

Research article

Use of spatial analytics and geographic information systems in childhood diarrhoea in rural areas of India: A protocol paper

Dmello Mackwin Kenwood¹, Badiger Sanjeev², Kumar Sudeep³, Kumar N.⁴, D'Souza N.⁵, Purushotam J.¹¹Department of Public Health, ⁴Department of Community Medicine, ⁵Department of Biostatistics, KS Hegde Medical Academy, Nitte (Deemed to be University), Mangaluru, Karnataka, India²Department of Community Medicine, AJ Institute of Medical Sciences and Research Centre, Mangaluru, Karnataka, India³Senior State Child Consultant, United Nations Children's Fund (UNICEF), Karnataka, India

Received: August 2021

Revised: November 2022

Accepted: December 2022

Corresponding author: Sanjeev Badiger. Email: badiger1971@gmail.com

ABSTRACT

Introduction and Aim: The most recent advancement in epidemiology is using Geographic Information Systems (GIS) to comprehend illness based on clustering and spatial analysis. Using spatial analysis has shown positive changes in sanitation, reducing diarrhoea's burden. This study put forwards the methodology used to analyze spatiotemporal clustering and hotspot areas of childhood diarrhea in Karkala taluk, Udupi during 2015 to 2018.

Materials and Methods: This study describes the methods used to analyze the space-time clustering and local indicators of spatial association (LISA) of childhood diarrhoea using the GeoDa and SaTScan programmes. At the Karkala taluk, data on under-five diarrhoea was gathered throughout time and space to develop the geospatial models. While the timing of the onset of the diarrhoea impacts the temporal characteristics, the child's position at the time of the event determines the spatial characteristics. The child's location and time of diarrheal episode was considered for the geospatial analysis.

Results: The Kulldorff scan statistic (SaTScan v10.1) was employed for geographical analysis, and GeoDa software used for LISA and hotspot analysis. The randomness distribution of childhood diarrhoea over time and space was determined using the Kulldorff scan statistic software. The SaTScan locates and evaluates the statistical significance of geographical clusters based on the number of cases inside and outside the cylinder window. The scanning window in the software uses a cylinder with an elliptical or circular base to study intervals for time, space, and space-time. The different combinations of High and Low categories of the spatial association are covered by the hotspot analysis utilizing the GeoDa software.

Conclusion: To build data-driven public health decision-making, it is required to link spatial and temporal patterns of diarrheal incidence among under-five children with predictive variables derived from routine data.

Keywords: SaTScan; LISA; hotspot analysis; space-time clustering.

INTRODUCTION

A significant waterborne illness that causes an enormous burden of mortality and morbidity in populations around the world is a diarrheal disease, or simply diarrhoea (1). In 2017, diarrhoea was the 9th most significant cause of mortality, accounting for more than 1.6 million deaths, most of which were in children under five (2,3). India's prevalence of diarrhoea was 9.2%, while in 2021, it was 7.3% (4). According to the NFHS 5 report, 9.5% boys and 8.9% girls under-five had diarrheal episodes before two weeks of the survey (5). In eight states and one union territory, the prevalence of diarrhoea was found to be greater than the national norm. The states with the highest rates of diarrhoea are Uttarakhand (17.1%), Uttar Pradesh (15%), Puducherry (11.3%), and Meghalaya (10.6%). Assam (2.9%), Kerala (3.4%), and Sikkim (1.8%), on the other hand, have recorded a lower prevalence of diarrheal illness among children aged 0 to 5 in India. Diarrhoea is more common in male

children, with rates ranging from 1.8% (Sikkim) to 17.5% (Uttarakhand) (5,6).

In India, childhood diarrheal illness has become a significant public health concern. Despite numerous interventions and control measures, the fatality rate from diarrhoea among children under five in India remains concerning (7). Many studies have identified risk factors at the child, maternal and household levels but have failed or overlooked the potential effect of geographical and environmental factors in explaining child survival. Hence GIS mapping and spatial analysis can significantly improve the assessment of ecological concerns in diseases like childhood diarrhoea.

