30.42 million. The country is divided into 16 administrative regions and 260 districts as of the year 2020 (Figure 1). Health infrastructure and implementation of health interventions are centred at the district level, with five teaching hospitals located in Volta, Greater Accra, Ashanti, Northern and Central regions, in addition to 16 regional hospitals provide tertiary care. These facilities serve as referral centers for the lower facilities at the district level. The national capital Accra is located in the south-eastern part of the country and hosts much of the genral infrastructure including the international airport.

Data acquisition

Anonymized data on reported COVID-19 cases and deaths in Ghana from March 2020 to February 2022 were obtained from GHS. The variables: date of sample taken, district of residence, outcome, such as recovery or death (including date of death) were extracted. Based on the date of each sample, the temporal pattern of the pandemic was constructed. It appeared as Wave 1 (W_1), from 12 March to 30 September 2020, Wave 2 (W_2) from 1 October

2020 to 22 April 2021, Wave 3 (W_3) from 23 April to 30 November 2021 and Wave 4 (W_4) from 1 December to 28 February 2022 (Figure 2). In line with Lawton *et al.* (2021) we considered the following variables affecting the distribution: the urban population, the proportion of males, the proportion of those aged ≥ 65 years and the distance from Greater Accra (the epicentre). The proportions of the population living in urban settings in the districts were obtained from the Ghana Statistical Service (2021).

Cluster detection

Flexible scan statistic with restricted likelihood ratio (FSRLR)

following the GHS disease surveillance report structure since specific locations are not routinely monitored (GHS, 2020), district-level reported cases $y_a(i,t)$ or deaths $y_b(i,t)$ were aggregated for each district $i = i, \ldots, 260$ to correspond to the four major waves $t = 1, \ldots, 4$. For each combination of district and wave, we represented our counts of COVID-19 cases and deaths as an offset based on realizations from the Poisson distribution according to an underlying population at risk, $N(i,t)$. Poisson-based spatial scan statistic is widely employed to detect areas with elevated risk (Alves *et al.* 2021; Franch-Pardo *et al.*, 2020; US Department of Health and Human, and Centers for Disease Control and Prevention, 2006; Islam *et al.,* 2021). Under the null hypothesis of no clustering, H_0, the expected null counts for incidence and death were then H_0: $E[y_a(i,t)] = E_a(i,t)$ and H_0: $E[y_b(i,t)] = E_b(i,t)$, respectively. In this scenario, spatial scan statistics detects potential clusters by passing a circular window, Z, over the entire district to detect clusters of variable sizes in proportion to the total population that might be at risk (Kulldorff 1997), recent applications to COVID-19 (Alves *et al.*, 2021; Islam *et al.,* 2021). Although powerful, this method falls short of correctly identifying non-circular clusters (Otani & Takahashi, 2021).

Since the 260 districts in Ghana all have an irregular shape, flexible scan statistic would be preferable due to the possibility of applying a large number of irregularly shaped windows on each district shaped by connection to adjacent districts (Toshiro & Kunihiko, 2005). To avoid absorbing surrounding areas with low risk, the likelihood ratio is restricted so that scanning is limited to districts with elevated risk (Tango & Takahashi, 2012). Using the *rflexscan* package in R, the upper limit is determined by specifying the maximum length, of its nearest neighbours (K=50). Let $\boldsymbol{Z}(i,k_i,t)$ be the set of windows for the i^{th} district composed of its k_i nearest neighbours for wave t. Let $Z_j(i,k_i,t)$ denote the j^{th} window for the district i, and the set of its nearest neighbours k_i, $k_i = 1, \ldots, K$. Here j satisfies $Z_j(i,k_i,t)$ Î $\boldsymbol{Z}(i,k_i,t)$. For each scanning window, the null hypothesis of no clustering within the window is evaluated and compared to that outside. A log-likelihood ratio test is performed and ranked to detect primary and secondary clusters (Otani and Takahashi 2021). For instance, for the COVID-19 incidences, the null hypothesis of no clusters is specified as H_0: $E[(i,t)] = e_a(Z_j)$ and the alternate hypothesis of clusters is H_i: $E[y_a(Z_j)] > e_a(Z_j)$. Here, the expected number of cases is computed based upon the fact that the risk of contracting COVID-19 is fixed for a given wave. Thus, the fixed risk for a given wave is given by the equation:

$$\hat{r}_a(t) = \frac{\sum_{i=1}^{N} y_\alpha(i,t)}{\sum_{i=1}^{N} N_\alpha(i,t)}$$

For a given window, Z_j, the expected number of cases equals:

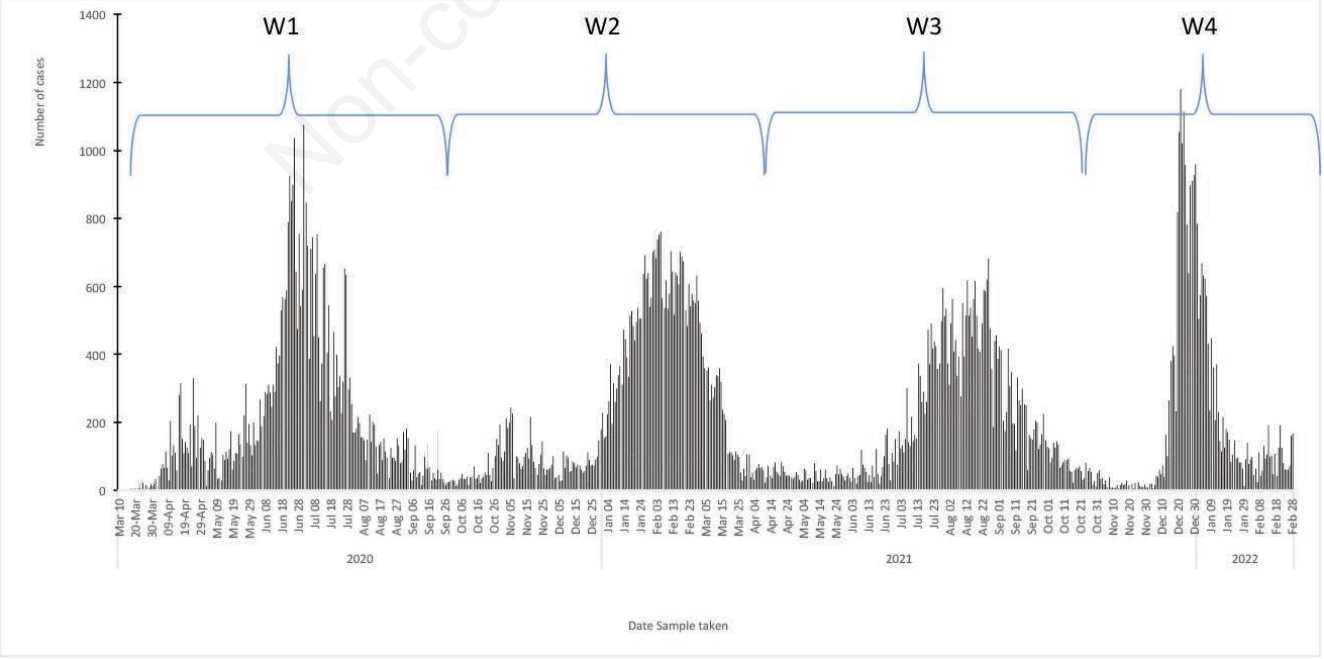


Figure 2. Distribution of COVID-19 cases in Ghana by date of a sample taken, identified as Waves W_1, W_2, W_3, and W_4.

