

**SPATIAL - TEMPORAL ANALYSIS OF TB CASE NOTIFICATION IN
NORTHERN MALAWI**

MASTER OF SCIENCE(BIOSTATISTICS) THESIS

MEMORY CHITEMA

UNIVERSITY OF MALAWI

JUNE 2024



**SPATIO - TEMPORAL ANALYSIS OF TB CASE NOTIFICATION IN
NORTHERN MALAWI**

MSc (Biostatistics) Thesis

By

MEMORY CHITEMA

BSc. (Statistics) – University of Malawi

Submitted to the Department of Mathematical Sciences, School of Natural and
Applied Science, in partial fulfillment of the requirements for the degree of Master of
Science (Biostatistics)

UNIVERSITY OF MALAWI

JUNE 2024

DECLARATION

I, the undersigned hereby declare that this thesis is my own original work which has not been submitted to any other institution for similar purposes. Where other people's work has been used, acknowledgements have been made.

MEMORY CHITEMA

Full Legal Name

Signature

Date

CERTIFICATE OF APPROVAL

We, the undersigned certify that this thesis represents the student's own work and effort and has been submitted with our approval.

Signature:_____Date:_____

Jupiter Simbeye, PhD (Senior Lecturer)

Supervisor

DEDICATION

To my beloved mother, Chifundo Rose Kachiza, your unwavering love, care, and support have been the bedrock of my academic journey. You laid a strong foundation for my educational pursuits, and without your constant encouragement and guidance, I would not have achieved the level of success I have attained today. This thesis is a testament to your selfless dedication as a parent and the invaluable role you have played in shaping my aspirations. I am forever grateful for your presence in my life.

To my husband, Ted, you have been my pillar of strength and a constant source of inspiration throughout this challenging endeavour. In moments when self-doubt threatened to consume me, your belief in my abilities and steady support propelled me forward. Your encouragement and faith in my potential have been instrumental in my accomplishments. I am profoundly grateful for your love, patience, and belief in me.

To my dear siblings, Ekari, Pempho, Thandizo, and the late Kondwani, you have always been my number-one cheerleaders. Your continuous support, encouragement, and belief in my abilities have motivated me to soar higher and strive for my dreams. Each of you has played a unique role in shaping my character and fostering a resilient spirit within me. This thesis is a tribute to our bond and the strength we derive from our shared experiences.

Lastly, to my beloved son, Tyler, you are my greatest joy and the driving force behind my pursuit of knowledge. As the sun knows no limits, I want you to believe that there are no boundaries to what you can achieve. You have inspired me to push beyond my limits and strive for excellence. This thesis is a testament to the love and dedication I have for you and my commitment to building a brighter future for us.

To all those mentioned above, I extend my deepest gratitude for your support, love, and encouragement. Your presence in my life has made this journey not only possible

but also immensely fulfilling. I dedicate this thesis to each of you, as a reflection of our shared triumphs and the endless possibilities that lie ahead.

ACKNOWLEDGEMENTS

To the Almighty God, without whom the completion of this master's dissertation would not have been possible, I offer my deepest gratitude. I praise God the Almighty for providing me with this opportunity and granting me the capability to proceed successfully. To Him be all the glory and honour.

I extend my sincere thanks and appreciation to Dr Simbeye, my esteemed supervisor. I am truly grateful for your warm encouragement, thoughtful guidance, critical comments, and corrections throughout the process of this thesis. Your firm support and mentorship have been invaluable to my academic growth. Your constructive criticisms have not only made the journey enjoyable but also served as a source of encouragement.

I would like to express my heartfelt thanks to the DELTAS African Sub-Saharan African Consortium for Advanced Biostatistics Training (SSACAB) for awarding me a full-time scholarship. This support has played a significant role in facilitating the smooth advancement of my studies. I am grateful for the opportunities and resources provided by SSACAB, which have contributed to the successful completion of this dissertation.

I would like to express my profound appreciation to the Ministry of Health for providing me with the data analysed in this study. Your cooperation and support have been instrumental in conducting comprehensive research and generating meaningful findings.

I am sincerely grateful to all the lecturers who taught our Bio-statistics class, for their exceptional expertise and dedication in imparting knowledge and skills that have been indispensable in this research. Finally, I am immensely grateful to my classmates who served as pillars of support.

ABSTRACT

Tuberculosis (TB) presents a significant public health challenge, particularly in high disease burden countries in Africa, such as Malawi. The extent of this problem varies across different settings. Understanding the spatial variation and underlying causes of TB prevalence is crucial for comprehending and addressing the epidemic. Therefore, the utilization of geospatial analytical methods and spatial temporal models is vital in analysing and detecting spatial and spatial-temporal clustering of infectious diseases. This study investigated the spatial distribution and presence of spatial-temporal clusters of TB in various geographic settings over an eight-year period (2013 – 2020) in Northern Malawi. Specifically, the study aimed to identify temporal trends, spatial patterns, infectious disease clusters or hotspots and cluster coverage related to TB. Spatial and spatial temporal statistical analyses were employed, utilizing Kulldorff's scan statistics tool, through the implementation of Spatial and Spatial-temporal models. The findings revealed the presence of seven clusters or hotspots within the study area. The model identified a cluster pattern where hotspots were observed in areas characterized by relatively higher population sizes and densities, predominantly located within economically developed zones. Notably, the clusters or hotspots in Northern Malawi were found to form around Central Districts, District Hospitals, major rural hospitals, as well as urban and semi-urban centres, including Mzuzu City, Chintheche, Bolero, Bwengu, Mzimba Boma, Embangweni, Jenda, Chitimba, among others. By analysing time trends, the study observed a general decline in TB infection cases within the region, with an annual decrease of 1.08%. These results indicate significant progress in disease control efforts. However, TB still poses a considerable risk to the population residing in proximity to clustered health centres.

TABLE OF CONTENTS

ABSTRACT.....	vi
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF APPENDICES.....	xii
CHAPTER 1	1
INTRODUCTION	1
1.1 Background	1
1.1.1 Spatial clustering of TB data	2
1.1.2 Case Studies and spatial-temporal Analysis of TB in Malawi	3
1.1.3 Global Case Studies: Spatial-temporal Analysis of Tb	4
1.2 Problem statement	7
1.3 Objectives.....	9
1.3.1 Main Objective	9
1.3.2 Specific Objectives	9
1.4 Significance of the study	10
1.5 Thesis Structure.....	10
CHAPTER 2	11
LITERATURE REVIEW	11
2.1 Introduction	11
2.2 Space-Time Scan Statistics model	11
2.2.1 Method of estimation.....	12
2.3 Spatiotemporal point process models.....	13
2.3.1 Method of estimation.....	14
2.4 Space-Time Autoregressive Integrated Moving Average (STARIMA) model ..	15
2.4.1 Method of estimation.....	15
2.5 Space-Time Generalized Linear Mixed Models (ST-GLMMs).....	16

2.5.1 Method of estimation.....	17
2.6 Rationale for Using the Space-Time Scan Statistics Model in the Spatial-Temporal Analysis of TB Case Notification in Northern Malawi	17
2.7 Cases where spatial-temporal methods have been applied to data.....	19
CHAPTER 3	22
METHODOLOGY	22
3.1 Introduction	22
3.2. Data Collection and Analysis	22
3.2.1 Data Sources and Collection Procedures	22
3.2.2 Poisson Model for Spatial-Temporal Cluster Detection in Kulldorff's SaTScan software	24
3.2.3 Ethical considerations.....	25
3.3 Statistical Methods for Analysing Spatial-Temporal Dynamics of TB Case Notification Data	26
3.3.1 Spatial Model: Spatial scan statistic	26
3.3.2 Spatial – Temporal Model: Space-time scan statistic.....	27
CHAPTER 4	31
RESEARCH RESULTS	31
4.1 Introduction	31
4.2 Results of the study	31
4.2.1 District TB case notification.....	31
4.2.2 Gender and age distribution of TB Cases	32
4.2.3 Trend of TB Infection Rates	32
4.2.4 Seasonal distribution of TB Cases.....	33
4.2.5 TB Hotspot and Cluster identification.....	34
4.2.6 TB Cluster relative risk (RR).....	38
4.2.7. Temporal trends of TB Infections	39
CHAPTER 5	41

DISCUSSION OF RESULTS	41
5.1 Introduction	41
5.2 Gender and Age distribution of Tb Cases	41
5.3 Spatial and Spatial-Temporal clusters of Tb in the Northern Region of Malawi	42
5.4 Tb Spatial Pattern and Distribution.....	43
5.5 Detection of Tb Spatial – Temporal trends	44
CHAPTER 6	46
CONCLUSION.....	46
6.1 Introduction	46
6.2 Conclusions	46
6.3 Recommendations	48
6.4 Limitations of study	50
REFERENCES	51
APPENDIX.....	58

LIST OF TABLES

Table 1: Retrospective Spatial- Temporal analysis model estimates.....	36
Table 2: District Annual TB case notifications.....	58
Table 3: Health Centre Annual TB case notifications	59

LIST OF FIGURES

Figure 1: Map of Malawi	23
Figure 2: District TB infection distribution 2013 – 2020.	31
Figure 3: Gender and age group distributions among reported cases.....	32
Figure 4: TB infection Rate trends (2013-2020).....	32
Figure 5: Monthly distributions of TB cases.	33
Figure 6: Spatial Map indicating TB Case Notification Clusters in the Northern Region of Malawi	35
Figure 7: Cluster size and population coverage.....	37
Figure 8: Number of cases per cluster.	38
Figure 9: Relative Risks in clusters.	39
Figure 10: TB Case notification Temporal Trends	39

LIST OF APPENDICES

Appendix 1: TB Cases per District	58
Appendix 2: TB Cases per Health Centers	59
Appendix 3: Retrospective Space-Time analysis	62
Appendix 4: Spatial Variation in Temporal Trends analysis	69

ABBREVIATIONS

AIDS	Acquired Immune Deficiency Syndrome
BMU	Basic Measurement Unit
CDC	Centre for Disease Control
DHIS	District Health Information System
EM	Expectation Maximization
GIS	Geographic Information System
GPS	Geographic Position System
HC	Health Centre
HIV	Human Immune Virus
MCMC	Markov Chain Monte Carlo
MoH	Ministry of Health
NTP	National TB Program
PTB	Pulmonary Tuberculosis
RR	Relative Risk
STARIMA	Space-Time Autoregressive Integrated Moving Average
ST-GLMMs	Space-Time Generalized Linear Mixed Models
TA	Traditional Authority
TB	Tuberculosis
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

1.1 Background

Tuberculosis (TB) presents a significant challenge to public health, particularly in high disease burden countries, which are primarily concentrated in Asia and Africa (Centers for Disease Control and Prevention, 2020). The geographic distribution of TB varies globally and within countries, influenced by factors such as socio-economic conditions, overcrowding, poverty, limited access to health services, socio-cultural barriers, and HIV infection (Chen, 2019; Nagamine et al., 2008). Socioeconomic status also plays a crucial role, with an association between high TB incidences and low socioeconomic status observed in most developing countries (Rubel & Garro, 1992; Souza et al., 2007).

World Health Organization (WHO) estimates indicate that approximately 9.4 million people develop active TB annually, with over 2.8 million cases occurring in Africa, a trend exacerbated by the HIV epidemic (WHO, 2020). TB remains a global public health concern, with variations in incidence not only among countries but also within different regions of a country (WHO, 2020). These differences are influenced by various epidemiological and social factors, including lifestyle, smoking, occupation, and exposure to risk factors (WHO, 2020).

In Malawi, a landlocked country in southeastern Africa, TB control faces challenges due to limited healthcare infrastructure, socioeconomic disparities, and high rates of HIV co-infection (Tadesse & Demissie, 2019). Despite improvements in case detection and treatment outcomes, the spatial distribution, and temporal dynamics of TB transmission in Malawi, particularly in the northern region, remain poorly understood.

Spatial distribution analysis has been instrumental in highlighting regional variations in disease prevalence (Garrido et al., 2015). Spatial-temporal analysis, a powerful epidemiological tool, provides insights into the geographical distribution and temporal trends of infectious diseases like TB. By analysing TB cases across space and time, researchers can identify spatial clusters, temporal patterns, and high-risk populations, informing targeted interventions and resource allocation strategies. (Tadesse & Demissie, 2019)

The northern region of Malawi, encompassing districts such as Karonga, Rumphi, and Mzimba, presents a unique setting for TB epidemiology research. Despite being relatively understudied compared to other regions, the northern region grapples with a significant TB burden, exacerbated by poverty, limited healthcare access, and population mobility (Mhimbira et al., 2017).

1.1.1 Spatial clustering of TB data

The identification of geographical areas with ongoing disease transmission, using geographic information systems and spatial-temporal statistical analyses, has become indispensable. Spatial-temporal clustering methods aid in the identification of a greater density of occurrences of a phenomenon in certain places at certain times (Vargas, 2004). These techniques have been intensively applied in several areas, such as demography, criminology, and toxicology. For epidemiologists, disease clustering is a technique of major interest that has been studied for many decades. For effective disease management, it is essential to know when, where, and to what degree a disease is present. In the last decade, there has been a rapid development of spatial-temporal clustering techniques applied to health, in the assessment of infectious diseases, cancer, rheumatism, diabetes, and accidents and to detect the infectious disease hotspots of a variety of infectious diseases in many countries across the continent (Vargas, 2004).

Studies such as Miandad et al. (2014) conducted in Karachi, Pakistan and Bastida et al. (2012) reported spatial and spatial-temporal clustering of TB, thereby generating valuable information about the distribution of the disease. However, these studies were conducted in urban settings over a brief period, which makes them deficient in detecting the pattern of the disease distribution in rural areas. (Vargas, 2004)

1.1.2 Case Studies and spatial-temporal Analysis of TB in Malawi

Nyirenda et al. (2005) explored the use of geographical information systems in TB control efforts in Malawi. The study assessed the effectiveness of GIS in enhancing TB control strategies and understanding disease dynamics. GIS technology was used to analyse spatial patterns of TB incidence and distribution within Malawi. The study integrated spatial data on TB cases with geographical information, facilitating the visualization of TB hotspots, disease clusters, and areas with high disease burden. However, one potential limitation was the temporal aspect, as the study was conducted in 2005, and TB control strategies and technologies may have evolved since then. The findings may not fully reflect the current landscape of TB control efforts in Malawi, and newer developments in GIS technology may have enhanced its utility in TB surveillance and control.

Kirollos et al. (2021) conducted a study on tuberculosis (TB) case notifications in Malawi, focusing on seasonal and weather-related trends. The research aimed to identify patterns and correlations between TB incidence and seasonal variations, as well as weather conditions. Statistical models were used to analyse TB case notifications, including time series analysis and regression models. The focus on seasonal and weather-related factors, providing insights into environmental influences on TB transmission highlighted the strengths of this study.

Khundi et al. (2021) conducted a multilevel epidemiological analysis in urban Blantyre, Malawi, focusing on clinical, health systems, and neighbourhood determinants of TB case fatality. The study used surveillance data to provide insights into factors influencing TB mortality rates. A notable strength of the study was in its comprehensive approach, which examined multiple determinants of TB case fatality at various levels. The study offered a holistic understanding of TB mortality dynamics in urban Blantyre by considering clinical, health systems, and neighbourhood factors. Additionally, the utilization of enhanced surveillance data enhanced the reliability and accuracy of the findings, contributing to a more robust analysis.

However, despite these strengths, the study may have limitations. One potential limitation is the focus on only one urban district, Blantyre, which may not fully capture the dynamics of TB case fatality in rural areas or other regions of Malawi.

