https://doi.org/10.1038/s43856-024-00699-1

The use of SatScan software to map spatiotemporal trends and detect disease clusters: a systematic review

Check for updates

Ahmed Taha Aboushady [1,2], Fatma Mansour[3], Moustafa El Maghraby[4,5], Bárbara Teixeira [6], Sandra Cunha [7], Maria Manuel Dantas [8], Ahmed Nawwar[9], Amira Hegazy [10] ✉ & José Chen-Xu [11,12]

Abstract

Background Low and Middle-Income Countries (LMICs) often experience a disproportionate burden in health issues. One public health, epidemiology, and spatial statistics software tool has emerged as a stalwart for detecting disease clusters, mapping spatiotemporal trends, and analyzing health-related data—SatScan.

Methods This systematic review aims to provide a comprehensive overview of the extent of the use of spatiotemporal analysis, namely the use of SatScan for understanding health inequalities within LMICs within space and time parameters, shedding light on its potential to inform evidence-based public health interventions and policies. A systematic search was conducted in six electronic databases: PubMed, ScienceDirect, Web of Science, Cochrane, Scopus, and Embase. It included all human health-related articles, looking into data from LMICs. A descriptive analysis and quality assessment of the articles was performed.

Results Out of 5215 articles from different databases, 719 are included. Over 516 articles include themes on communicable diseases and over 50% of the articles come from China, Ethiopia, and Brazil. The Poisson-based model is the most commonly used model type, and more than 85% use secondary data sources, with the Demographic Health Surveys datasets being the most used.

Conclusions This systematic review allows us to understand which areas have been studied and which LMICs have developed research. This helps us detect health issues that have been neglected and the countries which require additional resources to increase their research capacities in this domain.

Plain language summary

Low and Middle-Income Countries often experience a disproportionate burden in health issues. Specific software tools are useful to detect disease clusters (regions where many people experience a disease), mapping disease trends over time and region, and analyzing health-related data. This systematic review provides an overview of the use of SatScan, a specific analysis tool designed for these types of analyses, with the results showing its main use in studying infectious diseases in China, Ethiopia, and Brazil. This allows for detection of health issues which may have been overlooked in scientific research and identification of countries which require additional research efforts. Results can inform future funding opportunities for research, policies, and public health interventions.

Health inequalities are a persistent and concerning global challenge, with Low and Middle-Income Countries (LMICs) often experiencing a disproportionate burden[1,2]. These inequalities materialize as disparities in access to healthcare services, differences in the prevalence of diseases, and variations in health outcomes among different socioeconomic groups and populations[3]. More than eighty percent of the world's population lives in LMICs, facing unique and complex dynamics contributing to health disparities[4]. Many interconnected factors, including economic disparities, limited healthcare infrastructure, inadequate sanitation, education gaps, and cultural norms, influence health inequalities in LMICs[5]. These nations often grapple with the double burden of infectious diseases and non-communicable diseases, further compounding health disparities[6,7]. Addressing health inequalities requires a comprehensive approach that considers the social determinants of health, healthcare access, and the

[1]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [2]Mohammed Bin Rashid School of Government, Dubai, United Arab Emirates. [3]Faculty of Medicine, Alexandria University, Alexandria, Egypt. [4]Faculty of Dentistry, Minia University, Al Minya, Egypt. [5]Liverpool John Moores University, Liverpool, United Kingdom. [6]Local Health Unit Entre o Douro e Vouga, Santa Maria da Feira, Portugal. [7]USF As Gândras, Local Health Unit Coimbra, Coimbra, Portugal. [8]Public Health Unit, Local Health Unit Coimbra, Coimbra, Portugal. [9]Department of Global Health and Population, Harvard T.H Chan School of Public Health, Boston, MA, USA. [10]Department of Community Medicine and Public Health, Kasr Al Ainy Faculty of Medicine, Cairo University, Cairo, Egypt. [11]NOVA National School of Public Health, Public Health Research Centre, Comprehensive Health Research Center, CHRC, REAL, CCAL, NOVA University Lisbon, Lisbon, Portugal. [12]Public Health Unit, Local Health Unit Baixo Mondego, Figueira da Foz, Portugal. ✉e-mail: amirahegazy@kasralainy.edu.eg

broader economic and social context in which these inequalities emerge[3]. By acknowledging the complexities of health disparities in LMICs, policymakers, researchers, and healthcare professionals can work together to develop targeted interventions and policies to reduce these gaps and promote equitable access to healthcare, ultimately improving the well-being and quality of life for the most vulnerable populations[3,5].

The interplay between space and time has revealed a critical field of study - spatiotemporal analysis[8]. This interdisciplinary approach allows us to examine how phenomena evolve and interact in geographical space and over varying time intervals[9]. Spatiotemporal analysis methodologies have become essential tools in numerous domains, including epidemiology, environmental science, urban planning, transportation, and beyond, offering insights that conventional spatial or temporal analyses alone cannot provide[8,9]. The essence of spatiotemporal analysis lies in its ability to uncover hidden patterns, identify trends, and detect anomalies within dynamic datasets[8]. By integrating spatial and temporal dimensions, researchers gain a more holistic understanding of how geographical locations and temporal sequences influence events, behaviors, and phenomena. This approach is particularly valuable in addressing questions related to the spread of diseases, the impact of environmental changes, the optimization of transportation networks, and many other complex, real-world challenges[9].

