Spatial Analysis of Identifying the Association between Risk Factors and Tuberculosis Cases: A Review

Nur Adibah Mohidem[1,3]*, Malina Osman[2], Zailina Hashim[3], Farrah Melissa Muharam[4], Saliza Mohd Elias[3], Aminuddin Ma'pol[5]

[1]Public Health Unit, Department of Primary Health Care, Faculty of Medicine and Health Sciences, Universiti Sains Islam Malaysia, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia
[2]Department of Community Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[3]Department of Environmental and Occupational Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[4]Department of Agriculture Technology, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[5]Dungun Health District Office, Jalan Yahya Ahmad, 23000, Dungun, Terengganu, Malaysia

*Corresponding author: nuradibahmohidem@usim.edu.my

Abstract

Tuberculosis (TB) transmission frequently occurs in a household or group within a population, resulting in a variety of spatial patterns. However, the apparent spatial clustering of TB may represent the ongoing transmission or co-location of associated risk factors, which can vary significantly based on the type of data available, the analysis methods used, and the dynamics of the underlying population. This study aims to review the spatial analyses used for monitoring the trends involving and associations between risk factors and TB cases by applying the concept of spatial epidemiology. The role of the Geographic Information System in spatial epidemiology is discussed. Previous studies involving spatial analysis of TB cases - which include kriging, spatial autocorrelation, kernel density estimation, hotspot analysis, and regression analysis - are reviewed. The type of analysis was chosen based on the purpose of each study, which could explain the role of the transmission to reactivation of the disease as a driver of TB spatial distribution. In diverse situations, a number of different spatial analysis techniques were used, with all the studies demonstrating significant heterogeneity in terms of the spatial distribution of TB. Future research is needed to determine the best methods to use in different situations and, where possible, consider unreported cases when using notification data. A combination of genotypic, molecular, and geospatial approaches to examine epidemiologically related cases could improve TB control and provide significant contributions to the current knowledge.

Keywords: risk factor, trend, spatial, tuberculosis

INTRODUCTION

Tuberculosis (TB) is a chronic infectious disease caused by *Mycobacterium tuberculosis*. TB primarily attacks the lungs, when it is termed pulmonary TB, but it can also spread to other organs of the body, when it is called extra-pulmonary TB. The disease spreads from person to person through the air. Contact with a person who has active TB may result in infection, with the disease being in either an active or a latent state. Droplets of nuclei containing tubercle bacilli are exhaled into the air whenever a patient with active TB

coughs, talks, or spits [1]. People who inhale these droplets will experience symptoms including fever, weakness, night sweats, coughing, fatigue, loss of appetite, and weight loss [2, 3].

TB is one of the leading ten infectious disease-related deaths worldwide, ranking above the human immunodeficiency virus (HIV). In 2019, there were approximately 10 million cases of and 1.4 million deaths due to TB. Although the incidence has decreased by about 2% per year, the End TB strategy introduced by the WHO did not achieve its 20% reduction target between 2015 and 2019, attaining only a 9% reduction. Although the number of deaths worldwide is decreasing each year, the 2020 target to reduce deaths by 35% was not on track because only a 14% reduction was achieved between 2015 and 2019 [4]. According to the Practice Guideline for the Control and Management of Tuberculosis (2012), screening for TB should be conducted in high-risk groups, but more studies are needed to assess the efficacy and feasibility of preventive measures among high-risk groups.

Many studies have concentrated on the biomedical aspects of TB, with the goal being to find a cure [5, 6]. The findings of these studies are limited because the statistical independence of the variables was derived from a molecular analysis of disease risk that did not take into account geographic referencing. Previous studies used non-geospatial data that did not consider patient locations when determining the relationship between risk factors and TB cases [7-9]. Forthcoming studies should focus on using geospatial data to improve the estimation of disease risk in high-risk population groups and disease transmission from one area to another.

A comprehensive literature search and research reports have utilised molecular epidemiology studies [10-12]. Spatial epidemiology studies, on the other hand, have been recognised as effective approaches in monitoring disease transmission over time and space. Analysis of TB cases in large areas is difficult because each case in every single location is small, making it difficult to produce accurate estimates of the underlying disease risk without linking the characteristics of another case from a neighbouring location. Therefore, the trends and spatial patterns of TB cases could be evaluated to identify high-risk areas.

When a TB case is reported, the local health authority investigates the area where the case occurs within seven days and conducts contact tracing to control the spread of the disease. This is accomplished by locating people who had close contact with patients who became infected with TB so that the former can receive testing and diagnosis if necessary [13]. However, these investigations do not collect geographical information about the locations of reported TB cases. Space is important in describing the epidemic and the distribution of the disease. Recently, there has been greater interest in the field of spatial epidemiology [14-15], a continuation of ecological studies that involve investigating disease distribution in different areas to better comprehend disease ecology.

The integration of spatial and epidemiological approaches has received attention as researchers have used these methods to undertake more sophisticated statistical analysis instead of just visual representations of observations in the form of maps [16-18]. In contrast, geospatial TB clusters not only assist in identifying ongoing person-to-person transmission but could also monitor latent infection reactivation in groups of people who share risk factors. Spatial analysis and identification of high-TB-rate areas (clusters), as well as characterisations of the dynamics in these clusters, have been supported for targeted TB control and increased use of the existing TB control techniques [19].

Local health authorities can monitor locations near reported cases using the Geographic Information System (GIS) before the respective district health office is notified of new cases. Hence, with its ability to handle, manipulate, and analyse data, GIS serves as an enhanced warning system in identifying high-risk areas. The ability of GIS to integrate and manipulate complex data has become an important tool in epidemiological studies, allowing disease occurrence to be investigated from a geospatial perspective. GIS tools allow researchers to better understand the geographic variations of TB, identify hotspot areas, and assess disease risk factors, which then become the basis for predicting future occurrences of the disease in order to control them [20, 21]. This gap means the increasing rate of TB cases may never diminish. Therefore, there is a need for improved intervention by combining screening and contact tracing with a geospatial approach, which could be a better solution for TB control programmes in high-risk areas compared to depending on only traditional clinic-based screening.