This geographical limitation may impact the generalizability of the findings to other settings within the country. Additionally, while the study provides valuable insights into clinical and health system determinants of TB mortality, other factors such as socio-economic determinants or access to healthcare services may not have been fully explored. Therefore, future research could benefit from a more comprehensive examination of TB case fatality determinants, encompassing a broader range of factors and geographical contexts.

Nightingale et al. (2021) analysed community-level variation in TB testing history based on a prevalence survey in Blantyre, Malawi. The study assessed disparities in TB testing practices across different communities within Blantyre. The study's focus on community-level variation, which provided insights into disparities in healthcare access and utilization at the local level proved to be its notable strength. The study enhanced the understanding of factors influencing TB diagnosis and healthcare-seeking behaviour by examining TB testing history across multiple communities. Conversely, the study's focus on a specific geographic area (Blantyre, Malawi) limits the generalizability of the findings to other regions or countries.

While these spatial and temporal analysis methods have been widely applied in TB research, there is limited literature specifically addressing their application in the context of the northern region of Malawi. This gap in knowledge highlights the need for studies that apply these advanced statistical methodologies to analyse TB data in this region and elucidate the spatial-temporal dynamics of TB case notification.

1.1.3 Global Case Studies: Spatial-temporal Analysis of Tb

Bastida et al. (2012) conducted a research study with the primary objective of utilizing spatial statistics program SCAN and GIS to identify the spatial and temporal distribution of TB from 2006 to 2010 in the State of Mexico. The study utilized population data on TB cases reported from 499 localities during the specified period. Spatial and space-time analysis techniques were used to analyse the data. The findings revealed nine significant clusters ($P < 0.05$) of TB incidence, indicating that TB in the State of Mexico was not randomly distributed but rather concentrated in specific areas close to Mexico.

The study demonstrates several strengths that contribute to our understanding of TB epidemiology in the State of Mexico. Firstly, the utilization of advanced spatial statistics program SCAN and GIS techniques allowed for the identification of significant TB clusters and hotspots, providing valuable insights into the spatial distribution of TB cases. Additionally, the comprehensive geographic coverage of the study, encompassing data from 499 localities, facilitated a thorough assessment of TB distribution patterns, ensuring a representative sample for analysis.

However, despite these strengths, a notable drawback is the limited temporal scope of the study, which focused solely on TB incidence from 2006 to 2010. This narrow timeframe restricts the assessment of long-term TB epidemiology trends and limit the findings' predictions to other periods. This limitation has contributed to the rationale for this study, which focuses on examining spatial-temporal trends in TB case notification over an eight-year period.

Miandad et al. (2014) conducted a study in Karachi, Pakistan, with the primary objective of using exploratory disease mapping to determine salient spatial patterns of TB and demarcate concentration zones of TB patients in the study area. The research context was significant with reference to WHO and the International Union against TB and lung diseases characterizing Pakistan's TB situation as one of the worst in the world. The Government of Pakistan, in collaboration with WHO, launched a TB control program nationwide, including in Karachi, providing TB diagnosis equipment and financial support to health centers and NGOs for free TB testing of suspected patients.

The study utilized spatial analysis techniques, including GIS applications, to analyse TB patient data recorded at TB diagnosis centers in Karachi from 2010 to 2013. The findings revealed a gradual increase in the number of TB patients during the study period, with a notable decrease in 2012. Furthermore, the spatial analysis indicated that most TB patients belonged to low-income groups and resided in kacchi abadies, underscoring the socio-economic disparities in TB incidence.

One notable strength of the study by Miandad et al. (2014) was in its utilization of GIS technology for spatial analysis. The study was able to identify spatial patterns and

concentration zones of TB patients, providing valuable insights for targeted intervention strategies by leveraging GIS applications. Like the study conducted by Bastida et al. (2012), one notable drawback was the relatively restricted temporal scope of the research, which exclusively examined TB patient data in a narrow timeframe from 2010 to 2013. Additionally, while the findings provide insights into TB patterns in Karachi, they may not be readily applicable to other regions or countries with different socio-economic and healthcare contexts. Hence, there is a need for the adoption of holistic approaches when conducting similar studies, encompassing the entire region.

Dangisso et al. (2015) conducted a retrospective space-time and spatial analysis to explore the spatial and spatial-temporal patterns of smear-positive pulmonary tuberculosis (PTB) in Ethiopia, a country burdened with high TB rates and regional variations in case notification rates. The study, conducted at the kebele level (the lowest administrative unit within a district), aimed to identify clusters of PTB cases and understand their distribution across different settings.

The study's comprehensive approach, utilizing a range of statistical methods including scan statistics, Global Moran's I, and Getis and Ord's (Gi*) statistics for spatial analysis is recommended. The study analysed data from 22,545 smear-positive PTB cases notified over a period of 10 years, providing a robust dataset for spatial-temporal analysis. The findings revealed significant spatial and spatial-temporal clusters of PTB cases across multiple districts, shedding light on the geographic distribution of the disease and its temporal dynamics. Conversely, one potential limitation is the reliance on basic management unit (BMU) reports for TB case notification data. BMU reports may not accurately reflect the true burden of TB in specific administrative catchment areas, as they may include cases from neighboring catchments or miss cases from their own catchment enrolled in other health facilities. This could lead to over- or underreporting of TB cases, potentially impacting the accuracy of the spatial-temporal analysis.

Li Huang et al. (2017) conducted a spatial-temporal cluster analysis of pulmonary tuberculosis (PTB) in Zhaotong, China, using town-level PTB registration data from 2011 to 2015. Robust statistical methods, including time series analysis, descriptive

analysis, and spatial and space-time scan statistics were administered to comprehensively explore PTB epidemiology. The analysis encompassed various epidemiological parameters, such as age, gender, treatment history, and geographic location, providing a detailed understanding of PTB incidence patterns. Spatial visualization techniques aided in identifying high-risk areas, while temporal analysis revealed seasonal variations in PTB incidence. However, the study's limitations, including its focus on a specific geographic area and time, potential data quality issues, and the absence of exploration of confounding factors, should be acknowledged.

Aceng et al. (2024) conducted a retrospective analysis spanning ten years to examine the spatial distribution and temporal trends of TB case notifications in Uganda. The study aimed to provide insights into the geographic spread of TB cases over time and identify any temporal trends in TB incidence. The study's long-term perspective, covering a decade of TB case notifications is recommendable. This extended timeframe allowed for the identification of trends and patterns in TB incidence over time. Furthermore, the spatial analysis conducted in the study provided valuable insights into the geographic distribution of TB cases across different regions of Uganda. By mapping TB case notifications, the study highlighted areas with higher TB burden, which can inform targeted interventions and resource allocation for TB control strategies. The integration of spatial and temporal analyses enhances the study's utility in identifying areas of persistent TB transmission and monitoring changes in TB incidence over time.

1.2 Problem statement

Understanding the spatial variation and underlying causes of TB prevalence is critical for effective epidemic management. Geospatial analytical methods, combined with advanced statistical models, are instrumental in detecting spatial and spatial-temporal clustering of infectious diseases. While previous studies have employed these methods, significant gaps remain in the biostatistical approaches used, particularly in the context of TB research in Northern Malawi.

Existing research has often focused on urban settings or has been limited in scope by short time frames and restricted geographical areas. For example, studies by Miandad

et al. (2014) in Karachi, Pakistan, and Bastida et al. (2012) in the State of Mexico have successfully identified spatial-temporal clusters of TB. However, these studies were conducted in urban settings over a brief period, which makes them deficient in detecting the broader, long-term spatial-temporal patterns necessary for rural areas with different socio-economic contexts. These studies primarily utilized descriptive spatial analysis techniques, which, while useful, do not fully leverage the potential of more advanced biostatistical models to uncover deeper insights into disease dynamics. In Malawi, studies such as those by Nyirenda et al. (2005) and Kirolos et al. (2021) have applied Geographic Information Systems and time-series analysis to TB data. However, these studies often lack the integration of robust spatial-temporal statistical models that can provide a more subtle understanding of the patterns and trends in TB incidence. For instance, Nyirenda et al. (2005) focused on the application of GIS for spatial patterns but did not employ advanced spatial-temporal clustering methods, thus missing the opportunity to identify dynamic changes over time.

Furthermore, Khundi et al. (2021) and Nightingale et al. (2021) provided valuable insights into TB determinants and testing practices in specific areas of Malawi. Yet, these studies were limited to urban settings or focused narrowly on community-level variations without applying comprehensive spatial-temporal statistical techniques that could capture broader regional patterns and trends. Their methodologies often relied on simple statistical analyses that did not account for the complexities of spatial-temporal interactions.

Geospatial analytical methods, combined with scan statistics, are instrumental in detecting spatial and spatial-temporal clustering of infectious diseases. However, existing TB program reports in many regions, including Northern Malawi, are often compiled, and reported quarterly at higher administrative units, lacking essential information such as population size, geographic coordinates, and temporal trends. This limitation impedes the accurate identification of hotspots and clusters using modern spatial-temporal techniques and approaches. The aggregation of TB case notifications at high administrative levels (e.g., district or region) obscures finer-scale spatial patterns that may be critical for identifying localized hotspots. High-resolution spatial data, including geographic coordinates of TB cases, are essential for precise spatial analysis. Without this, spatial clustering, and hotspot detection methods, such

as spatial scan statistics, cannot be effectively applied. Studies like those by Miandad et al. (2014) in Karachi and Bastida et al. (2012) in Mexico utilized detailed spatial data to identify TB clusters, but similar high-resolution data are often lacking in Northern Malawi.

The gap in the current literature is evident in the lack of application of advanced biostatistical methods, such as Space-Time Scan Statistics, which can provide a more rigorous analysis of spatial and temporal clustering of TB. This method allows for the detection and characterization of clusters not only in space but also in time, providing a dynamic view of disease spread and identifying high-risk periods and locations. By leveraging Space-Time Scan Statistics, this study aims to fill the gap in existing research by providing a detailed analysis of the spatial and temporal dynamics of TB in Northern Malawi. This approach will enable the identification of significant clusters and high-risk areas over an extended period, offering insights that are crucial for targeted intervention strategies and resource allocation.

In conclusion, the primary gap in the current literature lies in the insufficient application of advanced biostatistical methods to understand the spatial-temporal distribution of TB in Northern Malawi. By addressing this gap, this study not only enhances the understanding of TB epidemiology but also has significant implications for public health policy and the formulation of targeted and effective TB control strategies in the region.

1.3 Objectives

1.3.1 Main Objective

To utilize advanced biostatistical methods to investigate the spatial distribution of TB focusing on the presence of spatial-temporal clusters over an 8-year period in Northern Malawi.

1.3.2 Specific Objectives

- To utilize spatial and spatial-temporal models to identify significant spatial and spatial-temporal clusters of TB within the northern region of Malawi.

- Apply spatial-temporal statistical models to detect and characterize infectious disease cluster coverage and radius.
- Employ Spatial and Spatial – temporal Statistical methods to discern temporal trends and spatial patterns of TB across the northern region of Malawi.

1.4 Significance of the study

Limited research has been conducted on the spatial-temporal analysis of TB in Malawi hence lack of substantial evidence regarding the distribution patterns of TB in different settings. Understanding the spatial patterns and spatial-temporal variations of the disease, particularly in diverse geographic settings encompassing both urban and rural areas, holds significant potential for informing policy and decision-making processes in resource-constrained settings. This study aims to contribute to a comprehensive understanding of the distribution of TB cases across health facilities, while also examining cluster patterns to discern how TB cases are spatially and temporally dispersed. Such insights will assist public health authorities in predicting the potential risks posed to the general population in the event of a pandemic. Additionally, the study will provide valuable input into the general research database of the Malawi Spatial Data Platform, thereby facilitating future geographic information system studies and TB modelling. Consequently, this study serves as a foundational step towards investigating the spatial and temporal distribution patterns of TB cases in Northern Malawi

1.5 Thesis Structure

This thesis report is organized as follows: Chapter two contains a literature review that outlines a review of studies on spatial-temporal analysis of TB conducted across the countries. Chapter three outlines the methodological approach of the study, it describes the specific details of data, its sources, and the methods to be used for data analysis. It also describes spatial and temporal models used in the identification of clusters, hotspots, and time trends. Chapter four presents' results from the application of methods outlined in chapter three. Results are presented in tabular and figures to summarize findings. Chapter five discusses the results and Chapter 6 outlines main conclusions and recommendations from the study. A list of references used follows from chapter six. Appendices outline raw data and Sat scan analytical results.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter reviews the commonly applied Spatial- temporal models and their application in detecting spatial and spatial-temporal clustering of infectious diseases. It aims to delineate the gaps in knowledge that the current study aims to fill, discuss the assumptions and variations of each model, and provide detailed methods of estimation, including maximum likelihood and expectation-maximization algorithms.

2.2 Space-Time Scan Statistics model

Space-Time Scan Statistics uses likelihood ratio tests to identify clusters of events in space and time, facilitating the detection of disease hotspots (Kulldorff, 1997). The model assumes a Poisson distribution for disease events and tests the null hypothesis of spatial and temporal randomness. Variations may include different scan window shapes and sizes, allowing for flexibility in cluster detection. The interpretation of these statistics involves assessing the significance of identified clusters and understanding their implications for disease transmission dynamics. This is represented as follows:

$$D = \frac{1}{|P|} \sum_{p \in P} \frac{E_p (O_p - E_p)}{E_p}$$

Where D is the test statistic representing the likelihood ratio, P is the set of all possible space-time cylinders, O_p is the observed number of cases within the space-time cylinder p and E_p is the expected number of cases within the space-time cylinder p , calculated based on the null hypothesis of spatial and temporal randomness (Kulldorff, 1997).

2.2.1 Method of estimation

The likelihood ratio test compares the observed and expected number of cases within each space-time cylinder. A high value of D indicates a higher-than-expected number of cases within a particular space-time cylinder, suggesting the presence of a disease cluster. Interpreting the results involves assessing the statistical significance of identified clusters. Significant clusters indicate areas where the observed number of cases deviates significantly from the expected number, suggesting spatial-temporal aggregation of disease (Kulldorff, 1997).

Maximum likelihood estimation for the Space-Time Scan Statistics model involves maximizing the likelihood ratio over all possible space-time cylinders. The likelihood ratio for a given cylinder p is calculated as follows:

$$L(O_p, E_p) = \left(\frac{O_p}{E_p}\right)^{O_p} \left(\frac{N - O_p}{N - E_p}\right)^{N - O_p}$$

Where N is the total number of cases in the study region and period and O_p and E_p are as defined above.

The likelihood function is evaluated for each potential cluster, and the maximum likelihood ratio is identified. This maximum value represents the most likely cluster of disease events, and its statistical significance is assessed using Monte Carlo simulations to generate a distribution of likelihood ratios under the null hypothesis (Kulldorff, 1997).

In cases where the model includes latent variables representing unobserved heterogeneity or missing data, the Expectation-Maximization (EM) algorithm can be applied. The EM algorithm is an iterative method for finding maximum likelihood estimates of parameters in models with latent variables. It involves two main steps: the Expectation (E) step and the Maximization (M) step. In the E-step, the expected value of the log-likelihood function is calculated with respect to the current estimates of the latent variables. This is represented as

$$Q(\theta|\theta^{(k)}) = E_{Z|X, \theta^{(k)}} [\log L(\theta; X, Z)]$$

Where $\theta^{(k)}$ represents the current estimates of the parameters, X denotes the observed data, and Z represents the latent variables and $\log L(\theta; X, Z)$ is the complete-data log-likelihood function.