One public health, epidemiology, and spatial statistics software tool has emerged as a stalwart for detecting disease clusters, mapping spatiotemporal trends, and analyzing health-related data—SatScan[10,11]. Satscan is the first free software for cluster detection and the most used for implementing the spatial scan statistic method, incorporating temporal variation, and offering a wide variety of scanning models, with adjustable parameters[12,13]. SaTScan provides the location, size, and p-value for the clusters. SatScan is a versatile and powerful software that is easy to use and interpret playing a pivotal role in unraveling patterns of disease occurrences and uncovering hidden insights within complex data sets[11,13]. Developed by Martin Kulldorff and his colleagues, in 1997, SatScan has evolved into an indispensable resource for epidemiologists and researchers seeking to understand the distribution of health events, including disease outbreaks and health disparities[10]. By combining spatial and temporal information, SatScan offers a unique capacity to identify clusters of diseases or health outcomes that may otherwise go unnoticed. This software has been applied to various health-related challenges, from tracking infectious diseases to studying the spatial distribution of chronic conditions[14,15]. Given its widespread use, SatScan has been integrated into other software, such as WHONET for Hospital Acquired Infections cluster detection and R for statistical computing. Furthermore, SatScan has received several awards and endorsements from the US Center for Disease Control and Prevention, the National Cancer Institute, and others[16,17]. It also surpasses other software, such as GeoDa and QGIS, which do not provide a specialized spatial and temporal overview in analyzing public health issues, which makes SatScan a preferable software in public health research. There are also other similar software, such as FlexScan, which is used for non-parametric data, and has a different definition and algorithm for clusters, and Tree-Scan, by the same developers but more focused on drug and vaccine surveillance[18]. Despite the development of other software, SatScan is still being used for spatiotemporal analysis more than other software.

SatScan can detect and characterize statistically non-random spatial event clusters, including insights into their timing, duration, and precise locations. It accomplishes this by employing a model featuring a cylindrical window, where the base of the cylinder represents spatial dimensions, and its height represents time. This dynamic window systematically traverses space and time, investigating various locations, sizes, and time intervals[19]. A likelihood ratio is computed for each window position, which measures the deviation between observed and expected cases while considering the relevant population size and time period. Furthermore, the software evaluates the relationship between observed and expected cases inside and outside the cylindrical window, calculating the likelihood ratio and relative risk (RR)[20,21]. To effectively utilize SatScan, three crucial inputs are required: the observed variable quantity over a defined period (e.g., daily case counts), the population size for each spatial unit (e.g., population per health region), and a shapefile containing the necessary geographical coordinates (in this analysis, Cartesian coordinates were utilized)[10,22].

SatScan provides flexibility by offering various model types, including a Poisson-based model suitable for areas where the number of events follows a Poisson distribution relative to a known underlying population at risk, a Bernoulli model tailored to analyze 0/1 event data, such as cases and controls, a space-time permutation model designed for case data exclusively, an ordinal model suitable for ordered categorical data, an exponential model for analyzing survival time data with or without censored variables, a normal model for other types of continuous data, and finally a multinomial statistical model for assessing geographical variations in age-specific populations[23–25].

With the unprecedented availability of data and advancements in spatial and temporal analytics, understanding health inequalities in Low- and Middle-Income Countries (LMIC) has taken on a new dimension. The intricate interplay between geography, time, and public health outcomes has made it imperative to employ innovative tools and techniques for comprehensive assessment and effective intervention. Among the analytical tools at researchers' disposal, the SatScan software is powerful for detecting disease clusters and exploring spatiotemporal patterns in health disparities.

This systematic review aims to provide a comprehensive overview of the extent of the use of spatiotemporal analysis, using SatScan to analyze health issues within LMICs. This tool provides information about the distribution of health issues through space and time, uncovering the most affected areas, shedding a light on its potential to inform evidence-based public health interventions and policies. In addition, the results could inform researchers utilizing spatiotemporal analyses and cluster detection to enhance outbreak detection and resource allocation. We aim to offer valuable insights into the evolving landscape of spatiotemporal analysis in public health, improve the decision-making processes, and contribute to the ongoing efforts to mitigate health disparities in LMICs.

Methods

Search strategy

This study analysed papers published until 2023 with the following characteristics: Population: Low and Middle-Income Countries; Methodology utilized: SatScan software, at any study stage; Field of Study: Human Health. The following search string was utilized: "Satscan" OR "Sat-Scan" OR "Sat Scan".

We utilized PRISMA Guidelines (Fig. 1) to define the search strategy. Studies were identified through a systematic search of six electronic databases: PubMed, ScienceDirect, Web of Science, Cochrane, Scopus, and Embase. The search was conducted in all databases on February 2nd, 2024.

Regarding inclusion criteria, we included any article that approached health or health determinants in humans. Furthermore, we limited articles to the following languages: English, Spanish, Portuguese, and French. Studies in animals, categorized as non-human, were excluded. We also excluded articles due to their study design, namely reviews, such as scoping or systematic reviews, viewpoints or perspectives, conference proceedings, abstracts only, and editorials.

Reviewers screened articles in three phases: title, abstract, and full text, and data were extracted from each study. The protocol of this review was registered in PROSPERO (ID: CRD42023411382)[26] and the complete PRISMA Checklist is available in the Supplementary Data 1.

Selection process and data extraction

The retrieved articles were analysed independently in pairs by authors in line with the pre-defined criteria to determine eligibility for inclusion. The screening phase comprises the articles' titles, abstract, and full-text analyses, and it was conducted in CADIMA, requiring two reviewers[27]. For the screening phase, the possible classifications for the inclusion of the studies are: "Yes," "Unclear," and "No". If the paper is classified as "No" by both researchers, the paper is removed from the database. If it receives an "Unclear" or "Yes," it moves to the next selection phase. Disagreements

