Spatial Patterns and Correlates of Lower Respiratory Tract Illnesses in Children from Drakenstein, Western Cape

Namhla Bhenxa

BHNNAM001

Submitted to the School of Public Health and Family Medicine

In partial fulfilment of the requirements for the degree of

Master of Public Health (Epidemiology)

Faculty of Health Sciences

University of Cape Town

Supervisor: Dr Kirsty Brittain

Co-Supervisors: Dr Jabulani Ncayiyana & Professor Heather J. Zar

Date: July 2021

Contents

Declaration

I, …Namhla Bhenxa…, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature:

Signed by candidate

Date: 06/07/2021

Acknowledgments

This research is my own work, I was responsible for data analysis and writing up of the manuscript. This study forms part of lung study from the Drakesntein Child Health Study (DCHS) would like to first thank God for the wisdom and courage to undertake this project to completion. This body of work would not have been possible without the substantial contribution guidance and support of my former supervisor Dr Jabulani Ncayiyana. I am also grateful to Dr Kirsty Brittain who stepped up and supported and guided me to finish this dissertation. This work would not have been possible without the help and input of Prof Maia Lesosky. I am grateful to the DCHS staff for recruitment, data collection and allowing the use of the data for the purpose of this project. This study was funded by national research foundation (NRF) Thuthuka fund. I would also like to acknowledge the financial assistance provided by Tiso Foundation which has funded my postgraduate studies. My grandmother Alicia Ntombi Benxa and father Bathandwa Benxa have supported my academic career financially, emotionally and mentally, they have given the best gift parents can ever gift a child. I am grateful to my sisters and friends for the support. Finally, I would like to thank the community of Drakenstein for participating in the study because without their participation and perseverance, none of this would have happened.

List of abbreviations

GWR – geographical weighed regression

LRTIs – Lower respiratory tract infections

URTIs – Upper respiratory tract infections

KDE – Kernel density estimation

DCHS - Drakenstein Child Health Study

GIS -geographic information system

GPS – Global positioning system

TB – Tuberculosis

ARI – Acute respiratory infections

RSV – Respiratory syncytial virus

GAM – Generalized additive model

SES – socio-economic status

LMIC – Low-middle income countries

HIC – High income countries

RR – Risk ratio

OR – Odds ratio

AIDS – Acquired immune-deficiency syndrome.

SaTScan – A software that analyses spatial data and detects space-time disease clusters

Morans I index – Correlation coefficient that measures the overall spatial autocorrelation

STATA – A statistical software

R – A statistical software

PART A: THE PROTOCOL

A.1 Background: Infections of the respiratory tract are the most common infections of the human host and account for a substantial proportion of patients seen by paediatricians (Madhi et al.,2000). These are classified as lower respiratory tract infections (LRTIs) and upper respiratory tract infections (URTIs) (Simoes *et al*.,2006). While URTIs consist of airways of nostrils to the larynx, LRTIs covers the continuation of airways to bronchi and bronchioles to the alveoli and the most common LRTIs are pneumonia, bronchiolitis, and influenza (Simoes *et al*.,2006). These are important causes of morbidity and mortality especially in developing countries with pneumonia being the biggest contributor of cases (Madhi et al.,2000; Troeger et al.,2017; Ye et al., 2009). In 2016, it was estimated that there were approximately 68 million episodes of LRTIs with an estimated 5 million hospital admissions among children and of those, about 652572 deaths globally were attributed to LRTIs among children (GBD 2016,2018). Despite the global advances to strengthen health systems like intensive immunization programmes, prevention of HIV/AIDS, improvement of socioeconomic status (SES), improved childhood nutrition and improved access to health care, the burden of LRTIs is still alarmingly high (Savitha et al.,2007; Mehta et al.,2013; le Roux et al.,2019).

Global health data also showed pneumonia as the leading cause of mortality and morbidity from LRTIs, and respiratory syncytial virus (RSV) as the second aetiology to contribute high proportion of cases and mortality, with up to 54% of LRTIs caused by RSV in children under 5 years (GBD 2016,2018). While influenza contributed the least to mortality, it is still the most common aetiology of LRTIs and results in most hospitalisations (GBD 2016,2018). It was suggested that there is a synergistic effect between pneumonia, bronchiolitis and other LRTIs with asthma thus increasing the burden (Puig et al.,2010). LRTIs are not only a public health concern because of the burden, but studies have suggested that LRTIs may predispose individuals to other chronic diseases in

childhood and adulthood such as chronic obstructive pulmonary disease (COPD), asthma and wheezing (Martinez, 2005). RSV has been strongly associated with chronic obstruction of airways among infants aged 2 to 3 years (Martinez et al,1998). The association between RSV and subsequent development of asthma has not been concisely elucidated however it is suggested that a third of children with LRTIs develop persistent episodes of wheezing by the age of 6 (Sigurs et al., 2010; Piedimonte et al.,2015). Studies show that respiratory infections of viral origin lead to acute wheezing in children and thus may be a risk factor for childhood asthma (Puig et al.,2010). It has been hypothesized that infection by viral pathogens at an early age might lead to alteration of immune response and interfere with lung development, thereby leading to inception of asthma (Yoshida et al.,2013).

Studies have shown that there has been a great reduction in LRTI episodes and mortality over the past decade although this varies geographically, with Asia and Africa being regions with a high prevalence (GBD 2016,2018; Murdoch and Howie, 2018). Although there has been a substantial global decrease in LTRIs, some countries like South Africa still failed to achieve the child millennium development goals (MDG) set for 2015. To achieve post 2015 MDG, collaborative efforts to combat the residual burden of LRTIs and other childhood illnesses should be made such as strong immunisation programmes, increase in health care access in disadvantaged areas, improving health care seeking behaviour and timely specific resource allocation. As part of the efforts to decrease LRTI-related mortality, focused and translational research needs to be conducted to further understand the epidemiology, aetiology and factors influencing the burden of LRTIs.

While a lot is known about the aetiology, risk factors and prevalence of LRTI, very few studies have been conducted on the spatial analysis of LRTIs clustering. It is important to incorporate space-time clustering in analysis because the risk factors for diseases such as climate, pollution, land use, and other factors tend to cluster together and are spatially dependant and therefore

analysis that does not account for spatial dependence may overestimate or underestimate the effect (Congdon , 2016 ; Poortinga, 2007). This means risk factors are not randomly distributed across populations. Studies suggest that lifestyle risk factors or behaviours such as smoking, unhealthy diets, malnutrition, and lack of exercise cluster within certain socio-demographic groups and are typically driven by socioeconomic status, education levels and availability of resources (Poortinga, 2007).

Analysis of space -time clustering of diseases could help identify disease patterns in a population, and high incidence areas or "hotspots" of disease in order to inform policy and help improve resource allocation, as well as target public health interventions directly to high burden areas to alleviate the burden (Crighton et al.,2007). It could also provide clues about the local risk factors. LRTIs follow certain patterns because some infectious agents like RSV and influenza occur seasonally, where cases peak around winter or colder temperature and LRTIs-associated risk factors such as low socio-economic status, smoking rates and health care services are spatially dependant (Vieira et al.,2001; Crighton et al.,2007; Mäkinen et al.,2009). This study helped uncover these disparities in populations and demonstrate high prevalence and incidence areas and facilitate implementation of prevention strategies (Vieira et al.,2001).

A.1.1 Problem statement: There is a high global burden of LRTIs such as bronchiolitis and pneumonia in children under the age of 2 years and these are principal causes of mortality and morbidity in children globally (Nair et a.,2013; Beck at al., 2015; Thomas et al.,2015). The high burden can be attributed to lack of access to the health care system and misallocation of resources and un-timely implementation of interventions (Thomas et al., 2015). To our knowledge, there is no published data on spatial clustering of LRTIs at the national, provincial, district and sub-district level in south Africa. While under - 5 mortalities are collected annually in the Western Cape province since 2011, routine data has shown a decrease in pneumonia deaths and hospital

admissions however the burden is still high and the district specific distribution of LRTIs is unknown (Bamford et al.,2018).