METHODOLOGY

This study is a conventional review paper. Sources of information were identified from several databases such as Google Scholar, Scopus, ScienceDirect, Semantic Scholar, and ResearchGate. Primary keywords such as 'spatial' and 'spatiotemporal' were paired with the secondary keyword, 'tuberculosis', and the third keyword, 'risk factors', which were screened as the sources of content exploration. Each database search utilised this combination of keywords. The search was carried out in 2021. All the search results were filtered using the following criteria: (i) the information source must come from an original research article, (ii) the article must use spatial analysis techniques in the methodology, (iii) spatial analysis techniques must be employed in monitoring the spatial patterns and distributions of TB cases with its associated risk factors, (iv) the article must be written in English. The articles were then reviewed based on their titles and abstracts to exclude those that did not meet the scope of this review. Then, the complete texts of the remaining articles were scrutinised to determine whether they corresponded to the aims of this review. Finally, the findings from the documents were extracted and transferred into a Microsoft Excel spreadsheet.

The citation information, study objectives, findings from spatial analysis employed on TB cases such as interpolation techniques, spatial autocorrelation, kernel density estimation, hotspot analysis, and spatial regression analysis that associated TB with its risk factors are discussed. This paper is organised into three sections. The first describes the concept of spatial epidemiology, the history of disease mapping, the categories of spatial data, and the implementation of the GIS database. The second part explains the role of GIS in spatial epidemiology, referring to the types of GIS techniques used for public and environmental health issues. The third part highlights the various spatial analysis techniques used to determine the trends of and associations between risk factors and TB cases.

SPATIAL EPIDEMIOLOGY

Epidemiology is the study of the interrelationships between individuals, places, and times in the distribution of diseases in the population. Beginning in the early 1800s, the spatial epidemiology approach consists of three major components: (i) the visualisation of spatial data using disease mapping, (ii) the identification of spatial patterns and relationships between risk factors and disease, and (iii) the modelling of spatial data to test hypotheses [22,23]. Spatial epidemiology is the study of geographically indexed health data and the risk factors associated with it. This field involves the use of spatial statistics to analyse geo-referenced health data in the form of maps [24]. Therefore, spatial information is used to expand the analysis of epidemic disease. Spatial epidemiology studies the spatial distribution model that predicts disease spatial-temporal trends, as well as the relationship between a disease and its risk factors.

The Lancet published the first example of disease mapping in 1831, showing a cholera outbreak map that successfully predicted the disease's progression in England [25]. The most popular theory was that the cholera cases were spread by contaminated air. Meanwhile, a case map of the Baltic city of Gdansk revealed that the disease began in filthy alleys and spread irregularly. Then, in 1849, John Snow demonstrated for the first time that a cholera outbreak had a waterborne source [26]. His theory was based on the logic that diseases were most commonly caused by something swallowed rather than inhaled. However, his theory was not initially accepted, with critics pointing out that the disease was not universal among people who drank contaminated water and that people in some places became ill despite the assumption that the water was safe.

William Farr published an alternative theory about the spread of cholera in 1852 [(Britain, G., & Farr [27]), cited by Smith [28]. He observed seasonality in disease transmission, which led him to believe that cholera was generated in warm airs that evaporated from polluted water during the summer. In London in 1854, John Snow would eventually have more credible evidence of his theory of waterborne transmission. Using a map, Snow illustrated the relationship between cholera cases and a contaminated water supply. He discovered that the majority of those who died had lived near a specific water pump on Broad Street, with only a few cases living closer to another pump (Snow [29], cited by Smith [28]). This led him to believe that the Broad Street pump was the source of the outbreak and inspired him to create his diagram by mapping the areas where cholera cases had occurred (Figure 1).
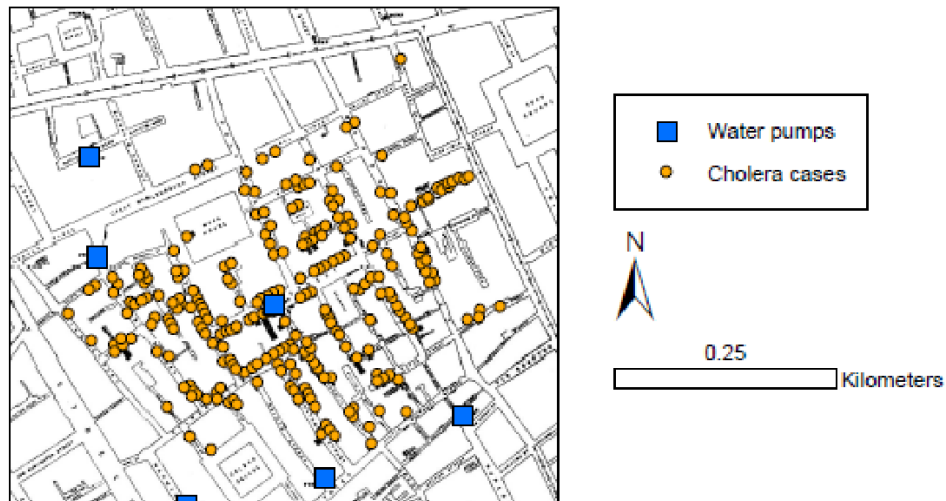
Figure 1 John Snow's map of cholera cases [29]

Mapping is a visual representation of geographical distribution, whereas disease mapping is an illustration of outbreak locations and a summary of the measurements or statistics applied to a specific group of infected individuals and their geographical linkage. The advantage of disease mapping is that it can visualise a spatial pattern that cannot be analysed in any other data presentation such as a table, graph, or chart [30]. It is also useful for making an initial evaluation of the relationship between exposure and disease, particularly the acute health effects. Interpretations of disease mapping should take into account a number of factors, including vector representation and spatial resolution [31].

