RESEARCH ARTICLE

Spatial patterns of pulmonary tuberculosis (TB) cases in Zimbabwe from 2015 to 2018

Isaiah Gwitira[ID][1]*, Norbert Karumazondo[1], Munyaradzi Davis Shekede[1], Charles Sandy[2], Nicolas Siziba[2], Joconiah Chirenda[3]

1 Department of Geography Geospatial Sciences and Earth Observation, Faculty of Science, University of Zimbabwe, Harare, Zimbabwe, 2 National TB Control Program, Ministry of Health and Child Care, Harare, Zimbabwe, 3 Department of Community Medicine, Faculty of Medicine and Health Sciences, Parirenyatwa Hospital, University of Zimbabwe, Harare, Zimbabwe

* gwitsakuely@gmail.com

Abstract

Introduction

Accurate mapping of spatial heterogeneity in tuberculosis (TB) cases is critical for achieving high impact control as well as guide resource allocation in most developing countries. The main aim of this study was to explore the spatial patterns of TB occurrence at district level in Zimbabwe from 2015 to 2018 using GIS and spatial statistics as a preamble to identifying areas with elevated risk for prioritisation of control and intervention measures.

Methods

In this study Getis-Ord G_i* statistics together with SaTscan were used to characterise TB hotspots and clusters in Zimbabwe at district level from 2015 to 2018. GIS software was used to map and visualise the results of cluster analysis.

Results

Results show that TB occurrence exhibits spatial heterogeneity across the country. The TB hotspots were detected in the central, western and southern part of the country. These areas are characterised by artisanal mining activities as well as high poverty levels.

Conclusions and recommendations

Results of this study are useful to guide TB control programs and design effective strategies which are important in achieving the United Nations Sustainable Development goals (UNSDGs).

Background

Most disease-related deaths recorded in developing countries are associated with commonly occurring communicable diseases such as TB and Human Immunodeficiency Virus (HIV) [1,

2]. The WHO African region contributed 24% of the estimated 10 million *Mycobacterium tuberculosis* (*M.tuberculosis*) infections recorded globally in 2018 [3]. TB is an infectious bacterial airborne disease caused by *M. tuberculosis* complex and ranks among the top 10 causes of deaths in the world [3–5]. The disease is transmitted from person to person through breathing air contaminated by the bacteria [1]. TB is ranked above HIV/AIDs as the leading cause of death from a single infectious agent [6]. Most countries in sub-Saharan Africa (SSA) are characterised by high TB incidence rates of up to 275 cases per 100,000 of the population [7]. In an effort to reduce TB incidence, Zimbabwe adopted the global World Health Organization (WHO) '*End TB Strategy*' which aims to reduce annual TB-related deaths by 95% and TB incidence by 90% at the end of 2035 compared to 2015 [3, 8]. This strategy is in line with United Nations Sustainable Development Goals (UNSDGs) particularly SDG 3 which targets to end the global TB epidemic by 2030 [9].

The success of national TB control programmes (NTP) in most parts of the developing world particularly in Africa is hindered by failure to consider spatial heterogeneity in the distribution of the disease [10]. Without evidence of fine scale spatial heterogeneity in the occurrence of the disease [11], uniform interventions are implemented across different settings by NTP. Thus, a better understanding of the spatial epidemiology of TB may guide policy makers in formulating effective prevention and control strategies [12–14]. For instance, the identification of TB clusters and hotspots provides the basis for targeted control of the disease as well optimizing resource allocation [10]. This is particularly important in resource limited countries such as Zimbabwe where there is need to prioritise allocation of resources by focusing on areas with the highest disease burden as they have the greatest need. Therefore, this study hypothesises that effective management of communicable diseases such as TB primarily depends on accurate detection and mapping of spatial heterogeneity in the occurrence of the disease. Achieving TB targets and goals at national or global scales require an understanding of the spatial pattern of the disease which acts as a preamble to developing effective strategies aimed at reducing new infections.

In the context of Zimbabwe, most research on TB has been largely limited to trend analysis of prevalence of the disease. For example, studies have focused on the prevalence of TB in Zimbabwe resulting from the economic crisis of 2008 [15] including the prevalence of drug resistant TB using non-spatial statistical analysis [16]. Other studies have focused on developing strategies to control TB such as the use of directly observed treatment (DOT) [17–19] while others assessed factors influencing adherence to TB treatment [20]. Although these studies have provided valuable insights in understanding risk factors of TB occurrence, there is limited application of spatial analysis particularly in detecting hotspots and cold spots of TB especially in Zimbabwe. This is despite the fact that GIS-based spatial analysis combined with spatial statistics are indispensable tools for supporting surveillance and control of most diseases in the world [21–23].

To date, GIS and spatial statistics have been applied to understand patterns of TB occurrence in several countries with different environmental settings [6, 7, 13, 22, 24, 25]. These studies have generated important information about the distribution of the disease and its transmission patterns. However, the major limitation of these studies is that they are biased towards smaller spatial scales such as a single urban area over short temporal durations [14, 26]. By focussing on smaller spatial scales, these studies fail to capture the spatial pattern of TB at a spatial scale that is relevant for TB programming, for example district scale. The district is the spatial epidemiological administrative unit at which TB interventions and control are planned in Zimbabwe [27]. In this regard, spatial analysis becomes relevant when the spatial unit at which TB data is analysed represents the expected epidemiological dynamics in the country. Thus, it is important to fully understand the spatial heterogeneity in the occurrence

of TB at large spatial scales such as the district level to inform targeted public health response, which is regarded as one of the most effective approaches in disease control [28].

Although GIS and spatial statistics have been successfully applied to identify TB hotspots and clusters, most of these studies have applied either of these techniques in isolation despite their complementarity as well as their potential to reduce uncertainties associated with adoption of a single method to guide control interventions [29]. This limitation calls for the adoption of multiple cluster detection methods as this is critical in the detection of truly representative high-risk areas. The application of multiple techniques is increasingly recommended in spatial epidemiological studies to eliminate the limitations of a single method in disease hotspot detection [30–32]. This study therefore aimed at characterising the spatial pattern of TB hotspots and clusters using geospatial techniques and spatial statistical tools based on TB notification data recorded at district level from 2015 to 2018 in Zimbabwe.