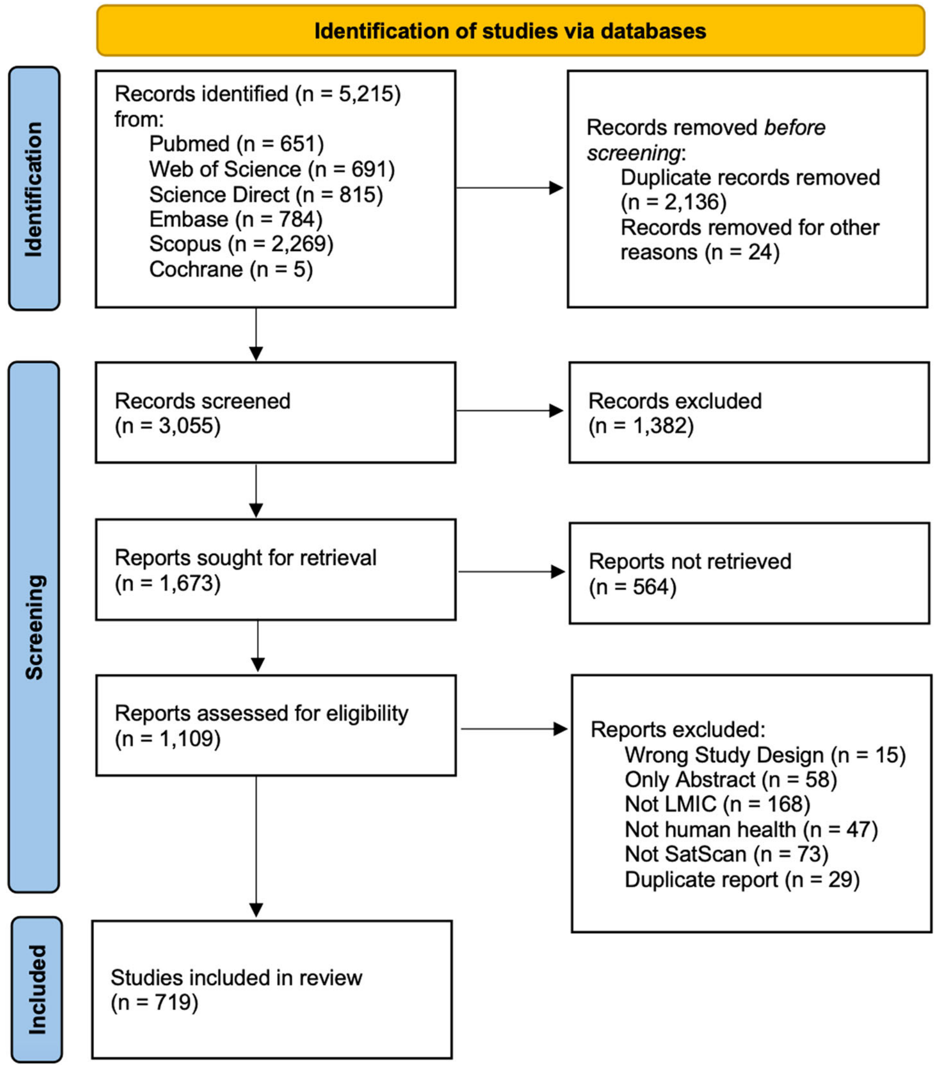

Fig. 1 | PRISMA flow diagram. The scheme shows the number of articles excluded in the identification and screening processes, with a final number of 719 articles included in this review.

between the reviewers in independent assessments were discussed in a specific meeting, resolved by consensus, or after discussion with another researcher.

The data was extracted independently by two people. Conflicts were solved by reconciliation between reviewers. Data collection for this study was conducted through an extensive review of articles incorporated into the CADIMA database. The collected data encompassed a range of characteristics, including the journal in which the articles were published, the year of publication, the country under study, income category based on the World Bank Group classification (comprising low-income, lower-middle-income, and upper-middle income), the title of the article, the study's stated objectives, study design (prospective or retrospective), study timeframe, the main data source employed, the utilization of secondary data sources, population size, age groups within the population, specific target groups, the disease or diagnosis investigated, the purpose of implementing SatScan, the SatScan model used, the unit of spatial analysis, the unit of temporal analysis, and any additional software used for statistical analysis.

Risk of bias assessment

The risk of bias assessment phase was carried out by all reviewers, with revision by a second reviewer. This step allowed for a comprehensive assessment of the methodological rigor in all incorporated studies, employing a quality evaluation instrument for evaluating spatial and modeling studies; our methodology was adopted from the Risk of bias assessment used in other similar articles[15,28,29]. It features an eight-point scoring system that has been revised and adapted to gauge the quality of each study considering its defined objectives, model validity, overarching findings, and research conclusions. A standardized item list was used to grade the included studies' quality and risk of bias, with a maximum total of 16 points. The overall scores were categorized into distinct quality tiers, encompassing low (for scores less than 8), medium (for scores ranging from 8 to 10), high (for scores between 11 and 13), and very high (for scores exceeding 13). Detailed information about the methodology is available in Supplementary Table T1.

Data synthesis and statistical analysis

The qualitative synthesis was performed as a narrative review of the results of the studies, describing the summary of the studies which utilized SatScan and main health outcomes. In the analysis of the countries included in the multi-country studies, we accounted for each individual LMIC.

Data synthesis and statistical analysis of data were performed using OpenEpi, Microsoft Excel, and IBM SPSS Statistics 28 ®, mainly through descriptive statistics. The map presented was produced using R 4.3.2.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Geographical distribution

A total of 5215 articles were identified through the search of databases, of which 651 were from PubMed, 815 from ScienceDirect, 691 from Web of