A.1.2. Rationale: Numerous studies have investigated risk factors for LRTIs. Spatial statistics have the potential to add important insights to this body of knowledge. Geospatial analysis would assist in identifying hotspots of LRTIs in children and the identification of spatiotemporal clusters could provide clues about local demographic, environmental, behavioural, socioeconomic, and genetic risk factors (Bamford et al.,2018). Geospatial data must be collected at the local level to inform policy makers on where to target health interventions and resource allocation for implementation of cost effective and timely control of diseases.

A.1.3. Aim: This study aims to investigate the spatial clustering of child LRTIs and associated individual-household- and community-level risk factors and correlates in the Drakenstein area of the Western Cape province, South Africa.

A.1.4 Research questions:

1. Do LRTIs cluster spatially among children under the age of 5 residing in Drakenstein?
2. What are factors that explain the observed spatial clustering of LRTIs?

1.1.6 Hypothesis: We hypothesize that the patterns of LRTIs outcomes will differ by geographic small areas in Drakenstein.

A.1.5 Objectives:

1. To analyse spatial patterns of childhood LRTIs in terms of geographic small areas in Drakenstein.

2. To identify the sociodemographic and environmental covariates that might best explain the observed patterns of childhood LRTIs in Drakenstein.

A.2 Literature review

A.2.1.1 Introduction: Respiratory tract infections which, depending on their location can be classified as lower respiratory infections and upper respiratory infections are the major cause of mortality especially in children under 5 years of age (Man et al.,2019; Mehta et al.,2013). LRTIs are a leading cause of morbidity and mortality. In 1990, there were an estimated 1.7 million LRTI-associated deaths in children, but over the years there has been a substantial decrease in LRTI mortality, with an estimated 700 000 deaths in 2015 (Le Roux et al.,2019; Feldman et al., 2019; Troeger et al., 2017; Mehta et al.,2013). Although there has been substantial progress in reducing morbidity, the burden is still high (Mehta et al.,2013). The residual LRTI-associated mortality undermines the global efforts made to combat the diseases. Infections that are considered as LRTIs are pneumonia, bronchiolitis, and influenza (Troeger et al.,2017).

A global study in 2016 has reported LRTIs as the sixth leading cause of disease. In total, LRTIs resulted in around 2.4 million cases and high number of fatalities in children younger than 5 years old (around 652,000 deaths) and in elderly people who are at least 70 years old (around 1.1 million deaths) (Troeger et al.,2017). Pneumonia in particular accounts for approximately 16% of LRTI related deaths (Troeger et al.,2017).

A.2.1.2 Pneumonia: Of all LRTIs, pneumonia is the biggest contributor to global LRTI mortality (Bénet et al., 2015; Zar and Ferkol, 2014; McAllister et al.,2019). Studies have established that Streptococcus pneumoniae and Haemophilus influenzae type b are the most common bacterial causes of pneumonia. A conjugate against Streptococcus pneumoniae and Haemophilus influenzae type b vaccine is available and has helped decrease the burden globally (Bénet et al., 2015; McAllister et al.,2019). However, pneumonia is one of the diseases that have been neglected and, as a result, there is limited up-to-date data on the burden and mortality. Several reasons for the neglect of pneumonia have been outlined, including the lack of

intervention programmes specific for vulnerable groups such as deprived communities, multiple aetiologies, and lack of consensus on how to best control pneumonia (Bénet et al., 2015; McAllister et al.,2019).

A.2.1.5. Wheezing: Wheezing is a common problem seen by physicians in children and it was estimated that at least 20 – 30 % of children have at least one wheezing episode during their childhood (Verwey et al.,2020; Ngwarai, 2019). The most documented causes of wheezing are asthma, allergies, infections, and gastroesophageal influx diseases (Piedimonte, 2015). Wheezing in children is known to follow seasonal patterns and may be persistent from birth. Seasonal wheezing may be associated with environmental factors while persistent wheezing may be caused by congenital abnormalities. Seasonal wheezing is mostly associated with infection by RSV which is common in winter (Verwey et al.,2020). Although there is an extensive literature on risk factors, studies often do not discriminate between wheezing and asthma (Barreto et al., 2010). Because of this, there are discrepancies in the current burden of wheeze in children (Barreto et al., 2010). Furthermore, wheezing remains underdiagnosed in LMIC and this also results in limited data on the burden and wheeze related mortality. It is suggested that wheezing is associated with socio-economic status, parental education, pollution, and smoking (Barreto et al., 2010).

A.2.1.7 The burden of LRTIS in South Africa: In South Africa, there is limited data published recently on the prevalence of LRTIS in children. According to WHO recent published data, in 2017 pneumonia and influenza accounted for 7.04% of all mortalities and was on 46th rank worldwide (Boyles *et al.,*2017). Since South Africa is a developing country with limited resources, inequality and low SES and insufficient vaccine coverage could potentially drive the burden in remote areas. According to stats SA, there were 1716 (8.3%) pneumonia and influenza related deaths in children within their first year of life. This is followed by 7.2 % of mortality

between the ages of 1-14. This shows that the estimates have heavily declined from the early 90's following the introduction of PCV into routine programmes in 2009 (Boyles *et al.,*2017). However, in 2016, these were a leading underlying cause of mortality in infants with Cape Town and Johannesburg among the top cities that contributed the highest burden (Stats SA). Lack of data indicates that these diseases have been neglected and there is no surveillance and research in South Africa.

A. 2.2 Risk factors for Lower respiratory tract illnesses

Previous studies have suggested that the determinants of LRTIs include age, parental and caregiver status, education, comorbidities (e.g., measles, diarrhoea, malaria), and environmental factors (Merera, 2021). Our study will focus on age, socio-economic status, parental smoking status and breastfeeding (Asri et al., 2021; Amsalu et al.,2016).

A.2.2.2 Household and interpersonal risk factors: COPD, including coughing and wheezing among other diseases, is normally instigated by inflammation due to smoking, exposure to fumes and burning (Tian et al.,2019). Involuntary exposure to tobacco smoke has been suspected to cause an array of respiratory problems. A systematic review has suggested that prenatal smoking by a mother increases the odds of acquiring LRTI by 1.24 while postnatal maternal smoking increases the odds of childhood LRTI by 1.58 (Jones et al.,2011). It was further shown that smoking by both parents increased the odds of childhood LRTI by 1.62 and paternal smoking increased the odds by 1.22 (Jones et al.,2011).

A cross sectional study by Imran et al. uncovered an association between sex of a child, and socio-economic status with acute respiratory infections. The findings of this study suggest that boys are more vulnerable to acute respiratory infection (ARI) than girls hence the high prevalence was apparent in boys (Imran et al., 2019).

A.2.2.3 Socio-economic risk factors: People living in low and middle socio-economic status are more vulnerable to LRTIs (Imran et al.,2019; Thörn et al.,2011; Mukai et al.,2009). Some evidence suggests that lack of exclusive breastfeeding within the first four months of life might lead to development of severe cases of LRTIs (Jackson et al.,2013; Savitha et al.,2007). Children under the age of 2 years are more susceptible to LRTIs than those who are above 2 years. This can be further divided, with those aged 0-11 months being highly susceptible, and 12-24 months being moderately susceptible to LRTIs (Imran et al.,2019).