The medical sector uses geospatial data more frequently to support decision-making. However, these data are primarily used in thematic map format in the health sector, which restricts the complete insight and utility of geospatial information (8). More detailed information on the place and time to map the disease risk will allow the creation of a platform for multi-

disease surveillance activities and the generation of several etiological theories (9). Many scientific disciplines use spatial science, also known as geographic information science, to study, analyze, and display real-world phenomena concerning their locations. GIS databases include both spatial and non-spatial data to provide a deeper comprehension of their relationships through several geography-related thematic aspects. Information not based on geometric considerations is referred to as non-spatial data, often known as attribute or characteristic data (10). Spatial scientists analyze spatial data using technologies like geographic information systems (GIS) and remote sensing to identify and decipher spatial patterns. Real-time development of big spatial data is connected to the present significant data era. It is now widely accessible from environmental sensors collecting meteorological data to geo-tagged social media posts (11).

The current research explores the mechanisms for using routine data to generate evidence-based decision-making in public health. This paper presents the method used to analyse the hotspot and spatiotemporal cluster of childhood diarrhoea in Karkala taluk, Udupi, from 2015 to 2018. By using representative data from the villages and the subcenter, the current study intends to demonstrate the prevalence of the diarrheal disease among children under the age of 5 in the villages of Karkala Taluk, Karnataka. By adding numerous related risk factors that have not been published, we intend to employ spatial methods to demonstrate diarrhoea's prevalence and spatiotemporal and hotspot clustering.

METHODOLOGY

Study location

The study will take place in the Karkala Taluk, in the southernmost district region, around 13011'60" N latitude and 74058'48" E longitude. There are 19 health facility centres and 49 villages in this taluk. Winter, summer, and monsoon are the three seasons in the Karkala taluk. Beginning in September and lasting through February, the winter season is followed by the summer season (March through May) and the monsoon season (June through August). The study area's highest and lowest average temperatures are 260 Celsius and 110 Celsius. (Fig. 1).

Data gathering

Participants in the study will be children under five years who have experienced diarrhoeal episodes. This study will cover every child under five who had diarrhoea in the three years prior (2015 - 2018). Information on child diarrhoea will be collected from the district health authority. The data will be verified at the household level. After being retrieved from the healthcare institution, a qualified healthcare expert will validate the demographic information of 49 villages at each village panchayath. QGIS software will be used to build the shape file for the Karkala Taluk. Geographic location will be used to indicate the child's location when the diarrheal episode occurred.