$$e_a(Z_j) = N_a(Z_j)\hat{r}_a(t)$$

where $N_a(Z_j)$ is the population within that window. The likelihood of observing the recorded number of cases within and outside the window Z_j is given by the formula:

$$\sup_{Z_j \epsilon \boldsymbol{Z}} \left(\frac{y_a(Z_j)}{e_a(Z_j)}\right)^{y_a(Z_j)} \left(\frac{y_a(Z_j^c)}{e_a(Z_j^c)}\right) I\left(\frac{y_a(Z_j)}{e_a(Z_j)} > \frac{y_a(Z_j^c)}{e_a(Z_j^c)}\right)$$

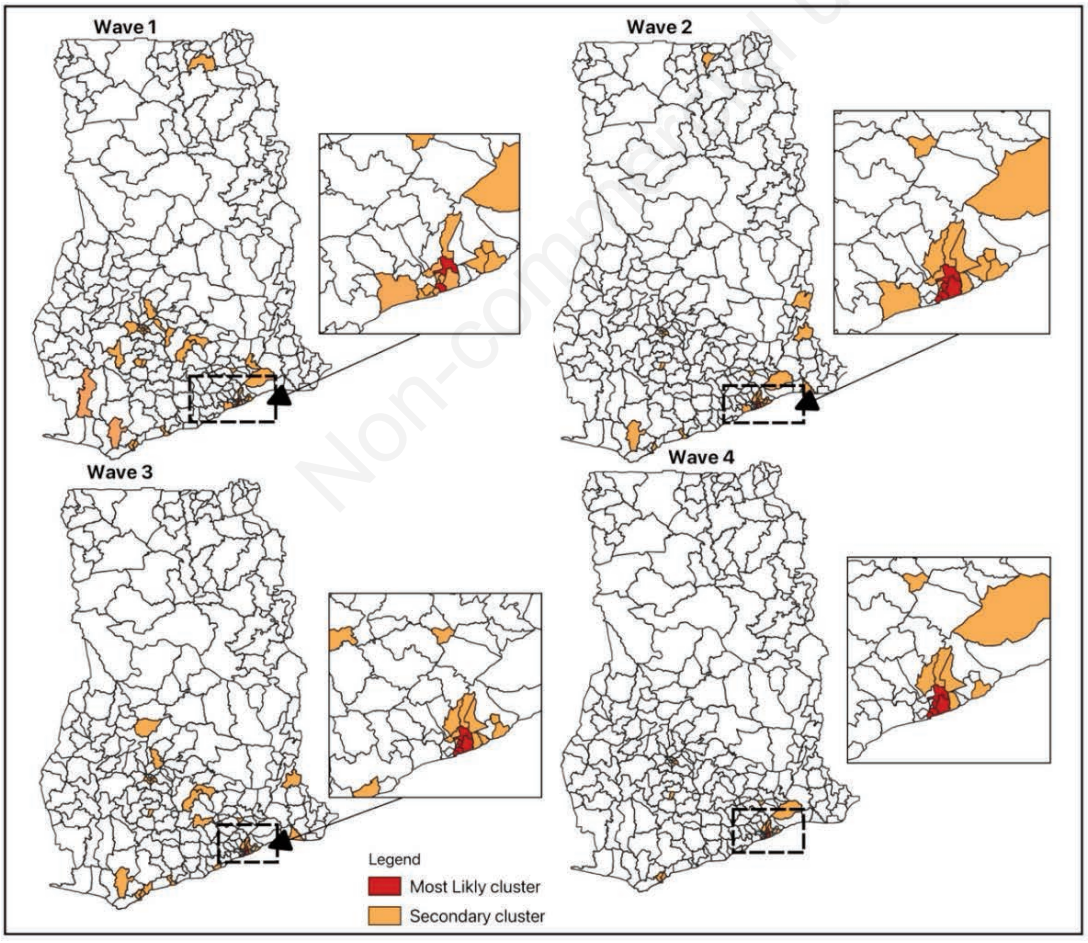
where Z^c_j represent the complement of Z_j, $y_a(Z_j)$ the observed number of cases within the specified window Z_j, $e_a(Z_j)$ is the null expected number of cases within the specified window; and $I(x)$ is the indicator function, with $I(x)$ for $x = TRUE$ and 0 otherwise. The window ($\boldsymbol{Z}$) that reaches the maximum likelihood is defined as the Most Likely Cluster (MLC). The significance of the clusters was evaluated using 999 random Monte Carlo simulations to generate the corresponding p-values. The null hypothesis of no clustering was rejected if p <0.05. A detailed description of the FSRLR is provided here (Tango & Takahashi 2012; Otani & Takahashi 2021). As shown in the Results section, the FSRLR can be visualized using choropleth maps. Again, a matrix was developed to visualise the detected cluster along the four waves. In the matrix, the distribution of the incidence and CFR were classified into three groups. A district with a cluster in all four waves was classified as 'emerged', a district detected as a cluster in more than one wave but not directly following each other was classified as 're-emerged', while a district with a cluster continuing into the subsequent wave was classified as 'persisted'.