A study showed that children from deprived SES were more vulnerable to pneumonia and should be targeted for intervention (Thörn et al.,2011). It was proposed that educating mothers about nutritional needs of their children could help reduce malnutrition which in turn would reduce the incidence of LRTIs (Imran et al.,2019; Jackson et al.,2013; Thörn et al.,2011). Parents and caregiver's literacy are important because studies show that consistent handwashing and hygiene can reduce community acquired pneumonia by 30% (Thörn et al.,2011). A study conducted in Korea, reported high incidence of pneumonia and Korea is a developed country with good infrastructure, good health system, economic growth and improved human living conditions (Kim et al.,2019). Therefore, it is very crucial to investigate factors that keep pneumonia prevalent even though it is preventable by vaccines and other precautions (Kim et al.,2019).

A.2.3 Spatial clustering of LRTIs

According to the report released by United Nations International Emergency Fund released in 2016, a substantial number of deaths occur in Sub-Saharan Africa (Asri et al., 2021). The incidence of LRTIs varies by countries and the differences can be attributed to child, maternal, and environmental related factors. While several studies on LRTIs have been conducted in Africa, more emphasis is still required on the geographic disparities. Therefore, spatial epidemiology remains an indispensable tool for identifying geographic risk factors for various diseases (Asri et

al., 2021). Previous studies have used spatial analysis to determine health outcomes hotspots and have confirmed that these techniques can be used to identify hotspots of infectious diseases (Asri et al., 2021).

Since public health is concerned with monitoring, identification and maintenance of health in the population and these are achieved by monitoring diseases, identifying health hazards and designing and implementing interventions to ameliorate those risks (Preim & Lawonn,2020; Chowell & Rothenberg,2018). The efforts aiming at further reducing LRTI mortality in children should strive to improve understanding of risk factors, increased equity to access of healthcare to vulnerable populations and identification of vulnerable populations. Exploration of spatial patterns has the potential to identify key populations and guide the implementation of interventions. Spatial analytical techniques have emerged and proved to be useful in understanding the epidemiology of infectious diseases, for example identifying novel risk factors as well as high risk areas which might benefit from public health intervention. In the past, knowledge of spatial variation of infectious diseases contributed to the understanding of disease distribution, diffusion, causal factors and mechanisms associated with risk of infection or developing a disease (Crighton et al.,2007).

Spatial analysis of COPD, TB and lung cancer in Beijing confirmed environmental and atmospheric factors as major risk factors for lung diseases and linked those factors to an increase in hospitalisation rates and spatial variations in incidence (Tian et al.,2019). It was also found that those regions with high burden were more exposed to risk factors, and these led to heterogeneity in hospitalisation and burden of disease (Tian et al.,2019). A study also found space time clusters of RSV cases in regions with high rainfall, confirming that moisture increases viability and transmissibility of RSV (Omer et al.,2008). In the Philippines, spatial analysis observed a high burden of pneumonia in densely populated areas and clusters of pneumonia in urban areas and the possible explanation for this observation is the health seeking behaviour where most people

in rural areas do not seek healthcare while those in rural areas have poor access to health care (Thomas et al.,2015). It was shown that space-time clustering of childhood pneumonia varies across different regions in the Netherlands and interestingly the incidence was different in 2013 and 2014, this can be explained by different climatic and environmental activities occurring at those locations (Benincà et al.,2017). These yearly variations need to be considered as they may disrupt resource allocation and targeting intervention since unexpected changes (Benincà et al.,2017).

Across all studies, spatial analysis has consistently shown that resource deprivation and low SES regions are hotspots of pneumonia and LRTIs in children. Since LTRIs are preventable diseases, interventions should be directed at the location-based burden because geography and ethnicity can lead to failure to access healthcare and therefore compromise child health in the population (Sartorius et al.,2011). A study that assessed child mortality patterns in South Africa showed that mortality was highest in impoverished areas and in children living with HIV/AIDS, with pneumonia among the biggest cause of child mortality (Sartorius et al.,2011). These studies show that geostatistical models can quantify health disparities in the population and may help predict disease patterns and child mortality. Space -time variation for diseases like LRTIs need more consideration as we have observed that influenza and pneumonia are season dependant where they peak during winter. This information is important for effective resource and service allocation.

A.2.4 Conclusions: Previous research has focused intensively on risk factors, the vaccine candidate, and the aetiology of LRTIs and not much on the geographic variation of the burden of LRTIs in children. This information has been an important cornerstone providing clues about the burden of LRTIs and led to vaccine development. Although the health system and health resources in developing countries have been decentralised, it is still difficult to allocate health resources and interventions to relevant people. Along with the existing research and information,

spatial analysis is therefore required as a blueprint to direct health resources, health workers and together with child vaccination and other interventions to prevent and control childhood diseases. Although many studies have conducted research on spatial clustering of LRTIs, they focus on hospitalisation rates using hospitalisation data which may be biased by health seeking behaviour, socio-economic status, and private health care. As can be noticed there are emerging infectious diseases including respiratory infections that have taken the world's public health system and has led to a shift in attention of researchers and neglect of LRTIs research. There is a growing need to collect spatial health data for LRTIs in South Africa because risk factors cluster and occur concurrently, thus geostatistical data would help to generate hypotheses and guide surveillance and direct interventions.

A.3 Methods

A.3.1 Ethics: The primary study on which this research will draw was approved by the Human Research Ethics Committee of the University of Cape Town's Faculty of Health Sciences and the Stellenbosch University Ethics Committee. Written informed consent was obtained from mothers prior to enrolment. The informed consent is renewed annually, and participants are free to discontinue from the study any time. This study will conduct a secondary data analysis, and separate ethics approval will be sought from the University of Cape Town's Faculty of Health Sciences Human Research Ethics Committee (HREC).

A.3.2 Study design: This study will include secondary analysis of data collected from an ongoing population-based birth cohort of ~1000 mother-child pairs. The details of the population-based birth cohort and study procedures have been previously published (Zar et al.,2015). Comprehensive clinical, socioeconomic, behavioural, and demographic data have already been collected in this cohort of ~1000 mother-child pairs, with state-of-the-art measures of child lung health.

A.3.3 Study setting: Drakenstein is one of the five local municipalities located within the Cape Winelands District of the Western Cape Province, South Africa. The study setting covers an area of 1538 km^2, with a population of ~200 000 people. Unemployment in the area is high, with many people of low socio-economic status. The birth cohort was recruited from two former township communities namely Mbekweni and TC Newman. The burden of infectious diseases such as TB and HIV is high in these communities (Zar et al.,2015).

A.3.4 Study participants: The participants for this study will be children up to the age of five years in the DCHS. Mothers received compensation for their time and the children received health benefits, free health care and regular check-ups. The communities were under an ongoing surveillance for emerging diseases in children and prevention interventions are implemented. Parents receive education on risk factors for lung diseases and ways to prevent them.

A.3.5 Data collection and analysis

A.3.5.1 Childhood pneumonia and wheeze: Data on LRTIs (pneumonia and wheezing) will be obtained from active surveillance systems for LRTIs in the DCHS. LRTIs are measured using the World Health Organization (WHO) case definition criteria. Data on wheezing illnesses will be obtained from active surveillance database for LRTIs in the DCHS and maternal self-reports at each of the study visits and medically ascertained visits for wheezing (Zar et al.,2015). Wheezing is categorised into wheezing if the child experienced at least one episode of wheeze during the first 5 years of life (MacGinty et al.,2018).

A.3.5.2 Geospatial data collection: Two research assistants were recruited and trained to collect latitude and longitude coordinates using a hand-held geographic positioning system (GPS) device (eTrex 10, Garmin). The research assistants visited all the households of the DCHS participants and recorded latitude and longitude coordinates of each household.

A.3.6 Variables to be included in the analysis.

Exposures: Exposures were measured at study measurement visits using standardised questionnaires by trained study nurses. Study variables measured are listed in Table 1.

Table 1: Baseline characteristics and exposures related to LRTIs.