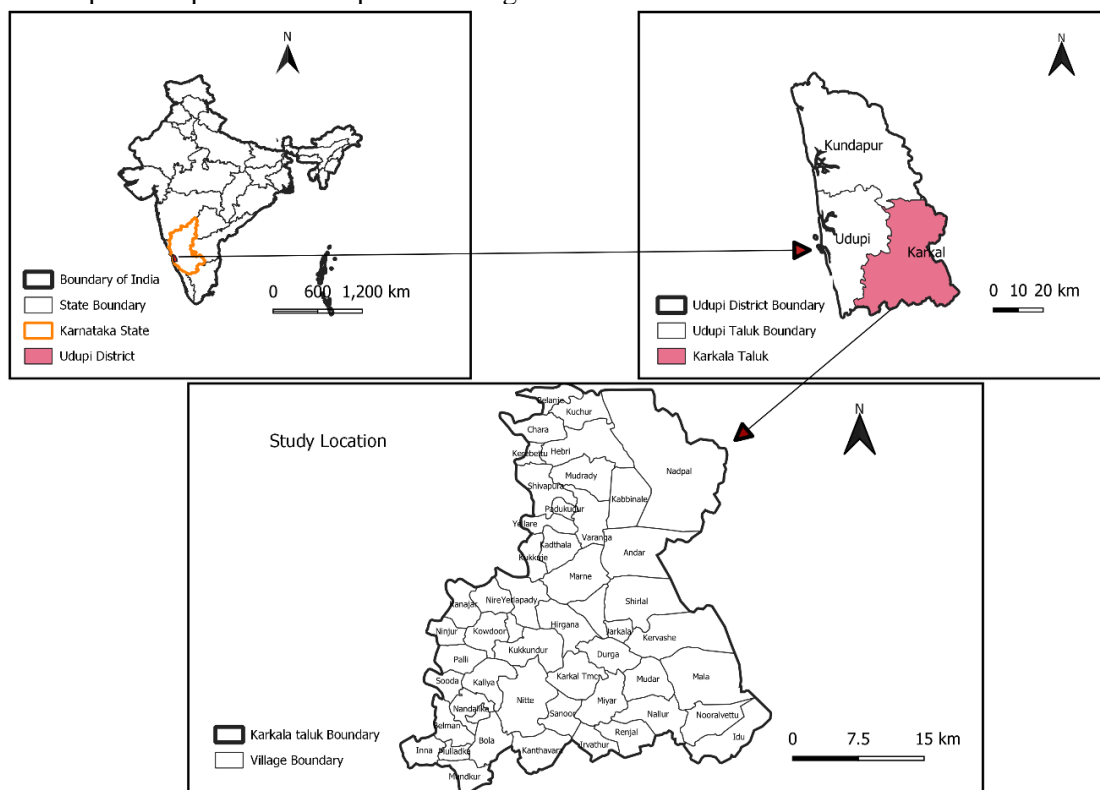


Fig. 1: Location of study area

After the data validation in the sub-centre, the researcher will assist in collecting data at the field level. Information on childhood diarrhoea at the home level will be gathered with assistance from Accredited Social Health Activists (ASHA) workers. The geo-coordinates will be mapped at the field level using an Etrex 10 GPS instrument. The spatiotemporal cluster analysis will be performed using SaTScan software (Fig. 2).

Ethical approval

The Central Ethics Committee approves this study (Ref: NU/CEC/2019/0212). The Declaration of Helsinki's guiding principles will be followed in the conduct of this study.

Plan of analysis

Descriptive statistics and geospatial analysis will be carried out to summarize the data. The geographical information system (GIS) and spatial statistics will be

conducted to determine geospatial analysis. The geographic information system will give statistical evidence of diarrheal clusters and predictive modeling. At the same time, trends and randomness in the spatial distribution of childhood diarrhoea will be carried out by spatial statistics using SaTScan and GeoDa softwares (Fig. 3).

Descriptive statistics

The socio-demographic information will be tabulated using frequency, proportion, average and range. Each village's annual incidence of childhood diarrhoea will be summarised using excel. The standard morbidity ratio for each village will be calculated with observed cases in the numerator and expected cases in the denominator.

Mapping of the incidence rate of childhood diarrhea

The mapping of the health centre will be completed using an Etrex GPS instrument. The incidence rate per 1000 children at risk will be calculated and plotted using GeoDa software.