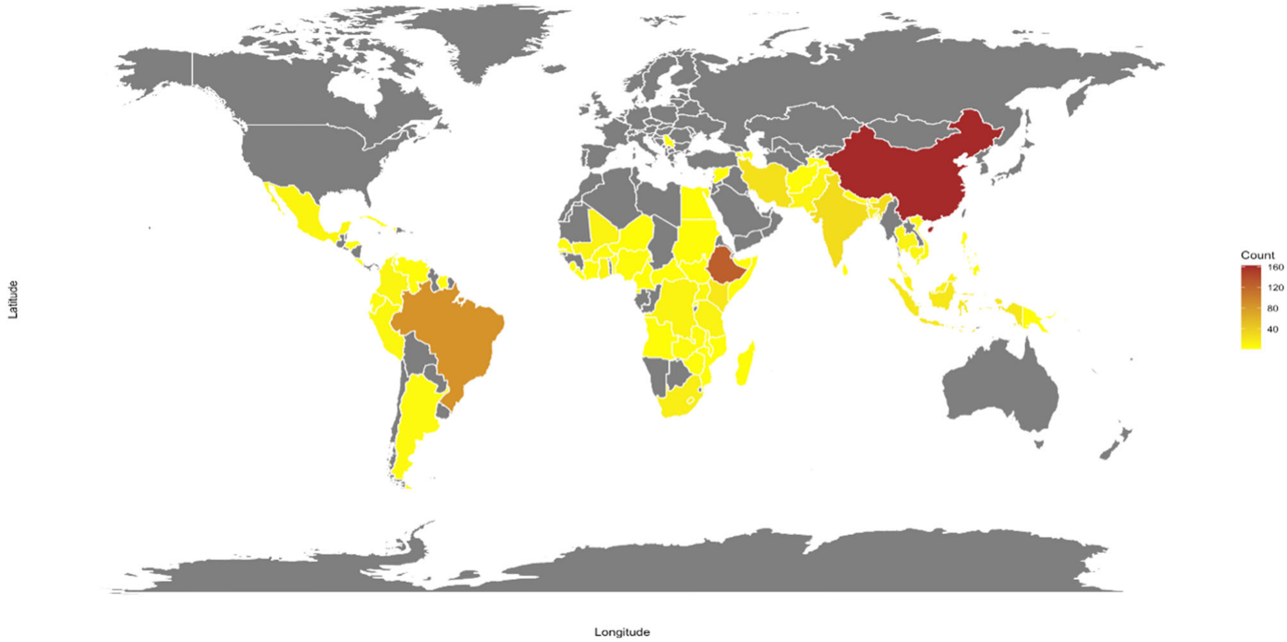

Fig. 2 | Distribution of papers analysed through low- and middle-income countries. The map shows the number of publications, ranging from the lowest (in yellow) to the highest (in red) frequency. High-income countries and countries with no publications using SatScan are marked in grey.

Science, 2269 from Scopus, 5 from Cochrane, and 784 from Embase. After removing duplicates, 3055 articles remained for title screening. Of these, 1382 were excluded, resulting in 1673 selected for abstract screening. The full-text screening included 1109 articles selected for review. Ultimately, 719 studies met the inclusion criteria and were included in the review. The complete list of articles included can be found in the Supplementary Note 1.

The analysis of the studies showed that 68 countries presented papers with a geospatial analysis of human health issues using SatScan (Fig. 2). Studies were mostly conducted in three main countries: China ($n = 163$, 21.6%), Ethiopia ($n = 128$, 16.9%), and Brazil ($n = 92$, 12.2%). The continent most represented was Asia, with 301 papers (39.8%), followed by Africa ($n = 290$, 38.4%). A further breakdown by income categories showed that Upper Middle-Income Countries present most of the papers ($n = 368$, 48.7%), followed by Low-Income Countries ($n = 1198$, 26.2%) and closely by Lower Middle-Income Countries ($n = 190$, 25.1%). Supplementary Data 2 presents the data by country and by income category.

Content analysis

The distribution of studies in this meta-synthesisis categorized into various topics related to communicable and non-communicable diseases and other health-related subjects.

Most studies focus on communicable diseases ($n = 516$, 71.7%) with a special focus on "Neglected Tropical Diseases," with 125 studies indicating a significant research emphasis in this area. Following this, "Malaria" and "Tuberculosis" have substantial representation with 101 and 51 studies, respectively. COVID-19 is also a prominent topic, with 23 dedicated studies likely reflecting the global interest in the pandemic in recent years.

58 studies represent non-communicable diseases, with cancer being the most frequent topic, with 36 articles. Additionally, there are studies covering various health-related behaviors and practices, maternal and sexual health, with 26, 40, and 19 studies, respectively. This diverse distribution of topics reflects the multifaceted nature of public health research and the importance of addressing a wide range of health issues.

When analyzing the three most prevalent countries, the most common health outcomes were communicable diseases in Brazil and China (around 85% of all outcomes reported for each country), whereas in Ethiopia the focus was in communicable diseases (29%), maternal health (24%), and nutrition (18%).

Regarding data sources, a relatively small proportion of the reviewed articles, specifically 100 papers (13.9%), relied on primary data sources. In contrast, a majority of the articles ($n = 619$) used secondary data sources. Notably, the Demographic Health Survey datasets emerged as the most frequently employed secondary data source, used in 115 articles, constituting approximately 16.0% of all articles that utilized secondary data. Additionally, out of these 115 articles, a significant majority, precisely 95 articles, originated from research conducted in Ethiopia.

Publications over time

Figure 3 presents the variation of the number of articles included, categorized by their publication year. The data reveals a steady increase in the quantity of articles over time. The earliest article was from 2005 and reports just one article, but as time progresses, the numbers demonstrate a notable surge in research publications. 2021 stands out as the peak, with 97 articles, followed closely by 2022 with 94, and 2023 with 84. When analysing disaggregated data (and excluding East Asia), it is noticeable that the increase in published papers is mostly driven by the African region ($n = 257$), with a peak in 2021 with 44 articles.

Methodological aspects

From the studies included, the most common study design was an ecological design ($n = 272$, 37.8%), followed closely by cross-sectional studies ($n = 235$, 32.7%) and cohorts ($n = 94$, 13.1%). Regarding their time characteristics, the majority of the studies included in the review were retrospective ($n = 637$, 88.6%), against 76 prospective studies (10.6%) and six papers categorized as conducting both prospective and retrospective or not mentioned.