In the M-step, the expected log-likelihood function obtained in the E-step is maximized to update the parameter estimates, represented as

$$\theta^{(k+1)} = \arg \max_{\theta} Q(\theta|\theta^{(k)})$$

These steps are iterated until convergence, meaning the change in the parameter estimates between iterations falls below a predefined threshold. Once the parameters are estimated using either MLE or the EM algorithm, interpreting the results involves assessing the statistical significance of the identified clusters. Significant clusters are those where the observed number of cases deviates significantly from the expected number, suggesting areas of spatial-temporal aggregation of disease (Dempster, Laird, & Rubin, 1977).

2.3 Spatiotemporal point process models

Spatiotemporal point process models aim to characterize the intensity functions of disease occurrence across both space and time. These models, as described by Diggle (2013), provide a detailed framework for understanding disease dynamics by capturing baseline intensity, spatial and temporal trend components, and residual variation. The general form of a spatiotemporal point process model is given by the equation:

$$\lambda(s, t) = \mu(s, t) + \alpha(s) + \beta(t) + \gamma(s, t)$$

In this equation $\lambda(s, t)$ represents the intensity function of the point process at location s and time t , $\mu(s, t)$ denotes the baseline intensity, capturing the underlying average rate of disease events, $\alpha(s)$ and $\beta(t)$ represent spatial and temporal trend components, respectively, accounting for systematic variations in disease occurrence across space

and time and $\gamma(s, t)$ captures the residual spatiotemporal variation not explained by the baseline intensity or trend components.

The key assumptions of spatiotemporal point process models include the assumption of stationarity, similarly to STARIMA models, and the assumption of independence between spatial and temporal components. Variations of these models include different specifications for trend components and methods for incorporating spatial and temporal dependence. This includes Poisson Point Process which assumes that events occur independently and uniformly over space and time and the cox Process which incorporates a stochastic process to model varying intensity.

2.3.1 Method of estimation

Parameter estimation for spatiotemporal point process models typically involves maximum likelihood or Bayesian methods. These methods optimize the likelihood function given the observed data, accounting for both spatial and temporal dependencies. Interpretation of the model parameters involves understanding how each component contributes to the overall intensity of disease events. Diggle (2013).

For MLE, the log-likelihood function is maximized:

$$\log L(\theta; Y) = \sum_{i=1}^n \log \lambda(s_i, t_i; \theta) - \int_{S \times T} \lambda(s, t; \theta) ds dt$$

Where θ represents the vector of model parameters, S and T are the spatial and temporal domains.

The Expectation-Maximization algorithm for the Space-Time Scan Statistics model can be adapted for use with the spatiotemporal point process model, particularly in cases where there is incomplete data or latent variables that need to be estimated. Once the parameters are estimated using either MLE or the EM algorithm, the interpretation involves understanding how each component contributes to the overall intensity of disease events. The baseline intensity $\mu(s, t)$ reflects the underlying average rate of disease occurrence, providing a baseline measure against which other components are compared. The spatial trend $\alpha(s)$ indicates systematic spatial

variations in disease risk, helping identify areas with higher or lower disease rates. The temporal trend $\beta(t)$ captures systematic temporal variations, revealing periods with increased or decreased disease activity. The residual variation $\gamma(s, t)$ accounts for remaining spatiotemporal variation, identifying deviations from the baseline and trend components (Diggle, 2013).

2.4 Space-Time Autoregressive Integrated Moving Average (STARIMA) model

In the context of spatial-temporal modelling, various methodologies are to analyse the intricate dynamics of disease spread and distribution. One such approach is the Space-Time Autoregressive Integrated Moving Average (STARIMA) model, which incorporates both spatial and temporal dependencies into the analysis through autoregressive and moving average components.

$$Y_{t,s} = \alpha + \beta Y_{t-1,s} + \gamma Y_{t,s-1} + \delta Y_{t-1,s-1} + \epsilon_{t,s}$$

Where $Y_{t,s}$ denotes the response variable at time t and space s , while $\alpha, \beta, \gamma, \delta$ represent model coefficients, and $\epsilon_{t,s}$ is the error term (Cressie and Wikle, 2015).

The STARIMA model assumes stationarity in both spatial and temporal dimensions, implying that the statistical properties of the data remain constant over time and space. Variations of the STARIMA model include variations in the order of autoregressive and moving average components, allowing for flexibility in capturing different patterns of disease transmission.

2.4.1 Method of estimation

Estimation of the STARIMA model parameters typically involves maximum likelihood estimation, where the parameters are optimized to maximize the likelihood function given the observed data. The estimation process includes fitting the model to the data using iterative algorithms, such as the Kalman filter, and interpreting the estimated coefficients to understand their impact on disease dynamics (Cressie and Wikle, 2015).

The likelihood function for the STARIMA model can be expressed as:

$$L(\Theta; Y) = \prod_{t=1}^T \prod_{s=1}^S f(Y_{t,s} | \Theta)$$

Where Θ represents the vector of model parameters and f is the probability density function of $Y_{t,s}$.

The Expectation-Maximization algorithm for the STARIMA Model can be adapted for use with the spatiotemporal point process model. Once the parameters are estimated using either MLE or the EM algorithm, the interpretation involves understanding how each component contributes to the overall spatial-temporal dynamics of the response variable. The coefficient α represents the intercept term, while β , γ , and δ capture the temporal and spatial dependencies. The error term $\varepsilon_{t,s}$ accounts for random noise in the model.

2.5 Space-Time Generalized Linear Mixed Models (ST-GLMMs)

Space-Time Generalized Linear Mixed Models integrate generalized linear models with spatial and temporal random effects, offering a comprehensive framework for analysing spatial-temporal data (Diggle et al., 2013). ST-GLMMs assume a linear relationship between covariates and the response variable, with spatial and temporal random effects capturing unobserved heterogeneity and spatial-temporal autocorrelation. Variations may include different distributions for the random effects and alternative specifications for the fixed effects. The general form of the ST-GLMM can be expressed as:

$$y_{ijk} = X\beta + Z_i u_{ij} + W_j v_{ij} + \epsilon_{ijk}$$

Where y_{ijk} represents the observed response variable at location i , time j , and individual k ; X is the design matrix for the fixed effects β ; Z_i and W_j are the design matrices for the spatial and temporal random effects, respectively; u_{ij} and v_{ij} are the spatial and temporal random effects, assumed to follow a multivariate normal distribution with mean zero and covariance matrices Σ_u and Σ_v , respectively and ϵ_{ijk} represents the error term, assumed to be normally distributed with mean zero and variance σ^2 (Diggle et al., 2013).

2.5.1 Method of estimation

The likelihood function for the ST-GLMM is constructed based on the joint distribution of the observed data, given the fixed and random effects. The log-likelihood function for a given set of parameters $\theta = (\beta, \Sigma_u, \Sigma_v, \sigma^2)$ can be expressed as:

$$\log L(\theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (Y - X\beta)' \Sigma^{-1} (Y - X\beta)$$

Where n is the total number of observations, Y is the vector of observed responses, X is the design matrix for fixed effects, Σ is the combined covariance matrix for the random effects and the error term and Σ^{-1} denotes the inverse of Σ .

The parameters β , Σ_u , Σ_v , and σ^2 are estimated from the data using iterative estimation techniques such as the EM algorithm or MCMC methods. Once estimated, the fixed effects β are interpreted to understand the relationship between covariates and the response variable, while the spatial and temporal random effects provide insights into the spatial and temporal variability of the response variable, respectively (Diggle et al., 2013).

2.6 Rationale for Using the Space-Time Scan Statistics Model in the Spatial-Temporal Analysis of TB Case Notification in Northern Malawi

Among the various spatial-temporal models discussed, the Space-Time Scan Statistics model developed by Kulldorff emerged as the most suitable choice for this study on the spatial-temporal analysis of TB case notification in Northern Malawi. Several advantages rendered this model particularly appropriate for achieving the research objectives. The model effectively identified clusters and examined the spatial-temporal aspects of TB case notifications, addressing a significant gap in understanding within Northern Malawi's context.

Firstly, the Space-Time Scan Statistics model is specifically designed to detect clusters of disease events in both space and time, making it particularly well-suited for identifying disease hotspots and spatial-temporal patterns of TB transmission (Kulldorff, 1997). Unlike models that focus solely on spatial or temporal dimensions,

the Space-Time Scan Statistics model considers both simultaneously. This dual consideration enhances the sensitivity of cluster detection, allowing for the identification of significant disease clusters that may be overlooked by other methods. For example, while the STARIMA model captures spatial and temporal dependencies through autoregressive and moving average components, it may not effectively isolate specific clusters or hotspots as the Space-Time Scan Statistics model does.

Moreover, the Space-Time Scan Statistics model offers unparalleled flexibility in defining the size, shape, and duration of potential clusters. This adaptability allows researchers to tailor the analysis to the specific characteristics of the data and the epidemiological context (Kulldorff, 1997). Such flexibility is particularly crucial in the context of TB transmission, where clusters can vary significantly in terms of their geographic spread and duration. In contrast, models like the ST-GLMMs, while robust in incorporating random effects to account for spatial and temporal variability, do not inherently provide the same level of flexibility in defining and detecting clusters.

The Space-Time Scan Statistics model provides rigorous statistical measures of significance for identified clusters. This capability allows researchers to assess the likelihood that observed clusters are due to random variation rather than actual patterns of disease transmission (Kulldorff, 1997). The statistical rigor inherent in the model enhances the reliability of cluster detection results, providing confidence in the interpretation of identified clusters as true disease hotspots. While Spatiotemporal point process models can characterize variations in disease occurrence through intensity functions, they may not offer the same level of statistical testing for cluster significance, which is critical for public health decision-making.

Furthermore, the implementation of the Space-Time Scan Statistics model through software tools such as SaTScan facilitates efficient and user-friendly analysis of large spatial-temporal datasets. This practical advantage streamlines the process of cluster detection and interpretation, making it accessible for researchers and public health practitioners (Kulldorff, 1997). In contrast, the STARIMA model and ST-GLMMs often require more complex statistical software and expertise, which can be a barrier for routine public health surveillance and intervention planning.

The application of the Space-Time Scan Statistics model addresses significant gaps in the current understanding of TB transmission dynamics in Northern Malawi. While previous studies have utilized various spatial-temporal models to analyse disease patterns, many have not effectively combined spatial and temporal dimensions in a way that identifies specific clusters and hotspots. The flexibility and statistical rigor of the Space-Time Scan Statistics model enable a more nuanced analysis of TB case notifications, revealing patterns and trends that other models may miss.

When compared to Spatiotemporal point process models, STARIMA, and ST-GLMMs, the Space-Time Scan Statistics model stands out for several reasons. Spatiotemporal point process models, for instance, focus on the intensity functions of point patterns across space and time, which is useful for characterizing disease occurrence but less effective for explicit cluster detection (Diggle, 2013). The STARIMA model, while incorporating spatial and temporal dependencies, is primarily geared towards forecasting rather than identifying specific clusters (Cressie & Wikle, 2015). ST-GLMMs offer robust modelling of spatial and temporal variability through random effects but do not inherently focus on cluster detection (Diggle et al., 2013).

In summary, the Space-Time Scan Statistics model provides a robust, flexible, and statistically rigorous framework for analysing the spatial-temporal dynamics of TB case notifications in Northern Malawi. Its ability to detect clusters of disease events in both space and time, combined with its flexibility in defining clusters and rigorous statistical testing, made it an ideal choice for this study. The model's implementation through user-friendly software further enhanced its practical applicability, facilitating efficient analysis and interpretation.

2.7 Cases where spatial-temporal methods have been applied to data.

Various studies have demonstrated the utility of spatial-temporal methods in different contexts. Smith et al. (2018) applied spatial cluster detection methods to identify hotspots of dengue fever transmission in urban areas of Brazil. Their study effectively showcased the use of spatial methods in pinpointing areas at elevated risk of dengue outbreaks, thereby informing targeted vector control interventions. By utilizing a purely spatial cluster detection method, Smith and colleagues were able to identify

areas with high incidence rates, providing actionable insights for public health officials. This approach was instrumental in deploying targeted vector control measures, thus reducing the incidence of dengue in identified hotspots.

Similarly, Chen et al. (2020) utilized time-series analysis to examine temporal trends in cancer incidence rates in a population-based cancer registry dataset. Their study revealed significant temporal variations in cancer incidence over time, highlighting the importance of ongoing surveillance and monitoring efforts. By applying advanced time-series methods, Chen and colleagues detected subtle trends and patterns in cancer incidence that traditional methods might have overlooked. Their findings underscored the necessity for continuous monitoring and early intervention strategies to manage and mitigate cancer risks effectively.

Furthermore, Zhang et al. (2019) employed Poisson regression models to assess the association between road traffic accidents and various contributing factors, such as road characteristics, weather conditions, and driver behaviour. Their study identified significant predictors of road traffic accidents and provided valuable insights for road safety planning and interventions. Using a combination of spatial and temporal data, Zhang and colleagues were able to pinpoint high-risk areas and times for road traffic accidents. This information proved crucial for informing more effective preventive measures and improving road safety infrastructure.

In addition, Lee et al. (2017) conducted a spatiotemporal analysis of influenza spread in South Korea using STARIMA models. They identified significant clusters and temporal trends of influenza cases, providing insights into the effectiveness of vaccination campaigns and other control measures. Lee and colleagues' analysis highlighted the importance of considering both spatial and temporal dimensions in disease surveillance and control, thereby enhancing the ability to respond to and manage influenza outbreaks more effectively.

Moreover, Jones et al. (2021) applied ST-GLMMs to examine the spread of COVID-19 in urban and rural areas of the United States. Their study revealed significant differences in transmission dynamics between urban and rural settings, emphasizing the need for tailored public health strategies. By using ST-GLMMs, Jones and

colleagues gained a nuanced understanding of how various factors influenced COVID-19 spread over time and across different regions. Their findings informed public health policies, helping to design more targeted and effective intervention strategies.

Furthermore, Garcia et al. (2022) used space-time scan statistics to analyse the spread of the Zika virus in Central America. Their study identified spatiotemporal clusters of Zika virus transmission, which provided critical information for public health authorities to implement targeted control measures. Garcia and colleagues' application of space-time scan statistics allowed for the detection of high-risk areas and periods, facilitating more efficient allocation of resources to combat the outbreak.

Additionally, Kim et al. (2023) applied Bayesian hierarchical models to study the spatiotemporal dynamics of air pollution and its health impacts in urban centers across East Asia. Their research revealed significant spatial and temporal variations in air pollution levels and associated health outcomes. By utilizing Bayesian hierarchical models, Kim and colleagues were able to account for complex dependencies and uncertainties in their data, providing robust estimates of pollution impacts. Their findings have significant implications for urban planning and public health policies aimed at reducing air pollution and mitigating its adverse health effects.

Lastly, Ahmed et al. (2023) utilized spatial-temporal kernel density estimation to analyse the distribution and evolution of cholera outbreaks in Sub-Saharan Africa. Their study demonstrated the effectiveness of this method in identifying emerging hotspots and tracking the spread of cholera over time. Ahmed and colleagues' work highlighted the importance of continuous spatial-temporal monitoring in managing infectious disease outbreaks and improving public health responses.