Methods

Study area

The study was conducted in Zimbabwe located in southern Africa between latitudes 15.5° and 22.5° S and longitudes 25° and 33°E [33]. Zimbabwe is characterized by a subtropical climate with distinct seasons [34]. The country is divided into ten administrative provinces and sixty-six (66) districts (Fig 1A). The total human population was estimated at 14.64 million in 2019 *https://data.worldbank.org* [35]. Of the 14.64 million, 67% of the population resides in rural areas. The greater proportion of the population is at risk from the three top killer diseases i.e., HIV, TB and malaria. A network of health infrastructure has been constructed to provide health services across the country (Fig 1B)

Data sources

TB notification data. Aggregated data on TB notification from 2015 to 2018 were obtained from the Ministry of Health and Child Care (MOHCC) head office in Zimbabwe. The country has a well-established WHO-recommended directly observed treatment, short-course (DOTS)-based national TB programme (NTP) in which TB treatment services are integrated within the general primary health care delivery system [36]. The health care system includes rural hospitals, district, provincial and mission hospitals (Fig 1B). TB is a notifiable disease in Zimbabwe hence a person with TB is required by law to report to the nearest health centre for treatment due to the infectiousness of the disease. The assumption in this study was that every health facility is manned by well trained health personnel who can accurately test for TB. The data used in this study are aggregated district totals from the rural health centres (RHC/clinics) and thus did not include the physical address of the patients [37]. The diagnosis of TB in Zimbabwe follows national guidelines and is mostly based on chronic cough, or suggestive clinical symptoms (fever, night sweats, weight loss), and positive sputum smears and/
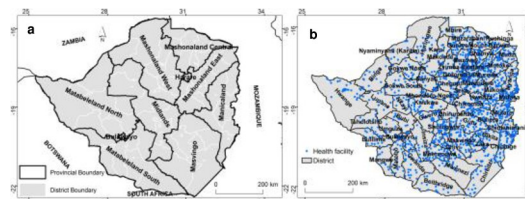
Fig 1. a) Location of the study area including the administrative Provinces and districts and b) Spatial distribution of health facilities where TB cases were recorded.

https://doi.org/10.1371/journal.pone.0249523.g001

or chest X-Ray [15] including bacteriological confirmation using Gene Xpert [38]. After TB diagnosis is completed at each health facility, heath workers complete notification forms and patient treatment cards for all recurrent patients. The notification information includes diagnosis, patient category, treatment regimen and monitoring indices [39]. This information is then entered into the DOTS register at each facility and later uploaded into the District Health Information Management System (DHIS 2) [38]. In this study, a patient with TB relapse was considered to belong to any of the following categories: treatment after failure, treatment after default or retreatment based on the WHO TB case definition [40]. A relapse may also be a patient who completed treatment but shows smear-positive symptoms while patients treated after failure represent those who failed first line treatment. In addition, patients who default from treatment and present again with smear-positive pulmonary TB (PTB) are classified as treatment after default while retreatment other refers to all other recurrent TB cases [40]. Since the study used anonymized data abstracted from district registers there was no need for ethics approval or informed patient consent. However, permission to use the routinely collected data was obtained from the NTP under the Ministry of Health and Child Care.

Population data

Population data used in this study were obtained from the Zimbabwe Statistics Agency (ZIM-STAT) based on the 2012 population census as well as from projections for the years 2015 and 2018. The 2015–2018 population figures were obtained using a medium scenario based on the projected annual growth rate of 1.1% [41]. Population projection data were used for the years 2015–2018 because in Zimbabwe, a national population census is conducted after every 10 years hence the population of intercensal years is estimated based on the projected growth rate. The population projections were carried out assuming geometric growth in the population [41].

Data analysis

Global autocorrelation of TB cases. The aggregated TB notification data at district level were used to calculate TB incidence by dividing the total TB notification for each district for a particular year by the projected population of that given year before multiplying by 100, 000. This gives a TB incidence per 100,000 of the population. The level of spatial clustering of TB cases at district level was examined by applying spatial autocorrelation using global Moran's I [42] statistic with row standardized inverse distance weight matrices. The technique tests whether there is any systematic pattern in the distribution of TB notifications among districts of Zimbabwe as opposed to being randomly distributed. The index ranges from -1 to +1 with a score of zero indicating no clustering while a positive value suggests that the distribution of TB cases in neighbouring districts is more spatially aggregated than a random pattern [43, 44]. In contrast, a negative value suggests dispersion of TB cases at the district level. Dispersion of TB cases means that districts with similar TB cases are actually less likely to be located near each other [45]. The TB cases and the projected population were spatially joined to the geometry of districts in a GIS for mapping using the spatial join function in Arcmap 10.3 [46].

Local spatial clusters of TB cases. Global techniques of spatial autocorrelation such as the Moran's I can successfully detect spatial clustering, but they fail to specify the location of the clusters [47]. Therefore Getis-Ord G_i^* was used to examine where TB cases formed statistically significant local aggregation in geographic space [48, 49]. The scan statistic method by Kulldorff (SaTScan 9.6) [50] was used to further detect the location as well as extent of TB clustering across districts of Zimbabwe. The following sections outline how each of these methods were implemented to detect clustering of TB cases at district level.

Detecting local spatial clusters using Getis-Ord G_i^*. The Getis-Ord G_i^* was used to obtain additional information on the spatial pattern of TB such as the intensity and stability of core hotspot or coldspot areas [48]. The Getis-Ord Gi^* method considers each district in the context of neighbouring districts which suggests likelihood of spatial autocorrelation among neighbouring districts [48, 49]. The conceptualization of spatial relationships was based on contiguity edges corners (*also referred to as the Queen's case*) available in ArcGIS 10.1.3 software where districts that share an edge or corner were considered as neighbours. The *Queen's* case was adopted to analyse spatial adjacency relationships since it compensates for the irregular size and shape of the districts [51] and is most ideal when modelling contagious diseases such as TB [46]. Based on the neighbours of a district, the ratio of the local sum of TB cases within neigbouring districts was compared to the sum of TB cases in the whole study area. Where the local sum of TB cases was significantly different from the expected local sum, and that difference was too large to be the result of random chance, the result was regarded as statistically significant based on the Z-score [52].

