Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Epidemiological investigation of the COVID-19 outbreak in Vellore district in South India using Geographic Information Surveillance (GIS)

Malathi Murugesan 1, Padmanaban Venkatesan 2, Senthil Kumar 3, Premkumar Thangavelu 4, Winsley Rose 5, Jacob John 3, Marx Castro 6, T. Manivannan 6, Venkata Raghava Mohan 3,4,†, Priscilla Rupali 4,7,∗

1 Department of Clinical Microbiology & Hospital Infection Control Committee, Christian Medical College, Vellore, Tamil Nadu, India
2 Department of Biochemistry, Christian Medical College, Vellore, Tamil Nadu, India
3 Department of Community Health, Christian Medical College, Vellore, Tamil Nadu, India
4 Department of Infectious Diseases, Christian Medical College, Vellore, Tamil Nadu, India
5 Department of Pediatrics, Christian Medical College, Vellore, Tamil Nadu, India
6 Deputy Director of Health Services, Vellore, Tamil Nadu, India

A R T I C L E   I N F O

Article history:
Received 21 April 2022
Revised 12 June 2022
Accepted 2 July 2022

Keywords:
COVID-19
GIS
Spatial analysis
Mapping

A B S T R A C T

Objectives: Geographical Information Surveillance (GIS) is an advanced digital technology tool that maps location-based data and helps in epidemiological modeling. We applied GIS to analyze patterns of spread and hotspots of COVID-19 cases in the Vellore district in South India.

Methods: Laboratory-confirmed COVID-19 cases from the Vellore district and neighboring taluks from March 2020 to June 2021 were geocoded and spatial maps were generated. Time trends exploring urban-rural burden with an age-sex distribution of cases and other variables were correlated with outcomes.

Results: A total of 45,401 cases of COVID-19 were detected, with 20,730 cases during the first wave and 24,671 cases during the second wave. The overall incidence rates of COVID-19 were 462.8 and 588.6 per 100,000 population during the first and second waves, respectively. The spread pattern revealed epicenters in densely populated urban areas with radial spread sparring rural areas in the first wave. The case fatality rate was 1.85% and 1.6% during the first and second waves, which increased with advancing age.

Conclusions: Modern surveillance systems like GIS can accurately predict the trends and spread patterns during future pandemics. In addition, real-time mapping can help design risk mitigation strategies, thereby preventing the spread to rural areas.

© 2022 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Background

Various epidemiological mapping systems have been employed in the past to track the spatial and temporal patterns of infectious diseases like cholera, influenza, and even plague (Sarfo and Karuppannan, 2020). In 1854, John Snow plotted the location of cholera patients in a hand-drawn map, which linked the cases to the source (broad street pump), confirming that cholera was likely waterborne (Gilbert, 1958). Over the years, there has been tremendous advancement in Geographical Information Surveillance (GIS) technologies, which have provided accurate information in a real-time manner. During the COVID-19 pandemic, integrated and innovative surveillance methods using new technologies were crucial to identify the hotspots because of the dynamic and evolving nature of the infection (Ibrahim, 2020). Geospatial technology helps alleviate this by using different methods, namely GIS, global positioning systems, and other satellite-mediated technology systems.

Therefore, GIS-assisted data capture based on location can identify clusters, link social and epidemiological risk factors, predict the susceptible population that will get infected in the near future, and estimate the disease’s incidence rates (Musa et al., 2013). This identification of high-risk populations can help formulate risk mitigation strategies by creating new healthcare infrastructure and improving existing facilities. Because there were many uncertainties regarding transmission in the initial stages of the COVID-19 pandemic (World Health Organization, 2020) with the possibility...
of transmission from pre-asymptomatic and asymptomatic individuals (Zou et al., 2020), we felt that it would be prudent to track the epidemiology and patterns of spread in the Vellore district. In this study, we present a spatial mapping of COVID-19 infected cases in the Vellore district in Tamil Nadu, South India, and adjacent taluks to identify epidemic clusters or hotspots in the community, track the pattern of spread, and study the demographic characteristics of the study population.