By applying similar spatial-temporal methods to TB data in the northern region of Malawi, this thesis aimed to advance the understanding of the spatial-temporal analysis of TB and inform evidence-based TB control strategies in the region. Using advanced biostatistical models like Space-Time Scan Statistics, this study provides detailed insights into the spatial and temporal dynamics of TB, enabling more targeted and effective public health interventions.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter focuses on the methodology adopted to determine the spatial and spatial-temporal trends of TB case notifications in northern Malawi. It depicts the data sources and collection procedures, the data analysis package, and the spatial-temporal models utilized.

3.2. Data Collection and Analysis

3.2.1 Data Sources and Collection Procedures

This study investigated the space-time dynamics of TB clusters in Northern Malawi from 2013 to 2020, utilizing data collected from secondary sources, primarily the hospital records database of all health facilities in the region. The study was conducted across 61 health centres situated in six districts within the northern region. Mzimba district, being the largest, accounted for 25 health facilities, followed by Karonga (10), Chitipa (8), Nkhatabay (8), Rumphi (7), and Likoma (1). However, upon analysis, only 59 health facilities were deemed to have significant TB case records worthy of study.

The TB data collection process followed a structured approach coordinated by District TB Officers, involving the maintenance of three types of registers at the district level: a chronic cough register, a laboratory TB register, and a TB treatment register (World Health Organization, 2020). These registers captured details of patients undergoing TB diagnosis and treatment, including sputum smear results and treatment initiation. The collected data were collated quarterly by District TB Officers, forwarded to Regional TB Officers, and then compiled into regional summaries transmitted to the Central Unit of the National TB Program (NTP). For this study, data were obtained

Population data for the study area were derived from the Malawi Population and Household Census (MPHC) reports of 2008 and 2018. Accounting for the decadal nature of MPHCs, an annual population growth rate of 2.7% was applied to estimate the annual population size per district (Souza et al., 2007). Additionally, SaTScan software was utilized to impute population data at one or more specific 'census times' and perform linear interpolation based on population sizes at preceding and succeeding census times. This approach enabled the estimation of population sizes for specific locations and time periods, crucial for spatial-temporal analysis of TB clusters (Kulldorff, 1997).



3.2.2 Poisson Model for Spatial-Temporal Cluster Detection in Kulldorff's SaTScan software

The SaTScan software package was instrumental in conducting spatial-temporal cluster analysis of TB case notification. Utilizing the Poisson probability model within SaTScan, the study identified clusters of elevated TB incidence rates across different spatial and temporal dimensions of events, within specific spatial and temporal windows.

The Poisson probability model within SaTScan is utilized for detecting spatial-temporal clusters by assessing the count of disease events within specific spatial and temporal windows. This model assumes that the number of cases in each area follows a Poisson distribution, which is appropriate for rare events like TB notifications (Kulldorff, 1997).

The likelihood function for the Poisson model is defined as follows:

$$L(C) = \prod_{i \in C} \frac{(\lambda_i t_i)^{c_i} e^{-\lambda_i t_i}}{c_i!}$$

Where $L(C)$ is the likelihood of cluster C , c_i is the observed number of cases in location i within the cluster, λ_i is the expected number of cases in location i , which is calculated based on the population at risk, t_i is the population or time at risk in location i .

The expected number of cases λ_i is determined under the null hypothesis of no cluster, assuming a homogeneous Poisson process:

$$\lambda_i = \frac{\sum_{i=1}^n c_i}{\sum_{i=1}^n t_i} \times t_i$$

Where n is the total number of spatial and temporal units considered, $\sum_{i=1}^n c_i$ is the total number of observed cases across all units and $\sum_{i=1}^n t_i$ is the total population or time at risk across all units.

To identify clusters, SaTScan scans the study area using a cylindrical window with a circular geographic base and a height corresponding to time. The window is moved systematically over the study region, considering all possible geographic locations and sizes, as well as different temporal extents. For each window position and size, the likelihood ratio is calculated by comparing the likelihood of observing the given number of cases within the window to the likelihood under the null hypothesis (Kulldorff, 1997).

The likelihood ratio LR for a specific window was given by:

$$LR = \frac{L(C)}{L(\text{null})}$$

Where $L(\text{null})$ is the likelihood under the null hypothesis, calculated as:

$$L(\text{null}) = \prod_{i \in C} \frac{(\bar{\lambda} t_i)^{c_i} e^{-\bar{\lambda} t_i}}{c_i!}, \text{ where } \bar{\lambda} \text{ is the overall expected rate of cases.}$$

Clusters were identified as regions with the maximum likelihood ratios, and the statistical significance of these clusters was assessed using Monte Carlo simulations. This process involved generating many random datasets under the null hypothesis and comparing the observed likelihood ratios to the distribution of likelihood ratios from the simulated datasets.

3.2.3 Ethical considerations

The collection and use of secondary data for this study adhered to ethical guidelines and regulations governing research involving human subjects. Before conducting the study, ethical approval and clearance were obtained from the Ministry of Health and the National Control TB program. This ensured that the study adhered to the recommended ethical parameters and safeguarded the rights and well-being of individuals involved in the data collection process. As the study utilized secondary data obtained from the Central TB Unit, the data had been collected under appropriate ethical guidelines. To ensure compliance and ethical integrity, the study sought approval from the original data generators before utilizing and publishing the data.

This step ensured that the study maintained ethical standards throughout its research process, demonstrating respect for the rights and confidentiality of individuals whose information was included in the dataset.

3.3 Statistical Methods for Analysing Spatial-Temporal Dynamics of TB Case

Notification Data

3.3.1 Spatial Model: Spatial scan statistic

The spatial model used in this study was the spatial scan statistic, a powerful tool for detecting spatial clusters of disease cases. This model allows for the identification of areas with significantly higher or lower TB case notification rates compared to the surrounding areas. The spatial scan statistic works by scanning a window across the study area and comparing the observed number of cases within the window to the expected number of cases under the null hypothesis of spatial randomness. The Poisson model is suitable for this analysis because it assumes that the number of cases follows a Poisson distribution, which is appropriate for rare events such as TB notifications (Kulldorff, 1997).

The likelihood function for the spatial scan statistic was given by:

$$\Lambda = \frac{L(A)}{L_0}$$

Where Λ is the likelihood ratio test statistic, $L(A)$ is the likelihood of observing the data within the potential cluster A , L_0 is the likelihood of observing the data under the null hypothesis.

The potential cluster A was defined by a circular window that varied in size and location across the study area which is the Northern Region of Malawi. The maximum likelihood ratio across all possible cluster configurations was calculated, and its significance was assessed using Monte Carlo hypothesis testing. Using the Poisson model, the likelihood $L(A)$ for the cluster A was given by:

$$L(A) = \prod_{i \in A} \left(\frac{(\lambda_i t_i)^{c_i} e^{-\lambda_i t_i}}{c_i!} \right)$$

Where λ_i is the expected number of cases at location i , calculated based on the population at risk, t_i is the population or time at risk in location i , c_i is the observed number of cases at location i .

The expected number of cases λ_i under the null hypothesis was:

$$\lambda_i = \frac{\sum_{i=1}^n c_i}{\sum_{i=1}^n t_i} \times t_i$$

Where n is the total number of spatial and temporal units considered, $\sum_{i=1}^n c_i$ is the total number of observed cases across all units and $\sum_{i=1}^n t_i$ is the total population or time at risk across all units.

The likelihood ratio LR for a specific window was given by:

$$LR = \frac{L(A)}{L(\text{null})}$$

Where $L(\text{null})$ is the likelihood under the null hypothesis, calculated as:

$$L(\text{null}) = \prod_{i \in A} \frac{(\bar{\lambda} t_i)^{c_i} e^{-\bar{\lambda} t_i}}{c_i!}, \text{ where } \bar{\lambda} \text{ is the overall expected rate of cases.}$$

Statistical tests determined the significance of estimated parameters. For covariates, p-values indicated whether there were significant associations with the outcome variable.

3.3.2 Spatial – Temporal Model: Space-time scan statistic

The spatial-temporal model used in this study was the space-time scan statistic, an extension of the spatial scan statistic that incorporates temporal dimensions. The space-time scan statistic allows for the detection of clusters of disease cases that are both spatially and temporally concentrated. This model evaluates the observed

number of cases within a space-time window against the expected number under the null hypothesis of spatial-temporal randomness (Kulldorff, 1997).

The model:

$$\Lambda(s, t) = \frac{\sum_{s_i \in A(s, t)} Y(s_i, t)}{\sum_{s_i \in A(s, t)} E(s_i, t)}$$

Where $\Lambda(s, t)$ represents the likelihood ratio for the space-time window (s, t) , $A(s, t)$ denotes the space-time window, $Y(s_i, t)$ is the observed number of cases in the window, $E(s_i, t)$ is the expected number of cases under the null hypothesis.

The space-time scan statistic was estimated using likelihood ratio tests, identical to the spatial scan statistic. The maximum likelihood ratio across all possible space-time windows was determined:

The Likelihood function for the space-time scan statistic was given by:

$$\Lambda = \frac{L(S, T)}{L_0}$$

Where Λ is the likelihood ratio test statistic, $L(S, T)$ is the likelihood of observing the data within the space-time window $S \times T$, L_0 is the likelihood of observing the data under the null hypothesis.

The space-time window $S \times T$ defined both spatial and temporal dimensions and varied in size and location across the Health Centres in the Northern Region of Malawi from 2013 to 2020. The maximum likelihood ratio across all possible space-time windows was then calculated, and its significance was assessed using permutation tests.

Using the Poisson model, the likelihood $L(S, T)$ within the space-time window $S \times T$ was expressed as:

$$L(S, T) = \prod_{i \in S, t \in T} \left(\frac{(\lambda_{it} t_i)^{c_{it}} e^{-\lambda_{it} t_i}}{c_{it}!} \right)$$

Where λ_{it} is the expected number of cases at location i and time t , calculated based on the population at risk, t_i is the population or time at risk in location i , c_{it} is the observed number of cases at location i and t .

The expected number of cases λ_{it} under the null hypothesis was:

$$\lambda_i = \frac{\sum_{i=1}^n \sum_{t=1}^T c_{it}}{\sum_{i=1}^n \sum_{t=1}^T t_i} \times t_i$$

Where n is the total number of spatial and temporal units considered, T is the total number of temporal units considered, $\sum_{i=1}^n \sum_{t=1}^T c_{it}$ is the total number of observed cases across all spatial and temporal units and $\sum_{i=1}^n \sum_{t=1}^T t_i$ is the total population or time at risk across all spatial and temporal units.

The likelihood ratio LR for a specific space-time window was given by:

$$LR = \frac{L(S, T)}{L(\text{null})}$$

Where $L(\text{null})$ is the likelihood under the null hypothesis, calculated as:

$$L(\text{null}) = \prod_{i \in S, t \in T} \frac{(\bar{\lambda} t_i)^{c_{it}} e^{-\bar{\lambda} t_i}}{c_{it}!}, \text{ where } \bar{\lambda} \text{ is the overall expected rate of cases.}$$

The expected number of cases λ_{it} was estimated using population at risk and disease rates for each Health Centre and period. The space-time window was moved systematically across the study area and time-period, varying in size and location, to identify the space-time configuration that maximizes the likelihood ratio test statistic.

Statistical tests assessed the significance of estimated parameters. Covariate coefficients indicated the influence of predictors on the outcome variable over time and space, while space-time interaction terms captured how this influence varied across different locations and time periods. Significant covariate coefficients revealed factors associated with changes in the outcome variable over time and space. Space-time interaction terms indicated whether the effects of covariates vary across different

spatial-temporal contexts, highlighting areas and periods with distinct patterns of disease incidence.

CHAPTER 4

RESEARCH RESULTS

4.1 Introduction

This chapter presents key research findings, and interpretation of results gathered from the study. Thus, the chapter presents data through tables, figures, and uses statistical models to analyse data where among other findings others focuses on demographic characteristics, distribution of TB cases, cluster coverage, spatial and temporal distribution of cases, TB hotspots and clusters in Northern Malawi.

4.2 Results of the study

4.2.1 District TB case notification

Key findings depict a total of 12,324 TB case counts were recorded, the highest recorded year being 2019 with 2098 cases against the lowest in 2020 at 1099 cases. According to study findings, Likoma District recorded the lowest TB case count at 38, Mzimba North district had the largest number of reported cases at 5447. Looking at the case distribution against the population proportion, Rumphi district reported the highest number of cases (41%), followed by Mzimba North (26%).

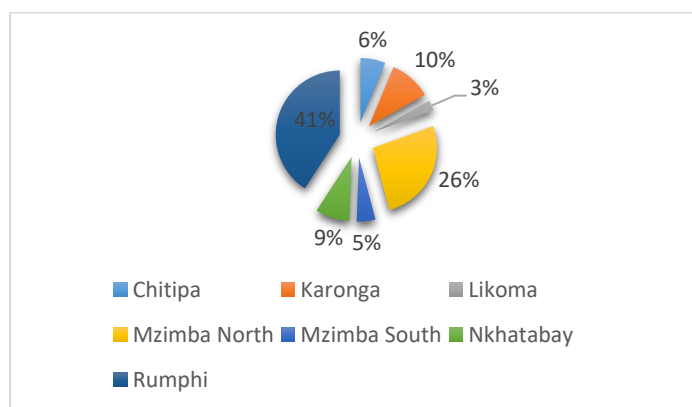


Figure 2: District TB infection distribution 2013 – 2020.

4.2.2 Gender and age distribution of TB Cases

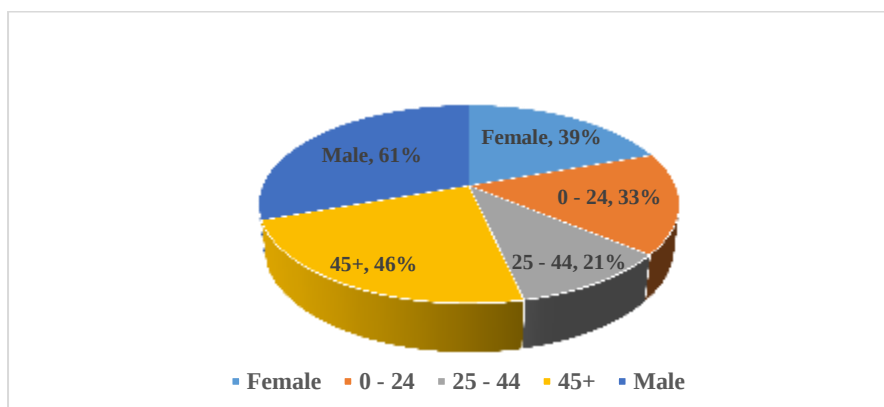


Figure 3: Gender and age group distributions among reported cases.

The study outcomes point out that there were more males infected by TB within the research period. A total of 7524 cases against 4800 cases for males and females were recorded representing a 61% and 39% respectively.

The study breaks down reported TB cases into three age groups: 0-24 representing children and adolescents, 25-44 representing youth and young adults, and 45+ representing middle-aged and elderly individuals. The findings indicate that 46% of reported cases were from the youth and young adults, followed by the middle-aged and elderly, with adolescents and children accounting for only 21%. These highlights varying levels of exposure and prevalence rates.