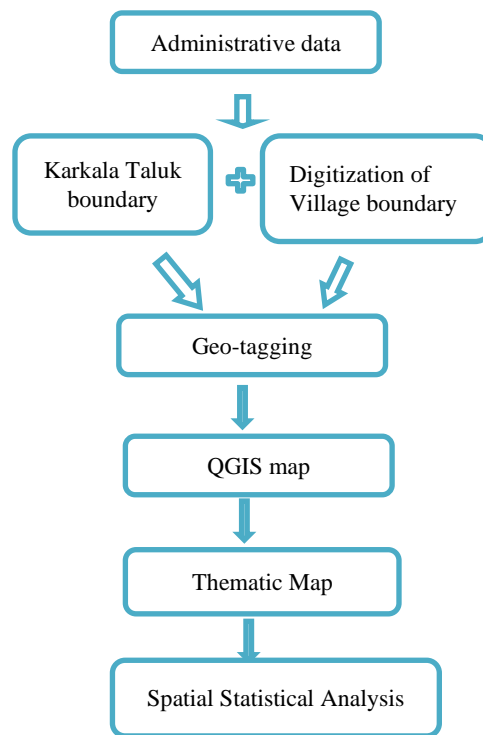


Fig. 2: Thematic map development

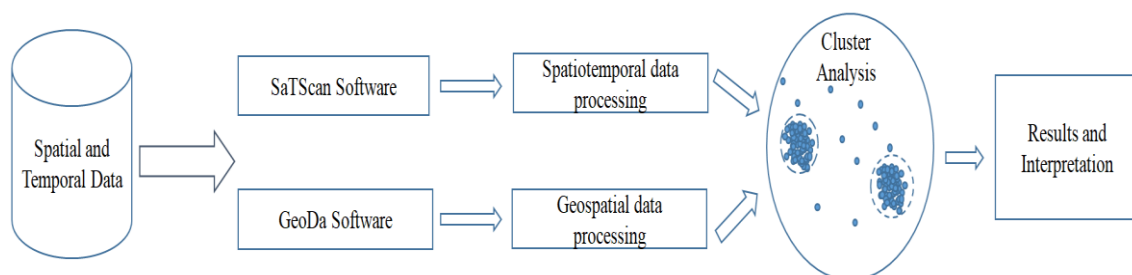


Fig. 3: Data processing for spatiotemporal cluster analysis

Geospatial mapping

The traditional Cartesian coordinate system will be used to determine the geographic coordinates. Modelling and geographic autocorrelation will be carried out Using the GeoDa programme. Global Moran's I and local Moran's I (LISA) will be utilized to determine spatial clustering over the block and regional clustering across the villages. LISA offers statistics for each site, evaluates their importance, and establishes the connection to develop a relative association between the total of the local data and a corresponding global statistic. The high and low categories of the spatial relationships are determined by the local Moran's I scatter plot (12).

Cluster analysis

SaTScan software will be used to do the cluster analysis. The software will identify the randomness in under-five diarrhoea's spatial and temporal distribution. The statistical significance of the spatial cluster is calculated using the sum of the observed and predicted cases inside the window. The circular and cylindrical scanning window of the SaTScan software considers time (interval), space (circle/ellipse), and space-time (a cylinder with an elliptical or circular base). The monthly reported diarrheal cases were examined using a discrete Poisson-based model with a Poisson distribution (13; Fig 4 A)

Spatio-temporal clusters

The cylindrical scan window of the scan statistic will be used to identify the spatial and time cluster. The

software uses a cylindrical scan window with a circular bottom for area and elevation proportional to time to find spatiotemporal clusters. The cylindrical window's geographical bottom represents the spatial scan statistic, while the cylinder-shaped window's size represents the time. The cylindrical window swings across time and space, pausing at all possible points in time (Fig 4 B). The probability ratio is used to calculate both the observed and anticipated case rates for each space-time window. The p-value is measured using Monte Carlo simulations 999 (14).

RESULTS

We expect that the distribution of childhood diarrhoea cases is not random over space and time across Karkala taluk. There will be a clustering and hotspot of diarrheal cases in some villages across seasons and years.

As of Aug 2021, 4042 under-five children with diarrheal episodes had been enrolled in the study since recruitment, enrolment, and data collection got underway in February 2019. The data collection was delayed due to COVID-19 restrictions. After the data collection is finished, data analysis will begin. Future findings will be presented at academic and professional events and published in peer-reviewed health-related research journals. This article is based on the protocol version, which was released on August 3, 2021. The distribution of diarrheal cases according to age, gender, primary health centre and months is given in Table 1.