Since the Getis-Ord G_i^* considers neighbouring districts, the results are likely to be correlated hence the need to account for multiple testing and spatial dependency to accurately detect the pattern of TB [53]. Spatial dependence arises from the geometric relationship among districts that share a common boundary in addition to the TB cases recorded in neigbouring districts. Although the Bonferron method is most widely used to account for multiple testing and spatial dependency, it has a limitation of being conservative [53]. To overcome this limitation, the false discovery rate (FDR) method available in ArcGIS 10.3.1 was applied in this study. The FDR correction accounts for multiple testing and spatial dependency by controlling for the average rate that identified hotspots are truly significant [54]. This method has been shown to accurately evaluate occurrence of hotspots [55].

The detected significant hotspots/coldspots imply high/low TB cases in a particular district are surrounded by districts with high/low TB cases, respectively [48, 56]. To classify a district as a hotspot, a threshold Z-score of > 1.96 and a P-value of < 0.05 were considered [12]. Districts with Z-scores between -1.96 and +1.96 were considered as having non-significant clusters while those with Z-scores $<$-196 were classified as coldspots.

The Getis-Ord G_i^* takes the form;

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)X_j - w_i^* \overline{X}^*}{S^*\left\{[(nS_{1i}^*) - w_i^{*2}]/(n-1)\right\}^{1/2}} \tag{1}$$

where:

X_j = TB Case for district j

W_{ij} = spatial weights between district i and j

n = total number of districts

$W_{ij}(d)$ = spatial weights vector with values for all districts j within distance d of district i

W_i^* = sum of weights

$S1_i^*$ = sum of squared weights

s^* = standard deviation of data cells

Detecting local spatial clusters using SaTscan. The scan statistic method developed by Kulldorff (SaTScan 9.6) [50] was applied to detect the spatial pattern of TB at the district level. The method has been widely used for detecting spatial clusters of diseases in different environmental settings [57–59]. Purely spatial analysis based on the discrete poisson method was applied to detect clusters of varying sizes at different locations including their relative risk (RR) and a significance value generated using Monte Carlo Simulations [60]. The relative risk for each cluster was calculated by comparing the observed and expected TB cases in each

circular window resulting in a log likelihood ratio [61]. In this study, the number of TB cases aggregated at district level, the projected population and the centroid coordinates of each district were used as input files. The discrete Poisson model which assumes that the number of TB cases in each district followed a Poisson distribution with a known population at risk was applied to detect significant TB clusters. The population at risk for each of the four years was derived from population projections based on the 2012 census. The Gini coefficient [62] was used to determine the maximum reported cluster size as a percentage of the population for the respective years (Table 1). For the four years, the maximum reported cluster size ranged from 6–20% of the population at risk.

The Gini coefficient is an intuitive way to evaluate the degree of heterogeneity among a collection of clusters [62].

Results

The spatial distribution of TB notification rates in Zimbabwe at district level from 2015 to 2018 is illustrated in Fig 2. The TB notification is relatively higher in the central and southern part of the country than in the east and north western regions. In fact, across all the years considered in this study, TB notification was consistently high in the central and southern regions. The TB notification was high in Beitbridge, Buhera, Chirumhanzu, Gwanda, Hwange, Mwenezi and Sanyati districts (Fig 2).

The results of Moran's I analysis showed a strong positive spatial autocorrelation (P< 0.05) across all the four years from 2015 to 2018. Specifically, results show that the Moran's index was 0.173, 0.140, 0.243 and 0.032 for 2015, 2016, 2017 and 2018 respectively. The significant positive spatial autocorrelation indicates that the distribution of TB cases was more spatially clustered than would be expected in a random process for each year across the districts of Zimbabwe.

### Results of hotspot analysis

Results of hotspots analyses using local Getis Ord G_i^* after applying the false discovery rate are illustrated in Fig 3. The results show that significant TB hotspots (Z-score>1.96) are characteristic of the central, southern and western part of the country. These hotspots covered most urban areas and districts characterised by both formal and informal mining activities such as Kwekwe, Mhondoro-Ngezi and Chegutu (Fig 3). Over the study period, Chegutu, Kwekwe and Mhondoro-Ngezi and were persistently characterised by significant hotspots (Fig 3). Most of these districts in the centre of the country lie along the Great Dyke of Zimbabwe where small scale artisanal miners predominate. While hotspots of TB were observed in the central and southern part of the country the majority of the county was characterised by non-significant TB hotspots (Z-score<1.96) across the four years (Fig 3).

Table 1. Maximum reported cluster size (MRCS) from 2015–2018 in Zimbabwe determined using the Gini coefficient.

Year	% of the Population	Optimal Gini Coefficient
2015	6	0.2784
2016	10	0.1956
2017	10	0.1949
2018	20	0.2179

https://doi.org/10.1371/journal.pone.0249523.t001
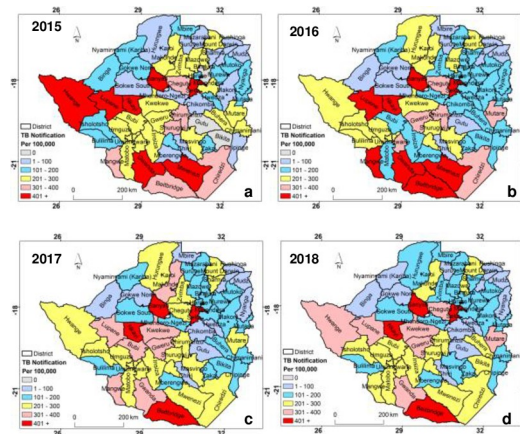
Fig 2. Spatial distribution of TB notification per 100,000 of the population at district level in Zimbabwe for (a) 2015, (b) 2016, (c) 2017 and (d) 2018.

Results of cluster detection using SaTscan

The TB epidemic in Zimbabwe is characterised by geographically distinct clusters across the study area (Fig 4). Using spatial analysis, different clusters were identified for each year with the year 2015 having the highest number of clusters centred on specific districts (Fig 4). In general, the pattern of TB occurrence exhibits stability or persistence for some regions e.g., central and southern regions which are characterised by significant clusters in each year considered in this study. Overall, 25 significant clusters (P < 0.001) were detected and their characteristics are summarised in Table 2.