The usage of SatScan in these articles entailed a Spatiotemporal analysis. Figure 4 below provides an insight into the usage of SatScan within the articles included in the systematic review, which are categorized according to the model types employed in the studies. The most prevalent model used among these studies is the Poisson-based model, with a substantial count of 393 articles, indicating its popularity in the context of SatScan research. The Bernoulli model is the second most utilized, with 162 articles dedicated to it. Notably, there are 99 articles where the model type is not mentioned. Other model types like Permutation, Normal, and Ordinal also play a role in this body of literature, with varying, though smaller, counts. Additionally, seventeen articles used more than one model in their analysis. Table 1

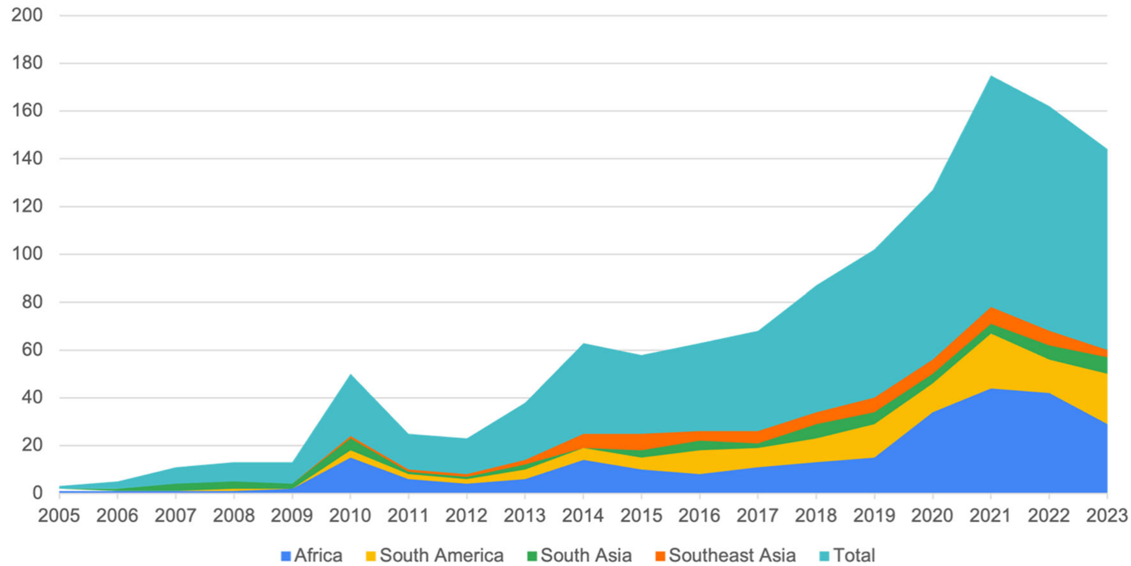

Fig. 3 | Absolute variation of the year of publishing of studies included. The graphic illustrates the total variation of articles which used SatScan from 2005 to 2023, in all regions and with disaggregation for Africa, South America, South Asia, and Southeast Asia.

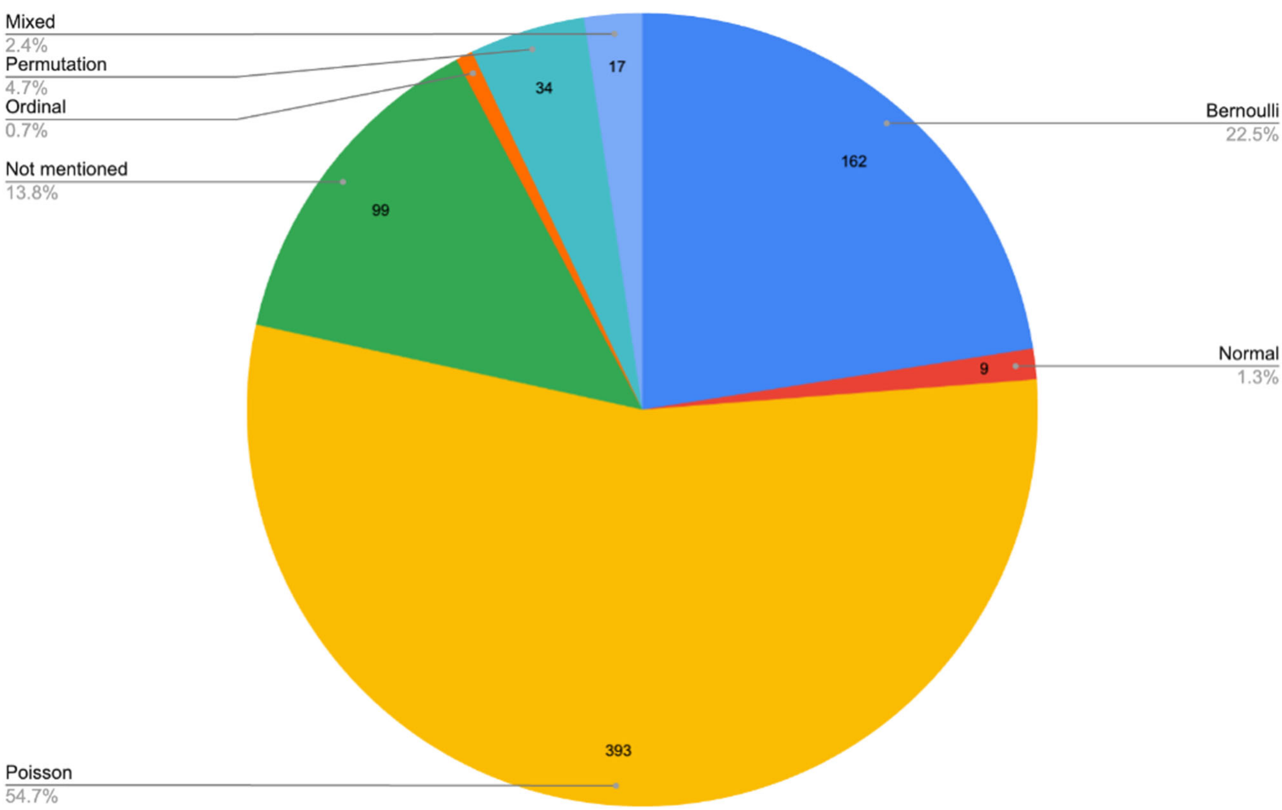

Fig. 4 | Distribution of articles by Model utilized in the SatScan software. The pie chart presents the relative distribution of the papers by model utilized, with most articles adopting the Poisson model, followed by the Bernoulli and the Permutation models.