Methodology

Study area

In this observational cohort study, mapping the burden of disease and further analyses have been performed using the earlier undivided Vellore district region. Study regions depicting different taluks, rural administrative areas, urban town panchayats, and city corporation limits are shown in Supplementary Figure 1. The undivided Vellore district (12°55′13″N 79°08′00″E) spanned over 392.62 sq. km and had a population of 906,745 as per the census 2011 data by the Government of India (Census 2011, 2022). In 2020, the Vellore district was divided into three districts: Vellore, Tirupattur, and Kanipet.

Study population

We included documented, laboratory-confirmed patients with COVID-19 from the Vellore district from March 28, 2020, to June 30, 2021, in the study analyses.

Data collection

The data used for the analysis in this study were obtained from the Vellore district surveillance portal after obtaining necessary permission from the district authorities. De-identified data of all the laboratory-confirmed COVID-19–positive patients were collected from the district portal with patient address, unique epidemiological identification number, symptoms, co-morbidities, and outcome.

Spatial mapping of COVID-19 cases

The documented addresses from the district portal were geocoded to the closest neighborhoods using Google Earth Pro (Google Earth Versions, 2022) by trained GIS technicians at the Christian Medical College, Vellore. Spatial data on geocordinates were linked to the available attribute data of COVID-19 cases, and personal identifiers were removed from the database used for final analyses. Village and taluk-level vector layers of the study region, including population-level data sourced from the Survey of India, the National Survey and Mapping Organization under the Government of India, were used as base layers for spatial mapping. All cases were mapped onto the base layer using ArcMap version 10.8.1 software (ArcGIS, 2021) using the local neighborhood level spatial information.

Statistical analysis

Descriptive analysis was performed to study the characteristics of the study population for age, gender, locality (urban or rural), presentation of COVID-19 symptoms and survival, and was calculated as a percentage among the study population. We analyzed the age-sex distribution of COVID-19 cases and positivity rates by age groups, gender, and area of residence (rural/urban) with 95% confidence interval (CI). In addition, time trends exploring urban-rural burden and other variables were correlated with age and mortality and compared between the first wave (March 28, 2020, to March 31, 2021) and the second wave (April 1, 2021, to June 30, 2021).

Spatial analysis

Spatial maps depicting the distribution of COVID-19 case counts by villages and urban wards in the Vellore district were generated using the geocoded addresses and layered as points on the base map to illustrate the distribution of all COVID-19 cases from March 2020 to June 2021. A total of 42,921 cases of COVID-19 were available for spatial mapping based on the completeness of documented addresses in the study area. There were 938 villages, urban localities, and reserved forest areas available as individual polygons for the Vellore district. The total number of cases falling under each of these individual area polygons was estimated by spatially joining the case layer with the polygon layer and sub-block level maps depicting caseloads for the overall period, and waves 1 and 2 were generated using ArcMap. Densities of COVID-19 cases around each documented case were estimated using the point density tool within the Spatial Analyst toolbox in ArcMap, and quarterly heat maps were generated to demonstrate the spread of the disease for March 2020 to June 2021. The spatial distribution of cases among different age groups was mapped. The earlier undivided Vellore district was subdivided into eight taluks, including Arckanam, Wallajah, Arcot, Katpadi, Vellore, Gudiyattam, Vanjambadi, and Tirupattur, spanning from the west to east (Vellore District, Government of Tamil Nadu, 2022). Taluk-level population counts were obtained from the recently published figures on the taluk websites and were used as denominators to estimate and map incidence rates of COVID-19 disease during the two waves.

Results

A total of 45,401 cases of COVID-19 were detected between March 28, 2020, to June 31, 2021, with 20,730 cases during the first wave (March 28, 2020, to March 31, 2021) and 24,671 cases during the second wave (April 1, 2021, to June 30, 2021). Data were available in 42,921 cases for spatial mapping based on the completeness of documented addresses during both waves, of which 19,021 (44.3%) and 23,900 (55.7%) had occurred during the first and second waves, respectively. A higher number of cases resided in Vellore, Katpadi, and Gudiyattam taluks (Supplementary Figure 2). We mapped weekly and monthly trends in the epidemiological spread of infection and identified geographical clusters, hotspots, and spatial-temporal trends using GIS.