The summary of TB clusters detected using the Gini coefficient and their characteristics illustrate that the most likely cluster with the highest relative risk was detected in 2015 while the lowest relative risk was recorded in 2018 (Table 2).

Discussion

Results of this study indicated that TB exhibits spatial heterogeneity as indicated by the occurrence of spatial clusters as well as hotspots of the disease in specific districts. The identification of clusters and hotspots in similar locations by the two approaches adopted here i.e., SaTscan
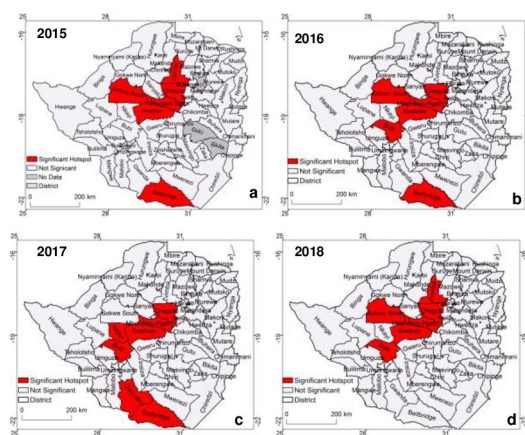
Fig 3. Spatial distribution of TB hotspots in Zimbabwe at district level for (a) 2015, (b) 2016, (c) 2017 and (d) 2018.
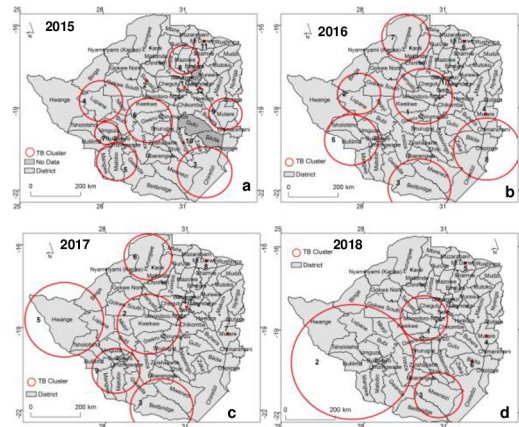
Fig 4. Spatial distribution of statistically significant primary and secondary TB clusters detected by purely spatial analysis based on the Gini coefficient for a) 2015, b) 2016, c) 2017 and d) 2018.

https://doi.org/10.1371/journal.pone.0249523.g004

and Getis Ord G_i^* statistic suggest that they are viable and robust options for identifying and detecting areas of unusually high TB occurrence [12]. This demonstrates the advantages of utilising more than one spatial technique to understand spatial variability in disease occurrence which increases confidence in the validity and application of the results [43]. In studies where Getis Ord G_i^* statistic and SaTscan have been applied, the results have been used to guide TB interventions [24]. The use of GIS and spatial statistics in spatial epidemiological is an important approach for investigating infectious diseases and has been used to study diseases such malaria [60], breast cancer [57] and foot-mouth disease [61]. In these studies, the two techniques yielded useful information for guiding disease control.

The results of this study indicate that statistically significant hotspots and spatial clusters were common in central and southern parts of Zimbabwe. These regions are dominated by large urban areas such as Bulawayo, Gweru and Kwekwe and other rural districts characterised by formal and illegal mining (predominantly artisanal) activities. The occurrence of TB hotspots and clusters in these areas may be associated with high population resulting in overcrowding hence ease transmission of the disease. These results are similar to findings by Chirenda [14] who found that TB hotspots and clusters are common in urban areas due to large populations and inadequate health care services. Consistently across the four years, and especially in 2017, the districts along the Great Dyke exhibited TB hotspots. The Great Dyke region is rich in minerals and has a high concentration of illegal and commercial mining activities. As illustrated by previous studies [26, 63], mining activities expose miners to silica dust which increases the vulnerability of the populations to TB. Characteristics that increase the risk of small scale artisanal miners are reduced access to health care services, inadequate personal protective equipment and the relatively young age. Other health challenges like sexually transmitted infections and HIV infection have been found to be common among this population. All these further confound the increased risk of hotspot and clustering observed in this study.

The tendency of TB cases to cluster in specific localities is of interest to public health response. Significant interventions such as targeted active case finding, early treatment and DOTS could focus on these hotspots. As artisanal miners are a highly mobile population, interventions such as the Zimbabwe National TB Programme could prioritise its targeted active case finding using mobile clinics along the Great Dyke area. Although this study identified critical areas for TB interventions, it did not explore factors explaining the observed patterns.

Table 2. Significant spatial clusters of TB in Zimbabwe based on maximum reported cluster size from 2015–2018.

Year	Cluster #	District	Radius	# of Locations	LLR	Pvalue	Observed	Expected	RR
	1	Seke	0	1	904.05	0.001	1075	210	5.29
	2	Sanyati	0	1	771.97	0.001	1024	223	4.74
	3	Chiredzi	102	2	687.02	0.001	2295	970	2.50
	4	Lupane	82	2	269.10	0.001	968	418	2.37
	5	Matobo	76	3	175.84	0.001	1135	621	1.87
2015	6	Gweru	89	4	159.26	0.001	2147	1443	1.53
	7	Umguza	45	2	131.78	0.001	2137	1491	1.47
	8	Mazowe	58	3	67.12	0.001	1766	1335	1.35
	9	Mutare	63	2	40.85	0.001	1165	888	1.33
	10	Masvingo	0	1	17.57	0.001	766	615	1.25
	11	Mt Darwin	0	1	3.76	0.02	494	436	1.14
	12	Marondera	0	1	2	0.01	410	371	1.10
	1	Mhondoro-Ngezi	112	8	620.61	0.001	4434	2579	1.86
	2	Lupane	82	2	405.73	0.001	1143	435	2.70
	3	Beitbridge	117	3	380.88	0.001	1822	899	2.10
2016	4	Mutare	0	1	132.36	0.001	767	403	1.93
	5	Bulima	102	5	100.36	0.001	2800	2144	1.34
	6	Mt Darwin	0	1	15.37	0.001	580	458	1.27
	7	Hurungwe	92	2	13.75	0.001	1207	1037	1.17
	8	Chipinge	125	5	6.62	0.001	2496	2326	1.08
	2	Kwekwe	120	9	714.10	0.001	4561	2569	1.94
	3	Beitbridge	117	3	305.04	0.001	1649	852	2.00
	4	Seke	0	1	267.41	0.001	626	212	3.00
2017	5	Hwange	150	3	70.46	0.001	931	618	1.53
	6	Hurungwe	92	2	50.08	0.001	1316	992	1.34
	7	Mutare	0	1	34.13	0.001	554	383	1.46
	8	Mt Darwin	0	1	33.94	0.001	617	436	1.43
	9	Insiza	90	5	24.28	0.001	2546	2225	1.16
	1	Mhondoro-Ngezi	90	6	1254.75	0.001	4415	1953	2.49
	2	Bulilima	231	14	665.80	0.001	6961	4559	1.69
2018	3	Mwenezi	90	2	127.49	0.001	1099	656	1.70
	4	Mutare	0	1	111.54	0.001	774	432	1.81
	5	Mt Darwin	0	1	24.45	0.001	653	492	1.34
	6	Bikita	0	1	18.12	0.001	494	373	1.33