Risk of bias and quality of papers included

Most papers were rated with the highest value ($n = 358$, 49.8%), with the majority making it into the Very High classification ($n = 630$, 87.6%). Around 10.8% ($n = 78$) of articles presented a high quality, whereas 1.5% ($n = 11$) were rated as medium quality. We have still included the results of these articles in our analysis. The results of the risk of bias are presented in Table 2. The full scoring can be found in Supplementary Data 3.

Discussion

This article presents a systematic literature review on geospatial analysis and clustering using SatScan, particularly on Low- and Middle-Income Countries (LMICs). This study represents the first review of its kind in this context, aimed at enhancing our understanding of utilizing spatio-temporal analysis and clustering methods to further understand health inequalities.

Table 1 | Diseases and other health topics analysed in the articles

Group	Disease	N	Percentage
Communicable diseases	Neglected tropical diseases	125	17.4%
	Malaria	101	14.0%
	Tuberculosis	51	7.1%
	Other communicable diseases	61	8.5%
	Sexually transmitted infections	34	4.7%
	Vaccine-preventable communicable diseases	40	5.6%
	COVID-19	23	3.2%
	Diarrheal infections	23	3.2%
	Acute Respiratory infections	14	1.9%
	Hemorrhagic fever	12	1.7%
	Leprosy	13	1.8%
	Hand, foot, and mouth disease	10	1.4%
	Hepatitis	9	1.3%
Non-communicable disease	Cancer	36	5.0%
	Mental health	8	1.1%
	Congenital diseases	5	0.7%
	Asthma	2	0.3%
	Diabetes	2	0.3%
	Others	5	0.7%
Health-related behaviors and practices		26	3.6%
Maternal Health		40	5.6%
Sexual Health		19	2.6%
Health Services Indicators		15	2.1%
Nutrition		37	5.1%
Other		6	0.8%

Table 2 | Classifications derived from the risk of bias assessment

Classification	Rating	Count	Percentage
Medium	8	7	1.0%
	9	1	0.1%
	10	3	0.4%
High	11	5	0.7%
	12	26	3.6%
	13	47	6.5%
Very High	14	92	12.8%
	15	180	25.0%
	16	358	49.8%
Total		719	100.0%

In the studies included in this review, the Poisson and Bernoulli models were the most used in the literature. These two models are particularly suitable for geographical disease surveillance and cluster detection[25]; this aligns with the most common diseases and topics we found. Also, as shown in Fig. 3, there has been a substantial growth in the number of articles over time, which suggests a growing interest and research activity in spatial clustering and SatScan. Furthermore, this rise is aligned with the general increase in publications with the COVID-19 pandemic[30,31].

Each model requires parameters including the maximum size for the spatial scanning window in the population at risk, the maximum size of the temporal window, the Monte Carlo replications, the acceptable alpha level, and a null hypothesis of complete spatial randomness[22,32]. The team also tried to include the different parameters used in the studies. However, most papers didn't include model variables and information, highlighting the need for more comprehensive reporting in research publications.

The demographic health surveys, funded by the USAID, were the most common dataset included in the studies. Such population-based surveys that are made openly accessible do enable public health research in LMICs, particularly since accessing national data could be challenging and sometimes impossible, and conducting primary data would be costly for many institutions in resource-limited settings[33].

Despite its usefulness, studies using SatScan in LMICs fail to cover almost half of the countries in these categories, with 47.4% of LMICs not presenting a sole study in our systematic review. Several countries in Africa, the Middle East, and Southeastern Asia, as well as countries located in the Caribbean islands, fail to conduct geospatial studies, which proves to be resourceful for decision-making. This could be due to the lack of capacities, the use of other software[34], and/or publication in other avenues not included in the covered databases. Joint efforts are required for these countries and neighboring ones for additional funding to strengthen research capacity for health policy and intervention.

China, Ethiopia, and Brazil were the most producers, which can be explained by having China and Brazil being the top countries worldwide spending on research. The health outcomes focus in the three most prominent countries evidenced the different levels of economic development of these countries, which reflect on the definition of health priorities. This is ultimately influenced by the existence of health surveillance systems, as well as capable human resources dedicated to these topics. Along with data availability for researchers, most of their studies were based on national surveillance data from health authorities[35]. Particularly China, nationwide routinely collected health datasets are kept in public databases, with regular monitoring of communicable diseases[36]. Likewise, Brazil also implemented a monitoring system for notifiable diseases, including tuberculosis[37]. While for Ethiopia, research output can be attributed to a confluence of factors, including the national emphasis on developing a research culture, evidenced by initiatives like Research and Community Service Directorates and availability of Demographic Health Survey data, a very popular data source in Ethiopia funded by multilateral agreements, which represents around 75% of all Ethiopian studies[18,38].

The distribution of studies provides valuable insight into the extensive and diverse utilization of SatScan in various health fields. The data reveal that this method has been extensively applied in the domain of communicable diseases, with a special emphasis on Neglected Tropical Diseases, which indicates the method's effectiveness in identifying spatial patterns in diseases that predominantly affect specific regions or populations. The substantial representation of Malaria and Tuberculosis studies underscores its relevance in the context of diseases with global significance. Moreover, the presence of studies related to COVID-19 signifies the adaptability of SatScan in addressing emerging and dynamic health crises.

In addition to communicable diseases, the data also demonstrate the utility of SatScan in the analysis of non-communicable diseases, including cancer and obesity. Studies covered various health-related behaviors and practices, maternal health, and sexual health highlighting the versatility of SatScan in understanding and monitoring spatial patterns of health-related issues beyond traditional infectious diseases.

This diverse distribution of topics underscores the multifaceted nature of public health research and the importance of utilizing innovative spatial analysis techniques like SatScan to address various health issues. SatScan's adaptability to various health fields reflects its potential as a valuable tool for policymakers, epidemiologists, and researchers to better understand the geographical distribution of health outcomes and tailor intervention strategies accordingly. Despite its funding and endorsement by the National Cancer Institute[24], SatScan remains less explored for non-communicable diseases and their health determinants in LMICs, and such analyses might prove valuable for targeted public health interventions.