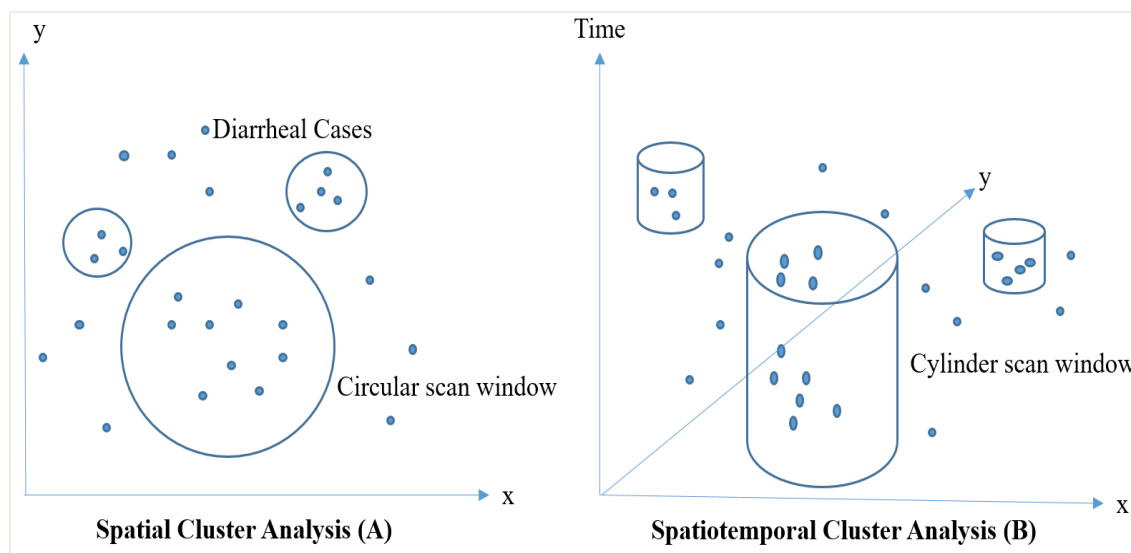


Fig 4: Space (circular) and Space-time (cylindrical) scan window for cluster detection

Table 1: Time, place-wise distribution of diarrhea among children under-five

Characteristics (n=4042)	n (%)
Age (in months)	
0-6	121 (3.0)
7-12	588 (14.5)
13-24	1094 (27.1)
25-36	949 (23.5)
37-48	774 (19.1)
49-60	516 (12.8)
Gender	
Male	2034 (50.3)
Female	2008 (49.7)
Primary Health Centre	
Ajekar	163 (4.0)
Bailur	184 (4.6)
Bajagoli	370 (9.2)
Belma	98 (2.4)
Donderangadi	102 (2.5)
Durga	46 (1.1)
Hebri	1135 (28.1)
Hirgan	226 (5.6)
Idu	259 (6.4)
Inna	40 (1.0)
Irvathur	152 (3.8)
Karkala	199 (4.9)
Kukundoor	128 (3.2)
Mala	112 (2.8)
Nandalike	150 (3.7)
Nitte	451 (11.2)
Palii	138 (3.4)
Sacherepette	89 (2.2)
Month	
January	267 (6.6)
February	267 (6.6)
March	416 (10.3)
April	445 (11.0)
May	351 (8.7)
June	355 (8.8)
July	445 (11.0)
August	363 (9.0)
September	314 (7.8)
October	277 (6.9)
November	258 (6.4)
December	284 (7.0)

DISCUSSION

The constraints on experts and inequitable access have challenged the healthcare sector. Developing spatial research to understand the environmental and spatial role in disease incidence and build-up algorithms for addressing unmet needs is essential to achieving Universal Health Coverage. Voluminous health-related data are generated, stored and validated. The value of the data is limited, especially in an Asian country like India (15). Analyzing standard data sets requires knowledge and evidence from multiple sectors. Data from multi-sectors is necessary to develop public health decision support tools and early warning signs (16,17). The various data sets available in different

formats require pre-processing to amalgamate them for spatial and temporal analysis.

The primary health centre's routine line list data on diarrheal cases among children under five are included in the current study. Clear guidelines on case definitions and identifications are employed for various studies, notwithstanding complaints about the quality and reporting system (17). The present study will provide insights into the temporal cluster, spatial distribution, and spatial cluster of childhood diarrhoea in Karkala taluk. The study results will help planners and policymakers create effective diarrheal management strategies for the community under the supervision of primary health centres.

Although useful information can be obtained using spatial analytic approaches for multi-factorial studies, methodology, software, and test choice are still essential. It will take some time to gather the data and geocode each home in the village and any local water sources. Additional training for data collectors is required to operate Garmin Etrex 10 (GPS) device. Software tools like QGIS and SaTScan, as well as the acquisition of village satellite pictures, required specialised knowledge. Log points will be manually collected using an Etrex 10 GPS.

Scan statistics' new combination of field and geographical data will inspire ideas for enhancing children's health in the remotest parts of the community. For resource allocation, the "hotspot" and healthcare accessibility zones can be used as a guide. Although this study explores demographic, climatic, spatial and temporal aspects of diarrheal incidence among under-five children, this study has certain limitations. In the future, for robust scientific support, developing data sets that include these parameters is necessary.

Limitations

The limitations include the unavailability of routine spatial data, poor mechanism of reporting system by the private sectors, and lack of coordination between public and private health players in reporting diarrheal disease.