RR: Relative risk, LLR: likelihood ratio.

Thus, further research on the possible environmental or social determinants of spatial pattern of TB in Zimbabwe is strongly recommended.

An important result from this study is an observed decrease in relative risk associated with TB clusters over the four year period. The result implies that there has been a general decrease in number of TB notifications in Zimbabwe which suggests the effectiveness of interventions implemented in the general population to reduce TB transmission. The declining TB case notification rates also coincided with the scale-up of antiretroviral therapy in Zimbabwe [64]. The scaling up of antiretroviral therapy under the UNAIDS 90-90-90 targets is likely to result in continued decline in TB case notification rates [39, 64].

This study showed that southern districts such as Chiredzi and Beitbridge had high clustering but relatively low hotspot occurrence compared to its neighbours. Several plausible explanations could be proffered to explain this observation. First, Beitbridge is a border town characterised by high population movements between South Africa and Zimbabwe and is considered the gateway to South Africa. High migration activities might explain the volume of cases reported in Beitbridge district but may not necessarily explain transmission. Secondly, Chiredzi and Mwenezi districts which consistently showed high TB clusters throughout the four years and these are sugarcane growing districts. Studies describing high TB prevalence in agricultural settings are few and the findings of this study require further interrogation to assess the contribution of the agricultural sector to the burden of TB. However, because of the high economic activities from the agricultural business, there could be high local transmission and immigration associated with overcrowding in these areas [65]. Although the potential risk factors that can explain the clustering of TB were not explored in this study, previous studies demonstrated that the TB epidemic in Zimbabwe is largely driven by HIV [64]. Interestingly, areas where TB clusters were detected coincide with areas where clusters of HIV prevalence were also detected [66] suggesting possible co-infection by these two diseases [39].

The use of spatial statistics to explore the distribution and aggregation of TB may be useful in understanding the pattern of TB hotspots [12]. Therefore, TB interventions may not be able to use the "*one size fits all*" approach across different districts. Districts with high disease rates as a result of the social determinants of disease will require more focussed attention than areas with a low risk [21]. Areas where hotspots and clusters of TB cases were detected may be used as priority areas for targeted control to achieve high impact [44, 29]. Thus, to achieve the set targets under the UNSDGs there is need to implement more effective and stronger measures to control TB transmission in these clusters. Policymakers and health authorities will need to strengthen resource mobilization for improved TB prevention and control measures.

The strengths of this study is that the detection of TB hotspots and clusters were performed at district level which is the spatial epidemiological unit at which interventions are planned hence they are likely to be more relevant to policymakers. Performing analysis at district level improves our understanding of the spatial pattern of TB at a wider geographic scale in contrast to most studies which mostly focussed on urban areas [67]. The four years considered in this study may also be sufficient to assess the spatial pattern and stability of hotspots and clusters of the disease in the country as previous studies have focussed on a short time period of about a year. The utility of SaTscan and Getis-Ord G_i^* in differentiating districts with high incidence of TB cases, from those districts where active transmission may actually be common is significant. This may indicate that focussing resources in districts showing high incidence of TB disease may not address transmission with the associated risk factors. We therefore recommend that the NTP complements the surveillance data with geospatial techniques to identify districts with active transmission from those with cold spots.

Although scan statistics and Getis were successfully used to understand TB clusters and hotspots, there are limitations associated with this study. The TB cases reported in each district may be affected by under reporting which influences the detection of hotspots and clusters [12]. Under reporting is a major limitation for all studies that use TB notification data particularly in low income countries without health information management systems [68]. TB notification is dependent on programmatic responses such as ability to test, diagnose and treat individuals including accessibility of heath care services. Some previous studies have found that high TB notification rates in some areas may be associated with better access to TB diagnostic services rather than increased burden or transmission [68]. This may be common in areas with sparse health services. Under reporting impacts negatively on the capacity to accurately evaluate the epidemiology of the disease [68]. However, TB notifications are an

important source of information in the context of local health systems as they can be used as a basis for intervention or programmatic efforts. The information obtained from spatial analysis of TB notifications can be used to prioritise areas requiring further supervision and to tailor interventions to local needs.

A potential drawback in the detection of hotspots in this study is data timeliness. Most of the heath data takes time to compile making it difficult for relevant information to be extracted on time to guide interventions. However, through this retrospective study useful insights on the spatial dynamics of TB occurrence were generated across districts of Zimbabwe. Another potential limitation of this study is that the district TB notification rates are based on data recorded at each health facility. It is therefore possible that a patient can visit more than one health facility resulting in double counting which introduces bias and subsequently affects the detection of TB hotspots and clusters. This is likely to happen in instances where health facilities are in close proximity to each other as well as in referral cases. This means the data recorded in such areas are unlikely to be independent of each other which increase the risk of over-reporting.

Furthermore, TB clusters detected in this study were assumed to be circular and given the irregularity of the administrative boundaries of districts in Zimbabwe, this could result in the exclusion of districts with excess risk. Moreso the results of this study are population-based making it difficult to apply them at individual level. Despite these limitations, this study applied GIS and spatial statistics to identify TB hotspots and clusters at the district level which is important for decision making and resource allocation. Future research need to focus on relationship between TB occurrence and various socio-economic and environmental risk factors to enhance understanding of the pattern of the disease.