Of the 938 smaller geographical areas within the study region, 188 were classified as reserved forest areas and the rest comprised villages, urban localities, town panchayat, and a city corporation located in Vellore taluk. Throughout both waves, COVID-19 cases were documented from 636 inhabited villages and urban areas in the Vellore district. The mean (SD) number of cases in these smaller units was 67 (278), with the minimum being one case documented in one of the villages, and the maximum documented caseload was 3323 in the Konavattam urban ward in Vellore city. Ten urban areas had recorded over 1000 cases each, namely areas numbered one, four, and five, i.e., Gudiyattam, Katpadi, and Vellore taluks in the region. Village and urban ward-wise caseloads during both the waves and overall study period are presented in Figure 1, and a similar trend was observed with the same urban and peri-urban neighborhoods experiencing a higher burden during both the waves with relative sparing of rural areas.

Heatmaps generated estimating the point density around geolocations of cases in the initial stages of the pandemic from April 2020 to June 2020 revealed that the epicenters were in the middle of Vellore city, a densely populated city corporation, and the
Figure 1. Distribution of COVID-19 cases across different geographical subunits in the study area.

Figure 2. Heatmaps depicting the spread of COVID-19 cases during the first and second waves in the study area.
The age distribution among the COVID-19 cases during both waves is listed in Supplementary Table 1. Most of them belonged to the age group 21–60 years. Among the study population, 3.1% in the first wave and 1.99% in the second wave were children younger than 10 years. Older adults (>60 years) were 16% in the first wave and 17.3% in the second wave. Overall, COVID-19 incidence was higher in men (first wave 58.71%; second wave 56.86%) than in women (first wave 41.29%; second wave 42.90%). The density of children (under 15 years) with COVID-19 was higher in and around two urban localities (Vellore and Gudiyatham). Higher densities of cases in the other age groups were noted mainly in urban and peri-urban areas with relatively lesser numbers from the rural areas located between Vellore and Gudiyatham. The rest of the villages and smaller towns in the district experienced relatively fewer cases during the two waves (Figure 4). Figure 5 shows the overall time trends of daily cases and deaths during the outbreak in this location. The first case documented as COVID-19–positive was on March 28, 2020. The outcome details were documented for 20,730 cases in the first wave and 21,672 cases in the second wave. The case fatality rate was 1.89% during the first wave and 1.6% during the second wave and steadily increased with advancing age, i.e., 7.38% were older than 60 years in the first wave and 5.02% in the second wave (Supplementary Table 2). Case fatality rates were higher in men (first wave 2.40% and 1.76% in the second wave) than in women (first wave 1.16% and second wave 1.38%). When correlating co-morbidities with mortality, data from the first wave revealed that the fatality rates were highest among those with ≥2 co-morbidities (9.52%).

Figure 3. Subdistrict level incidence of COVID-19 during the first and second waves
The data on risk factors, co-morbidities, clinical symptoms, and outcomes were available only for 17,748 of 19,092 cases during the first wave of the pandemic but were not available for the second wave as recording this was not mandatory during the second wave. On analyzing risk factors for acquiring COVID-19 infection, 4% had a recent history of travel outside India, and 2.1% reported a history of travel within India, but 62.9% had a history of contact with a COVID-19 infected individual. The co-morbidity profile data available for 17,748 cases showed that 14.4% of the patients had at least one co-morbidity, but 7% had more than one co-morbidity. Diabetes mellitus, hypertension, and asthma were the most seen co-morbidities in these patients (Supplementary Table 3). Among the 61.93% of symptomatic patients, fever (39.7%), cough (24.8%), and sore throat (15.9%) were the most common symptoms. Breathlessness was reported in 9% of all cases (Supplementary Table 4). Only 99 of 17,748 (0.6%) were pregnant among the study population during the first wave of the pandemic.