Disease mapping includes the creation of maps, their transformation, the symbolic representation of results, data processing, and map interpretations that can be used in disease investigations. Information about disease boundaries - such as those of the county, district, municipal, or province - and scale should be considered when creating a disease map. The geographic boundaries of a disease are derived on an administrative level, and the spatial analysis is limited to a polygonal map.

The symbolic representation of spatial data is designated as point data (exact location) or count data (aggregated level), in which the data is mutually dependent and has the same characteristics in the specified geographic proximity [32]. Because of the geographical coverage of individual cases of the disease itself, point data is more precise. Although the use of count data is simpler than that of point data, it is susceptible to misclassification and bias [31]. As a result, it is appropriate to justify the use of TB data (Figure 2).
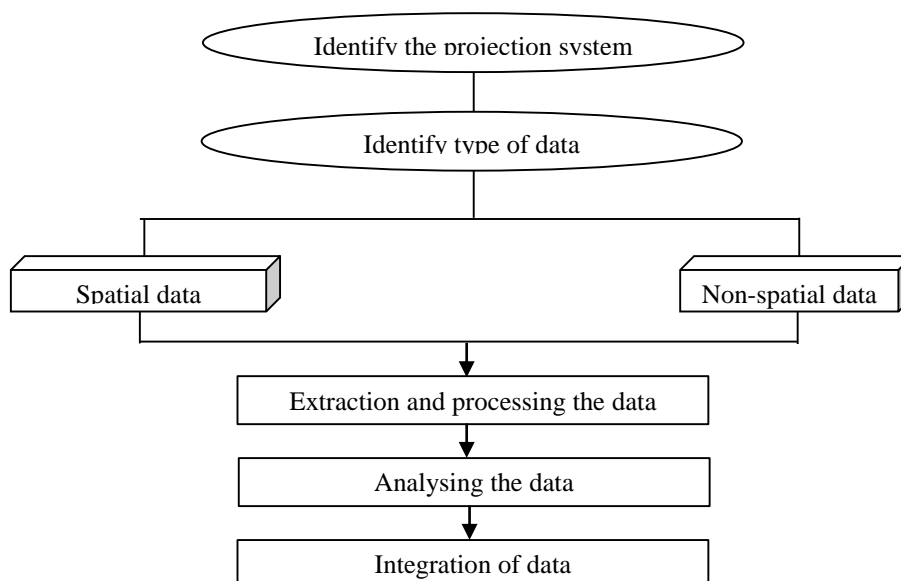
Figure 2 GIS database implementation

THE ROLE OF THE GEOGRAPHIC INFORMATION SYSTEM IN SPATIAL EPIDEMIOLOGY

Kulldroff [33] and Elliot et al. [34] reported in the late 1800s that the use of GIS in spatial epidemiological studies might include communicable and non-communicable diseases. According to Gatrell et al. [35], three types of GIS techniques can be used for public and environmental health issues: visualisation, exploratory, and modelling. Data visualisation entails displaying disease locations and influential variables on maps, as well as presenting disease variability in space and time; meanwhile, exploratory data analysis enables the discovery and investigation of disease spatial trends and the generation of hypotheses. The development of statistical models includes regression models of disease relationships that vary spatially. Collaborative efforts are needed among statisticians, epidemiologists, and geographers to ensure the connectivity of multilevel or hierarchical initiatives between the health and spatial elements of the GIS platform [36].

The application of spatial statistics is likely to increase each year, particularly since researchers have begun using GIS for purposes other than map-based visualisation. GIS software is continually updated to produce new versions from year to year in response to the growing interest in the relationship between space and health [37]. The software includes an embedded relational database component, which organises data management for disease surveillance across time and space. The advantage of using GIS in public and environmental health studies is its capacity to predict the location of an outbreak and estimate risk levels, making it a useful decision-making tool for investigation, monitoring, simulation, and resource allocation.

SPATIAL ANALYSIS IN TUBERCULOSIS STUDIES

Kriging

Kriging has been used in a large and growing body of literature to analyse the spatial association between environmental factors and TB cases [38,39]. Bonell et al. [40] conducted a time series analysis of the spatial and temporal clustering of TB in Vietnam, reporting that relative humidity, temperature, and hours of sunlight were associated with the incidence of TB, while kriging was employed to interpolate climate data. Tadesse et al. [41] discovered that kriging was the best model for estimating the risk of TB distribution in the Gurage zone of southern Ethiopia using the same technique. Their study examined how true anisotropy, altitude, and latitude covariates reduced total error and improved model predictive accuracy. This finding is supported by the work of Shojaei [42], who discovered a higher incidence rate of TB in the east of Iran.

In a Malaysian context, Hassan et al. [43] discovered that Petaling and Gombak had the highest risk of dengue infection in a study that interpolated rainfall data and generated a risk map of dengue infection using kriging, in which rainfall and population density were significant factors. Interpolation is another technique for gathering data that uses previously known information. In the context of mapping, this technique involves estimating values in unstructured or measured regions to create build maps or distribute values across regions. In the interpolation process, error and bias are unavoidable. Errors generated prior to interpolation could be the result of errors in the data sampling method, measurement error, or field analysis error [44]. Several interpolation methods are commonly used in data processing, each with its own set of characteristics.