Conclusion

GIS and scan statistics were successfully applied to determine the extent to which TB cases were clustered in space at district level in Zimbabwe. The observed hotspots and clustering in districts with intense small scale mining activities may indicate social determinants of TB disease aiding high transmission. The information generated from this study is useful in providing detailed knowledge on the spatial pattern of TB occurrence which is critical for targeted TB interventions. This is particularly important in developing countries where there is need to prioritise allocation of limited resources by focussing on areas with greatest need which results in high impact in terms of disease control.

Supporting information

S1 Data.
(XLS)

S1 Annex. Districts of Zimbabwe showing the projected population, recoded TB cases and the notification rate from 2015 to 2018.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Isaiah Gwitira, Norbert Karumazondo.

Data curation: Charles Sandy, Nicolas Siziba.

Formal analysis: Isaiah Gwitira, Norbert Karumazondo.

Investigation: Norbert Karumazondo.

Methodology: Isaiah Gwitira, Munyaradzi Davis Shekede.

Software: Isaiah Gwitira.

Supervision: Charles Sandy, Joconiah Chirenda.

Writing – original draft: Isaiah Gwitira.

Writing – review & editing: Norbert Karumazondo, Munyaradzi Davis Shekede, Joconiah Chirenda.

References

1. Izumi K., Ohkado A., Uchimura K., Murase Y., Tatsumi Y., and Kayebeta A., Detection of Tuberculosis Infection Hotspots Using Activity Spaces Based Spatial Approach in an Urban Tokyo, from 2003 to 2011. PloS one, 2015. 10(9): p. https://doi.org/10.1371/journal.pone.0138831 PMID: 26382251

2. Alene K.A., Viney K., Moore H.C., Wagaw M., and Clements A.C.A., Spatial patterns of tuberculosis and HIV co-infection in Ethiopia. PLoS ONE, 2019. 14(12): p. e0226127https://doi.org/10.1371/journal.pone.0226127 PMID: 31805149

3. WHO, World Tuberculosis Report. 2019, WHO: Geneva.

4. Tiwari N., Kandpal V., Tewari A., Rao K.R.M., and Tolia V.S., Investigation of tuberculosis clusters in Dehradun city of India. Asian Pacific Journal of Tropical Medicine, 2010: p. 486–490.

5. Shaweno D., Karmakar M., Alene K.A., Ragonnet R., Clements A.C.A., Trauer J.M., et al., Methods used in the spatial analysis of tuberculosis epidemiology: a systematic Review. BMC Medicine 2018. 16(193): p. https://doi.org/10.1186/s12916-018-1178-4 PMID: 30333043

6. Tiwari N., Adhikari C.M.S., Tewari A., and Kandpal V., Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic. International Journal of Health Geographics, 2006. 5(33): p. https://doi.org/10.1186/1476-072X-5-33 PMID: 16901341

7. Touray K., Adetifa I.M., Jallow A., Rigby J., Jeffries D., and Cheung Y.B., Spatial analysis of tuberculosis in an urban west African setting: is there evidence of clustering? Tropical medicine & international health:TM & IH, 2010. 15(6): p. 664–72.

8. WHO, The End TB Strategy. 2015: Geneva.

9. UNDESA, Sustainable Development Goal 3: Ensuring Health Lives and PromoteWell-Being for All at All Ages. 2015: Available online: (accessed on 15 September 2019). p. https://sustainabledevelopment.un.org/sdg3.

10. Dowdy D.W., Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. Proceedings of the National Academy of Sciences, 2012. 109(24): p. 9557–9562. https://doi.org/10.1073/pnas.1203517109 PMID: 22645356

11. Meehan C.J., Moris P., Kohl T.A., Pečerska J., Akter S., Merker M., et al., The relationship between transmission time and clustering methods in Mycobacterium tuberculosis epidemiology. EBioMedicine 2018. 37: p. 410–416. https://doi.org/10.1016/j.ebiom.2018.10.013 PMID: 30341041

12. Dangisso M.H., Datiko D.G., and Lindtjørn B., Spatio-temporal analysis of smear-positive tuberculosis in the sidama zone, Southern Ethiopia. PLos One, 2015. 10(6): p. e126369. https://doi.org/10.1371/journal.pone.0126369 PMID: 26030162

13. Liu Y., Li X., Wang W., Li Z., Hou M., He Y., et al., Investigation of space-time clusters and geospatial hot spots for the occurrence of tuberculosis in Beijing. Int J Tuberc Lung Dis, 2012. 16(4): p. 486–91. https://doi.org/10.5588/ijtld.11.0255 PMID: 22325066

14. Chirenda J., Gwitira I., Warren R.M., Sampson S.L., Murwira A., and Masimirembwa C., Spatial distribution of Mycobacterium Tuberculosis in metropolitan Harare, Zimbabwe. PLoS ONE, 2020. 15 (e0231637). https://doi.org/10.1371/journal.pone.0231637 PMID: 32315335

15. Burke S.J., Lass E., Thistle P., Katumbe L., and Jetha A., Increased Incidence of Tuberculosis in Zimbabwe, in Association with Food Insecurity, and Economic Collapse: An Ecological Analysis. PLoS ONE, 2014. 9(2): p. e83387. https://doi.org/10.1371/journal.pone.0083387 PMID: 24505245

16. Timire C., Metcalfe J.Z., Chirenda J., Scholten J.N., Manyame-Murwira B., Ngwenya M., et al., Prevalence of drug-resistant tuberculosis in Zimbabwe: A health facility-based cross-sectional survey. International Journal of Infectious Diseases, 2019. 87: p. 119–125. https://doi.org/10.1016/j.ijid.2019.07.021 PMID: 31357057

17. Metcalfe J.Z., Makumbirofa S., Makamure B., Sandy C., Bara W., and Mungofa S., Drug-Resistant Tuberculosis in High-Risk Groups, Zimbabwe. Emerg Infect Dis, 2014. 20(1): p. 135–7. https://doi.org/10.3201/eid2001.130732 PMID: 24377879

18. Guo C., Du Y., Shen S.Q., Lao X.Q., Qian J., and Ou C.Q., Spatiotemporal analysis of tuberculosis incidence and its associated factors in mainland China. Epidemiol Infect, 2017. 145(12): p. 2510–9. https://doi.org/10.1017/S0950268817001133 PMID: 28595668

19. Davies A.G.R., Squire S.B., Bmj S., Medical B., and Mar N., Doubts about DOTS. BMC Infect Diseases, 2018. 336(7642): p. 457–8.