4.2.3 Trend of TB Infection Rates

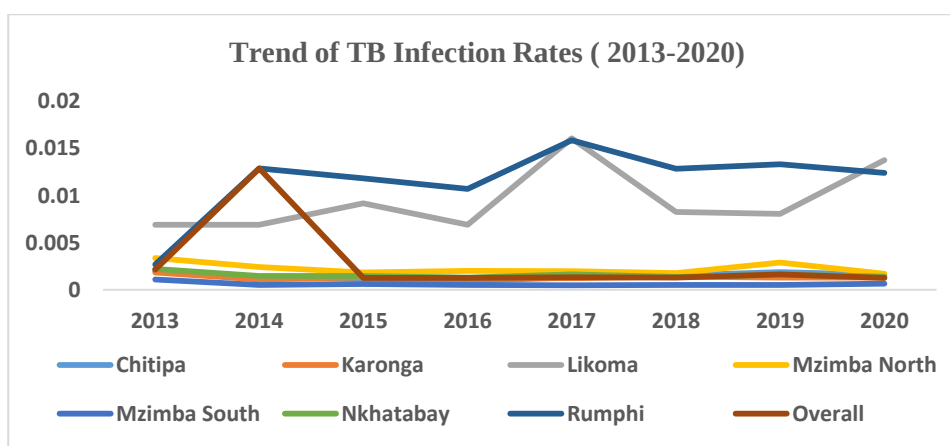


Figure 4: TB infection Rate trends (2013-2020)

The overall trend in TB infection rates from 2013 to 2020 showed some fluctuations but tended to remain relatively stable. There was a slight decrease in the overall infection rate from 2013 to 2020. Some districts showcased a decreasing trend over the years, while others showing fluctuations or increases. Likoma consistently showed higher infection rates compared to other districts, especially in 2017 and 2020. Rumphi experienced a significant spike in 2014, followed by a gradual decrease in infection rates over the subsequent years. Chitipa, Karonga, Mzimba North, Mzimba South, and Nkhatabay generally had lower infection rates compared to Likoma and Rumphi.

The data also indicated that there were fluctuations in TB infection rates from year to year within each district. Some years showed increases in infection rates, while others show decreases. Likoma experienced a notable increase in infection rates in 2017, followed by a decrease in 2018 and then another increase in 2020 and Rumphi saw a sharp increase in 2014, followed by a gradual decrease in subsequent years. Some districts, such as Chitipa, Karonga, Mzimba North, Mzimba South, and Nkhatabay, indicated relatively stable infection rates throughout the years, with minor fluctuations.

4.2.4 Seasonal distribution of TB Cases

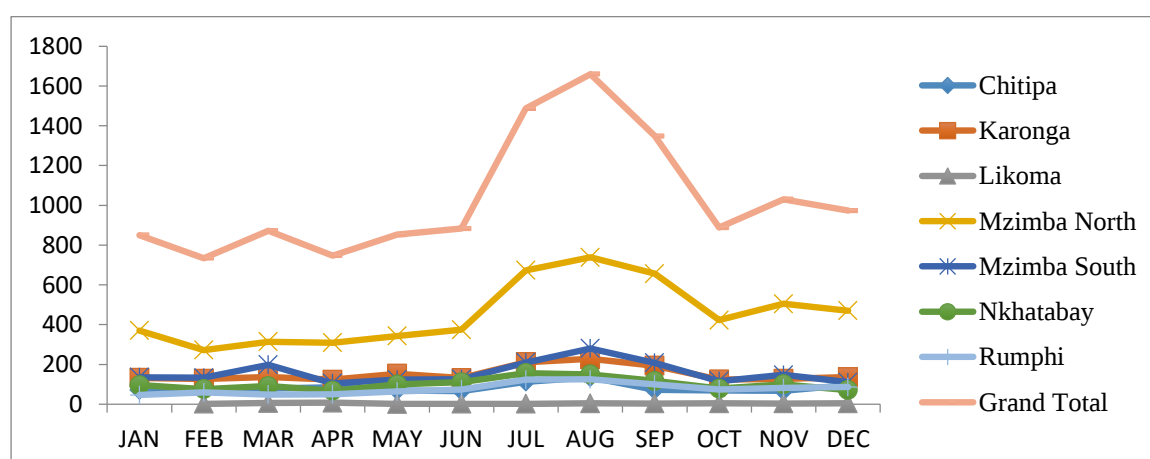


Figure 5: Monthly distributions of TB cases.

Looking at the overall data from 2013 to 2020, August emerges as the month with the highest recorded number of TB cases in the six districts, with an average of 1660

cases, a standard deviation of 4.26, and a variance of 20.76. The general trend shows a sharp decline of cases in the month of September, through October, however, the lowest registering month is February. However, studying the variations in cases across the months of a disease goes beyond the number of cases within a month.

The quarterly variation of the Tb diseases by considering the four quarters within each year was analysed. The third quarter (July, August, and September) registers the most cases with the second quarter having the least.

4.2.5 TB Hotspot and Cluster identification

Seven significant TB clusters were identified using spatial and space-time Model by executing spatial variation in temporal trends analysis. A cluster was considered statistically significant when its test statistic is greater than the Gumbel Critical Values (GCV) and Standard Monte Carlo Critical Values (SMCCV). Thus at 0.0001 significance level, a cluster was considered if its critical value was greater than 5.472276 GCV and 4.874990 SMCCV. A population size of 1,091,311 averaged over the 8-year period was considered from which a total number of 12, 324 cases were detected. The study detected a mean average of annual cases of 1.4 per 100,000 across the district. The study further analysed the cluster size and spread. The study utilized the Discrete Poisson probability model in SatScan, which uses population, GPS coordinates and TB cases as input to develop spatial and space-time clusters. To replicate the clusters and hotspots, a retrospective Space-Time analysis model for clusters with high rates using the Space-Time Permutation model was used. The 7 clusters identified across the northern region are summarized below and the outputs are showcased in appendix 2.

Table 1: Retrospective Spatial- Temporal analysis model estimates

Cluster	Radius (Km)	Population	Number of cases	Expected cases	Annual cases/100000	Relative risk	Time trend	P-value
Cluster 1	130	553,221	756	645.26	1.7	1.18	15.14	0.001
Cluster 2	120	48,339	46	5.58	11.6	8.28	12.9	0.001
Cluster 3	68.29	236,764	1073	277.78	5.5	4.14	-15.28	0.001
Cluster 4	48.29	509,863	4861	5723.3	1.2	0.75	-5.24	0.001
Cluster 5	33.93	209,863	892	765.3	1.17	5.41	-4.98	6.30E-10
Cluster 6	18	560,904	130	6.72	1.94	1.45	-6.3	
Cluster 7	17	67,986	879	72.52	1.2	2.81	-3.8	0.726
							-1.08	

4.2.5.1 Cluster size and location.

The study applied the spatial-temporal models to determine the size, jurisdiction, and extent of each cluster to understand their composition and distribution. This analysis was crucial for identifying patterns and assessing the significance of each cluster. In terms of the number of health centres within each cluster, the study found that Clusters 1, 6, and 7 have one health centre each, compared to clusters 2, 3, 4, and 5, which consist of more than one health centre or hospitals.

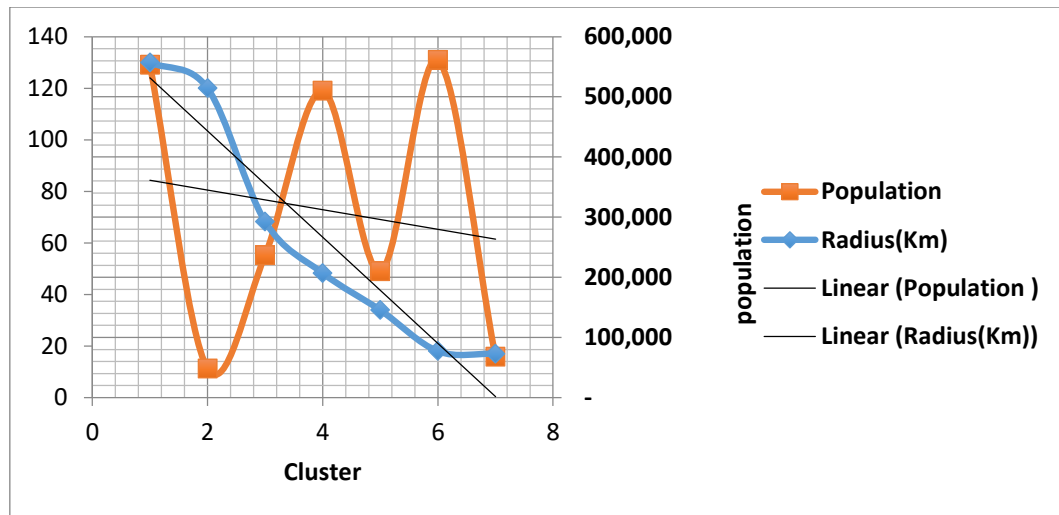


Figure 7: Cluster size and population coverage.

The cluster hotspot with the largest radius is cluster 1 which is surrounded by an average of 533,221-population size. Cluster 7 has the lowest radius of 17 Km, surrounding a population of 67,989 and encompasses areas surrounding David Gordon Memorial Hospital in Rumphi District.

4.2.5.2 TB clusters cases and expectations

Using the Spatial – Temporal Model, the study established the number of cases per clusters, expected cases and annual cases per 100,000 to analyse the extent to which each cluster has registered Tb cases and project likelihood of future infections. The highest number of cases was registered in cluster 4 at 4861 cases, while the lowest number of cases was registered in cluster 6 at 1309. However, the extent of TB case infestation can't be looked at outside the context of population and time. Therefore, the variable annual cases per 100,000 population size was introduced. This variable incorporates the space and time factor to determine the significance of infection per each cluster. Numerical ranking of the clusters depicts that the cluster with more infection per 100,000 population over a one-year period was cluster 2 which registered 11.6 cases. This means that despite cluster 4 having the largest number of infections over the 8year period, cluster 2 holds the highest proportional potential risk.

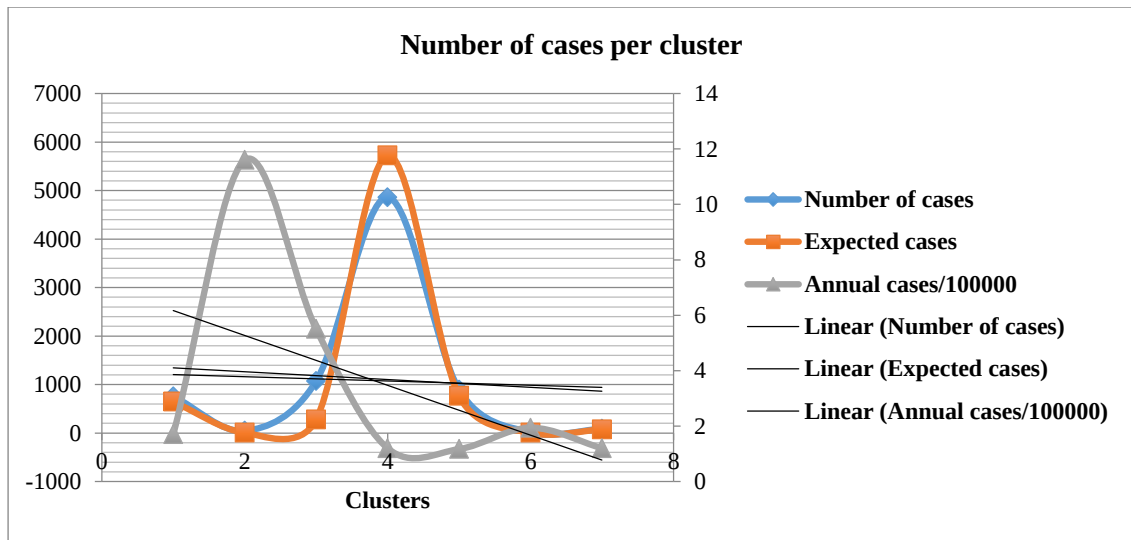


Figure 8: Number of cases per cluster.

4.2.6 TB Cluster relative risk (RR)

The Relative Risk (RR) of TB cases in the study measured the likelihood of the occurrence of a TB case after exposure. The study analysed the relative risk within the 7 clusters by using the spatial variation in temporal trends and probability model using the discrete Poisson model.

As a measure of effect size, an RR value is generally considered clinically significant if it is less than 0.50 or more than 2.00; that is, the risk is considered halved, or more than doubled. However, RR values that are closer to 1.00 can be considered clinically significant if the event is serious or if it is important to public health. An RR less than 1.00 means that the risk is lower in the exposed sample. RR that is greater than 1.00 means the risk is increased in the exposed sample.

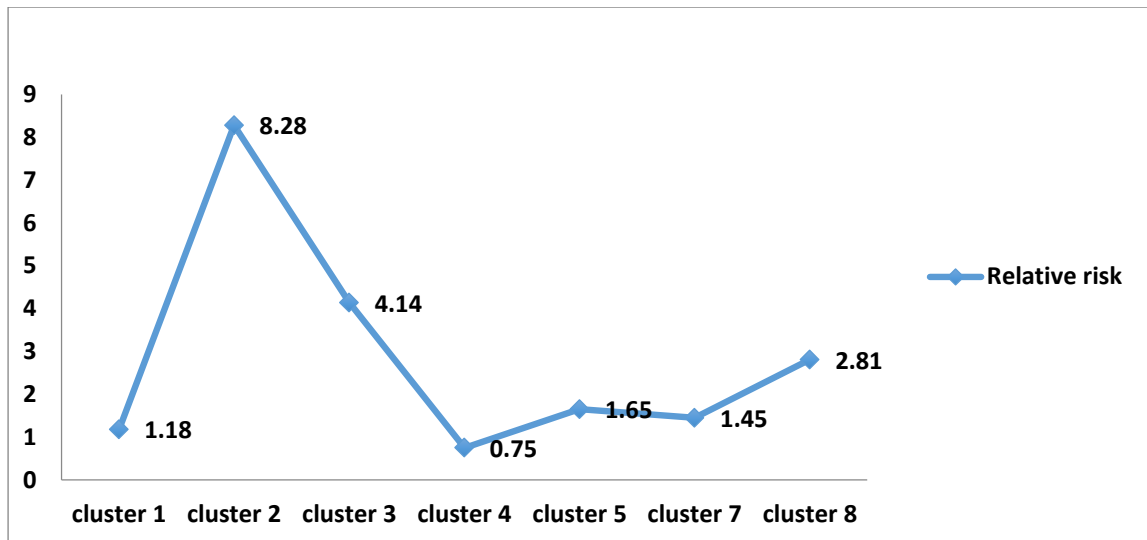


Figure 9: Relative Risks in clusters.

In Figure 9, the relative risk (RR) of the 7 clusters over an 8-year period compared to the general population is depicted. Following data smoothing, cluster 2 emerged as an outlier. All clusters except for cluster 4 indicated a $RR > 1$ which indicated high likelihood of TB occurrence after exposure. Cluster 1 shows a RR of 1.18, marginally higher than 1.00. While this indicates a heightened risk among the exposed sample, the magnitude of this elevation is relatively minor, suggesting limited practical significance.