Exposure	Variable Type	Categories
Sex	Categorical	• Female • Male
Ethnicity	Categorical	• Black • Mixed ancestry
Feeding practices at 6 weeks postpartum	Categorical	• Exclusive breastfeeding • Mixed breastfeeding
Study site	Categorical	• Mbekweni • TC Newman
SES-quartile	Categorical	• Lowest SES • Low-moderate-SES • Moderately high SES • High SES
Electricity	Categorical	• Yes • No
Caregiver Smoking categories	Categorical	• Active smoker • Passive smoker • Non-smoker
Caregiver Education	Categorical	• Tertiary education • Completed secondary • Some secondary • Primary school

A.3.6 Data analysis

A.3.6.1 Objective 1: Spatial distribution of LRTI (pneumonia and wheezing) will be depicted in kernel density maps using the R Studio 4.0.2 (2020-06-22). The quadrat analysis, kernel estimation, nearest neighbour distances and k functions for point pattern analysis will be used to assess spatial variations and to count number of cases per area/zone. The regular shaped Kulldorf spatial-temporal and irregular shaped Flexible scan statistics will be used to detect clusters of higher-than-expected cases of LRTI and secondary clusters if there is geographic

overlap. Clusters can be circular and irregular depending on geographic position and landscape. A cluster will be detected if the highest likelihood ratio is statistically significant. The statistical significance of the clusters will be determined by Monte Carlo simulation. The scan statistic will be performed in the SaTScan software.

A.3.6.2 Objective 2: The association between location and childhood pneumonia and wheeze will be investigated using generalized additive models (GAM) from the MapGAM package in R Studio to identify the geospatial and sociodemographic covariates associated with the observed patterns of these LRTIs. GAMs are semi-parametric extensions of GLMs; the only underlying assumption made is that the functions are additive and that the components are smooth (Hastie and Tibshirani, 1986, 1990).

A.3.6.3 Risks and benefits: This is secondary data analysis and poses no risks to the study participants. Confidentiality will be maintained through the secure storage of data and use of unique patient identifiers. Paper-based and electronic study case report forms were securely stored and only accessible to authorised personnel.

A.3.6.4 Dissemination of Results

The results of this study will be submitted in partial fulfillment of the requirements for the Master of Public Health degree at the University of Cape Town. This study will be using geographic location (coordinates) of study participants for geospatial analysis and the coordinates will not be published to maintain the confidentiality of the study participants.

A.4 References

Asher, I. and Pearce, N., 2014. Global burden of asthma among children. *The international journal of tuberculosis and lung disease*, *18*(11), pp.1269-1278.

Asri AK, Pan WC, Lee HY, Su HJ, Wu CD, Spengler JD. Spatial patterns of lower respiratory tract infections and their association with fine particulate matter. Scientific reports. 2021 Mar 1;11(1):1-2.

Barreto, M.L., Cunha, S.S., Fiaccone, R., Esquivel, R., Amorim, L.D., Alvim, S., Prado, M., Cruz, A.A., Cooper, P.J., Santos, D.N. and Strina, A., 2010. Poverty, dirt, infections and non-atopic wheezing in children from a Brazilian urban center. *Respiratory research*, *11*(1), pp.1-10.

Bamford, L., Barron, P., Kauchali, S. and Dlamini, N., 2018. Inpatient case fatality rates improvements in children under 5: Diarrhoeal disease, pneumonia and severe acute malnutrition. *South African Medical Journal*, *108*(3), pp.33-37.

Beck, A.F., Florin, T.A., Campanella, S. and Shah, S.S., 2015. Geographic variation in hospitalization for lower respiratory tract infections across one county. *JAMA pediatrics*, *169*(9), pp.846-854.

Benincà, E., van Boven, M., Hagenaars, T. and van der Hoek, W., 2017. Space-time analysis of pneumonia hospitalisations in the Netherlands. *PloS one*, *12*(7).

Bénet T, Sylla M, Messaoudi M, Sánchez Picot V, Telles JN, Diakite AA, Komurian-Pradel F, Endtz H, Diallo S, Paranhos-Baccalà G, Vanhems P. Etiology and factors associated with pneumonia in children under 5 years of age in Mali: a prospective case-control study. PloS one. 2015 Dec 22;10(12):e0145447.

Boyles, T.H., Brink, A., Calligaro, G.L., Cohen, C., Dheda, K., Maartens, G., Richards, G.A., van Zyl Smit, R., Smith, C., Wasserman, S. and Whitelaw, A.C., 2017. South African guideline for the management of community-acquired pneumonia in adults. *Journal of thoracic disease*, *9*(6), p.1469.

Cliff, A.D., Haggett, P. and Ord, J.K., 1986. *Spatial aspects of influenza epidemics*. Routledge.

Cliff, A.D. and Smallman-Raynor, M.R., 1992. The AIDS pandemic: Global geographical patterns and local spatial processes. *Geographical Journal*, pp.182-198.

Crighton, E.J., Elliott, S.J., Moineddin, R., Kanaroglou, P. and Upshur, R.E.G., 2007. An exploratory spatial analysis of pneumonia and influenza hospitalizations in Ontario by age and gender. *Epidemiology & Infection*, *135*(2), pp.253-261.

Congdon, P., 2016. Spatiotemporal frameworks for infectious disease diffusion and epidemiology.

Chowell G, Rothenberg R. Spatial infectious disease epidemiology: on the cusp. BMC medicine. 2018 Dec;16(1):1-5.

Flory, J.H., Joffe, M., Fishman, N.O., Edelstein, P.H. and Metlay, J.P., 2009. Socioeconomic risk factors for bacteraemic pneumococcal pneumonia in adults. *Epidemiology & Infection*, *137*(5), pp.717-726.

Everard, M.M., Bara, A., Kurian, M., Elliott, T.M. and Ducharme, F.M., 2002. Anticholinergic drugs for wheeze in children under the age of two years. *Cochrane Database of Systematic Reviews*, (1).

GBD 2016 Lower Respiratory Infections Collaborators, 2018. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet. Infectious diseases*, *18*(11), p.1191.

Gray, D.M., Turkovic, L., Willemse, L., Visagie, A., Vanker, A., Stein, D.J., Sly, P.D., Hall, G.L. and Zar, H.J., 2017. Lung function in African infants in the Drakenstein child health study. Impact of lower respiratory tract illness. *American journal of respiratory and critical care medicine*, *195*(2), pp.212-220.

Graham, A.J., Atkinson, P.M. and Danson, F.M., 2004. Spatial analysis for epidemiology.

Gehring, U., Cyrys, J., Sedlmeir, G., Brunekreef, B., Bellander, T., Fischer, P., Bauer, C.P., Reinhardt, D., Wichmann, H.E. and Heinrich, J., 2002. Traffic-related air pollution and respiratory health during the first 2 yrs of life. *European respiratory journal*, *19*(4), pp.690-698.

Gray, D.M., Turkovic, L., Willemse, L., Visagie, A., Vanker, A., Stein, D.J., Sly, P.D., Hall, G.L. and Zar, H.J., 2017. Lung function in African infants in the Drakenstein child health study. Impact of lower respiratory tract illness. *American journal of respiratory and critical care medicine*, *195*(2), pp.212-220.

Imran, M.I.K., Inshaf, M.U.A.,Chowdhury, M.A.B, Uddin,M.J.,2019. Risk factors for acute respiratory infection in children younger than five years in Bangladesh. Elsevier Public Health 173(2019) 117-119.

Jackson, S., Mathews, K.H., Pulanić, D., Falconer, R., Rudan, I., Campbell, H. and Nair, H., 2013. Risk factors for severe acute lower respiratory infections in children–a systematic review and meta-analysis. *Croatian medical journal*, *54*(2), pp.110-121.