Covariate adjustment

Adjustment for covariate effects is not straightforward. Kulldorff proposed the use of a covariate-adjusted expected number of cases rather than those from the null hypothesis (Kulldorff 2021). We considered the distance to the epicentre (Greater Accra Region), the proportion of the male population, the proportion of the population aged ≥ 65 years and the percentage of the urban population as possible factors influencing the distribution pattern. To do so, we fitted a Poisson log-linear model

$$y_a(i,t) \sim Pois(\lambda_a(i,t))$$

where the mean equals $\lambda_a(i,t) = N_a(i,t)\, r_{a,adj}(i,t)$ and the unknown



DID	W1	W2	W3	W4	classification
1					emerged
3					emerged
7					persisted
8					persisted
14					persisted
33					emerged
37					emerged
40					persisted
43					persisted
59					emerged
62					re-emerged
63					persisted
64					persisted
65					persisted
79					persisted
84					emerged
91					persisted
96					emerged
102					persisted
103					emerged
112					persisted
121					persisted
123					emerged
129					emerged
141					persisted
148					emerged
149					persisted
152					persisted
158					persisted
159					persisted
160					persisted
163					persisted
164					emerged
177					persisted
180					emerged
189					persisted
190					emerged
193					persisted
194					persisted
196					emerged
207					persisted
212					persisted
215					persisted
216					emerged
227					emerged
231					persisted
235					persisted
236					re-emerged
250					emerged
252					re-emerged

Figure 3. Spatial distribution of District level detected clusters for COVID-19 incidence in Ghana.

covariate-adjusted risk on the log scale is modelled as

$$\log r_{\alpha,adj}(i,t) = \beta_{0\alpha} + \beta_{dist,\alpha}(t)x_{dist}(i,t) + \beta_{up,\alpha}(t)x_{up}(i,t) + \beta_{aged\,\square\,65,\alpha}(t)x_{aged\,\square\,65}(i,t) + \beta_{male,\alpha}(t)x_{male}(i,t)$$

where $\beta_{0\alpha}$ is the intercept; $\beta_{dist,a}(t)$ a time-varying effect of the distance from the epicentre; $x_{dist}(i,t)$, $\beta_{upa}(t)$ the time-varying effects of the percentage of urban population; $x_{up}(i,t)$, $\beta_{aged \geq 65,a}(t)$ the time-varying effect of the proportion of male population; and $x_{male}(i,t)$, $\beta_{aged \geq 65,a}(t)$ the time-varying effect of the proportion of the population aged ≥ years $x_{aged \geq 65}(i,t)$. This model was fitted using the *glm* function in the R statistical software. For each district *i* and wave *t*, the covariate-adjusted risk is predicted by the formula:

$$\hat{r}_{\alpha,adj}(i,t) = \exp\left(\hat{\beta}_{0\alpha} + \hat{\beta}_{dist,\alpha}(t)x_{dist}(i,t) + \hat{\beta}_{up,\alpha}(t)x_{up}(i,t) + \hat{\beta}_{aged\,\square\,65,\alpha}(t)x_{aged\,\square\,65,\alpha}(i,t) + \hat{\beta}_{male,\alpha}(t)x_{male}(i,t)\right)$$

where $\hat{\beta}_{0\alpha}$, $\hat{\beta}_{dist,a}(t)$, $\hat{\beta}_{malea}(t)$, $\hat{\beta}_{aged \geq 65,a}(t)$, and $\hat{\beta}_{upa}(t)$ are the estimates for the parameters $\beta_{0\alpha}$, $\beta_{dist,a}(t)$, $\beta_{malea}(t)$, $\beta_{aged \geq 65,a}(t)$ and $\beta_{upa}(t)$, respectively. The adjustment for the expected number of cases is straightforward; for a given window Z_j, the covariates-adjusted expected cases $e_{a,adj}(Z_j)$ $\hat{r} = N_a(Z_j)$ $_{a,adj}(i,t)$
We demonstrated this using the COVID-19 cases $y_a(i,t)$ and repeating the same approach for the COVID-19 deaths. Next, we implemented these adjustments in the FSRLR to control for the covariates. To implement these adjustments for the scan statistic, we replaced the population variable with the adjusted expected counts.

Results

COVID-19: spatial distribution of district-level clusters and relative risk

The FSRLR detected significant clusters for all four waves (W_1 (8), W_4 (7), W_3 (6), and W_4 (4)). The MLCs were detected in the Southeast (Greater Accra region) and persisted throughout the four waves (Figure 3). For the MLCs, the relative risk (RR) of COVID-19 incidence was 8.85 (p<0.001), 22.54 (p<0.001), 20.78 (p<0.001) and 32.10 (p<0.001) for waves W_1, …, W_4, respectively (Supplementary materials Table1).

At the next level, the study identified SCs with RRs ranging from 1.1 to 6.23 (p<0.001)in the southern, central and north-eastern parts. An SC persisted in the centre (Ashanti region) throughout the whole period (Figure 3). Several persistent SCs were identified in the Southwest and the East. Persistent SCs were identified in the Southeast during W_2 and W_3, which shared a border with Togo (Volta region).

CFR: spatial distribution of district-level clusters

District-level clusters of CFR at p<0.05 were detected during all four waves (W_1 (3), W_4 (3), W_3 (3) and W_4 (3)). The spatial patterns of MLCs took relative risk values equal to RR = 2.12 for W_1, and for W_2, persisted with values at 5.87 and 8.50 during W_3 and W_4, respectively, (Supplementary materials Table 2). During W_1, the MLC was detected in the centre (Ashanti Region) whereas in the 2nd wave, it was in the northern part of Ghana (Upper East, Upper West, Savana, Northern and North-East regions). During W_3 and W_4, the MLC persisted in the central part of Ghana towards the West, involving regions like Ashanti and Bono.

FSRLR also detected SCs during all four waves. During W_1, SC was detected in the centre, at just one district distance from the MLC. During W_2, the SCs persisted mainly in the centre (Ashanti, Bono, Bono East and Ahafo). During W_3, areas in the Northwest (Upper Western region) and centre (Ashanti) persisted as SCs. In the western part of the centre (Bono and Bono East region), an SC during W_2 developed into an MLC during W_3 with increased RRs 3.74 (p<0.001) to 5.87 (p<0.001)) (Supplementary materials Table 2). These detected clusters persisted into W_4 with the re-emergence of SCs in the Northeast and Northwest.

GLM performance of variables explaining variation in incidence and CFR distributions

Our initial scan statistics results show that the south was particularly susceptible to the COVID-19 infections (Figure 3). Considering the urban population, the proportion of the male population, the proportion of the population aged ≥ 65 years and distance to the epicentre as explanatory variables, we fitted a Poisson-GLM to explain the observed distribution of the incidence and case fatalities. For all epidemic waves, the incidence rate was positively related to the urban population, while it was negatively related to the epicentre distance (Greater Accra Region) and the proportion of population aged ≥ 65. During W_1, we observed that one unit increase in the urban population, the proportion of male population became associated with an increased RR of 2.5346 and 2.35 (p<0.001), whereas the proportion of population aged ≥ 65 years reduced the incidence rate by 1.28 (p<0.001) (Table 1). During W_2, W_3 and W_4, a similar relationship was observed between the incidence rate and urban population with their estimated coefficients increasing as the epidemic wave progressed. The proportion of population aged ≥ 65 years was negatively associated with COVID-19 incidence in W_2-7.09 (p<0.001), W_3 -2.0 (p<0.001) and W_4 -1.10 (p<0.001). In the 2nd and 4th wave, an increase in the proportion of male population was found to reduce the COVID-19 incidence.