CONCLUSION

The spatiotemporal patterns of any infection are an inherent characteristic of routine health data. Also, voluminous data on health are generated from multiple sectors, which can be used as indicators for risk profiling. Our comprehension of the distribution of diarrhoea over geography and time will be improved using Geographic Information Systems (GIS) and spatial analysis methods. Planning public health interventions and strategies for reducing the diarrhoea burden by identifying potential spatial risk factors, thereby maximising the utilisation of current management methods, would be supported by

conducting geospatial studies. The data-guided public health policy advocacy and decision-making must link regional and temporal patterns of diarrheal incidence with predictive variables derived from data that is accessible and available at the health department.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Seboka, B.T., Hailegebreal, S., Yehualashet, D.E., Kabthyer, R.H., Negas, B., Kanno, G.G., *et al.*, Methods used in the spatial analysis of diarrhea: A protocol for a systematic review. *Medicine: Case Reports and Study Protocols*. 2022;3(4): e0228.
2. Dadonaite, B., Ritchie, H., Roser, M. Diarrheal diseases. Published online at OurWorlddata.org. (2018): Retrieved from: <https://ourworldindata.org/diarrheal-diseases>.
3. UNICEF: Diarrhea remains a leading killer of young children, despite the availability of simple treatment solutions, updated. 2016: UNICEF.org/child-health/ diarrheal disease.html.
4. International Institute for Population Sciences (IIPS) and ICF. 2017. National Family Health Survey (NFHS-4), 2015-16: India. Mumbai: IIPS.
5. International Institute for Population Sciences (IIPS) and ICF. 2021. National Family Health Survey (NFHS-5), 2019-21: India. Mumbai: IIPS.
6. Ghosh, K., Chakraborty, A.S., Mog, M. Prevalence of diarrhoea among under five children in India and its contextual determinants: A geo-spatial analysis. *Clinical Epidemiology and Global Health*. 2021;12(4):1-7.
7. Sarkar, R., Prabhakar, A.T., Manickam, S., Selvapandian, D., Raghava, M.V., Kang, G., *et al.*, Epidemiological investigation of an outbreak of acute diarrhoeal disease using geographic information systems. *Trans R Soc Trop Med Hyg*. 2007;101(6):587-593.
8. Greenough, P.G., Nelson, E.L. Beyond mapping: a case for geospatial analytics in humanitarian health. *Conflict and Health*. 2019;13(1):37-50.
9. Saravana, K.V., Devika, S., George, S., Jeyaseelan, L. Spatial mapping of acute diarrheal disease using GIS and estimation of relative risk using empirical Bayes approach. *Clinical Epidemiology and Global Health*. 2017;5(2):87-96.
10. Shaw, N.T., McGuire, S.K. Understanding the use of geographical information systems (GISs) in health informatics research: a review. *J Innov Health Inform*. 2017;24(2):228-233.
11. VoPham, T., Hart, J.E., Laden, F., Chiang, Y.Y. Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. *Environmental Health*. 2018;17(1):34-40.
12. Anselin, Luc., Ibnu, S., Youngihn, K. GeoDa: An introduction to spatial data analysis. *Geographical Analysis*. 2006;38(1):5-22.
13. Oliveira, F.L.P., Caçado, A.L.F., deSouza, G., Moreira, G.J.P., Kulldorff, M. Border analysis for spatial clusters. *International Journal of Health Geographics*. 2018;17(1):1-5.
14. Silva, I., Assunção, R., Costa, M. Power of the sequential Monte Carlo test. *Sequential Analysis*. 2009;28(2):163-174.
15. Hung, Y.W., Hoxha, K., Irwin, B.R., Law, M.R., Grépin, K.A. Using routine health information data for research in low- and middle-income countries: a systematic review. *BMC Health Services Research*. 2020;20(1):775-790.
16. Roy, A. "National Strategy for Artificial Intelligence", NITI Aayog India. 2018. Available from: <https://indiaai.gov.in/research-reports/national-strategy-for-artificial-intelligence>.
17. Sakshaug, J.W., Couper, M.P., Ofstedal, M.B., Weir, D.R. Linking survey and administrative records: Mechanisms of Consent. *Social Methods Res*. 2012 Nov;41(4):535-569.