Jartti, T., Lehtinen, P., Vuorinen, T., Österback, R., van den Hoogen, B., Osterhaus, A.D. and Ruuskanen, O., 2004. Respiratory picornaviruses and respiratory syncytial virus as causative agents of acute expiratory wheezing in children. *Emerging infectious diseases*, *10*(6), p.1095.

Jacquez, G.M., 2000. Spatial analysis in epidemiology: Nascent science or a failure of GIS?. *Journal of Geographical Systems*, *2*(1), pp.91-97.

Johnson, C.P. and Johnson, J., 2001. GIS: a tool for monitoring and management of epidemics. *Proceedings of Map India*.

Jones, L.L., Hashim, A., McKeever, T., Cook, D.G., Britton, J. and Leonardi-Bee, J., 2011. Parental and household smoking and the increased risk of bronchitis, bronchiolitis and other lower respiratory infections in infancy: systematic review and meta-analysis. *Respiratory research*, *12*(1), p.5.

Karki S, Fitzpatrick AL, Shrestha S. Risk factors for pneumonia in children under 5 years in a teaching hospital in Nepal. Kathmandu University Medical Journal. 2014;12(4):247-52.

Kazembe, L.N., Muula, A.S., Appleton, C.C. and Kleinschmidt, I., 2007. Modelling the effect of malaria endemicity on spatial variations in childhood fever, diarrhoea, and pneumonia in Malawi. *International Journal of Health Geographics*, *6*(1), p.33.

Kim, A.M., Kang, S., Park, J.H., Yoon, T.H. and Kim, Y., 2019. A spatial analysis of geographic variation and factors associated with hospitalization for bacterial pneumonia in Korea. *BMC pulmonary medicine*, *19*(1), p.45.

Kistemann, T., Dangendorf, F. and Schweikart, J., 2002. New perspectives on the use of Geographical Information Systems (GIS) in environmental health sciences. *International journal of hygiene and environmental health*, *205*(3), pp.169-181.

le Roux, D.M., Nicol, M.P., Myer, L., Vanker, A., Stadler, J.A., von Delft, E. and Zar, H.J., 2019. Lower respiratory tract infections in children in a well-vaccinated South African birth cohort: spectrum of disease and risk factors. *Clinical Infectious Diseases*, *69*(9), pp.1588-1596.

Lovasi, G.S., O'Neil-Dunne, J.P., Lu, J.W., Sheehan, D., Perzanowski, M.S., MacFaden, S.W., King, K.L., Matte, T., Miller, R.L., Hoepner, L.A. and Perera, F.P., 2013. Urban tree canopy and asthma, wheeze, rhinitis, and allergic sensitization to tree pollen in a New York City birth cohort. *Environmental health perspectives*, *121*(4), pp.494-500.

Mukai, A.D.O., Alves, K.D.S.C. and Nascimento, L.F.C., 2009. Spatial analysis of hospitalizations for pneumonia in the Vale do Paraíba region of Brazil. *Jornal Brasileiro de Pneumologia*, *35*(8).

Martinez, L., le Roux, D.M., Barnett, W., Stadler, A., Nicol, M.P. and Zar, H.J., 2018. Tuberculin skin test conversion and primary progressive tuberculosis disease in the first 5 years of life: a birth cohort study from Cape Town, South Africa. *The Lancet Child & Adolescent Health*, *2*(1), pp.46-55.

Mäkinen, T.M., Juvonen, R., Jokelainen, J., Harju, T.H., Peitso, A., Bloigu, A., Silvennoinen-Kassinen, S., Leinonen, M. and Hassi, J., 2009. Cold temperature and low humidity are associated with increased occurrence of respiratory tract infections. *Respiratory medicine*, *103*(3), pp.456-462.

MacGinty, R.P., Lesosky, M., Barnett, W., Stein, D.J. and Zar, H.J., 2018. Associations between maternal mental health and early child wheezing in a South African birth cohort. *Pediatric pulmonology*, *53*(6), pp.741-754.

McAllister DA, Liu L, Shi T, Chu Y, Reed C, Burrows J, Adeloye D, Rudan I, Black RE, Campbell H, Nair H. Global, regional, and national estimates of pneumonia morbidity and mortality in children younger than 5 years between 2000 and 2015: a systematic analysis. The Lancet Global Health. 2019 Jan 1;7(1):e47-57.

Mehta, S., Shin, H., Burnett, R., North, T. and Cohen, A.J., 2013. Ambient particulate air pollution and acute lower respiratory infections: a systematic review and implications for estimating the global burden of disease. *Air Quality, Atmosphere & Health*, *6*(1), pp.69-83.

Merera AM. Determinants of acute respiratory infection among under-five children in rural Ethiopia. BMC infectious diseases. 2021 Dec;21(1):1-2.

Meliker, J.R., Slotnick, M.J., AvRuskin, G.A., Kaufmann, A., Jacquez, G.M. and Nriagu, J.O., 2005. Improving exposure assessment in environmental epidemiology: Application of spatio-temporal visualization tools. *Journal of Geographical Systems*, *7*(1), pp.49-66.

Middelkoop, K., Bekker, L.G., Morrow, C., Zwane, E. and Wood, R., 2009. Childhood tuberculosis infection and disease: a spatial and temporal transmission analysis in a South African township. *South African medical journal*, *99*(10).

Mirsaeidi, M., Motahari, H., Taghizadeh Khamesi, M., Sharifi, A., Campos, M. and Schraufnagel, D.E., 2016. Climate change and respiratory infections. *Annals of the American Thoracic Society*, *13*(8), pp.1223-1230.

Morgenstern, V., Zutavern, A., Cyrys, J., Brockow, I., Gehring, U., Koletzko, S., Bauer, C.P., Reinhardt, D., Wichmann, H.E. and Heinrich, J., 2007. Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children. *Occupational and environmental medicine*, *64*(1), pp.8-16.

Murdoch, D.R. and Howie, S.R., 2018. The global burden of lower respiratory infections: making progress, but we need to do better. *The Lancet Infectious Diseases*, *18*(11), pp.1162-1163.

Najafabadi, A.T., 2009. Applications of GIS in health sciences. *Shiraz E-Medical Journal*, *10*(4), pp.221-230.

Nair, H., Simões, E.A., Rudan, I., Gessner, B.D., Azziz-Baumgartner, E., Zhang, J.S.F., Feikin, D.R., Mackenzie, G.A., Moïïsi, J.C., Roca, A. and Baggett, H.C., 2013. Global and regional burden of hospital admissions for severe acute lower respiratory infections in young children in 2010: a systematic analysis. *The Lancet*, *381*(9875), pp.1380-1390.

Nelson, M.R., Orum, T.V., Jaime-Garcia, R. and Nadeem, A., 1999. Applications of geographic information systems and geostatistics in plant disease epidemiology and management. *Plant disease*, *83*(4), pp.308-319.

Ngwarai MR., 2019. The stool microbiota and infant wheezing illness-the Drakenstein child health study, *South Africa* (Master's thesis, Faculty of Health Sciences).

Nuckols, J.R., Ward, M.H. and Jarup, L., 2004. Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environmental health perspectives*, *112*(9), pp.1007-1015.

Nykiforuk, C.I. and Flaman, L.M., 2011. Geographic information systems (GIS) for health promotion and public health: a review. *Health promotion practice*, *12*(1), pp.63-73.

Ouédraogo AM, Crighton EJ, Sawada M, To T, Brand K, Lavigne E. Exploration of the spatial patterns and determinants of asthma prevalence and health services use in Ontario using a Bayesian approach. PloS one. 2018 Dec 10;13(12):e0208205

Omer, S.B., Sutanto, A., Sarwo, H., Linehan, M., Djelantik, I.G.G., Mercer, D., Moniaga, V., Moulton, L.H., Widjaya, A., Muljati, P. and Gessner, B.D., 2008. Climatic, temporal, and geographic characteristics of respiratory syncytial virus disease in a tropical island population. *Epidemiology & Infection*, *136*(10), pp.1319-1327.