Furthermore, despite the increasing trend and interest in geospatial methodologies and spatial clustering, it is important to acknowledge its dangers and limitations. The challenges in SatScan and disease cluster detection are multifaceted. A key issue is the common practice of reporting detected clusters without detailing the relative risk estimates for individual regions within these clusters. SatScan, which employs the likelihood ratio, may identify clusters that contain both statistically significant and non-elevated risk areas, risking misinterpretation. Neglecting the examination of relative risk estimates in individual regions can lead to unwarranted conclusions and potential criticism. Researchers must address this challenge to enhance cluster identification and interpretation's scientific rigor and accuracy[39].

The use of SatScan has implications for policy formulation and evaluation and the delineation of future public health initiatives. By highlighting the spatial distribution patterns of diseases and health-related phenomena, these insights empower policymakers to strategize resource allocation more effectively, directing interventions toward the most underserved regions. Specifically, the revelation of areas disproportionately burdened by communicable diseases, such as neglected tropical diseases and malaria, underscores the necessity for targeted disease management strategies, refined surveillance mechanisms, and the equitable distribution of healthcare resources. Additionally, identifying significant research capacity disparities among LMICs accentuates the critical need for policy interventions to bolster local research infrastructures, improve data availability, and foster international collaborations to mitigate these disparities. For future public health interventions, the utility of integrating geospatial analyses in the conceptualization and assessment phases is underscored, advocating for evidence-based and geographically nuanced interventions to augment efficacy. Furthermore, the discernment of neglected research domains and health conditions calls for a diversification of research focus and investment, urging the investigation of non-communicable diseases and novel health threats within these methodological frameworks. Such an approach aligns with the objective of fostering equitable health outcomes worldwide and enriches the global knowledge base, catalyzing innovations in public health policy and practice.

This systematic review identifies several critical limitations in the current research landscape on cluster detection in infectious diseases. Firstly, it highlights the need for a more explicit consideration of the temporal dimension in cluster detection, as it plays a pivotal role in revealing the cyclical and dynamic nature of diseases, providing essential insights into disease patterns. Secondly, there is a call for more comprehensive research that systematically compares the validity of clusters identified through repetitive spatial methods with those derived from true space-time clustering algorithms, as this methodology can significantly impact the risk of false alarms. Furthermore, underutilized clustering techniques such as space-time Bayesian modeling[40], local likelihood disease clustering[41], and FlexScan software[42], which delineates various levels of cluster boundaries, present promising avenues for research in the context of temporal processes and infectious diseases. It's also crucial to acknowledge the language limitation in the literature review, as focusing on English articles excluded a substantial body of relevant research in Chinese, as China had the highest number of articles in English. Future studies should aim to bridge this language gap for a more comprehensive understanding of the field.

Another limitation is the quality assessment tool. We utilized a proposed tool for risk of bias assessment which could be further improved, including a publication bias dimension. Nonetheless, our study did not focus on the results of the studies, rather analyzing structural concepts such as study design, methodologies, time and geographic aspects, and topics which the papers focused on.

Additionally, the paper's focus on specific space-time modeling algorithms run with SatScan excluded other software and approaches, warranting the exploration of a wider range of methodologies. Lastly, researchers should take measures to enhance the transparency and replicability of their studies, such as reporting all model parameters by sharing data and developing web-based visualization solutions. Finally, expanding the scope of research to encompass non-human studies is recommended, as it could offer valuable insights into disease clustering across various contexts regarding One Health. These limitations underscore the ongoing need for research and development in infectious disease cluster detection.

This systematic review allowed us to understand which areas have been studied and which LMICs have developed research capacities for Spatio-temporal analysis through SatScan. This helps us detect which health issues have been neglected and which countries require additional resources to increase their research capacities in this domain.

Data availability

The data that support the findings of this study are based on the studies included, available in the respective peer-reviewed journals. The list of included studies can be found in Supplementary Note 1. The data retrieved in this study is presented in the Supplementary Information and Supplementary Data 1–3. Any additional data can be made available upon reasonable request to the corresponding author.

Received: 31 October 2023; Accepted: 5 December 2024;
Published online: 21 March 2025

References

1. Aung, M. N., Koyanagi, Y. & Yuasa, M. Health inequality among different economies during early phase of COVID-19 pandemic. *J. Egypt Public Health Assoc.* **96**, 3 (2021).
2. Gwatkin, D. R. Trends in health inequalities in developing countries. *Lancet Glob. Health* **5**, e371–e372 (2017).
3. Riley, W. J. Health disparities: gaps in access, quality and affordability of medical care. *Trans. Am. Clin. Climatol. Assoc.* **123**, 167–174 (2012).
4. World Bank Open Data. World Bank Open Data https://data.worldbank.org.
5. National Academies of Sciences, E. et al. The Root Causes of Health Inequity. in *Communities in Action: Pathways to Health Equity* (National Academies Press (US), 2017).
6. Boutayeb, A. The burden of communicable and non-communicable diseases in developing countries. *Handbook of Disease Burdens and Quality of Life Measures* 531–546 https://doi.org/10.1007/978-0-387-78665-0_32 (2010).
7. Countries facing double burden with chronic and infectious diseases—*UN report | UN News*. https://news.un.org/en/story/2011/05/375072 (2011).
8. *Spatiotemporal Analysis—an overview | ScienceDirect Topics*. https://www.sciencedirect.com/topics/medicine-and-dentistry/spatiotemporal-analysis.
9. Sifaki-Pistolla, D. I. et al. *Geospatial and Spatio-Temporal Analysis in Health Research: GIS in Health*. https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-5225-8054-6.ch069 https://www.igi-global.com/gateway/chapter/www.igi-global.com/gateway/chapter/222961 (1AD).
10. M., K., Services, I. M. & Inc. SaTScanTM v.7.0: Software for the spatial and time-scan statistics.
11. Spatial disease clusters: detection and inference—PubMed. https://pubmed.ncbi.nlm.nih.gov/7644860/.
12. Chen, J., Roth, R. E., Naito, A. T., Lengerich, E. J. & MacEachren, A. M. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *Int. J. Health Geogr.* **7**, 57 (2008).
13. Block, R. Software review: scanning for clusters in space and time: a tutorial review of SaTScan. *Soc. Sci. Comput. Rev.* **25**, 272–278 (2007).
14. Byun, H. G., Lee, N. & Hwang, S. A systematic review of spatial and spatio-temporal analyses in public health research in Korea. *J. Prev. Med Public Health* **54**, 301–308 (2021).
15. Tesema, G. A., Tessema, Z. T., Heritier, S., Stirling, R. G. & Earnest, A. A systematic review of joint spatial and spatiotemporal models in health research. *Int. J. Environ. Res. Public Health* **20**, 5295 (2023).