For CFR, the urban population, the proportion of population aged ≥ 65, and the distance to the epicentre (Greater Accra Region) were positively associated with deaths (Table 1). Significant associations between urban population and CFR were detected during all waves (Table 1). The coefficients for urban population of 3.72 (p<0.001), 4.17 (p<0.001), 2.91 (p<0.001) and 3.88 (p<0.001) were observed for W_1, W_2, W_3 and W_4. A higher positive association was found between the proportion of population aged ≥ 65 years and the CFR in all pandemic waves (W_1 (2.00, p<0.001), W_2 (5.06, p<0.001), W_3 (3.86 p<0.001), W_4 (4.59 p<0.001)).

Spatial distribution of clusters after adjusting for covariates

The Poisson log-linear model showed a significant association between urban population, the proportion of population aged ≥ 65, the proportion of males, distance to the epicentre both incidence and CFR. Therefore, results from the FSRLR determining the distribution of incidence and CFR were obtained after adjusting for these covariates. Significant clusters (p<0.05) were detected during all four waves (W_1 (8), W_2 (6), W_3 (8) and W_4 (5)). MLCs were detected in two districts during W_1 (RR = 18.21, p<0.001), nine districts during W_2 (RR = 11.77, p<0.001), eight districts in W_3 (RR

= 11.17, p<0.001) and five districts during W_4 (RR = 18.52, p<0.001). The MLCs persisted in the South during all four waves similar to what we observed before covariate adjustment, but with a lower RR (11.17 – 18.52, p<0.001). The number of secondary clusters was also higher after covariate adjustment (Figures 3 and 5). Additional SCs were mostly located in the centre (Bono, Bono East and Ashanti districts) during W_1 and W_3. Most clusters persisted across the four epidemic waves with few emerging and re-emerging (Figure 5).

In total, 10 clusters were identified during all four waves of the pandemic after adjusting for urban population, the proportion of population aged ≥ 65, the proportion of males and distance to the epicentre. Out of these, four districts persisted as significant CFR clusters (Figure 6). During all pandemic waves, the MLC (RR = 7.04 – 9.03, p<0.001) were predominantly found in the in the South (Greater Accra, Volta and eastern regions). This was different from what was detected without covariate adjustment (Figures 2 and 4). The MLCs (RR = 7.04, p<0.001) during W_2 were found in the Southeast (Volta Region) after controlling for the covariates. Additional SCs were detected in the centre (Ashanti Region), which persisted throughout the pandemic waves (Figure 6). During W_4, the MLC was detected in the South (Eastern Region) whereas a SC persisted in the centre (Bono and Ashanti Regions).

Discussion

Despite the current low number of COVID-19 cases in Ghana, the risks of its severity and increase in cases remain eminent. Strong disease surveillance remains critical to understanding the evolution of the disease pattern (WHO Africa Region 2019; Kaburi *et al.* 2017). The call by WHO to integrate COVID-19 surveillance into the respiratory disease (WHO 2022) leaves countries like Ghana to identify possible high-risk areas for targeted surveillance and to improve sampling for genomic sequencing of the virus (Jacqui 2023). Application of the FSRLR before and after controlling for the covariate detected statistically significant clusters of high incidences during all four waves. Most of these clusters persisted in the South. During the first wave, Ghana instituted a partial lockdown together with other measures in Greater Kumasi and Accra in the South (Owusu *et al.*, 2020). These measures were expected to influence the distribution of the clusters during subsequent waves. Therefore, the persistent clusters most likely resulted from factors such as low level of adherence to the COVID-19 measures, and person-to-person contacts propagated by community interaction. In Ghana, there is a continuous migration from the poor North to the more developed South. This comes with challenges such as poor housing and increased risk of disease, which predominantly characterizes the South (Ghana Statistical Service 2021). Since COVID-19 is primarily transmitted in an enclosed place, the partial lockdown measures could have generated a unique transmission path such as increased contact within compound homes, possibly increasing the risk (Yi *et al.*, 2020; Zhang *et al.*, 2023; Alves *et al.*, 2021). Therefore, long-term disease response measures should take into consideration the unique Ghanaian setting when implemented. They should include improving housing in urban areas to reduce their vulnerability to infectious diseases. In the presence of these transmission dynamics, districts with persistently high disease incidence or CFR are particularly suited for targeted surveillance. Additionally, given the limited available resources, prioritization of health interventions such as surveillance, health education, restriction of movement and development of standards and disinfection, should be location-specific.

CFR is not only a disease severity measure (Suleiman *et al.*, 2021), but provides also a proxy estimate of the health system's

Table 1. GLM performance of covariates on COVID-19 outcomes.

Wave	Variable	Incidence estimate	p-value	CFR estimate	p-value
W1	(Intercept)	-1.90	0.001	-2.21	0.001
	Urban population	2.53	0.001	3.72	0.001
	Epicentre distance (km)	-1.70	0.001	1.46	0.604
	Male population	2.35	0.001	1.51	0.005
	Population ≥ 65 years	-1.28	0.001	2.00	0.003
W2	(Intercept)	-2.14	0.001	-1.67	0.001
	Urban population	3.67	0.001	4.17	0.001
	Epicentre distance (km)	-2.65	0.001	2.10	0.001
	Male population	2.64	0.001	-2.36	0.996
	Population ≥ 65 years	-7.09	0.001	5.06	0.001
W3	(Intercept)	-7.07	0.001	-1.08	0.001
	Urban population	3.02	0.001	2.91	0.001
	Epicentre distance (km)	-2.80	0.001	9.86	0.001
	Male population	-1.05	0.078	-8.84	0.114
	Population ≥ 65 years	-2.00	0.001	3.86	0.001
W4	(Intercept)	1.78	0.001	18.88	0.001
	Urban population	3.49	0.001	3.88	0.001
	Epicentre distance (km)	-3.58	0.001	1.76	0.001
	Male population	-4.70	0.001	3.31	0.001
	Population ≥ 65 years	-1.10	0.001	4.59	0.001

W, epidemic wave; CFR, case fatality rate.

ability to manage the prevailing situation (Zhang *et al.*, 2023). Unlike incidence, our findings indicate persistent clusters of higher RR in the central part of Ghana, during the first, third and fourth wave. The identified clusters occurred in the centre of the country and gradually moved towards the North during the second wave. The first case in Ghana was reported in the Greater Accra Region located in the South. The first wave spanned between 12 March and 30 September 2020 during which many districts in the North had zero reported cases likely accounting for low fatalities. Also, its progression to other parts of the country may have been slowed down by the partial lockdown implemented in the Greater Kumasi (in the centre) and the Greater Accra Region (in the South). As fatalities arise from incidences, it likely affected the distribution of the CFR clusters. Conversely, despite clusters of high incidences in the South (Greater Accra Region) there were no districts with higher RR for CFR. This could be due to the high number of reported cases in this region as CFR estimation uses the total number of reported cases as denominator. Additionally, most management resources were located in the Greater Accra Region in the South. Key health facilities such as Korlebu Teaching Hospital, University of Ghana Medical Centre, Ga East Hospital and Ghana Infectious Disease Centre are all located in the South. Therefore, the low CFRs could be due to better management of cases compared to other areas. Another factor affecting the CFR distribution is uneven testing rates, health-seeking behaviour and accessibility to health care that are better in the South. Actions aiming to reduce CFR during pandemics should consider improving measures regarding healthcare access. Higher CFR could also reflect changes in the virus's characteristics helping the virus evade the immune system and thus result in high fatalities. High CFRs should also be seen as a potential early warning for viral mutation and direct targeted surveillance towards areas where detected.