Piedimonte G. RSV infections: state of the art. Cleve Clin J Med. 2015 Nov 1;82(11 Suppl 1):S13-8.

Poortinga, W., 2007. The prevalence and clustering of four major lifestyle risk factors in an English adult population. *Preventive medicine*, *44*(2), pp.124-128.

Puig, C., Fríguls, B., Gómez, M., García-Algar, Ó., Sunyer, J. and Vall, O., 2010. Relationship between lower respiratory tract infections in the first year of life and the development of asthma and wheezing in children. *Archivos de Bronconeumología (English Edition)*, *46*(10), pp.514-521.

Preim B, Lawonn K. A survey of visual analytics for public health. InComputer Graphics Forum 2020 Feb (Vol. 39, No. 1, pp. 543-580).

Rytkönen, M.J., 2004. Not all maps are equal: GIS and spatial analysis in epidemiology. *International journal of circumpolar health*, *63*(1), pp.9-24.

Ruankaew, N., 2005. GIS and epidemiology. *Journal of the Medical Association of Thailand= Chotmaihet thangphaet*, *88*(11), pp.1735-1738.

Rutherford, S., Simpson, R., Williams, G., Mitchell, C. and McCall, B., 2000. Relationships between environmental factors and lung function of asthmatic subjects in south east Queensland, Australia. *Journal of occupational and environmental medicine*, *42*(9), pp.882-891.

Rutto, J.J. and Karuga, J.W., 2009. Temporal and spatial epidemiology of sleeping sickness and use of geographical information system (GIS) in Kenya. *Journal of Vector Borne Diseases*, *46*(1), p.18.

Ryan, P.H., LeMasters, G.K., Biswas, P., Levin, L., Hu, S., Lindsey, M., Bernstein, D.I., Lockey, J., Villareal, M., Khurana Hershey, G.K. and Grinshpun, S.A., 2007. A comparison of proximity and land use regression traffic exposure models and wheezing in infants. *Environmental health perspectives*, *115*(2), pp.278-284.

Sahsuvaroglu, T., Jerrett, M., Sears, M.R., McConnell, R., Finkelstein, N., Arain, A., Newbold, B. and Burnett, R., 2009. Spatial analysis of air pollution and childhood asthma in Hamilton, Canada: comparing exposure methods in sensitive subgroups. *Environmental Health*, *8*(1), p.14.

Sartorius, B., Kahn, K., Collinson, M.A., Vounatsou, P. and Tollman, S.M., 2011. Survived infancy but still vulnerable: spatial-temporal trends and risk factors for child mortality in rural South Africa (Agincourt), 1992-2007. *Geospatial health*, *5*(2), p.285.

Savitha, M.R., Nandeeshwara, S.B., Kumar, M.P. and Raju, C.K., 2007. Modifiable risk factors for acute lower respiratory tract infections. *The Indian Journal of Pediatrics*, *74*(5), pp.477-482.

Seng, S.B., Chong, A.K. and Moore, A., 2005, November. Geostatistical modelling, analysis and mapping of epidemiology of dengue fever in Johor State, Malaysia. In *The 17th Annual Colloquium of the Spatial Information Research Centre, University of Otago, Dunedin, New Zealand* (pp. 24-25).

Sigurs N, Aljassim F, Kjellman B, Robinson PD, Sigurbergsson F, Bjarnason R, Gustafsson PM. Asthma and allergy patterns over 18 years after severe RSV bronchiolitis in the first year of life. Thorax. 2010 Dec 1;65(12):1045-52.

Simoes, E.A., Cherian, T., Chow, J., Shahid-Salles, S.A., Laxminarayan, R. and John, T.J., 2006. Acute respiratory infections in children. In *Disease Control Priorities in Developing Countries. 2nd edition*. The International Bank for Reconstruction and Development/The World Bank.

Spengler, J.D., Jaakkola, J.J., Parise, H., Katsnelson, B.A., Privalova, L.I. and Kosheleva, A.A., 2004. Housing characteristics and children's respiratory health in the Russian Federation. *American journal of public health*, *94*(4), pp.657-662.

Thomas, D.S., Anthamatten, P., Root, E.D., Lucero, M., Nohynek, H., Tallo, V., Williams, G.M., Simões, E.A., ARIVAC Consortium, Mulholland, K. and Klugman, K., 2015. Disease mapping for informing targeted health interventions: childhood pneumonia in Bohol, Philippines. *Tropical Medicine & International Health*, *20*(11), pp.1525-1533.

Thörn, L.K., Minamisava, R., Nouer, S.S., Ribeiro, L.H. and Andrade, A.L., 2011. Pneumonia and poverty: a prospective population-based study among children in Brazil. *BMC infectious diseases*, *11*(1), p.180.

Tian, L., Yang, C., Zhou, Z., Wu, Z., Pan, X. and Clements, A.C., 2019. Spatial patterns and effects of air pollution and meteorological factors on hospitalization for chronic lung diseases in Beijing, China. *Science China Life Sciences*, pp.1-8.

Troeger, C.E., Blacker, B.F., Khalil, I.A., Zimsen, S.R., Albertson, S.B., Abate, D., Abdela, J., Adhikari, T.B., Aghayan, S.A., Agrawal, S. and Ahmadi, A., 2019. Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: an analysis for the Global Burden of Disease Study 2017. *The Lancet Respiratory Medicine*, *7*(1), pp.69-89.

Omer, S.B., Sutanto, A., Sarwo, H., Linehan, M., Djelantik, I.G.G., Mercer, D., Moniaga, V., Moulton, L.H., Widjaya, A., Muljati, P. and Gessner, B.D., 2008. Climatic, temporal, and geographic characteristics of respiratory syncytial virus disease in a tropical island population. *Epidemiology & Infection*, *136*(10), pp.1319-1327.

Vanker, A., Gie, R.P. and Zar, H.J., 2018. Early-life exposures to environmental tobacco smoke and indoor air pollution in the Drakenstein Child Health Study: Impact on child health. *SAMJ: South African Medical Journal*, *108*(2), pp.71-72.

Verwey C, Nunes MC. RSV lower respiratory tract infection and lung health in the first 2 years of life. The Lancet Global Health. 2020 Oct 1;8(10):e1247-8.

Vieira, S.E., Stewien, K.E., Queiroz, D.A., Durigon, E.L., Török, T.J., Anderson, L.J., Miyao, C.R., Hein, N., Botosso, V.F., Pahl, M.M. and Gilio, A.E., 2001. Clinical patterns and seasonal trends in respiratory syncytial virus hospitalizations in São Paulo, Brazil. *Revista do Instituto de Medicina Tropical de Sao Paulo*, *43*(3), pp.125-131.

Wang, J., Xiong, J., Yang, K., Peng, S. and Xu, Q., 2010, June. Use of GIS and agent-based modeling to simulate the spread of influenza. In *2010 18th International Conference on Geoinformatics* (pp. 1-6).

Weiss, L.N., 2008. The diagnosis of wheezing in children. *American family physician*, *77*(8), pp.1109-1114.

Wichmann, J., Wolvaardt, J.E., Maritz, C. and Voyi, K.V., 2009. Household conditions, eczema symptoms and rhinitis symptoms: relationship with wheeze and severe wheeze in children living in the Polokwane area, South Africa. *Maternal and child health journal*, *13*(1), pp.107-118.

Ye, Y., Zulu, E., Mutisya, M., Orindi, B., Emina, J. and Kyobutungi, C., 2009. Seasonal pattern of pneumonia mortality among under-five children in Nairobi's informal settlements. *The American journal of tropical medicine and hygiene*, *81*(5), pp.770-775.