Kriging is a powerful geostatistical process that estimates a surface from a scattering of z-valued points. Furthermore, it is a generalised least-squares regression method of spatial prediction that takes into account the spatial dependence between observations, as indicated by the variogram [45]. The kriging method differs from other interpolation methods in that it computes a variance value for each estimated point or area [46]. This demonstrates how reliable the predicted value is. As a result, in kriging, the weight is determined by a fitted model to the measured points; the distance to the prediction location; and spatial correlations between the measured values in the surrounding area of the prediction position [47].

The Inverse Distance Weighted tool uses an interpolation method to estimate cell values by taking the average of sample data points in the area near each processing cell. The higher a point's influence, or weight, in the averaging process, the closer it is to the centre of the cell being estimated. Natural Neighbour interpolation reveals the closest group of input samples to a query point and interpolates a value using proportionate area weights [48]. The algorithm takes a weighted average of adjacent observations, with weights proportional to the "borrowed area". The Natural Neighbour technique does not extrapolate

contours beyond the convex hull of the data locations (i.e., the outline of the Thiessen polygons). The Spline tool estimates values using a mathematical function that minimises overall surface curvature, resulting in a smooth surface that moves exactly through the input points. Every interpolation method - such as kriging, Inverse Distance Weighted, Natural Neighbour, and spline - has advantages and disadvantages (Table 1).

Table 1 Comparison between interpolated methods

Methods	Strength	Weakness
Kriging	• The ability to assess the quality of modelling executed. • Process parameter control. • Determined by not only the distance between the measured points and the predicted location but also the overall spatial arrangement of the measured points. • Provides error estimation via kriging variance as well as estimation of the own variable. • Assigns weights based on a data-driven strategy rather than an arbitrary function, giving isolated points more weight than points within a cluster.	• Requires significantly more computing time and user input. • Error assessment is influenced by the variogram, the distribution of data points, and the size of interpolated blocks. • Caution is needed when modelling spatial correlation structures.
Inverse Distance Weighted	• Intuitive and efficient, with interpolation functioning best with evenly distributed points.	• Sensitive to outliers with no error indication. • The selected weighting parameter and size of the search window influence the results. When used with high data densities, including digitised data, a poor choice of window can result in artefacts.
Natural Neighbour	• Fast operation speed. • Calculation process accessibility.	• The distance covered by data dependence is limited. • No assessment of error. • Each polygon can only have one data point. • Tessellation pattern is determined by data distribution.
Spline	• Can generate relatively precise and visually appealing surfaces from a small number of sample points.	• The resulting surface may vary from the input dataset in terms of minimum and maximum values. • Sensitive to outliers. • No indication of errors.

Spatial Autocorrelation

Much of the current literature on identifying the high-risk regions of outbreaks focuses on disease clustering [49-51]. Previous findings have demonstrated that Global Moran's I could be applied to assess clusters of TB cases in Ethiopia's East Gojjam Zone [52], whereas the same pattern of TB cases, PM2.5, and poverty level was found to be associated in Lima, Peru [53]. Similarly, Liao [54] discovered an increasing trend in

the spatial autocorrelation of sputum smear positive TB in China from 2005 to 2007, 2008 to 2009, and 2010 to 2015. These findings emphasise the importance of spatial autocorrelation and the non-random distribution of the disease. In Malaysia, Rajab et al. [55] proved that Bandar Tun Razak is the main clustered area of TB cases in Kuala Lumpur using the concept of spatial autocorrelation analysis. Not only did this area have the highest population compared to other cities, but it had also experienced rapid physical urban development. The authors discovered many crowded flats and terrace units, which could have led to the poor environmental conditions in the area, facilitating disease transmission.

Kernel Density Estimation

Recently, there has been a focus on identifying high-density areas of infectious diseases like TB [56-58]. Surveys, such as those conducted by Chun et al. in Hangzhou, China [59], have revealed that many high-density regions of TB cases were primarily located in the populated centres of the cities or counties of a state, with only a few high-density regions found in remote and rural mountainous areas far from TB-designated hospitals. Unlike Chun, Jiang et al. [60] tried to differentiate between districts in China with high population density and those with low population density. The authors discovered high-density TB cases in high-population-density areas in the Baoan and Luohu districts, as well as low-density areas in the Longgang district. Nevertheless, other researchers [61] investigating kernel density estimation in a different period, 2008 to 2015, discovered little variation in the high-density regions of TB deaths during this period.

In Malaysia, most research on demonstrating high-density regions has emphasised the use of kernel density estimation for many diseases over the last decade. Rahim et al. [62], for example, investigated high-density regions of cholera cases in the western areas of the districts of Kota Kinabalu, Putatan, Penampang, and Papar in Sabah, which they claimed could be due to the increased density of refuse dumps and a breakdown in sanitation, whereas typhoid and viral hepatitis A cases were evident in minimal high-density regions and seemed dispersed.

Hotspot Analysis

A number of authors have published analyses of TB hotspots using Getis-Ord Gi* statistics [63-65]. Recent evidence indicates that the hotspot areas identified in Harare's west-southwest zone result from an increased peri-urban population with insufficient social services, including health facilities [66]. Yu et al. [67] took a different approach, indicating that poor areas with characteristics such as low GDP, a shortage of health resources, a lack of funding, a majority of ethnic minorities, and higher non-permanent incomes among patients could be the primary factor of clustering hotspots of TB cases in the northeast and southeast regions of Chongqing, China, from 2011 to 2018. Furthermore, the authors conclusively demonstrated that hotspot locations in the northern zone gradually decreased over time, eventually disappearing in 2017. Several studies involving hotspot analysis of various diseases have been conducted. Seman et al. [68] observed that hotspot areas of hand, foot, and mouth disease were restricted to western areas of Kuching North City Council, primarily Rampangi Fasa II and Semariang Pinggir villages, and Kuching South City Council, primarily Kampung Stampin and Kampung Stutong Baru. An extensive literature search revealed a lack of previous studies on hotspot locations of TB cases, which could provide valuable insights into the spatial distribution of the disease.