Discussion

In our spatio-temporal analysis, we noted that the first wave of the pandemic started with a returned traveler from the UK in March 2020, after which a peak was noted in travelers who had attended a religious meet in another part of the country (BBC News, 2020). Subsequently, there was a steady rise in cases over five months (June 2020 to October 2020), with continuing sporadic cases reported after that until March 2021. Contrary to the first wave, the epidemic noted a sharp spike in April 2020, with a rapid decline over two months (April 2020 to June 2021) during the second wave of the pandemic. This was presumed to be mainly because of the Delta variant B.1.617.2, which had higher transmissibility and immune evasion than the original strain (Hwang et al., 2022; Mlochova et al., 2021). The reproduction number (R0) for COVID-19 based on the incidence data of European countries was 2.2 (95% CI 1.9-2.6) for the ancestral strain (Locatelli et al., 2021). However, China reported a higher R0 of 3.32 (95% CI 2.81-3.82) (Almohamadi et al., 2020). Subsequent waves caused by the Delta wave revealed an R0 of 5.08 (3.2-8) (Liu and Rocklöv, 2021) and the Omicron 9.5 (5.5-20) (Liu and Rocklöv, 2022). The rapid transmission left many countries and states unprepared. Therefore, using digital methods to map the cluster areas and predict hotspots would have enabled contact tracing in real-time, helping to cordon off affected households to prevent ongoing transmission.

The geospatial mapping in our study population showed that the spread pattern of the cases is mainly in highly populated urban areas followed by less populated semi-urban areas and then rural areas. In the first and the second waves, the initial cluster of cases is seen in the urban population. However, the peri-urban and rural areas were also affected in the second wave. Because 60% of the population in India is rural, it is imperative to curb the spread from urban to rural areas. Similar trends have been shown worldwide where the metropolitan cities and urban areas were affected earlier because of a higher population density, indoor crowding, and high usage of crowded public transport systems (Lee et al., 2020). This has also been supported by GIS mapping conducted in Jammu and Kashmir, which showed that the urban areas are affected first (Meer and Mishra, 2021). A study conducted by Gangwar and Ray (2021) analyzed COVID-19 cases in India using geospatial technology from statistics available in the public domain. It showed similar results suggesting that 80% of the confirmed cases during the first wave occurred in highly populated areas as cities underwent gradual unlocking after a stringent lockdown phase. Another study by Soni et al. (2022) that used an an-
alytical hierarchy process and geographical information system indicated that the states with a more urban population (Maharashtra and Uttar Pradesh) had more coronavirus-infected zones than other parts of India.

The case fatality rate in our study population was 1.89% and 1.6% during the first wave and second wave, respectively. A study published by Shah et al. found that the case fatality rate (CFR) for South Asia Association for Regional Cooperation (SAARC) countries was lower than the developed countries, and the estimated crude and adjusted CFR for India was 1.542% and 1.601% (95% CI 1.548-1.684) during the first wave (Shah et al., 2021). Another study by Bogam et al. estimated that the fatality rate was 47% lower in the Delta variant-driven second wave than in the first wave (adjusted CFR 0.43; 95% CI 0.41-0.45) (Bogam et al., 2021) despite the large numbers. This suggested that although the transmission was higher during the second wave, the CFR is probably lower because of pre-existing immunity obtained from either previous COVID-19 infection, COVID-19 vaccination, or easy access to health care services as all hospitals were directed to accept patients infected with COVID-19 under the Epidemic Act of 1897 (Ministry of Law and Justice, 2020). Even though the country managed well during the first wave of the pandemic, the rapid hit of the second wave decimated the health care system.