The influence of urbanization, proximity to infectious sources, age, and sex distribution on COVID-19 have been documented elsewhere (Bashir *et al.*, 2020; Huang *et al.*, 2021; Aral and Bakir 2022). Our analysis indicated a significant association between the COVID-19 distribution in Ghana and these covariates. We identified a reduction in the number of reported cases with an increase in the distance to the epicentre (Greater Accra Region) located in the south, together with a reduction in the RR, with regard to the proportion of the population aged ≥ 65 years and the urban population. Our study also detected a persistent MLC and SC in the centre (Bono and Ashanti regions) that in most districts persisted from W_1 into W_4. Such a pattern could be influenced by low capacity to manage COVID-19 cases propagated by an uneven distribution of health resources. Hence, an increase in resource distribution to these areas is advocated.

COVID-19 is a systemic disease with preferential symptoms

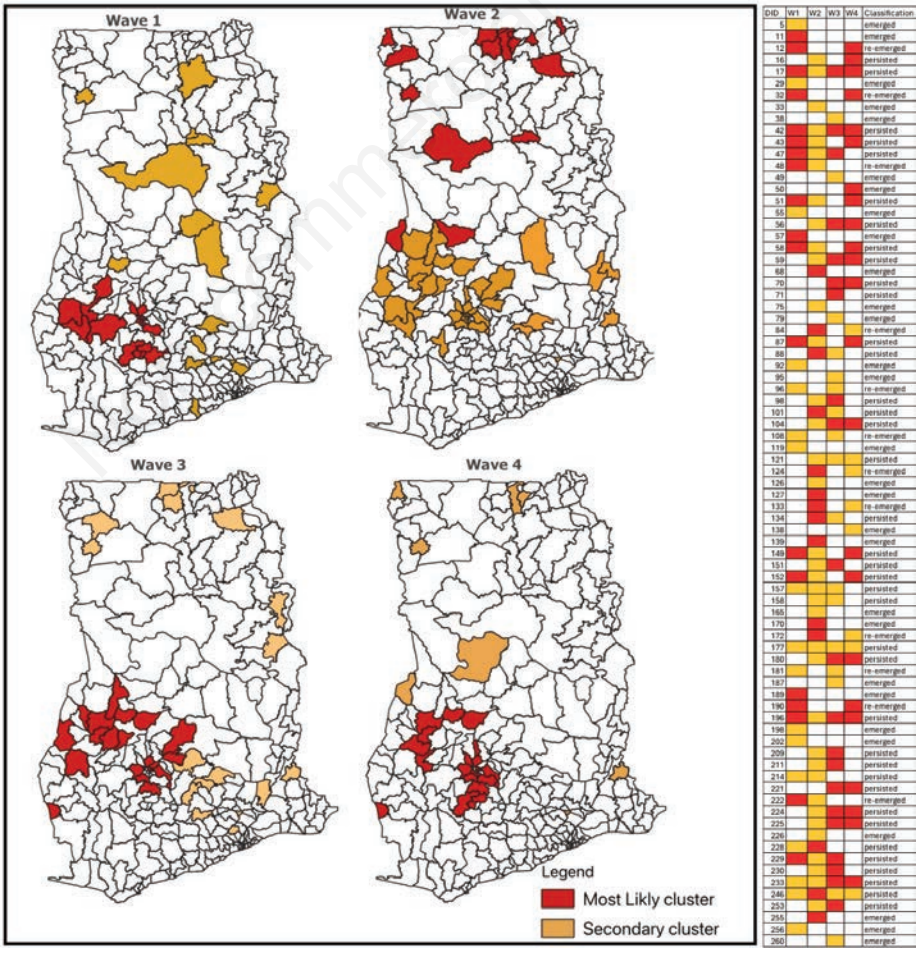


Figure 4. Spatial distribution of district-level detected cluster for COVID-19 CFR.

from the respiratory tract whose reproduction rate rests on direct human interactions. Therefore, high urbanization and close distance to the epicentre increase the risk of infection. Again, the observed reduction in cases when the proportion of the population aged ≥65 increases could be due to their low mobility and therefore less contact with infected individuals. For the CFR, the urban population, the proportion of aged ≥65 years and the distance from the epicentre increased this risk. The reason is that longer distances correspond to low healthcare infrastructure, whereas a higher urban population could mean higher competition for limited health resources (Fatima *et al.,* 2021). Therefore, programs to reduce rural-urban migration and decentralization of health resources should include long-term measures to reduce CFR during pandemics. Again, people aged ≥ 65 years have lower immunity and higher susceptibility to comorbidities, such as diabetes and cardiovascular disease that both increase the risk for fatality due to COVID-19 (Owusu *et al.*, 2020). Therefore, future interventions should consider program that protect these vulnerable populations.

The study found that, more districts with significantly high incidence were detected after adjusting for covariates. These additional districts included some districts in the North that were not detected. This means that transmission in these areas could be higher if new infections occur in closer proximity. During the 2nd wave, for instance, more secondary clusters were identified in the Southeast (Volta Region) which persisted during the 3rd Wave. These districts share a border with Togo with many approved and unapproved entry routes. Consequently, the district with a persistent secondary cluster after taking out the covariates could also be a pointer to the district's interaction with bordering countries. Therefore, such districts could be good sites for sampling to monitor viral mutations and imported COVID-19 cases. It also points to taking more international measures, than country-specific ones. Concerning to CFRs, fewer districts with higher RR were identified in the northern part after adjustment. During the 2nd wave, for instance, 15 districts detected as MLCs with regard to CFR in the North disappeared with few SCs after accounting for the covariates. This highlights the possible effect of low-quality healthcare accessibility in these districts. Improved resource allocation could limit the high number of deaths in these areas (Zhang *et al.*, 2023; Bermudi *et al.*, 2021b; Rodriguez Velásquez *et al.*, 2021). The cluster detected in the Northeast shares a border with Burkina Faso; therefore, the high CFR could also be due to the transmission dynamic along the border, which calls for international rather than national measures. In addition, such districts should be considered as surveillance sites for genomic representative sampling which could help identifying any changes and imported cases early.