Regression Analysis

There has been a recent increase in the amount of literature involving the use of OLS and GWR to evaluate the spatial association between risk factors and TB cases [69-71]. In the Dale district and Yirga Alem town, Ethiopia, the OLS model was found to indicate the relationship between TB cases and physical accessibility, i.e., population density, distance to TB control facilities, and altitude [72]. Using the GWR model, the same study discovered disease spatial heterogeneity in the central and northwestern borders of the regions, with population density and physical as risk factors. The GWR also addressed the fact that the model was installed between 2003 and 2009, rather than between 2010 and 2012, owing to improved access to TB control facilities.

Numerous studies have concluded that the GWR model performs better than the OLS model. After screening the significant explanatory variables with the OLS model, Wang et al. [73] used the GWR model to demonstrate the spatial heterogeneity of sociodemographic factors, including the proportion of illiterate people aged 15 and above, per capita disposable income, the number of health technicians per 1,000 population, and urban population density with TB cases. Unlike Wang, Zhang et al. [74] discovered only a connection between rainfall and tuberculosis cases, with no link to other meteorological factors like relative humidity, temperature, wind speed, atmospheric pressure, or sunshine duration.

Arroyo et al. [75] used the GWR model to evaluate the spatial dependence between cases of TB co-infected with mental disorders in municipalities in So Paulo State, using the same explanatory variables that remained in the OLS model. The most significant influences on the spatial variation of TB cases and mental disorders were mental illness, illicit drug use, alcoholism, smoking, social inequality, low income, a lack of primary care coverage, and the presence of prisons in municipalities.

CONCLUSION

A variety of spatial analyses have been used in various contexts, with most studies demonstrating significant heterogeneity in the spatial distribution of TB regardless of geographic resolution. Based on the review of the spatial analyses used in previous studies, kriging indicated by the variogram was able to display the trend related to the spatial and temporal distribution of TB by interpolating the environmental data, whereas spatial autocorrelation using Global Moran's I was able to indicate the trend involving the spatial cluster of the disease. Kernel density was able to identify the high-density regions of cases in different periods, while hotspot analysis using Getis-Ord Gi* could determine the spatial cluster for hotspot areas of the disease. In the spatial epidemiology field, spatial regression analysis - such as geographically weighted regression and ordinary least-squares - are the best methods of monitoring the spatial association between sociodemographic and environmental factors with TB cases. Various spatial cluster techniques have been used, but there is no firm agreement on how to interpret the inconsistencies in the findings obtained from the methods when they are applied to the same dataset. A combination of genotypic, genomic, and molecular analysis, along with spatial epidemiological linkage of cases, could improve the understanding of TB transmission.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Alsayed, S. S., & Gunosewoyo, H. (2023). Tuberculosis: pathogenesis, current treatment regimens and new drug targets. International journal of molecular sciences, 24(6), 5202.
[2]   Cioboata, R., Biciusca, V., Olteanu, M., & Vasile, C. M. (2023). COVID-19 and Tuberculosis: Unveiling the Dual Threat and Shared Solutions Perspective. Journal of Clinical Medicine, 12(14), 4784.
[3]   Sadowski, C., Belknap, R., Holland, D. P., Moro, R. N., Chen, M. P., Wright, A., ... & Gandhi, N. R. (2023). Symptoms and systemic drug reactions in persons receiving weekly Rifapentine plus isoniazid (3HP) treatment for latent Tuberculosis infection. Clinical Infectious Diseases, 76(12), 2090-2097.
[4]   World Health Organization (2019). Global tuberculosis report 2019: executive summary, 1- 6.
[5]   Cusack, D. A. (2020). COVID-19 pandemic: Coroner's database of death inquiries with clinical epidemiology and total and excess mortality analyses in the District of Kildare March to June 2020. Journal of Forensic and Legal Medicine, 76, 102072.
[6]   Hansen, M. A., Samannodi, M. S., Castelblanco, R. L., & Hasbun, R. (2020). Clinical epidemiology, risk factors, and outcomes of encephalitis in older adults. Clinical Infectious Diseases, 70(11), 2377-2385.

[7] de Jezus, S. V., do Prado, T. N., Arcêncio, R. A., Mascarello, K. C., Sales, C. M. M., Fauth, M. M., Terena, N. F. M., Amorim, R. F., Araujo, V. M. S., Aragón, M. A. L. & Maciel, E. L. N. (2021). Factors associated with latent tuberculosis among international migrants in Brazil: a cross-sectional study (2020). *BMC infectious diseases*, *21*(1), 1-9.

[8] Mirzazadeh, A., Kahn, J. G., Haddad, M. B., Hill, A. N., Marks, S. M., Readhead, A., Barry, P. M., Flood, J. Mermin, J. H. & Shete, P. B. (2021). State-level prevalence estimates of latent tuberculosis infection in the United States by medical risk factors, demographic characteristics and nativity. *PloS one*, *16*(4), e0249012.

[9] Taye, H., Alemu, K., Mihret, A., Wood, J. L., Shkedy, Z., Berg, S., & Aseffa, A. (2021). Factors associated with localization of tuberculosis disease among patients in a high burden country: A health facility-based comparative study in Ethiopia. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, *23*, 100231.

[10] Fakhruzzaman, M. N. N., Abidin, N. Z., Aziz, Z. A., Lim, W. F., Richard, J. J., Noorliza, M. N., Zaki, M. S., et al. (2019). Diversified lineages and drug-resistance profiles of clinical isolates of *Mycobacterium tuberculosis* complex in Malaysia. *International Journal of mycobacteriology*, 8(4), 320.