Zar, H.J., Barnett, W., Myer, L., Stein, D.J. and Nicol, M.P., 2015. Investigating the early-life determinants of illness in Africa: the Drakenstein Child Health Study. *Thorax*, *70*(6), pp.592-594.

Zar, H.J., Barnett, W., Myer, L. and Nicol, M.P., 2016. Childhood pneumonia: The Drakenstein Child Health Study. *SAMJ: South African Medical Journal*, *106*(7), pp.642-643.

Zar, H.J. and Ferkol, T.W., 2014. The global burden of respiratory disease—impact on child health. *Pediatric pulmonology*, *49*(5), pp.430-434.

Zhong, S., Xue, Y., Cao, C., Cao, W., Li, X., Guo, J. and Fang, L., 2005, July. The application of space/time analysis tools of GIS in spatial epidemiology: a case study of Hepatitis B in

China using GIS. In *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05.* (Vol. 3, pp. 1612-1615).

Part B: MANUSCRIPT

B.1 Abstract

Background: Lower respiratory tract infections (LRTIs) in children under 5 years of age are a major public health concern globally with high proportions of cases occurring in low- and middle-income countries. However, little is known about the spatial variation of LRTIs especially in developing countries such as South Africa. This study aimed to explore and map the spatial dependence and clustering of childhood pneumonia and wheeze and to identify factors associated with these diseases in two communities of the Western Cape province in South Africa.

Methods: This analysis used case-control data from the Drakenstein Child Health Study. Cases were children who developed pneumonia or wheeze, and controls were incidence density matched on date of birth, age, and enrolment site. Smooth maps were created using Kernel density estimation and cluster analysis was performed using Ripley's K function and Kullddorf spatial scan statistics. Finally, generalized additive models were used to identify socio-demographic factors associated with pneumonia and wheeze.

Results: A total of 947 children were included in analysis, with 406 cases of wheeze and 466 cases of pneumonia. Overall, cases and controls had a similar spatial concentration, although there was some variation in the density of wheezing cases in comparison to controls. In particular, there was evidence of a non-random difference in the distribution between cases and controls, and cases of wheeze were significantly more concentrated the TC Newman area than expected. Those who developed pneumonia and wheeze were more likely to be males as compared to females (both $p<0.05$). There were no statistically significant association between socioeconomic status, ethnicity, education, breastfeeding practices and maternal smoking (all $p>0.05$).

Conclusions: This study revealed a significant clustering of wheeze in TC Newman, which may be attributable to socio-demographic differences between communities, while pneumonia was spatially dispersed across the study areas. These findings suggest that focused public health interventions are needed to reduce the incidence of LRTIs in the area of TC Newman specifically.

Keywords: Cluster, Kernel density, pneumonia, wheezing, Drakenstein, spatial, LRTIs, mapping.

B.2. Background:

During the past decade, there has been a substantial decrease in the burden of lower respiratory tract infections (LRTIs), including pneumonia, in children under 5 years of age (Feldman and Shaddock, 2019). These decreases may be attributed to the roll out of the pneumococcal conjugate and Hib vaccines as well as improvements in socioeconomic status (SES) and improved access to health care (Knobbe et al., 2019). Despite these improvements, LRTIs remain a major public health concern and are a leading cause of morbidity especially in children under 5 years of age residing in low- and middle-income countries (LMICs) (Rahman and Shahidullah, 2001; Knobbe et al., 2019). It is estimated that over 90% of cases occur in LMICs, with about 70% of those being from Sub Saharan Africa (SSA) (Seidu et al.,2019). The World Health Organization (WHO) estimated that 800 000 deaths in children under 5 years of age were attributed to LRTIs in 2016 (Knobbe et al., 2019), with pneumonia remaining the leading cause of morbidity (Seidu et al., 2019; Hossain et al., 2019). The reduction in child mortality from preventable diseases is one of the sustainable development goals set for 2030 (Pina et al.,2020). However, it is estimated that pneumonia will be responsible for 735,000 deaths of children under 5 years of age annually by 2030, if uncontrolled (Pina et al., 2020).

Alongside the high burden of pneumonia, wheezing in early childhood is common. It is estimated that at least 50% of children in high-income settings experience an episode of wheezing before the age of 6 years (MacGinty et al., 2018). However, wheezing data in LMICs are limited due to poor diagnosis, limited healthcare service data, poor health seeking behavior, and poor research output. Consequently, the burden of wheezing in Africa is unknown. Wheezing is usually indicative of an underlying undiagnosed respiratory disease, poor lung function and asthma (Adeloye et al., 2013; MacGinty et al., 2018).

Commonly documented risk factors for LRTIs include malnutrition, non-exclusive breast feeding, overcrowding, air pollution and lack of immunization (Boloursaz et al., 2013; Seidu et al., 2019; Adane et al., 2020). Some studies have suggested that parental smoking and zinc deficiency may be risk factors for LRTIs (Boloursaz et al., 2013; Seidu et al., 2019). Moreover, parent's education has been found to be associated with a high burden of LRTIs (Boloursaz et al., 2013; Seidu et al., 2019). In parallel with these factors, urbanization, an increasing population in urban areas and the adoption of a Western lifestyle have been associated widely with wheezing and asthma (Adeloye et al., 2013). Furthermore, the issue of social inequalities continues to play a crucial role in the burden of LRTIs in developing countries because of deprivation of primary health care (Pina et al., 2020). Although asthma and wheezing were previously considered uncommon in children in LMICs, recent studies show that the prevalence is greater than previously thought in these settings and comparable to high income countries (Le Roux et al., 2015). Early viral LRTIs and genetic predisposition also elevate the risk of wheezing in children (MacGinty et al., 2018). In parallel with socio-demographics factors, environmental factors such as pollution, particulate matter (PM), other atmospheric gases and climatic conditions play a crucial role in the rise of the incidence of LRTIs (Gehring at al., 2002).

The distribution of LRTIs is not uniform across provinces or regions, and spatial analysis is critical to understand LRTIs in different geo-locations. Understanding LRTIs in a specific context requires knowledge of the social, environmental, health and geographical resources of a particular setting. Spatial analytical techniques and models are often used in epidemiology to identify spatial hotspots in disease surfaces and to understand spatial structure (Hinman et al., 2006). From John Snow's Victorian era map of cholera deaths to interactive maps tracking the spread of H1N1 Influenza, spatial point patterns have a long and rich history in the public health arena (Carlos et al., 2010). Spatial analysis in high-income countries has been used to identify regions with

significantly higher cases of LRTIs and regions with lower cases of LRTIs grouped near to each other in space that have a low probability of having occurred by chance (Beamer et al., 2016). Disease registries in high-income countries now include geolocation data, which allows the detection of clusters to enable public health practitioners to use point pattern analysis to quantify social determinants of health, but these data are not available in LMICs (Carlos et al., 2010). Studies conducted on the spatial patterns of LRTIs have identified distinct patterns of clusters for LRTI phenotypes, with those clusters associated with community-level risk factors including air pollution, poor housing and low SES (Mollalo et al., 2020; Beamer et al., 2016). To our knowledge, no study has been conducted on the spatial patterns of LRTIs in a small area in South Africa. Knowledge of spatial distribution and patterns would be the first step in identifying high-risk populations to help target interventions and resources and would assist in identifying hotspots of LRTIs in children. In addition, the identification of spatiotemporal clusters could provide clues about local demographic, environmental, behavioural, SES and genetic risk factors. To address these gaps, this study investigated the spatial clustering of childhood pneumonia and wheeze and the associated individual-, household- and community-level risk factors and correlates in Drakenstein in the Western Cape province of South Africa.

Specific objectives of this study were to (1) map the distribution of pneumonia and wheeze; (2) assess the spatial dependence and clustering of pneumonia and wheeze; and (3) identify risk factors for pneumonia and wheeze.