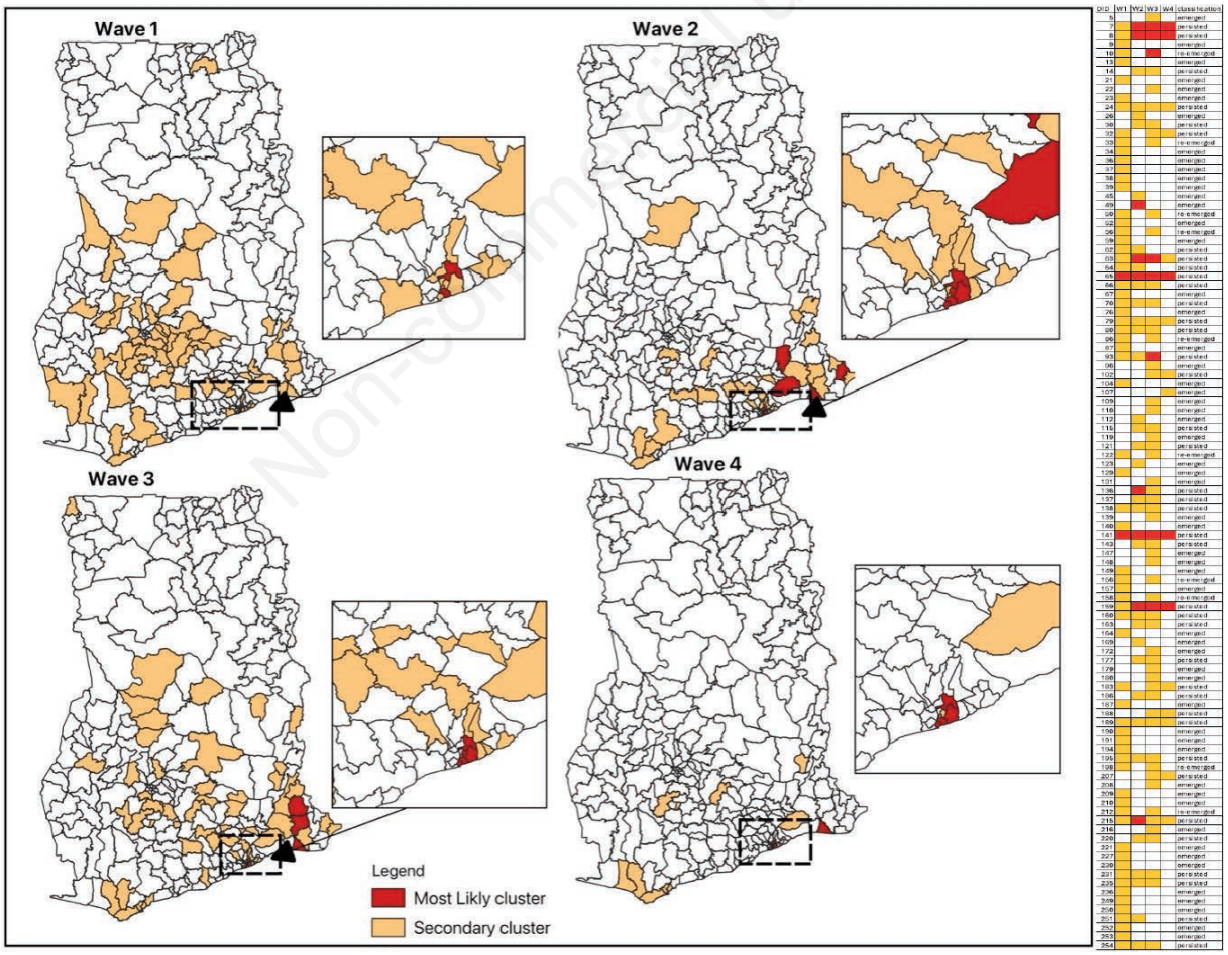
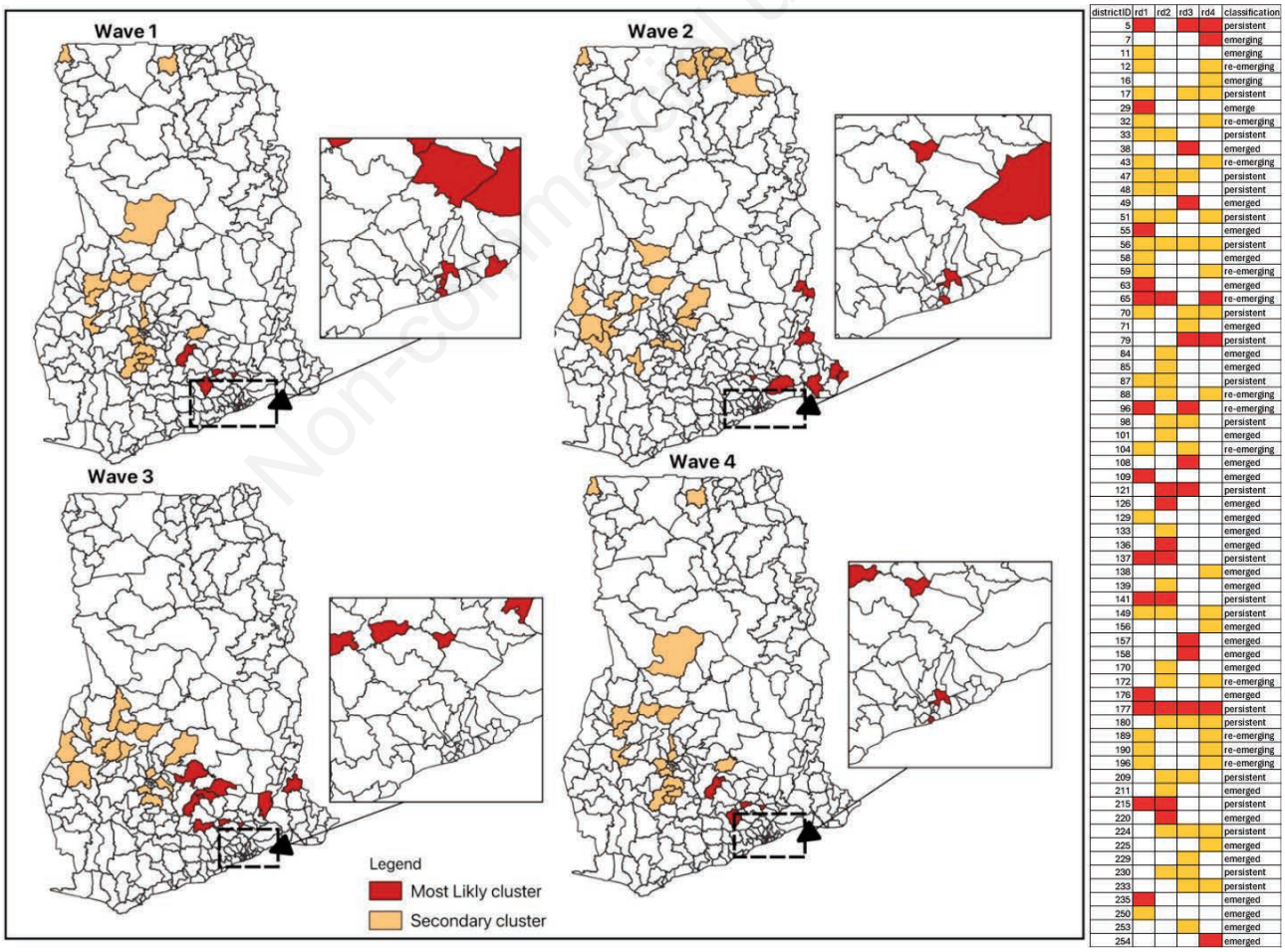