[11] Jani, J., Bakar, S. F. A., Mustapha, Z. A., Ling, C. K., Teo, R., & Ahmed, K. (2020). Identification and characterization of *Mycobacterium tuberculosis* Beijing genotype strain SBH163, Isolated in Sabah, Malaysia. *Microbiology Resource Announcements*, 9(2), e01322-19.

[12] Tan, J. L., Simbun, A., Chan, K. G., & Ngeow, Y. F. (2020). Genome sequence analysis of multidrug-resistant *Mycobacterium tuberculosis* from Malaysia. *Scientific Data*, 7(1), 1-4.

[13] Du, C. R., Wang, S. C., Yu, M. C., Chiu, T. F., Wang, J. Y., Chuang, P. C., ... & Fang, C. T. (2020). Effect of ventilation improvement during a tuberculosis outbreak in underventilated university buildings. *Indoor air*, *30*(3), 422-432.

[14] Atiq, M., Dosch, K., Miller, A., & Sudhagoni, R. G. (2020). Spatial Epidemiology of Pancreatic Cancer in South Dakota. *Journal of Gastrointestinal Cancer*, 51(1), 144-151.

[15] Corsi, D. J. (2020). Spatial Epidemiology of Diabetes and Tuberculosis in India. *JAMA Network Open*, 3(5), e203892-e203892.

[16] Qiao, M., & Huang, B. (2023). COVID-19 spread prediction using socio-demographic and mobility-related data. Cities, 138, 104360.

[17] Zhang, L., Wei, L., & Fang, Y. (2024). Spatial–temporal distribution patterns and influencing factors analysis of comorbidity prevalence of chronic diseases among middle-aged and elderly people in China: focusing on exposure to ambient fine particulate matter (PM2. 5). BMC Public Health, 24(1), 550.

[18] Zhou, C., Li, T., Du, J., Yin, D., Li, X., & Li, S. (2024). Toward tuberculosis elimination by understanding epidemiologic characteristics and risk factors in Hainan Province, China. Infectious Diseases of Poverty, 13(1), 20.

[19] Teibo, T. K. A., Andrade, R. L. D. P., Rosa, R. J., Tavares, R. B. V., Berra, T. Z., & Arcêncio, R. A. (2023). Geo-spatial high-risk clusters of Tuberculosis in the global general population: a systematic review. *BMC Public Health*, *23*(1), 1586.

[20] Lan, Y., & Delmelle, E. (2023). Space-time cluster detection techniques for infectious diseases: A systematic review. Spatial and Spatio-temporal Epidemiology, 44, 100563.

[21] Siddik, M. S. M., Ahmed, T. E., Awad Ahmed, F. R., Mokhtar, R. A., Ali, E. S., & Saeed, R. A. (2023). Development of Health Digital GIS Map for Tuberculosis Disease Distribution Analysis in Sudan. Journal of Healthcare Engineering, 2023.

[22] Walter, S. D. (2000). Disease mapping: A historical perspective. Spatial epidemiology: Methods and applications, 223-239.

[23] Roquette, R., Painho, M., & Nunes, B. (2017). Spatial epidemiology of cancer. *Geospatial health*, 12(1), 1-12.

[24] Meyer, S., Held, L., & Höhle, M. (2014). Spatio-temporal analysis of epidemic phenomena using the R package surveillance. *arXiv preprint arXiv:1411.0416*.

[25] Koch, T. (2004). The map as intent: Variations on the theme of John Snow. *Cartographica. The International Journal for Geographic Information and Geovisualization*, 39(4), 1-14.

[26] Bingham, S. A., Luben, R., Welch, A., Wareham, N., Khaw, K. T., & Day, N. (2003). Are imprecise methods obscuring a relation between fat and breast cancer?. *The Lancet*, 362(9379), 212-214.

[27] Britain, G., & Farr, W. (1852). *Report on the mortality of cholera in England, 1848-1849*. W. Clowes.

[28] Smith, A. M., Stull, J. W., Evason, M. D., Weese, J. S., Wittum, T. E., Szlosek, D., & Arruda, A. G. (2021). Investigation of spatio-temporal clusters of positive leptospirosis polymerase chain reaction test results in dogs in the United States, 2009 to 2016. *Journal of Veterinary Internal Medicine*, *35*(3), 1355-1360.

[29] Snow, J. (1855). *On the mode of communication of cholera*. John Churchill.

[30] Weng, D., Chen, R., Deng, Z., Wu, F., Chen, J., & Wu, Y. (2018). Srvis: Towards better spatial integration in ranking visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 459-469.

[31]   Elliott, P., & Wartenberg, D. (2004). Spatial Epidemiology : Current Approaches and Future Challenges. *Environmental Health Perspectives*, 112(9), 998–1006.

[32]   Diez-Roux, A. V. (2007). Neighborhoods and health: Where are we and where do we go from here?. *Revue d'Epidemiologie et de Sante Publique*, 55(1), 13-21.

[33]   Kulldorff, M. (1999). Geographic information systems (GIS) and community health: some statistical issues. *Journal of Public Health Management and Practice*, 5(2), 100-106.

[34]   Elliott, P., & Wartenberg, D. (2004). Spatial epidemiology: current approaches and future challenges. *Environmental health perspectives*, *112*(9), 998-1006.

[35]   Gatrell, A. C., & Löytönen, M. (1997). *GIS and health*. Taylor and Francis.

[36]   Langford, I. H., & Bentham, G. (1996). Regional variations in mortality rates in England and Wales: an analysis using multi-level modelling. *Social Science and Medicine*, 42(6), 897-908.

[37]   Manda, S., Haushona, N., & Bergquist, R. (2020). A scoping review of spatial analysis approaches using health survey data in Sub-Saharan Africa. *International Journal of Environmental Research and Public Health*, 17(9), 1-20.