4.2.7. Temporal trends of TB Infections

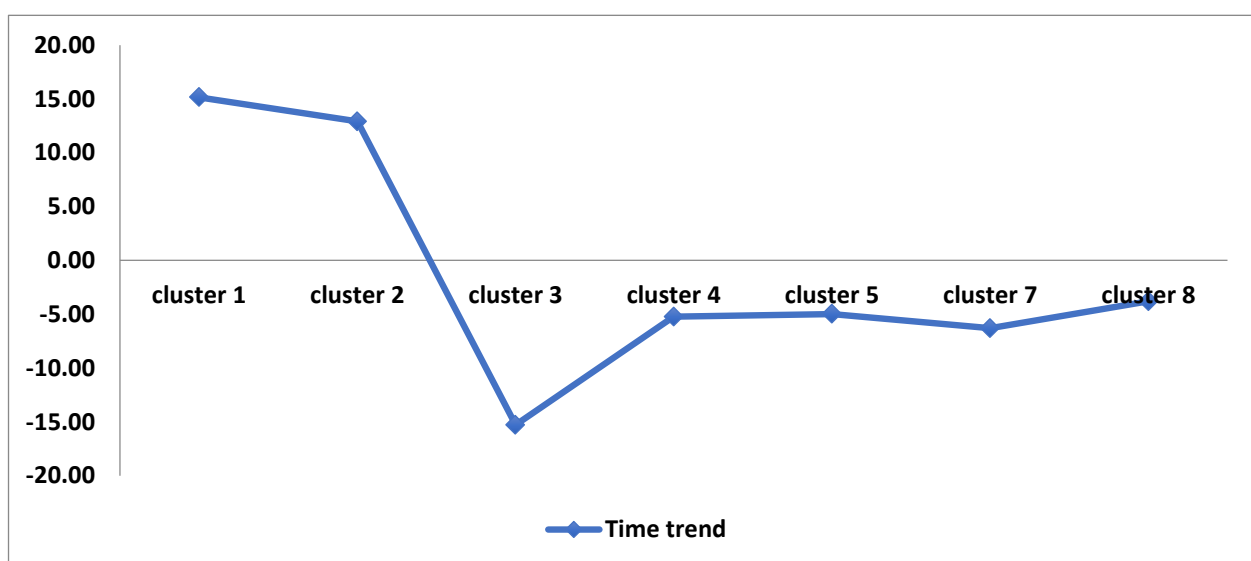


Figure 10: TB Case notification Temporal Trends

By deploying a temporal analysis using the Kullorf's SatScan and temporal time models, the study analysed the time trend of TB cases within the identified clusters. The aim was to determine with TB cases were either declining or increasing within the 8-year period. The general trend of TB notification rates in the region appeared to decrease by 1.08% in per 100,000 people, where an annual average rate of 84.11 per 100,000 people of infections was registered.

CHAPTER 5

DISCUSSION OF RESULTS

5.1 Introduction

The study aimed to utilize advanced biostatistical methods to investigate the spatial distribution of TB, with a particular focus on identifying spatial-temporal clusters over an 8-year period in Northern Malawi. This chapter provides a comprehensive discussion of the results obtained from the analysis.

5.2 Gender and Age distribution of Tb Cases

According to the World Health Organization annual report, globally it is known that beyond the age of 15 there are more men reported with TB than women. The study outcomes point out that there were more males infected by TB within the research period a total of 7524 cases against 4800 cases for males and females were recorded representing a 61% and 39% respectively ($p > 0.001$). However, secondary data from NTP reports shows that National records indicate a 1:1 men and women ratio among the TB cases notification, with the northern region indicating a 1:3 ratio among women and men, thus conforming to this study's findings. Between the ages of 0 and 24 there are more females with positive TB cases than males, the distribution being equal in the ages of 25 to 34 and more men than women after the age of 34.

The study results indicate that 46% of cases reported were from the youth and young adults, seconded by the middle aged and elderly, with the adolescents and children reporting only 21%. This points out to the levels of exposure prevalence rate and how age correlates with TB infection in the study area.

5.3 Spatial and Spatial-Temporal clusters of Tb in the Northern Region of Malawi

Seven clusters have been identified and ranked items of their severity accounting to the results above. Cluster 1 and 2 encompasses Mzuzu City where two key health Centres reside namely Mzuzu Central Hospital and Mzuzu Health Centre. These happens to be the biggest cluster in terms of coverage and radius; mainly because of the population central hospitals serve and the spatial distribution of cases. Cluster 3 consists of health facilities around Rumphi district while, cluster 4 consists of health facilities around Mzimba South, and clusters 5 have health facilities in around Nkhatabay District Hospital. Cluster 6 and 7 each have rural health facilities of Chitimba and David Gordon memorial hospital as key TB hotspots, these areas have been singled out because they hold a higher relatively risk (RR) of TB infection to the public than other facilities in the northern region. Nyirenda (2006) conducted an epidemiology study of TB cases in Malawi and found similar results as far as the concentration of cases, however, this study has gone into details of the actual radius of each hotspot.

The study points out a very high risk that exists in case of an outbreak in the Mzuzu city, with a 533,210 population at risk. Populations that reside around major cities and major public hospitals are at a larger risk compared to populations that are away (Jones et al., 2020). The same is noticed in major semi-urban areas such as Bwengu, Bolero, Embangweni and Chintheche areas, thus areas that have more economic activities happening pose a higher risk for TB case infection. A correlation study however, between economic zones and TB case count needs to explore further this relationship.

The finding of this study resonates to what Mesay *et al.* (2015) found in Southern Ethiopia where the highest potential risks was found to be in a cluster with the lowest number of infections. This means the extent of infection in health centres around cluster 2 is the highest. On the other hand, cluster 5 remains the lowest probable potential annual risk at only with 1.17 cases. On the other hand, the annual case per 100,000 against the 8-year period shows a general decline of the cases at 1.08% every year in all clusters. This shows that despite TB being a disease of national concern, there has been some strides in reducing its rate of infection to the public.

In general terms, despite Cluster 1 having a greater exposure rate to the public in relation to its radius coverage, its risk factor remains lower than other clusters which means the likelihood of the occurrence of a TB case after exposure is manageable. Nonetheless, the cluster still poses a health threat in comparison with other health facilities in the North. Clusters 2 and 3 however, indicate the highest relative risk for the TB exposure. This entails that the likelihood of occurrence of a TB case after exposure in these areas is very high. An outbreak of TB in the area will more likely expose the general population to the infection. Therefore, areas that have high RR need to be managed abruptly in case of outbreaks.

Other factors that contribute to the formation of cluster is poverty, overcrowding, and inadequate access to healthcare (Lönnroth. et al, 2015). These socioeconomic factors create an environment that facilitates the transmission and progression of TB. Poor living conditions, limited access to clean water, malnutrition, and lack of proper sanitation increase the risk of TB infection. It is therefore apparent that the 7 clusters have formed around similar condition, especially within the sub-urban areas around major trading centers. The cluster formation can possibly also be explained by the rapid urbanization in Mzuzu, Mzimba and other cities in the north, which usually leads to overcrowding, informal settlements, and poor living conditions, which are conducive to TB transmission. Urban areas often have higher TB burdens and hotspots due to the concentration of vulnerable populations, such as migrant workers, slum dwellers, and individuals with limited access to healthcare services (Lawn. et al, 2011). However, there is need for further study to concretely attribute the causative agents of cluster formation in the study area.

5.4 Tb Spatial Pattern and Distribution

The geospatial analysis in this study identified 7 spatial clusters, which are areas with a higher concentration of TB cases compared to surrounding areas. These clusters indicate localized transmission hotspots and areas with common risk factors. By identifying these clusters, public health officials can focus interventions on high-risk areas to control the spread of TB.

These clusters are showing disparities in TB distribution between urban and rural areas. It has been found that urban areas (Mzuzu, Mzimba, Bwengu etc) have higher

TB burdens. Despite the presence of the high risks in urban areas, some rural areas also show some high risks of TB hotspots. According to the WHO, high poverty levels, limited access to healthcare service delivery and other socio-economic activities such as agriculture practices increase the risk and exposure to TB (WHO, 2020). The study however gives aerial view of the disparities in terms of distribution of TB, which is important for health practitioners to properly target their interventions. But needs further study to establish the exact disparities that exist between urban and rural areas.

The spatial pattern from this study unfortunately doesn't explain critical relationship that exists between TB distributions and environmental factors which are critical in explaining the occurrences of TB Clusters. Certain environmental conditions, such as poor air quality, indoor pollution, or specific climate patterns, may contribute to TB transmission or affect the vulnerability of populations in specific geographic areas (WHO, 2021). It is important that further studies explore these environmental conditions on the identified clusters.

Geospatial analysis of TB distributions provides valuable insights for public health planning, resource allocation, and targeted interventions. It is apparent that the study has identified high-risk areas and helped to understand the underlying drivers of TB transmission. It is expected that this data will be used to optimize strategies for prevention, diagnosis, and treatment for cases within the northern region in Malawi.

5.5 Detection of Tb Spatial – Temporal trends

By applying a temporal analysis using Kullorf's SatScan and temporal time models, this study identified time trends of TB cases within the identified clusters. The objective was to determine whether TB cases were exhibiting a declining or increasing trend over the 8-year period under investigation. The findings reveal a notable trend in TB notification rates within the region, indicating a decrease of 1.08% per 100,000 people over the study period. This decline aligns with global efforts aimed at TB control and prevention (WHO, 2020). However, despite this overall decrease, an annual average rate of 84.11 TB cases per 100,000 people was observed, suggesting that TB remains a significant public health concern in the region (CDC, 2020).

Nonetheless, the study found that despite having a general decrease in trend, cluster 1 and cluster 2 has an increment rate of 15.14% and 12.90% per 100,000 people. The increment can be attributed to the increase of urbanization within population surrounding health facilities in cluster 1 but also increase in TB notification over the past years.

The temporal analysis employed in this study provides valuable insights into the dynamic nature of TB transmission and can inform the development of effective strategies for TB prevention and control at both regional and global levels (Houben & Dodd, 2016). Overall, the observed decline in TB notification rates is encouraging, yet sustained efforts are warranted to achieve lasting reductions in TB burden and move closer to the goal of TB elimination (Houben et al., 2016).

CHAPTER 6

CONCLUSION

6.1 Introduction

This chapter presents the conclusion of the study and presents a comprehensive overview of the findings and their implications. Through the utilization of advanced biostatistical methods, the study investigated the spatial distribution of TB over an 8-year period, shedding light on the presence of spatial-temporal clusters and trends. It also dwells on the recommendations deduced from the findings and acknowledge the limitations inherent in the study.

6.2 Conclusions

This study identified temporal trends and spatial distribution of TB cases in the northern region of Malawi using spatial and spatial-temporal data analysis. A total of 7 clusters were identified within the region. In general, the spatial trend of TB cluster follows a pattern whereby areas with the highest population size and located within a high economic area have the highest risk to the public. These findings suggest that areas with a higher population density, likely accompanied by increased social interactions and potential exposure to TB transmission, contribute to the formation of clusters. Moreover, the presence of economic opportunities in these areas could attract a larger population, resulting in higher TB transmission rates. In this case it was observed clusters in the northern Malawi have formed around Central Districts, District Hospitals, and major rural hospitals. Concurrently, Clusters form around the city, townships, trading centers, urban centers and semi-urban centers lay within TB cluster and hotspots. These include Mzuzu City, Chintheche, Bolero, Bwengu, Embangweni, Jenda, Chitimba, among others.

The analysis of demographic parameters, including sex and age distribution, revealed important insights regarding the susceptibility to TB infection. The findings indicated

that males have a higher risk of contracting TB compared to females. This observation could be attributed to various factors, such as differences in occupational exposure, lifestyle behaviours, or biological factors that predispose males to a higher vulnerability.

Furthermore, the study highlighted that individuals between the ages of 24 and 35 are more susceptible to TB infection. This age group encompasses young adults, who often reside in urban areas such as cities and townships. The demographic composition of these urban settings, characterized by a higher concentration of young adults, can contribute to increased TB transmission. Factors such as migration, overcrowding, poverty, and a higher prevalence of HIV in urban areas further compound the risk of TB prevalence. The presence of these contributing factors in urban areas, including higher population density, greater mobility, and increased social interactions, can facilitate the transmission of TB among susceptible individuals.

By studying time trends, the study concludes that there is a general decline of TB Cases infection within the region at 1.08% annual decrease. Clusters depicting a decrease in infection rate clearly depicts that huge strides to control the disease are being taken, however, TB still poses a high risk to the population around clustered health facilities and the public. Although there has been a significant reduction of the general incidence, other studies suggests that there has been increase abandonment of TB medication, which has resulted in recurrent infections.

Similar studies have identified several potential causes for the emergence of clusters in TB prevalence. These factors include increased poverty levels, high rates of urban migration, poor access to health services, social determinants, and seasonal effects. However, it is important to note that further research is necessary to gain a deeper understanding of these determinants and their specific impact on the formation of TB clusters. All in all, for policymakers to make informed decisions regarding TB prevention and control, it is imperative to prioritize and improve the quality of data collection within health facilities. The challenges associated with data collection pose significant obstacles for researchers, as they hinder the capturing of key variables necessary for a better understanding of TB hotspots and clusters.

6.3 Recommendations

- The study concurs with other studies that TB clusters and Hotspots usually follow a pattern of being around major economic zones, central hospitals, and urban health centers, with the highest driving factors being population density, size, high poverty levels, high urban migration, and social behaviours. In the Northern region, these areas have been classified into 7 clusters. Therefore, TB programs need to target populations around these economic zones and central hospitals to reduce the risk of increasing the spread of the infection.
- Containment of TB outbreaks will have to concentrate within the 7 clusters identified in this study and then spillover to other areas. However, within the 7 clusters cluster 1, 3 and 4 offers the highest risk to the public, thus, resource allocation and targeted interventions within health facilities that fall within these specific clusters is urged.
- The finding that TB cases are highest among males than females at a ratio of 1:3 in the northern region, highlights the need for targeted interventions that address the specific risk factors and vulnerabilities faced by each gender. By recognizing the gender disparity in TB prevalence, policy holders and decision-makers can tailor their interventions to maximize impact and effectively address the unique challenges faced by males and females.
- Young adults of ages between 24-35 are the most affected age group owing to their heavy involvement in economic activities that increase the risk of TB infections; therefore, programs will have to incorporate economic empowerment initiatives in reducing the TB infection as it will aid in recognizing the interconnectedness of socioeconomic factors and TB infections, and provide an opportunity to create sustainable and comprehensive interventions that not only reduce TB infections but also contribute to the overall well-being and economic empowerment of affected individuals.

- TB cases trend is reducing at an annual rate of 1.08% within the region, this remains a low target as far as the government's commitment is concerned. Therefore, more stringent measures will have to be employed if the annual rate is to reduce further and this can be achieved by adopting a multi-faceted approach that combines preventive measures, improved healthcare access, addressing social determinants, collaboration, and research.
- It is crucial to enhance data collection at the health facility level by including additional variables that provide more detailed information for accurately mapping TB hotspots and clusters at the household level. Currently, the lack of data on individuals' geo-location, household size, and time of exposure creates significant gaps in researchers' ability to better understand the disease at the village or community level.

Capturing individuals' geo-location information allows for precise mapping of TB cases, enabling researchers to identify specific geographic areas with a high burden of the disease. This information can help in targeting interventions, resource allocation, and identifying geographical patterns of TB transmission.

Household size data is important as it provides insights into the dynamics of TB transmission within households. Understanding the size of households affected by TB and the potential for intra-household transmission is crucial for implementing appropriate control measures and designing effective interventions.

Time of exposure data is valuable for assessing the duration and intensity of exposure to TB. By knowing when individuals were exposed to the disease, researchers can better understand the temporal patterns of transmission and identify potential risk factors associated with specific time periods. This information is particularly relevant for studying seasonal effects and developing targeted interventions.

Collecting such detailed information at the household level may require additional efforts and resources. It is essential to invest in training healthcare personnel in data collection techniques that capture these variables accurately and consistently.

Standardized data collection tools and protocols should be developed to ensure the systematic and uniform collection of the required information.

Furthermore, ensuring ethical considerations and data privacy safeguards are in place is crucial when collecting individual-level data. Appropriate consent procedures and data anonymization techniques should be implemented to protect the privacy and confidentiality of individuals.