Figure 5. Spatial distribution of district-level detected cluster for COVID-19 IR after covariate adjustment. DID, district identification number,W1,..,4:-pandemic wave 1 to 4.

Cluster detection from irregular shape polygons such as in our study is often the case in most health spatial data analysis (Abolhassani *et al.*, 2020). Detecting clusters using the spatial scan statistics presents a challenge due to non-standard shapes (Otani and Takahashi 2021; Toshiro and Kunihiko 2005; Tango and Takahashi 2012). Overall, the FSRLR used in the study offered a good opportunity to identify areas with elevated risk of COVID-19 incidence and CFR within the complex geographic boundaries.

Conclusion

The study investigated the spatial heterogeneity of district-level COVID-19 incidences and CFR in Ghana. FSRLR and a Poisson-GLM identified districts with elevated risks of COVID-19 incidence and CFR. Different patterns of clusters were observed during the four waves. The MLC of incidence was persistent in the South (Greater Accra Region) also after covariate adjustment. Districts with elevated CFR risks were found in the centre (Bono Region) and the North but the clusters in the northern part disappeared after adjustment. Considering the WHO surveillance strategies for COVID-19, the identified high-risk areas could aid in selecting sentinel sites for targeted surveillance. In genomic surveillance this study could guide representative sampling for sequencing and monitoring. The results further show how the distribution pattern of incidence and CFR changes if the urban population, the proportion of the male population, the proportion of the population aged ≥ 65, and the distance to the epicentre are taken into account. Therefore, long-term control measures for COVID-19 and other contagious diseases should consider improving quality healthcare access in peripheral districts. Additionally, measures aimed at preventing rural-urban migration could reduce incidence in Ghana. Finally, the study calls for cross border measures in addition to national measures.



districtID	rd1	rd2	rd3	rd4	classification
5					persistent
7					emerging
11					emerging
12					re-emerging
16					emerging
17					persistent
29					emerge
32					re-emerging
33					persistent
38					emerged
43					re-emerging
47					persistent
48					persistent
49					emerged
51					persistent
55					emerged
56					persistent
58					emerged
59					re-emerging
63					emerged
65					re-emerging
70					persistent
71					emerged
79					persistent
84					emerged
85					emerged
87					persistent
88					re-emerging
96					re-emerging
98					persistent
101					emerged
104					re-emerging
108					emerged
109					emerged
121					persistent
126					emerged
129					emerged
133					emerged
136					emerged
137					persistent
138					emerged
139					emerged
141					persistent
149					persistent
156					emerged
157					emerged
158					emerged
170					emerged
172					re-emerging
176					emerged
177					persistent
180					persistent
189					re-emerging
190					re-emerging
196					re-emerging
209					persistent
211					emerged
215					persistent
220					emerged
224					persistent
225					emerged
229					emerged
230					persistent
233					persistent
235					emerged
250					emerged
253					emerged
254					emerged

Figure 5. Spatial distribution of district-level detected cluster for COVID-19 IR after covariate adjustment. DID, district identification number,W1,..,4:-pandemic wave 1 to 4.

References

Abolhassani A, Prates MO, Castellares F, Mahmoodi S, 2020. Zero-inflated bell scan: a more flexible spatial scan statistic. Spat Stat 36:100433.

Adebowale AS, Fagbamigbe AF, Akinyemi JO, Obisesan KO, Awosanya EJ, Afolabi RF, Alarape, SA, Obabiyi, SO, 2021. Situation assessment and natural dynamics of COVID-19 pandemic in Nigeria, 31 May 2020. Scientific African 12: e00844.

Alves HJP, Fernandes FA, de Lima KP, Batista BDO, Fernandes TJ, 2021. Incidence and lethality of COVID-19 clusters in Brazil via circular scan method. Revista Brasileira de Biometria 39:556–70.

Arab-Mazar Z, Sah R, Rabaan AA, Dhama K, Rodriguez-Morales AJ, 2020. Mapping the incidence of the COVID-19 hotspot in Iran – Implications for travellers. Travel Med Infect Dis 34:101630.

Aral N, Bakir H, 2022. Spatiotemporal analysis of COVID 19 in Turkey. Sust Cities Soc 76:0–2.

Bashir M. F, Ma B, Bilal, Komal B, Bashir MA, Tan D, Bashir M, 2020. Correlation between climate indicators and COVID-19 pandemic in New York, USA. Sci Total Environment 728:138835.

Bermudi PMM, Lorenz C, Aguiar BS de, Failla MA, Barrozo LV, Chiaravalloti-Neto F, 2021a. Spatiotemporal ecological study of COVID-19 mortality in the City of São Paulo, Brazil: shifting of the high mortality risk from areas with the best to those with the worst socio-economic conditions. Travel Med Infect Dis 39:101945.

Chen Q, Toorop MMA, De Boer MGJ, Rosendaal FR, Lijfering WM, 2020. Why Crowding Matters in the Time of COVID-19 Pandemic- A Lesson from the Carnival Effect on the 2017/2018 Influenza Epidemic in the Netherlands. BMC Public Health 20:1–10.

European Centre for Disease Prevention and Control, 2021. "Guidance for Representative and Targeted Genomic SARS-CoV-2 Monitoring." Technical Report 1:1–18. Available from: https://www.who.int/publications/i/item/who-2019-nCoV-surveillanceguidance-2020.8%0Ahttps://www.ecdc.europa.eu/sites/default/files/documents/Guidance-for-representative-and-targeted-genomic-SARS-CoV-2-monitoring.pdf

Fatima M, O'Keefe KJ, Wei W, Arshad S, Gruebner O, 2021. Geospatial Analysis of Covid-19: A Scoping Review. Internat J Environ Res Public Health 18:1–14.

Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L, 2020. Spatial Analysis and GIS in the Study of COVID-19. A Review. Sci Total Environ 739:140033.