20. Chikwaiwa B.K., Dzingirai V., Nyikahadzoi K., and Rusakaniko S., Institutional and personal factors influencing adherence to tuberculosis treatment among patients in Harare,Zimbabwe. Afr J Soc Work, 2018. 8(2): p. 93–101.

21. Woldeyohannes S.M. and Abera S.Y., Worldwide Spatial and Temporal Distribution of Tuberculosis (TB) J AIDS Clin Res, 2015. 6(452): p. https://doi.org/10.4172/2155-6113.1000452

22. Onozuka D. and Hagihara A., Geographic prediction of tuberculosis clusters in Fukuoka, Japan, using the space-time scan statistic. BMC infectious diseases, 2007. 7(26). https://doi.org/10.1186/1471-2334-7-26 PMID: 17428326

23. Yates T.A., Khan P.Y., Knight G.M., Taylor J.G., McHugh T.D., Lipman M., et al., The transmission of Mycobacterium tuberculosis in high burden settings. Lancet Infect Dis, 2016. 16(2): p. 227–38. https://doi.org/10.1016/S1473-3099(15)00499-5 PMID: 26867464

24. Wang T., Xue F., Chen Y., Ma Y., and Liu Y., The spatial epidemiology of tuberculosis in Linyi City, China, 2005–2010. BMC public health, 2012. 12(885): p. https://doi.org/10.1186/1471-2458-12-885 PMID: 23083352

25. Randremanana R.V., Sabatier P., Rakotomanana F., Randriamanantena A., and Richard V., Spatial clustering of pulmonary tuberculosis and impact of the care factors in Antananarivo City. Tropical medicine & international health: TM & IH, 2009. 14(4): p. 429–37. https://doi.org/10.1111/j.1365-3156.2009.02239.x PMID: 19228353

26. Corbett E.L., C. S., and Moloi V.M., Human immunodeficiency virus and the prevalence of undiagnosed tuberculosis in African gold miners. Am J Respir Crit Care Med., 2004. 170(6): p. 673–679. https://doi.org/10.1164/rccm.200405-590OC PMID: 15191919

27. Gwitira I., Murwira A., Mberikunashe J., and Masocha M., Spatial overlaps in the distribution of HIV/AIDS and malaria in Zimbabwe. BMC Infect Diseases, 2018. 18(1): p. https://doi.org/10.1186/s12879-018-3513-y PMID: 30482166

28. Keshavjee S., Dowdy D., and Swaminathan S., Stopping the body count: a comprehensive approach to move towards zero tuberculosis deaths. Lancet, 2015. 386(10010): p. e46–7. https://doi.org/10.1016/S0140-6736(15)00320-7 PMID: 26515680

29. Coleman M., Coleman M., Mabuza A.M., Kok G., Coetzee M., and Durrheim D.N., Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. Malaria Journal, 2009. 8(68). https://doi.org/10.1186/1475-2875-8-68 PMID: 19374738

30. Nassel A.F., Root E.D., Haukoos J.S., McVaney K., Colwell C., Robinson J., et al., Multiple cluster analysis for the identification of high risk census tracts for out-of-hospital cardiac arrest (OHCA) in Denver, Colorado. Resuscitation, 2014. 85(12): p. 1667–73. https://doi.org/10.1016/j.resuscitation.2014.08.029 PMID: 25263511

31. Wartenberg D. and Greenberg M., Solving the cluster puzzle: clues to follow and pitfalls to avoid. Stat Med, 1993. 12 (19–20): p. 1763–70. https://doi.org/10.1002/sim.4780121905 PMID: 8272659

32. Burra T., Jerrett M., Burnett R.T., and Anderson M., Conceptual and practical issues in the detection of local disease clusters: a study of mortality in Hamilton, Ontario. Can Geographer/Le Géographe Canadien, 2002. 46(2): p. 160–71.

33. Unganai L.S., Historic and future climatic change in Zimbabwe. Climate Research, 1996. 6(2): p. 137–145.

34. Gwitira I., Murwira A., Masocha M., Zengeya F.M., Shekede M.D., Chirenda J., et al., GIS-based stratification of malaria risk zones for Zimbabwe. Geocarto International, 2019. 34(11): p. 1163–1176.

35. WorldBank. Population Growth in Zimbabwe. 2019.

36. Noppert G., Yang Z., Sandy C., and Chirenda J., Trends of sputum-smear positive tuberculosis in Zimbabwe: 2008–2011. BMC research notes, 2015. 8(1): p. 575. https://doi.org/10.1186/s13104-015-1568-z PMID: 26475610

37. MoHCC, Zimbabwe National Tuberculosis Control Programme Manual. 2007, Ministry of Health and Child Welfare: Harare.

38. MOHCC, Natonal Tuberculosis Program—Strategic Plan (2017–2020). 2017, MOHCC: Harare.

39. Takarinda K.C., Harries A.D., Srinath S., Mutasa-Apollo T., Sandy C., and Mugurungi O., Treatment outcomes of adult patients with recurrent tuberculosis in relation to HIV status in Zimbabwe: a retrospective record review. BMC Public Health, 2012. 12(1): p. 1–6. https://doi.org/10.1186/1471-2458-12-124 PMID: 22329930

40. Harries, A.D. and D. Maher, TB/HIV A clinical manual. 1996, World Health Organisation: Geneva.

41. ZIMSTAT, Zimbabwe Population Projections Thematic Report. 2015: Harare.

42. Moran P.A., Notes on continuous stochastic phenomena. Biometrika, 1950. 37: p. 17–23. PMID: 15420245

43. Zhao F., Cheng S., He G., Huang F., Zhang H., and Xu B., Space-time clustering characteristics of tuberculosis in China, 2005–2011. PLoS ONE, 2013. 8(12): p. 3–9.