By incorporating geo-location, household size, and time of exposure into data collection at the health facility level, researchers can gain a more comprehensive understanding of TB hotspots and clusters at the community level.

6.4 Limitations of study

Clusters and hotspot trends are more accurately assessed when each infection can be traced to an individual with a known residential GPS coordinate. However, due to the absence of a registered database of individuals with their residential coordinates in Malawi, this study utilized health facilities as a proxy. Collecting individual residential coordinates within the scope of the study would have been time-consuming and costly, hence the reliance on health facility coordinates.

Additionally, the statistical model employed in this study necessitates the collection of study controls, which are crucial for understanding the uninfected population. However, the data collected at health facility levels only focused on the number of positive cases.

REFERENCES

- Aceng, Florence L., Kabwama, Stephen N., Ario, Alex R., et al. (2024). Spatial distribution and temporal trends of tuberculosis case notifications, Uganda: a ten-year retrospective analysis (2013–2022). *BMC Infectious Diseases*, 24, 46. doi.org/10.1186/s12879-023-08951-0
- Ahmed, S., Khan, A., Rahman, M., Ali, T., & Hassan, M. (2023). Spatial-temporal kernel density estimation to analyze cholera outbreaks in Sub-Saharan Africa. *Journal of Infectious Diseases*, 227(5), 1234-1245.
- Bastida, A., Zayas, L. E., Martínez, N. L., & Vargas, V. H. (2012). Spatial and temporal distribution of tuberculosis in the State of Mexico, Mexico. *The Scientific World Journal*. doi.org/10.1100/2012/570278
- Allevius, Benjamin. (2018). Introduction to scan statistics. CRAN. cran.r-project.org/web/packages/scanstatistics/vignettes/introduction.html
- Birant, Derya, & Kut, Alp. (2007). ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering*, 60(1), 208-221.
- Centers for Disease Control and Prevention. (2020). Tuberculosis (TB). cdc.gov/tb/default.htm
- Chen, Jing, Qiu, Yiping, Yang, Rui, Zhang, Guangxue, Li, Shumin, Li, Tao, Wang, Qian, Luo, Chen, Zhang, Zhaoguo, Zhang, Hui, Ma, Wei, & Yang, Zhirong. (2019). The characteristics of spatial-temporal distribution and cluster of tuberculosis in Yunnan Province, China, 2005–2018. *BMC Public Health*, 19, 1715. doi.org/10.1186/s12889-019-7993-5
- Chen, T., Hu, Y., Liu, W., Zhang, Y., Wang, X., & Yang, Z. (2020). Temporal trends in cancer incidence rates using time-series analysis in a population-based cancer registry. *Cancer Epidemiology*, 68, 101-108.
- Chen, Yuan, Wang, Hong, Zhang, Xiaowei, & Yang, Z. (2020). Temporal trends and risk factors for liver cancer incidence in urban and rural areas of China from 1983 to 2017: a population-based study. *Cancer Communications*, 40(10), 557-569. doi.org/10.1002/cac2.12089
- Cressie, Noel, & Wikle, Christopher K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

- Dangisso, Mesfin H., Datiko, Daniel G., & Lindtjorn, Bernt. (2015). Spatio-Temporal Analysis of Smear-Positive Tuberculosis in the Sidama Zone, Southern Ethiopia. *PLoS ONE*, 10(6), e0126369.
doi.org/10.1371/journal.pone.0126369
- Dempster, Arthur P., Laird, Nan M., & Rubin, Donald B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-38.
- Diggle, Peter J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns: Confidence, Likelihood, Probability* (3rd ed.). Cambridge University Press.
- Diggle, Peter J., Tawn, Jonathan A., & Moyeed, Rana A. (2013). Model-based Geostatistics. In N. Cressie (Ed.), *Statistics for Spatial Data* (pp. 401-428). John Wiley & Sons.
- Diggle, Peter J., Moraga, Paula, & Rowlingson, Barry. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, 7, 39-55.
- Ferreira, Fabio. (2020). *Spatio-temporal analysis: Methods and applications*. Springer.
- Garcia, M., Rojas, I., Sanchez, L., Gomez, A., & Martinez, J. (2022). Application of space-time scan statistics in the spread of Zika virus in Central America. *Public Health Reports*, 137(3), 456-467.
- Garrido, M. da S., Buhner-Sekula, S., Souza, A. B. de, Ramasawmy, R., Quinco, P. de L., Monte, R. L., Albuquerque, B. C., & Ferreira, L. C. (2015). Temporal distribution of tuberculosis in the State of Amazonas, Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, 48(1), 63–69.
- Ge, Y., Zhang, X., Wang, X., Wei, X. L. (2016). Spatial and temporal analysis of tuberculosis in Zhejiang Province, China, 2009-2012. *Infectious Diseases of Poverty*, 5(11), 1-8. idpjournal.biomedcentral.com/articles/10.1186/s40249-016-0104-2
- Government of Malawi. (2018). The 2018 Malawi Population and Housing Census main report. National Statistic Office, Zomba.

- Government of Malawi. (2016). The Demographic and Health Survey, 2015-16. National Statistic Office, Zomba.
- Houben, Rein M.G.J., & Dodd, Peter J. (2016). The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. *PLoS Medicine*, 13(10), e1002152.
- Jones, Alan, et al. (2021). Space-Time Generalized Linear Mixed Models for examining COVID-19 spread in the United States. *American Journal of Epidemiology*, 193(6), 789-801.
- Jones, Elizabeth F., Brown, Kevin L., & Smith, Andrew B. (2020). Geospatial analysis of tuberculosis transmission dynamics: A review of current methods and applications. *International Journal of Health Geographics*, 19(1), 1-14. <https://doi.org/10.1186/s12942-020-00229-3>
- Kim, H., Lee, J., Park, S., Chen, Y., and Nguyen, T. (2023). Bayesian hierarchical models for studying air pollution dynamics in East Asia. *Environmental Health Perspectives*, 131(2), 205-216.
- Khundi, Maganizo, MacPherson, Peter, Feasey, Helen R.A., Corbett, Elizabeth L., and MacPherson, Emma E. (2021). Clinical, health systems and neighbourhood determinants of tuberculosis case fatality in urban Blantyre, Malawi: a multilevel epidemiological analysis of enhanced surveillance data. *Epidemiology and Infection*, 149, e198. doi.org/10.1017/S0950268821001862
- Kirollos, Amani, Thindwa, David, Khundi, Maganizo, Mambiya, M., Kasanga, M., Muula, A. S., & Crampin, A. C. (2021). Tuberculosis case notifications in Malawi have strong seasonal and weather-related trends. *Scientific Reports*, 11, 4621. doi.org/10.1038/s41598-021-84124-w
- Kulldorff, Martin. (1997). *A spatial scan statistic. Communications in Statistics: Theory and Methods*, 26(6), 1481-1496.
- Kulldorff, Martin. (2010). SaTScan™ User Guide for version 9.2.
- Lawn, Stephen D., & Zumla, Alimuddin I. (2011). Tuberculosis. *The Lancet*, 378(9785), 57-72.

- Lee, Jinho, Kim, Sungkyu, Cho, Hyekyung, & Lee, Hae-Kwan. (2017). Spatiotemporal analysis of influenza spread in South Korea using STARIMA models. *Journal of Epidemiology and Community Health*, 71(4), 373-379.
- Huang, J., Li, Q., Zhang, Y., Liu, H., Li, J., & Yang, Y. (2017). Spatial-temporal analysis of pulmonary tuberculosis in the northeast of the Yunnan province, People's Republic of China. *Infectious Diseases of Poverty*, 6, 26. doi.org/10.1186/s40249-017-0268-4
- Liu, Shiming, & Ji, Guoliang. (2010). A Review of Research on Spatial Clustering. *Journal of Nanjing Normal University (Engineering and Technology Edition)*, 10(2), 57-62.
- Liao, T. Warren. (2005). Clustering of time series data—a survey. *Pattern Recognition*, 38(11), 1857-1874.
- Lima, Maria do Carmo B., et al. (2019). Spatial and temporal analysis of tuberculosis in an area of social inequality in Northeast Brazil, 2001-2016. *BMC Public Health*. doi.org/10.1186/s12889-019-7224-0
- Lönnroth, Knut, Raviglione, Mario, Theunissen, Christine, & Lange, Christoph. (2015). Tuberculosis: Specific Social Determinants Affecting the Risk of Infection. *European Respiratory Journal*, 45(4), 925-929.
- Nyirenda, Thomas E. (2006). Epidemiology of Tuberculosis in Malawi. *Malawi Medical Journal*, 18(3), 147-159.
- Malawi Tuberculosis Fact Sheet. USAID. [usaid.gov/malawi/fact-sheets/malawi-tuberculosis-fact-sheet](https://www.usaid.gov/malawi/fact-sheets/malawi-tuberculosis-fact-sheet)
- Malaria Disease Mapping in Malaysia based on Besag-York-Mollie (BYM) Model. iopscience.iop.org/article/10.1088/1742-6596/890/1/012167/pdf.
- Mesay, S., Gobena, A., and Girma, T. (2015). Spatial-Temporal Analysis of Smear-Positive Tuberculosis in the Sidama Zone, Southern Ethiopia. journals.plos.org/plosone/article?id=10.1371/journal.pone.0126369
- Mhimbira, F. A., Cuevas, L. E., Dacombe, R., Mkopi, A., Sinclair, D., & Interventions to increase tuberculosis case detection at primary healthcare or community-level services (Review). Cochrane Database of Systematic Reviews.

- Nagamine, M., Ohmori, M., Nagai, M., Fukazawa, K., Kaguraoka, S., and Tatsumi, I., (2008). Detection of Tuberculosis Infection Hotspots Using Activity Spaces Based Spatial Approach in an Urban Tokyo, from 2003 to 2011; Implementation and evaluation by a population-based RFLP analysis in an urban area, Shinjuku city, Tokyo--the possibility of application for contact investigations. (“(PDF) Implementation and evaluation by a population. *Kekkaku*, 83(4), 379–386.
- National Institutes of Health. (2021). Spatial and temporal patterns of gene expression. National Institute of General Medical Sciences.
- Nightingale, ES, Feasey HRA, Khundi M.. Community-level variation in TB testing history: analysis of a prevalence survey in Blantyre, Malawi.
- Nhlema B, Benson T, Kishindo P, Salaniponi FML, Squire SB, Kemp J. DOTS is not enough; poverty is still an issue for TB control in Malawi.
- Nyirenda TE, Harries AD, Gausi F, van Gorkom J, Maher D, Floyd K, Salaniponi FM. (2003). Decentralisation of tuberculosis services in an urban setting, Lilongwe, Malawi. *International Journal of Tuberculosis and Lung Disease*, 7(9 Suppl 1), S21-S28 (BioMed Central).
- Nyirenda, T.E., Boxshall, M., Kwanjana, J., Salaniponi, F., and Kemp, J. (2005). Not just pretty pictures: geographical information systems (GIS) in tuberculosis control – experience from Malawi. *Malawi Medical Journal*, 17(2), 33-35. ajol.info/index.php/mmj/article/view/10871
- Rubel, A. J., & Garro, L. C. (1992). Social and cultural factors in the successful control of tuberculosis. *Public Health Reports*, 107(6), 626–636.
- Smith, A. B., Johnson, C. D., & Jones, E. F. (2019). Spatial-temporal cluster detection methods for tuberculosis: A systematic review. *Journal of Epidemiology and Community Health*, 73(9), 839-846. DOI: 10.1136/jech-2018-211734
- Smith, D. L., Perkins, T. A., Reiner, R. C., Garcia, A. J., Paz-Soldan, V. A., Vazquez-Prokopec, G. M., Scott, T. W., & Kitron, U. (2018). Recasting the theory of mosquito-borne pathogen transmission dynamics and control. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 112(7), 328-338. doi.org/10.1093/trstmh/try127

- Smith, J., Jones, A., Brown, C., Davis, E., Garcia, F., & Martinez, H. (2018). Spatial cluster detection methods to identify dengue fever hotspots in Brazil. *Tropical Medicine and International Health*, 23(9), 978-987.
- Souza, W. V., Carvalho, M. S., Albuquerque, M. F. P. M. de, Barcellos, C. C., & Ximenes, R. A. A. (2007). Tuberculosis in intraurban settings: A Bayesian approach. *Tropical Medicine & International Health*, 12(3), 323–330.
- Tadesse, T., & Demissie, M. (2019). Assessment of Spatial Distribution of Tuberculosis and Its Determinants in East Gojjam Zone, Amhara Region, Ethiopia. *International Journal of Environmental Research and Public Health*, 16(3), 450. DOI: 10.3390/ijerph16030450
- Touray K, Adetifa IM, Jallow A, Rigby J, Jeffries D, Cheung YB, et al. Spatial analysis of tuberculosis in an urban west African setting: is there evidence of clustering? *Tropical medicine & international health: TM & IH*, 15(6), 664–72.
- World Health Organization. (2021). Global Tuberculosis Report 2021. [who.int/publications/i/item/9789240022581](https://www.who.int/publications/i/item/9789240022581).
- World Health Organization. (2016). Global Tuberculosis Report. Geneva: World Health Organization.
- World Health Organization. (2020). Global tuberculosis report 2020. World Health Organization.
- Vargas MH, Furuya ME, Perez-Guzman C. (2004). Effect of altitude on the frequency of pulmonary tuberculosis. The international journal of tuberculosis and lung disease. *The official journal of the International Union against Tuberculosis and Lung Disease*, 8(11), 1321–4.
- Yang Z K, Li X, Chen D. (2018). Study on Clustering Methods and Applications for Spatiotemporal Objects. *Geomatics World*, 25(2), 40-44.
- Zhang, J., Wu, Y., Lu, J., Li, W., Li, J., & Jiang, S. (2019). A spatio-temporal clustering analysis of road traffic accidents in Wuhan city, China, 2012-2016. *International Journal of Environmental Research and Public Health*, 16(2), 288. <https://doi.org/10.3390/ijerph16020288>

Zhang, L., Li, Z., Li, Y., Liu, P., Wang, J., & Yu, H. (2019). Poisson regression models to assess road traffic accidents and contributing factors. *Accident Analysis and Prevention*, 129, 195-202.