Ghana Statistical Service, 2021. Ghana 2021 Population and Housing Census General Report3c. Available from: https://statsghana.gov.gh/gssmain/fileUpload/pressrelease/2021%20PHC%20General%20Report%20Vol%203A_Population%20of%20Regions%20and%20Districts_181121.pdf

Ghana Health Service, 2020. Techinical Guidelines for Integrated Disease Surveillance and Response in the WHO Africa Region.Section 4,5,6 and (3rd Edition). Available from: https://ghs.gov.gh/policy-document/

Ghana Health Service, 2022. COVID-19 Situation Dashboard _ Ghana. Ghana Health Service. 2022. Available from: https://www.ghs.gov.gh/covid19/dashboardm.php

Huang Q, Liu Q, Song C, Liu X, Shu H, Wang X, Liu Y, Chen X, Chen J, Pei T, 2021. Urban Spatial Epidemic Simulation Model: A Case Study of the Second COVID-19 Outbreak in Beijing, China. Trans GIS 26:297-316.

Lawton, R., Zheng, K., Zheng, D., & Huang, E. 2021. A longitudinal study of convergence betweenBlack and White COVID-19 mortality: A county fixed effects approach. Lancet Reg Health - Am 1:100011.

Islam A, Sayeed MA, Rahman MK, Ferdous J, Islam S, Hassan MM, 2021. Geospatial Dynamics of COVID-19 Clusters and Hotspots in Bangladesh. Transboundary and Emerging Diseases 68:3643–57.

Jacqui W, 2023. China Coronavirus: WHO Declares International Emergency as Death Toll Exceeds 200. BMJ (Clinical Research Ed.) 368:m408.

Kaburi BB, Kubio C, Kenu E, Ameme DK, Mahama JY, Sackey SO, Afari E A, 2017. Evaluation of Bacterial Meningitis Surveillance Data of the Northern Region, Ghana, 2010-2015. Pan Afr Med J 2017 Jun 30;27:164.

Kulldorff M, 2021. SaTScan User Guide 10.1(Issue July). Available from: https://www.satscan.org/cgi-bin/satscan/register.pl/SaTScan_Users_Guide.pdf?todo=process_userguide_download

Kulldorff M, 1997. Scan Statistic. Encyclopedia of GIS. https://doi.org/10.1007/978-3-319-17885-1_101147.

Otani T, Takahashi, K, 2021. Flexible Scan Statistics for Detecting Spatial Disease Clusters: The Rflexscan r Package. J Statist Software 99 :1–29.

Owusu M, Sylverken AA, Ankrah ST, El-Duah P, Ayisi-Boateng NK, Yeboah R, Gorman R, Asamoah J, Binger T, Acheampong G, Bekoe FA, Ohene SA, Larsen-Reindorf R, Awuah AA, Amuasi J, Owusu-Dabo E, Adu-Sarkodie Y, Phillips RO, 2020. Epidemiological profile of SARS-CoV-2 among selected regions in Ghana: A cross-sectional retrospective study. PLoS One 15(:e0243711.

Paul R, Adeyemi O, Ghosh S, Pokhrel K, Arif AA. 2021. Dynamics of Covid-19 Mortality and Social Determinants of Health: A Spatiotemporal Analysis of Exceedance Probabilities. Ann Epidemiol 62:51–8.

Rodriguez VS, Jacques L, Dalal J, Sestito P, Habibi Z, Venkatasubramanian A, Nguimbis B, Mesa SB, Chimbetete C, Keiser O, Impouma B, Mboussou F, William GS, Ngoy N, Talisuna A, Gueye AS, Hofer CB, Cabore JW, 2021. The Toll of COVID-19 on African Children: A Descriptive Analysis on COVID-19-Related Morbidity and Mortality among the Pediatric Population in Sub-Saharan Africa. Internat J Infect Dis 110:457–65.

US Department of Health and Human, and Centers for Disease Control and Prevention, 2006. Principles of Epidemiology in Public Health Practice, 3rd Edition, no. Cdc: 1–512.

Siljander M, Uusitalo R, Pellikka P, Isosomppi S, Vapalahti O, 2022. Spatiotemporal Clustering Patterns and Sociodemographic Determinants of COVID-19 (SARS-CoV-2) Infections in Helsinki, Finland. Spat Spatio-Temporal Epidemiol 41:100493.

Suleiman AA, Suleiman A, Abdullahi UA, Suleiman SA, 2021. Estimation of the case fatality rate of COVID-19 epidemiological data in Nigeria using statistical regression analysis. Biosafety and Health 3:4–7.

Tango T, Takahashi K, 2012. A flexible spatial scan statistic with a restricted likelihood ratio for detecting disease clusters. Statist Med 31:4207–18.

Tango T, Takahashi K, 2005. A Flexibly Shaped Spatial Scan

Statistic for Detecting Clusters. Internat J Health Geogr 13:1–13.

WHO, 2019. Techinical Guidelines for Integrated Disease Surveillance and Response in the WHO Africa Region.Section 4,5,6 and (3rd Edition). Available from: https://www.afro.who.int/publications/technical-guidelines-integrated-disease-surveillance-and-response-african-region-third

WHO, 2022. Public Health Surveillance for COVID-19. Interim Guidance, no. February: 253–78. Available from: https://www.who.int/publications/i/item/who-2019-nCoV-surveillanceguidance-2020.8.

Yi GY, Hu P, He W, 2020. Characterizing the Dynamic of COVID-19 with a New Epidemic Model: Susceptible-Exposed-Symptomatic-Asymptomatic-Active-Removed. MedRxiv, 2020.12.08.20246264. Available from: https://www.medrxiv.org/content/10.1101/2020.12.08.20246264v1%0Ahttps://www.medrxiv.org/content/10.1101/2020.12.08.20246264v1.abstract.

Zhang J, Dong X, Liu G, Gao Y, 2023. Risk and Protective Factors for COVID-19 Morbidity, Severity, and Mortality. Clin Rev Allergy Immunol 2023;64:90-107.

Zu J, Li ML, Li ZF, Shen MW, Xiao YN, Ji FP, 2020. Transmission Patterns of COVID-19 in the Mainland of China and the Efficacy of Different Control Strategies: A Data- And Model-Driven Study. Infect Dis Poverty 9:1–14.

Online supplementary materials

Table 1. Districts detected as the most likely cluster and secondary clusters for COVID-19 incidence in Ghana.

Table 2. Districts detected as the most likely cluster and secondary clusters for COVID-19 CFR in Ghana.

Table 3. Districts detected as the most likely cluster and secondary clusters for COVID-19 incidence in Ghana after adjusting for covariate.

Table 4. Districts detected as the most likely cluster and secondary clusters for COVID-19 CFR in Ghana after adjusting for adjusting for covariate.