[38]   Huang, K., Ding, K., Yang, X. J., Hu, C. Y., Jiang, W., Hua, X. G., Zhang, X. J., et al. (2020). Association between short-term exposure to ambient air pollutants and the risk of tuberculosis outpatient visits: A time-series study in Hefei, China. *Environmental Research*, 184, 109343.

[39]   Li, X. X., Ren, Z. P., Wang, L. X., Zhang, H., Jiang, S. W., Chen, J. X., ... & Zhou, X. N. (2016). Co-endemicity of pulmonary tuberculosis and intestinal helminth infection in the people's Republic of China. *PLoS neglected tropical diseases*, *10*(4), e0004580.

[40]   Bonell, A., Contamin, L., Thai, P. Q., Thuy, H. T. T., van Doorn, H. R., White, R., Choisy, M., et al. (2020). Does sunlight drive seasonality of TB in Vietnam? A retrospective environmental ecological study of tuberculosis seasonality in Vietnam from 2010 to 2015. *BMC infectious diseases*, 20(1), 1-11.

[41]   Tadesse, S., Enqueselassie, F., & Gebreyesus, S. H. (2018). Estimating the spatial risk of tuberculosis distribution in Gurage zone, southern Ethiopia: a geostatistical kriging approach. *BMC Public Health*, 18(1), 1-10.

[42]   Shojaei, S. R. H., Waghei, Y., & Mohammadzadeh, M. (2018). Geostatistical analysis of disease data: a case study of tuberculosis incidence in Iran. *Journal of Applied Statistics*, 45(8), 1476-1483.

[43]   Hassan, H., Shohaimi, S., & Hashim, N. R. (2012). Risk mapping of dengue in Selangor and Kuala Lumpur, Malaysia. *Geospatial Health*, 7(1), 21-25.

[44]   Lepot, M., Aubin, J. B., & Clemens, F. H. (2017). Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment. *Water*, *9*(10), 796.

[45]   Okunlola, O. A., Alobid, M., Olubusoye, O. E., Ayinde, K., Lukman, A. F., & Szűcs, I. (2021). Spatial regression and geostatistics discourse with empirical application to precipitation data in Nigeria. *Scientific Reports*, *11*(1), 16848.

[46]   Koutsos, T. M., Menexes, G. C., Eleftherohorinos, I. G., & Alexandridis, T. K. (2023). Using Block Kriging as a Spatial Smooth Interpolator to Address Missing Values and Reduce Variability in Maize Field Yield Data. *Agronomy*, *13*(7), 1685.

[47]   Hassim, M., Yuzir, A., Razali, M. N., Ros, F. C., Chow, M. F., & Othman, F. (2020). Comparison of Rainfall Interpolation Methods in Langat River Basin. *In IOP Conference Series: Earth and Environmental Science*, 479(1), 012018.

[48]   Etherington, T. R. (2020). Discrete natural neighbour interpolation with uncertainty using cross-validation error-distance fields. *PeerJ Computer Science*, *6*, e282.

[49]   AlQadi, H., Bani-Yaghoub, M., Balakumar, S., Wu, S., & Francisco, A. (2021). Assessment of retrospective COVID-19 spatial clusters with respect to demographic factors: Case Study of Kansas City, Missouri, United States. *International Journal of Environmental Research and Public Health*, 18(21), 11496.

[50]   Damasceno, D. M., da Paz, W. S., de Souza, C. D. F., Dos Santos, A. D., & Bezerra-Santos, M. (2021). High-risk transmission clusters of leprosy in an endemic area in the Northeastern Brazil: A retrospective spatiotemporal modelling (2001–2019). *Tropical Medicine & International Health*.

[51]   Smith, A. M., Stull, J. W., Evason, M. D., Weese, J. S., Wittum, T. E., Szlosek, D., & Arruda, A. G. (2021). Investigation of spatio-temporal clusters of positive leptospirosis polymerase chain reaction test results in dogs in the United States, 2009 to 2016. *Journal of Veterinary Internal Medicine*, *35*(3), 1355-1360.

[52]   Asemahagn, M. A., Alene, G. D., & Yimer, S. A. (2021). Spatial-temporal clustering of notified pulmonary tuberculosis and its predictors in East Gojjam Zone, Northwest Ethiopia. PLoS One, 16(1), e0245378.

[53]   Carrasco-Escobar, G., Schwalb, A., Tello-Lizarraga, K., Vega-Guerovich, P., & Ugarte-Gil, C. (2020). Spatio-temporal co-occurrence of hotspots of tuberculosis, poverty and air pollution in Lima, Peru. *Infectious Diseases of Poverty*, 9, 1-6.

[54]   Liao, W. B., Ju, K., Gao, Y. M., & Pan, J. (2020). The association between internal migration and pulmonary tuberculosis in China, 2005–2015: A spatial analysis. *Infectious Diseases of Poverty*, 9(1), 1-12.

[55]   Rajab, N. A., Hashim, N., & Rasam, A. R. A. (2020). Spatial Mapping and Analysis of Tuberculosis Cases in Kuala Lumpur, Malaysia. In 2020 IEEE 10th International Conference on System Engineering and Technology, 38-43.

[56]   Bui, D. P., Oren, E., Roe, D. J., Brown, H. E., Harris, R. B., Knight, G. M., ... & Grandjean, L. (2019). A case-control study to identify community venues associated with genetically-clustered, multidrug-resistant tuberculosis disease in Lima, Peru. *Clinical Infectious Diseases*, *68*(9), 1547-1555.

[57]   Ogbudebe, C., Jeong, D., Odume, B., Chukwuogo, O., Dim, C., Useni, S., ... & Gidado, M. (2023). Identifying hot spots of tuberculosis in Nigeria using an early warning outbreak recognition system: retrospective analysis of implications for active case finding interventions. *JMIR Public Health and Surveillance*, *9*(1), e40311.