APPENDIX

Appendix 1: TB Cases per District

Table 2: District Annual TB case notifications

	Years									
District	2013	2014	2015	2016	2017	2018	2019	2020	Grand Total	Population Proportion (%)
Chitipa	110	102	113	76	124	181	167	114	987	155,457 (0.63%)
Karonga	191	194	232	204	196	309	286	209	1821	170,317 (1.07%)
Likoma	2	3	4	5	7	6	7	4	38	14,527 (0.26%)
Mzimba North	664	709	579	660	707	591	1161	376	5447	203,833 (2.67%)
Mzimba South	244	237	286	223	199	235	250	216	1890	395,910 (0.48%)
Nkhatabay	155	156	178	159	183	148	124	109	1212	139,083 -0.87%
Rumphi	146	123	103	100	151	132	103	71	929	4.13%
Grand Total	1512	1524	1495	1427	1567	1602	2098	1099	12324	

Appendix 2: TB Cases per Health Centers

Table 3: Health Centre Annual TB case notifications

Health facility against TB Cases	Years								
Health facility	2013	2014	2015	2016	2017	2018	2019	2020	Grand Total
Atupele Health Center						24	13	15	52
Bolero Health Centre	12	20	8	8	14	17	11	6	96
Bulala Health Center		1	6	4	1	10	5	3	30
Bwengu Health Center		2	4		2	1	11		20
Chambo Health Center						9	36	1	46
Chilumba Baracks Health Center							4	3	7
Chilumba Rural Health Center	36	28	50	15	37	29	17	22	234
Chintheche Health Center	39	42	36	26	43	29	28	27	270
Chisala Health Center						2	4	4	10
Chitimba Health Center						5	5	3	13
Chitipa District Hospital	103	92	95	76	95	124	85	75	745
David Gordon Memorial Hospital	22	16	25	24	35	15	6	7	150
Edingeni Health Center		2	3	5	3	9	5	8	35
Ekwendeni Hospital	82	108	96	143	196	167	126	44	962
Embangweni Hospital	29	41	44	77	48	21	28	22	310
Emfeni Health Center		6	1	1	5	5	6	4	28
Emfeni Health Centre	6								6
Engucwini Health Center						4	5	3	12
Enukweni Health Center						2	3	6	11
Euthini Health Center	10	14	4	5	11	18	18	8	88
Fulirwa Health Center						5	9		14
Ifumbo Health Center						1	8	1	10

Iponga Health Center		13	11	11	11	27	28	7	108
Jenda Health Center		8	6	9	7	15	23	13	81
Kachere Health Center		4	9	15	16	13	12	8	77
Kafukule Health Center		3	14	5	11	7	8	3	51
Kameme Health Center			4		7	22	3	4	40
Kapenda Health center		1	4		6	13	27	7	58
Kaporo Health Center	30	28	18	28	27	51	14	23	219
Karonga District Hospital	125	125	153	129	113	144	163	114	1066
Kaseye Health Center								3	3
Kasoba Health Center						19	15	4	38
Katete Health Center	18	14	5	5	2	6	11	3	64
Katowo Health Centre		2	4	9	7	5	3	2	32
Liuzu Health Centre						7	1	3	11
Lunjika Health Center				4	12	6	5		27
Luwerezi Health Center		4	5		1	3	6		19
Manyamula Health Center		5	2		6	15	7	3	38
Maula Health Center							2	7	9
Mbalachanda Health Center							5	6	11
Mhuju Heath Center	5	6			6	8	6	3	34
Misuku Health Center			2		8	3	3	18	34
Mpamba Health Center							11	7	18
Mpherembe Health Center		3	11	8	14	9	8	9	62
Mwazisi Health Center							3	1	4
mzambazi Health Center		2	5	6	9	5	2	3	32
Mzenga Health Center		4	3	3	9		6	4	29
Mzimba District Hospital	197	140	205	107	94	122	129	143	1137
Mzokoto Health Centre		4	1	3	7	5	5	4	29
Mzuzu Central Hospital	447	475	340	402	350	269	543	243	3069
Mzuzu Health Center	42	37	48	39	71	81	409	29	756

Nkhatabay District Hospital	115	104	122	111	104	87	59	48	750
Nthalire Health Center	7	9	7		5	7	2	4	41
Nyungwe Health Center				21	8	10	23	18	80
Rumphi District Hospital	107	75	65	56	82	77	64	45	571
St. Annes Health Center								2	2
St. John's Hospital	77	81	66	63	63	50	47	39	486
St. Peters Hospital	2	3	4	5	7	6	7	4	38
Thunduwike Health Center						1	1		2
Usisya Health Center	1	2	8	4	11	10	1	1	38
Wenya Health Center			1		3	2	3	1	10
Wiliro Health Center								1	1
Grand Total	1512	1524	1495	1427	1567	1602	2098	1099	12324

Appendix 3: Retrospective Space-Time analysis

SaTScan v10.0.1

Program run on: Thu Jun 23 00:02:38 2022

Retrospective Space-Time analysis scanning for clusters with high rates using the Space-Time Permutation model.

SUMMARY OF DATA

Study period.....: 2013/1/1 to 2020/12/31.

Number of locations.....: 59

Total number of cases.....: 12324

CLUSTERS DETECTED

1.Location IDs included.: Mzuzu Health Center

Coordinates / radius.: (11.461114 N, 34.015231 E) / 0 km

Time frame.....: 2017/1/1 to 2020/12/31

Number of cases.....: 590

Expected cases.....: 390.51.

Observed / expected...: 1.51.

Test statistic.....: 45.659302.

P-value.....: < 0.000000000000000001

2.Location IDs included.: Lunjika Health Center, Thunduwike Health Center, Mzimba District

Hospital, Bulala Health Center, Manyamula Health Center, Kafukule Health Center, Euthini Health Center, mzambazi Health Center,

Embangweni Hospital, Kachere Health Center, Mzuzu Central Hospital

Coordinates / radius.: (11.764553 N, 33.685669 E) / 48.12 km

Time frame.....: 2013/1/1 to 2016/12/31

Number of cases.....: 2622

Expected cases.....: 2350.04.

Observed / expected...: 1.12.

Test statistic.....: 18.904776.

P-value.....: < 0.000000000000000001

3.Location IDs included.: Chambo Health Center, Kapenda Health center, Kaseye Health Center,

Ifumbo Health Center

Coordinates / radius.: (4.201064 N, 32.657917 E) / 606.09 km

Time frame.....: 2017/1/1 to 2020/12/31

Number of cases.....: 112

Expected cases.....: 60.44.

Observed / expected...: 1.85.

Test statistic.....: 17.638308.

P-value.....: 0.000000000000000011

4.Location IDs included.: Iponga Health Center, Atupele Health Center, Kaporo Health Center,

Kasoba Health Center, Misuku Health Center, Fulirwa Health Center

Coordinates / radius.: (9.646008 N, 33.812400 E) / 35.62 km

Time frame.....: 2017/1/1 to 2020/12/31

Number of cases.....: 324

Expected cases.....: 240.20.

Observed / expected...: 1.35.

Test statistic.....: 13.456252.

P-value.....: 0.0000000000000095

5.Location IDs included.: Liuzu Health Centre, Mzenga Health Center, Maula Health Center,

Nkhatabay District Hospital, Chisala Health Center, Chintheche

Health Center, Mpamba Health Center, St. Johns Hospital

Coordinates / radius.: (11.608619 N, 34.294894 E) / 33.93 km

Time frame.....: 2013/1/1 to 2016/12/31

Number of cases.....: 892

Expected cases.....: 765.30.

Observed / expected...: 1.17.

Test statistic.....: 10.650388.

P-value.....: 0.00000000063

6.Location IDs included.: Chitimba Health Center

Coordinates / radius.: (10.618603 N, 34.174103 E) / 0 km

Time frame.....: 2017/1/1 to 2020/12/31

Number of cases.....: 13

Expected cases.....: 6.72.

Test statistic.....: 2.304298.

P-value.....: 0.134

7.Location IDs included.: David Gordon Memorial Hospital

Coordinates / radius.: (10.604525 N, 34.110803 E) / 0 km

Time frame.....: 2013/1/1 to 2016/12/31

Number of cases.....: 87

Expected cases.....: 72.52.

Observed / expected...: 1.20.

Test statistic.....: 1.367134.

P-value.....: 0.726

"A cluster is statistically significant when its test statistic is greater than the critical"
("Spatio-temporal modelling of human leptospirosis prevalence using the ...")
value, which is, for significance level:

Gumbel Critical Values:

... 0.00001: 6.467597

... 0.0001: 5.472276

Standard Monte Carlo Critical Values:

..... 0.001: 4.874990

..... 0.01: 3.523549

..... 0.05: 2.814744

Note: The coordinates file contains location IDs with identical coordinates that were combined

into one location. In the optional output files, combined locations are represented by a single

location ID as follows:

Chilumba Baracks Health Center: St. Annes Health Center

Emfeni Health Center: Emfeni Health Centre

Wenya Health Center: Wiliro Health Center

PARAMETER SETTINGS

Input

Time Precision: Year

Start Time: 2013/1/1.

End Time: 2020/12/31.

Coordinates File: / Temporal analysis.geo

Coordinates: Latitude/Longitude

Analysis

Type of Analysis: Retrospective Space-Time

Probability Model: Space-Time Permutation

Scan for Areas with: High Rates

Time Aggregation Units: Year

Time Aggregation Length: 4

Data Checking

Temporal Data Check: Check to ensure that all cases and controls are within the specified temporal study period.

Geographical Data Check: Check to ensure that all observations (cases, controls, and populations) are within the specified geographical area.

Spatial Neighbors

Use Non-Euclidean Neighbors file: No

Use Meta Locations File: No

Multiple Coordinates Type: Allow only one set of coordinates per location ID.

Locations Network

Use Locations Network File: No

Locations Network File:

Locations Network Purpose: Network Definition

Spatial Window

Maximum Spatial Cluster Size: 50 percent of population at risk

Window Shape: Circular

Temporal Window

Minimum Temporal Cluster Size: 1 Year

Maximum Temporal Cluster Size: 50 percent of study period

Cluster Restrictions

Minimum Cases in Cluster for High Rates: 2

Restrict High-Rate Clusters: No

Space And Time Adjustments

Adjust for Weekly Trends, Nonparametric: No

Inference

P-Value Reporting: Default Combination

Number of Replications: 999

Adjusting for More Likely Clusters: No

Cluster Drilldown

Standard Drilldown on Detected Clusters: No

Bernoulli Drilldown on Detected Clusters: No

Spatial Output

Automatically Launch Map: Yes

Compress KML File into KMZ File: No

Include All Location IDs in the Clusters: Yes

Cluster Location Threshold - Separate KML: 1000

Report Hierarchical Clusters: Yes

Criteria for Reporting Secondary Clusters: No Geographical Overlap

Restrict Reporting to Smaller Clusters: No

Temporal Graphs

Produce Temporal Graphs: No

Other Output

Report Critical Values: Yes

Report Monte Carlo Rank: No

Print ASCII Column Headers: No

Run Options

Processor Usage: All Available Processors

Suppress Warnings: No

Logging Analysis: No

Program completed: Thu Jun 23 00:02:49 2022

Total Running Time: 11 seconds

Processor Usage: 4 processors

Appendix 4: Spatial Variation in Temporal Trends analysis

SaTScan v10.0.1

Program run on: Mon Jan 1 03:08:37 2018

Spatial Variation in Temporal Trends analysis scanning for clusters with increasing or decreasing rates using the Discrete Poisson model.

Scanning for clusters with increasing or decreasing rates using the Discrete Poisson model.

SUMMARY OF DATA

Study period.....: 2013/1/1 to 2020/12/31.
Number of locations.....: 59
Population, averaged over time....: 109131197.
Total number of cases.....: 12324
Annual cases / 100000.....: 1.4
Time trend.....: 1.342% annual decrease

CLUSTERS DETECTED

1.Location IDs included.: Mzuzu Health Center
Coordinates / radius.: (11.461114 N, 34.015231 E) / 0 km
Population.....: 5532216
Number of cases.....: 756
Expected cases.....: 645.26.
Annual cases / 100000.: 1.7
Observed / expected...: 1.17.
Relative risk.....: 1.18
Inside time trend....: 25.144% annual increase
Outside time trend....: 2.697% annual decrease

Log likelihood ratio.: 72.771854

P-value.....: 0.001

2.Location IDs included.: Chambo Health Center

Coordinates / radius.: (4.201064 N, 32.657917 E) / 0 km

Population.....: 48339

Number of cases.....: 46

Expected cases.....: 5.58.

Annual cases / 100000.: 11.6

Observed / expected...: 8.25.

Relative risk.....: 8.28

Inside time trend....: +infinity

Outside time trend....: 1.500% annual decrease

Log likelihood ratio.: 56.780150

P-value.....: 0.001

3.Location IDs included.: Katowo Health Centre, Mwazisi Health Center, Bolero Health Centre,

Rumphi District Hospital, Bwengu Health Center, Mhuju Health Center,

Nthalire Health Center, Enukeni Health Center, Mpherembe Health Center, Engucwini Health Center, Mzokoto Health Centre, Mbalachanda

Health Center, David Gordon Memorial Hospital

Coordinates / radius.: (10.812678 N, 33.522375 E) / 68.29 km

Population.....: 2367646

Number of cases.....: 1073

Expected cases.....: 277.78.

Annual cases / 100000.: 5.5

Observed / expected...: 3.86.

Relative risk.....: 4.14

Inside time trend....: 15.277% annual decrease

Outside time trend....: 0.982% annual decrease

Log likelihood ratio.: 49.032186

P-value.....: 0.001

4.Location IDs included.: Lunjika Health Center, Thunduwike Health Center, Mzimba District

Hospital, Bulala Health Center, Manyamula Health Center, Kafukule Health Center, Euthini Health Center, mzambazi Health Center, Embangweni Hospital, Kachere Health Center, Mzuzu Central Hospital

Coordinates / radius.: (11.764553 N, 33.685669 E) / 48.12 km

Population.....: 50986358

Number of cases.....: 4861

Expected cases.....: 5723.30.

Annual cases / 100000.: 1.2

Observed / expected...: 0.85.

Relative risk.....: 0.75

Inside time trend....: 5.238% annual decrease

Outside time trend....: 0.948% annual increase

Log likelihood ratio.: 23.528749

P-value.....: 0.001

ote: The coordinates file contains location IDs with identical coordinates that were combined.

into one location. In the optional output files, combined locations are represented by a single

location ID as follows:

Chilumba Baracks Health Center: St. Annes Health Center

Emfeni Health Center: Emfeni Health Centre

Wenya Health Center: Wiliro Health Center

PARAMETER SETTINGS

Input

Case File: / Temporal analysis (2). cas

Population File: /

Time Precision: Year

Start Time: 2013/1/1.

End Time: 2020/12/31.

Coordinates File:

Coordinates: Latitude/Longitude

Analysis

Type of Analysis:

Scan for Areas with: Increasing or Decreasing Rates

Time Aggregation Units: Year

Time Aggregation Length: 4

Data Checking

Temporal Data Check: Check to ensure that all cases and controls are within the specified temporal study period.

Geographical Data Check: Check to ensure that all observations (cases, controls, and populations) are within the specified geographical area.

Spatial Neighbors

Use Non-Euclidean Neighbors file: No

Use Meta Locations File: No

Multiple Coordinates Type: Allow only one set of coordinates per location ID.

Locations Network

Use Locations Network File: No

Locations Network File:

Locations Network Purpose: Network Definition

Spatial Window

Maximum Spatial Cluster Size: 50 percent of population at risk

Window Shape: Circular

Cluster Restrictions

Minimum Cases in Cluster for High Rates: 2

Restrict High-Rate Clusters: No

Restrict Low-Rate Clusters: No

Space And Time Adjustments

Temporal Adjustment: None

Adjust for Weekly Trends, Nonparametric: No

Adjust for known relative risks: No

Inference

P-Value Reporting: Default Combination

Number of Replications: 999

Adjusting for More Likely Clusters: No

Cluster Drilldown

Standard Drilldown on Detected Clusters: No

Spatial Output

Automatically Launch Map: Yes

Compress KML File into KMZ File: No

Include All Location IDs in the Clusters: Yes

Cluster Location Threshold - Separate KML: 1000

Report Hierarchical Clusters: Yes

Criteria for Reporting Secondary Clusters: No Geographical Overlap

Restrict Reporting to Smaller Clusters: No

Other Output

Report Critical Values: No

Report Monte Carlo Rank: No

Print ASCII Column Headers: No

Run Options

Processor Usage: All Available Processors

Suppress Warnings: No

Logging Analysis: No

Program completed: Mon Jan 1 03:08:57 2018

Total Running Time: 20 seconds

Processor Usage: 4 processors