44. Bekara M.E., Azizi L., Bénet J.J., and Durand B., Spatial-temporal Variations of Bovine Tuberculosis Incidence in France between 1965 and 2000. Transbound Emerg Dis, 2016. 63(1): p. 101–113. https://doi.org/10.1111/tbed.12224 PMID: 24735092

45. Lai P.C., So F.M., and Chan K.W., eds. Spatial Epidemiological Approaches in Disease Mapping and Analysis. 2009, CRC Press: New York.

46. ESRI, ArcGIS Desktop: Release 10.3. 2011, Environmental Systems Research Institute: Redlands CA.

47. Hassarangsee S., Tripathi N.K., and Souris M., Spatial Pattern Detection of Tuberculosis: A Case Study of Si Sa Ket Province, Thailand. Int J Env Res Public Health 2015, 2015. 12: p. 16005–16018. https://doi.org/10.3390/ijerph121215040 PMID: 26694437

48. Anselin L. and Getis O., Spatial statistical analysis and geographic information systems. Ann Reg Sci, 1992. 26: p. 19–33.

49. Getis, A. and K.J. Ord, eds. Local Spatial Statistics: An Overview." In Spatial Analysis:Modelling in a GIS Environment. ed. P. Longley and M. Batty. 1996, John Wiley: Cambridge.

50. Kulldorff M., A spatial scan statistic. Communications in Statistics -Theory and Methods, 1997. 26: p. 1481–1496.

51. Chowdhury A.I., Abdullah A.Y.M., Haider R., Alam A., Billah S.M., Bari S., et al., Analyzing spatial and space-time clustering of facility-based deliveries in Bangladesh. Tropical medicine and health, 2019. 47(1): p. 44. https://doi.org/10.1186/s41182-019-0170-9 PMID: 31346313

52. Getis O.J.K., A Local spatial autocorrelation statistics: Distributional issues and an application. Geogr. Anal, 1995. 27: p. 286–306.

53. de Castro M.C. and Singer B.H., Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Tests in Local Statistics of Spatial Association. Geographical Analysis, 2006. 38: p. 180–208

54. Benjamini Y. and Hochberg Y., Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society B, 1995. 57(1): p. 289–300.

55. Benjamini Y. and Hochberg Y., On the Adaptive Control of the False Discovery Rate in Multiple Testing with Independent Statistics. Journal of Educational and Behavioral Sciences, 2000. 25(1): p. 60–83.

56. Cromley, E.K. and S.L. McLafferty, GIS and Public Health. 2012, Guilford press: New York.

57. Kulldorff M., Feuer E.J., Miller B.A., and Freedman L.S., Breast cancer clusters in the northeast United States: A Geographic Analysis. Am. J. Epidemiology, 1997. 146(2): p. 161–70. https://doi.org/10.1093/oxfordjournals.aje.a009247 PMID: 9230778

58. Kulldorff M. and Nagarwalla N., Spatial disease clusters: Detection and inference. Stat Med, 1995. 14 (8): p. 799–810. https://doi.org/10.1002/sim.4780140809 PMID: 7644860

59. Wand H. and Ramjee G., Targeting the hotspots: investigating spatial and demographic variations in HIV infection in small communities in South Africa. Journal of the International AIDS Society, 2010. 13 (1): p. 41. https://doi.org/10.1186/1758-2652-13-41 PMID: 20979651

60. Xia J., Cai S., Zhang H., Lin W., Fan Y., Qiu J., et al., Spatial, temporal, and spatiotemporal analysis of malaria in Hubei Province, China from 2004–2011. Malaria Journal, 2015. 14(145): p. https://doi.org/10.1186/s12936-015-0650-2 PMID: 25879447

61. Liu Y., Wang X., Liu Y., Sun D., Ding S., Zhang B., et al., Detecting Spatial-Temporal Clusters of HFMD from 2007 to 2011 in Shandong Province, China. PLoS ONE, 2013. 8(5): p. e63447. https://doi.org/10.1371/journal.pone.0063447 PMID: 23704909

62. Han J., Zhu L., Kulldorff M., Hostovich S., Stinchcomb D.G., Tatalovich Z., et al., Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics. International Journal of Health Geographics, 2016. 15(27): p. https://doi.org/10.1186/s12942-016-0056-6 PMID: 27488416

63. Stuckler D., Basu S., McKee M., and Lurie M., Mining and risk of tuberculosis in sub-Saharan Africa. Am J Public Health, 2011. 101(3): p. 524–530. https://doi.org/10.2105/AJPH.2009.175646 PMID: 20516372

64. Takarinda K., Harries A., and Mutasa-Apollo T., Trend analysis of tuberculosis case notifications with scale-up of antiretroviral therapy and roll-out of isoniazid preventive therapy in Zimbabwe, 2000–2018. BMJ Open, 2020. 10 https://doi.org/10.1136/bmjopen-2019-034721 PMID: 32265241

65. Fenner L., Gagneux S., Helbling P., Battegay M., Rieder H.L., and Pfyffer G.E., Mycobacterium tuberculosis transmission in a country with low tuberculosis incidence: role of immigration and HIV infection. J Clin Microbiol, 2012. 50(2): p. 388–395. https://doi.org/10.1128/JCM.05392-11 PMID: 22116153

66. Cuadros D.F., Awad S.F., and Abu-Raddad L.J., Mapping HIV clustering: a strategy for identifying populations at high risk of HIV infection in sub-Saharan Africa. International Journal of Health Geographics, 2013. 12(28): p. 1–9. https://doi.org/10.1186/1476-072X-12-28 PMID: 23692994

67. Dietz N.A., Sherman R., Mackinnon J., Fleming L., Arheart K.L., Wohler B., et al., Toward the identification of communities with increased tobacco-associated cancer burden: application of spatial modeling techniques. J Carcinog, 2011. 10(22). https://doi.org/10.4103/1477-3163.85184 PMID: 22013392

68. Zhou D., Pender M., Jiang W., Mao W., and Tang S., Under-reporting of TB cases and associated factors: a case study in China. BMC public health, 2019. 19(1): p. 1664. https://doi.org/10.1186/s12889-019-8009-1 PMID: 31829147