[58]   Mohidem, N. A., Osman, M., Hashim, Z., Muharam, F. M., Mohd Elias, S., & Shaharudin, R. (2021). Association of sociodemographic and environmental factors with spatial distribution of tuberculosis cases in Gombak, Selangor, Malaysia. *PLoS One*, *16*(6), e0252146.

[59]   Chun, Z. M., Jun, J. Q., & Yan, H. Y. (2020). The Spatiotemporal dynamic distributions of new tuberculosis in Hangzhou, China. *Biomedical and Environmental Sciences*, 33(4), 277-281.

[60]   Jiang, Q., Liu, Q., Ji, L., Li, J., Zeng, Y., Meng, L., Gao, Q., et al. (2019). Citywide transmission of MDR-TB under China's rapid urbanization: a retrospective population-based genomic spatial epidemiological study. *Clinical Infectious Diseases: An official publication of the Infectious Diseases Society of America*.

[61]   Alves, L. S., Dos Santos, D. T., Arcoverde, M. A. M., Berra, T. Z., Arroyo, L. H., Ramos, A. C. V., Arcêncio, R. A., et al. (2019). Detection of risk clusters for deaths due to tuberculosis specifically in areas of southern Brazil where the disease was supposedly a non-problem. *BMC Infectious Diseases*, 19(1), 1-13.

[62]   Rahim, S. S. S. A., Shah, S. A., Idrus, S., & Azhar, Z. I. (2020). Spatial Analysis of Food and Waterborne Diseases in Sabah, Malaysia. *Sains Malaysiana*, 49(7), 1627-1638.

[63]   Alves, L. S., Dos Santos, D. T., Arcoverde, M. A. M., Berra, T. Z., Arroyo, L. H., Ramos, A. C. V., Arcêncio, R. A., et al. (2019). Detection of risk clusters for deaths due to tuberculosis specifically in areas of southern Brazil where the disease was supposedly a non-problem. *BMC Infectious Diseases*, 19(1), 1-13.

[64]   Chirenda, J., Gwitira, I., Warren, R. M., Sampson, S. L., Murwira, A., Masimirembwa, C., Streicher, E. M., et al. (2020). Spatial distribution of *Mycobacterium tuberculosis* in metropolitan Harare, Zimbabwe. *PLoS One*, 15(4), e0231637.

[65]   Gwitira, I., Karumazondo, N., Shekede, M. D., Sandy, C., Siziba, N., & Chirenda, J. (2021). Spatial patterns of pulmonary tuberculosis (TB) cases in Zimbabwe from 2015 to 2018. *Plos one*, *16*(4), e0249523.

[66]   Chirenda, J., Gwitira, I., Warren, R. M., Sampson, S. L., Murwira, A., Masimirembwa, C., Streicher, E. M., et al. (2020). Spatial distribution of *Mycobacterium tuberculosis* in metropolitan Harare, Zimbabwe. *PLoS One*, 15(4), e0231637.

[67]   Yu, Y., Wu, B., Wu, C., Wang, Q., Hu, D., & Chen, W. (2020). Spatial-temporal analysis of tuberculosis in Chongqing, China 2011-2018. *BMC Infectious Diseases*, 20(1), 1-12.

[68]   Seman, B. B., & Masron, T. (2019). Hotspot Analysis of Hand Foot and Mouth Disease (HFMD) Using GIS in Kuching, Sarawak, Malaysia. *Humanities & Social Sciences Reviews*, *7*(2), 36-44.

[69]   Bui, L. V., Mor, Z., Chemtob, D., Ha, S. T., & Levine, H. (2018). Use of geographically weighted poisson regression to examine the effect of distance on tuberculosis incidence: A case study in Nam Dinh, Vietnam. *PloS one*, *13*(11), e0207068.

[70]   Zhang, X., Ma, S. L., & He, J. (2019). Correlation analysis of rubella incidence and meteorological variables based on Chinese medicine theory of Yunqi. *Chinese Journal of integrative Medicine*, *25*(12), 911-916.

[70]   Makalew, L. A., Otok, B. W., & Barung, E. N. (2019). Spatio of lungs tuberculosis (TB Lungs) in East Java using geographically weighted poisson regression (GWPR). *Indian Journal of Public Health Research & Development*, *10*(8), 1757-1760

[72]   Dangisso, M. H., Datiko, D. G., & Lindtjørn, B. (2020). Identifying geographical heterogeneity of pulmonary tuberculosis in southern Ethiopia: A method to identify clustering for targeted interventions. *Global Health Action*, 13(1), 1785737.

[73]   Wang, Q., Guo, L., Wang, J., Zhang, L., Zhu, W., Yuan, Y., & Li, J. (2019). Spatial distribution of tuberculosis and its socioeconomic influencing factors in mainland China 2013–2016. *Tropical Medicine and International Health*, 24(9), 1104-1113.

[74]   Zhang, Y., Liu, M., Wu, S. S., Jiang, H., Zhang, J., Wang, S., Ma, W., Li, Q, Ma, Y., Liu, Y., Feng, W., Amsalu, E., Li, X., Wang, W., Li, W. & Guo, X. (2019). Spatial distribution of tuberculosis and its association with meteorological factors in mainland China. *BMC Infectious Diseases*, 19(1), 1-7.

[75]   Arroyo, L. A. H., Arcoverde, M. A. M., Alves, J. D., Fuentealba-Torres, M., Cartagena-Ramos, D., Scholze, A. R., Arcêncio, R. A. , et al. (2019). Spatial analysis of cases of Tuberculosis with Mental Disorders in São Paulo. *Revista brasileira de enfermagem*, 72(3), 654-662.