16. Natale, A., Stelling, J., Meledandri, M., Messenger, L. A. & D'Ancona, F. Use of WHONET-SaTScan system for simulated real-time detection of antimicrobial resistance clusters in a hospital in Italy, 2012 to 2014. *Eur. Surveill.* **22**, 30484 (2017).
17. Kleinman, K., Hostovich, S. & Moosa, A. *rsatscan: Tools, Classes, and Methods for Interfacing with 'SaTScan' Stand-Alone Software* (2023).
18. SaTScan—Related Software. https://www.satscan.org/related_software.html.
19. Azage, M., Kumie, A., Worku, A. & Bagtzoglou, A. C. Childhood diarrhea exhibits spatiotemporal variation in Northwest Ethiopia: a SaTScan spatial statistical analysis. *PLoS ONE* **10**, e0144690 (2015).
20. Kulldorff, M. A spatial scan statistic. *Commun. Stat. Theory Methods* **26**, 1481–1496 (1997).
21. Linton, S. L., Jennings, J. M., Latkin, C. A., Gomez, M. B. & Mehta, S. H. Application of space-time scan statistics to describe geographic and temporal clustering of visible drug activity. *J. Urban Health* **91**, 940–956 (2014).
22. SaTScan_Users_Guide.pdf.
23. Kulldorff, M., Huang, L. & Konty, K. A scan statistic for continuous data based on the normal probability model. *Int. J. Health Geogr.* **8**, 58 (2009).
24. Spatial and space-time scan statistics. https://surveillance.cancer.gov/satscan/.
25. Huang, L., Kulldorff, M. & Gregorio, D. A Spatial Scan Statistic for Survival Data. *Biometrics* **63**, 109–118 (2007).
26. Aboushady, A., Chen Xu, J. & Nawwar, A. *The use of spatiotemporal analysis for health inequalities in low and middle income countries: A systematic review*. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42023411382.
27. Kohl, C. et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. *Environ. Evid.* **7**, 8 (2018).
28. Fone, D. et al. Systematic review of the use and value of computer simulation modelling in population health and health care delivery. *J. Public Health Med.* **25**, 325–335 (2003).
29. Harris, R. C., Sumner, T., Knight, G. M. & White, R. G. Systematic review of mathematical models exploring the epidemiological impact of future TB vaccines. *Hum. Vaccine Immunother.* **12**, 2813–2832 (2016).
30. Ioannidis, J. P. A., Salholz-Hillel, M., Boyack, K. W. & Baas, J. The rapid, massive growth of COVID-19 authors in the scientific literature. *R. Soc. Open Sci.* **8**, 210389 (2021).
31. Ma, Y., Yin, F., Zhang, T., Zhou, X. A. & Li, X. Selection of the maximum spatial cluster size of the spatial scan statistic by using the maximum clustering set-proportion statistic. *PLoS ONE* **11**, e0147918 (2016).
32. Hohl, A., Delmelle, E. M., Desjardins, M. R. & Lan, Y. Daily Surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spat. Spatio-temporal Epidemiol.* **34**, 100354 (2020).
33. *The DHS Program—Quality information to plan, monitor and improve population, health, and nutrition programs*. https://dhsprogram.com/.
34. Zhou, J. et al. Determinants and spatial patterns of adult overweight and hypertension in a high HIV prevalence rural South African population. *Health Place* **18**, 1300–1306 (2012).
35. Top countries by R&D expenditure 2022. *Statista* https://www.statista.com/statistics/732247/worldwide-research-and-development-gross-expenditure-top-countries/.
36. Nation-Wide Routinely Collected Health Datasets in China: A Scoping Review—PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9532513/.
37. Rocha, M. S. et al. Notifiable Diseases Information System (SINAN): main features of tuberculosis notification and data analysis. *Epidemiologia e servicos de. Saude. Rev. do Sist. Unico de. Saude. do Bras.* **29**, e2019017 (2020).
38. Alemu, A. Factors determining research productivity in Ethiopian Universities: the case of Madda Walabu University. *Soc. Sci. Humanities Open* **8**, 100562 (2023).
39. Tango, T. Spatial scan statistics can be dangerous. *Stat. Methods Med. Res.* **30**, 75–86 (2021).
40. Hossain, M. M. & Lawson, A. B. Space-time Bayesian small area disease risk models: development and evaluation with a focus on cluster detection. *Environ. Ecol. Stat.* **17**, 73–95 (2010).
41. Hossain, M. D., Monir & Lawson, A. B. Local likelihood disease clustering: development and evaluation. *Environ. Ecol. Stat.* **12**, 259–273 (2005).
42. Tango, T. & Takahashi, K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J. Health Geogr.* **4**, 11 (2005).

Author contributions

A.T.A. coordinated the research project; A.T.A., A.N., and J.C-X. jointly conceived the study; A.T.A. and J.C.-X. conducted the database searches; A.T.A, A.N., J.C.-X., F.M., M.E.M., B.T., S.C., and M.M.D. contributed to the screening and data extraction; A.T.A. and J.C.-X. prepared the manuscript; A.T.A, A.N., J.C.-X., F.M., M.E.M., B.T., S.C., M.M.D and A.H. revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43856-024-00699-1.

Correspondence and requests for materials should be addressed to Amira Hegazy.

Peer review information *Communications Medicine* thanks Munazza Fatima, Oladimeji Mudele, and Chester Kalinda for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.