The COVID-19 pandemic caused an unprecedented global economic crisis because of lockdown, travel restrictions, border closures, and loss of employment. The United Nations International Children’s Emergency Fund estimated that 42-66 million children might face extreme poverty because of the COVID-19 crisis, in addition to the already existing poor population of 386 million children (2019) (United Nations Stable Development Group, 2020). Thus, an accurate mapping of the diseased population which could predict hotspots and patterns of spread of future respiratory pandemics, would help focus containment strategies, educational awareness, and targeted training based on the resources and capacity of each health setting. In addition, the less affected areas with susceptible populations can be targeted for vaccination by doorstep delivery of vaccines through mobile camps. Although the Epidemic Diseases Act 1897 was helpful in the initial containment of the disease, digitalized methods and real-time mapping of the clusters were not employed, and interventions that could have helped during the second wave could not be implemented effectively. Previous respiratory pandemics have led to enormous mortality and morbidity worldwide, like the measles epidemic in the 1960s (The Washington Post, 2019) and the plague in India in the 1900s, causing over 2-3 million deaths (Statista, 2022).

New diseases emerge, and old ones continue to re-emerge. Routine surveillance methods in public health include active, passive, categorical, integrated, syndromic, case-based, sentinel, or serological surveillance, and so forth. (Ibrahim, 2020; Nsubuga et al., 2006). However, because of the rapid spread of respiratory pandemics with a high R0 in the past, it makes sense to employ digital surveillance technology because it can save time, human resources, and resources. During the pandemic, many organizations, including the WHO, created various geospatial dashboards like the WHO Coronavirus (COVID-19) dashboard (World Health Organization, 2022), Johns Hopkins University COVID-19 Dashboard (Johns Hopkins University, 2022), and COVID-19 India (COVID-19 India, 2021), which proved extremely useful for epidemiologists and local health policymakers implement isolation and quarantine measures based on the geographic spread. Based on the lessons learned during this pandemic, budget allocation to more advanced information technology systems using satellite and geomapping could play an important role in future outbreak management.

Limitations

As the analysis is based on the data available in the public domain and conducted retrospectively, the geographic location of the study population (5.46%) and outcome details (6.60%) were unavailable for the entire study population. The clinical features and co-morbidity profiles were possible to retrieve only for 85.62% of the first wave population. This limited our comparison of the risk factors and clinical features during the first and second waves. The data captured are from the reported cases from the district portal and not from the state-wise data where there may be some patients from this district who were admitted to other places. However, because of lockdown and other restrictions, we believe that almost all of them were admitted within the district.
Conclusion

Respiratory pandemics with a high R0 seem to start in urban areas and then spread to rural areas. Therefore, GIS surveillance as a contact tracing digital tool would help rapid implementation of risk mitigation and containment strategies, preventing spread to rural areas where access to appropriate health care services is limited. The nation must build a catastrophe-resilient public health care system with well-equipped facilities under the existing National Rural Health Mission in India to manage pandemics in urban and rural areas.

Conflict of interest

The authors have no competing interests to declare.

Credit authorship contribution statement

Malathi Murugesan: Conceptualization, Methodology, Resources, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. Padmanaban Venkatesan: Methodology, Resources, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. Senthil Kumar: Methodology, Resources, Validation, Data curation.

Premkumar Thangavelu: Methodology, Resources, Validation, Data curation. Winsley Rose: Validation, Formal analysis, Data curation, Writing – review & editing. Jacob John: Validation, formal analysis, Data curation, Writing – review & editing. Marx Castro: Resources, Formal analysis, Writing – review & editing. T. Manivannan: Resources, Formal analysis, Writing – review & editing. Venkata Raghava Mohan: Conceptualization, Methodology, Resources, Validation, Formal analysis, Data curation, Writing – review & editing. Priscilla Rupali: Conceptualization, Methodology, Resources, Validation, Formal analysis, Data curation, Writing – review & editing.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical approval statement

This study was conducted jointly by the Christian Medical College, Vellore, and the district health authorities, approved by the institutional review board, Christian Medical College, Vellore, located in South India (Institutional Review Board no:13261 dated 26.08.2020).

Acknowledgment

We acknowledge Mr Varun Kumar from Department of Community Health, CMC Vellore for his assistance and technical support in geocoding of the cases during the study. We have attached the credit author statement to this query reply.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi: 10.1016/j.ijijd.2022.07.010.

References

Alimohammadi Y, Taghdiri M, Sepandi M. Estimate of the basic reproduction number for COVID-19: A systematic review and meta-analysis. J Prev Med Public Health 2020;53:151–7.