B.3 Methods:

B.3.1 Study setting and design:

These data arise from the Drakenstein Child Health Study (DCHS). Drakenstein is one of the five local municipalities located within the Cape Winelands District of the Western Cape Province in South Africa. The Drakenstein municipality covers an area of 1538 km^2, with a population of ~200 000 people. The study was conducted in two regions within the municipality, namely TC Newman and Mbekweni. Residents of these areas are ethnically, linguistically and culturally diverse, with TC Newman serving a population comprising of predominantly Afrikaans speaking individuals with mixed ancestry, and Mbekweni serving mainly IsiXhosa speaking black Africans.

Details of the birth cohort and study procedures have been previously published (Zar et al., 2015). Briefly, pregnant women were recruited from March 2012 into the cohort and were eligible to participate if they were aged 18 years or older and between 20-28 weeks gestation. Follow-up of ~1000 mother-child pairs is ongoing. The study was approved by the Human Research Ethics Committee of the University of Cape Town's Faculty of Health Sciences and the Stellenbosch University Ethics Committee. Written informed consent was obtained from mothers prior to enrolment.

Figure 1: Study area. Google maps showing the study area Drakenstein (circled in Red), Drakenstein is situated in the Western Cape, 75 km North East of Cape Town in the Western Cape province.

B.3.2 Data collection:

Antenatal and postnatal study measurement visits were conducted at primary healthcare clinics while birth and 6 week postnatal visits occurred at Paarl Hospital. Infants attended study visits at 6, 10 and 14 weeks, and 6, 9, 18, 30, 36, 42, 54 and 60 months of age. Trained interviewers collected data on demographic, SES and behavioural factors, and comprehensive clinical data on pneumonia, wheezing and other LRTIs were collected. Covariates considered for this study included sex, ethnicity, breastfeeding, household electricity, household SES, smoking exposure, and parental education, all assessed using questionnaires. Feeding practices at 6 weeks

postpartum and household smoking were reported by mothers while SES was based on a composite score consisting of four variables: level of education, employment status, household income, and number of household assets. Standardized scores were divided into 4 quartiles namely “low”, “moderate-low”, “moderate-high” and “high” groups using distribution-based cut-offs.

Primary health nurses and study staff were trained to recognize WHO-defined pneumonia, with an episode of pneumonia defined as cough or difficulty breathing and increased respiratory rate or lower chest wall in-drawing in a child older than 2 months of age. Mothers were counselled regarding these symptoms and were advised to contact study staff if these symptoms arose. In addition, study staff reviewed routine medical records at primary care clinics and at Paarl Hospital for missed pneumonia episodes. Wheeze was measured through maternal report at each of the study visits as well as through active surveillance for respiratory symptoms associated with LRTIs (MacGinty et al., 2018). For children with repeated episodes of pneumonia or wheezing, only the first episode was included in analysis. For this analysis, pneumonia and wheeze were examined separately. Of the 1000 children recruited and followed up in the original study, cases were children who developed pneumonia or wheezing from May 2012 to December 2018, and controls were those who did not develop pneumonia or wheezing during follow up. Controls were incidence density matched 1:1 on date of birth, age of presentation and enrolment site. Controls could be matched to both a pneumonia and/or a wheeze case. Many children (n=409) developed both pneumonia and wheezing during follow up. Due to loss to follow up and missing data, our analysis included a total of 946 participants, which includes both cases and controls.

Trained research assistants collected latitude and longitude coordinates of participants’ households using a hand-held geographic positioning system (GPS) device (eTrex 10, Garmin).

Shapefiles were taken from statistics South Africa (stats SA) and the Drakenstein map was subsetted.

B.3.3 Data analysis:

B.3.3.1 Descriptive statistics:

Data were analyzed using R studio version 4.0.2 (2020-06-22). Categorical variables were summarized using frequencies and percentages and were compared across episodes of (i) pneumonia and (ii) wheeze, with Chi-squared tests used to test for statistical significance. General child characteristics were described and compared across sites.

B.3.3.2 Kernel density estimation mapping:

To map individual point observations, we used an interpolation technique (kernel density estimation) which produces a spatially smooth intensity of a spatial point pattern. A kernel is placed over each point observation and intensity at each intersection of the superimposed grid is estimated upon a weighted sum of observed points that falls within the radius of the location grid (Vadrevu and Badarinath, 2009). Kernel density estimates were generated to describe the density of pneumonia and wheezing cases across space, identify high risk areas, and estimate the relative risk of pneumonia and wheezing cases. This analysis was performed in R studio using library metrics, spatstat, geoR and gtools libraries. The kernel density estimates were generated using bw.ppl bandwidth which uses likelihood cross-validation to select a smoothing bandwidth for the kernel estimation of point process intensity and, as a result, regions with the highest density of cases would have the brightest colour.

B.3.3.3 Spatial autocorrelation:

Ripley's K function was performed to estimate the spatial dependence of pneumonia and wheezing. This function compares the actual spatial location of points to a simulated random distribution of points using one or more distance bands. The value of the K-function determines

whether a given point pattern is more clustered or dispersed than a point pattern of complete spatial randomness. This analysis was performed in R Studio to test the hypothesis that the spatial point patterns for pneumonia and wheeze depart from complete spatial randomness using the spatstat package. Spatial dependency is present when the observed K value deviates significantly from the theoretical K value. The presence of spatial dependence suggests that there is clustering.

B.3.3.4 SatScan:

Kulldorff's spatial scan statistics was performed using SaTscan v9.6 to assess whether events cluster in space. This software applies multiple circular windows across the study area and each distinct circle represents a possible cluster. The purely spatial analysis was used to identify and locate statistically significant aggregations or clusters of pneumonia and wheezing in Drakenstein. Given that the data consist of binary variables, a Bernoulli model was used to scan for low and high-rate clusters. A likelihood ratio test was used to analyze the observed cases of pneumonia and wheeze within the circle to the expected cases across the full range to determine significant clusters of disease, giving relative risk and p-values for any clusters observed. In order to scan for small to large clusters, the maximum cluster size was set to 50% of the total population at risk and, to ensure sufficient statistical power, the number of Monte Carlo simulations was set to 999 replications and clusters with statistical significance of $p<0.05$ were reported. No geographic overlap was used as a default setting, therefore secondary clusters would not overlap the most significant cluster. A p-value of 0.05 and a recurrence interval of 365 days were used as the threshold to identify a statistical cluster.

B.3.3.5 Generalised additive models (GAM):

We investigated the association between location and childhood pneumonia and wheeze using generalized additive models (GAM) from the MapGAM package in R Studio to identify the

geospatial and sociodemographic covariates associated with the observed patterns of these LRTIs. GAMs are semi-parametric extensions of GLMs; the only underlying assumption made is that the functions are additive and that the components are smooth (Hastie and Tibshirani, 1986, 1990). A GAM uses a link function to establish a relationship between the mean of the response variable and a 'smoothed' function of the explanatory variable(s). To estimate the measure of association across space, we fit both bivariate and multivariate GAM using span of 0.15 for both models. To obtain a map, a rectangular grid was created based on the minimum and maximum latitude and longitude of the data set. Crude and covariate adjusted odds ratios (ORs) were obtained from each grid point using modgam to smooth by geolocations. We generated covariate adjusted heatmaps of ORs in R Studio. Smoking exposure, sex, ethnicity, SES and breastfeeding practices were included as covariates in multivariate models. Spatial confounding was considered present if the covariate was a risk factor and varied spatially. We tested the global null hypothesis of whether the diseases status is not dependent on location i.e. the map is flat.

B.4 Results:

B.4.1 Descriptive statistics:

A total of 946 children (51% male and 49% female) were included in analysis, including 54% black African children and 46% children with mixed ancestry. At 6 weeks postpartum, 451 mothers (48%) were breastfeeding their children exclusively and 495 (52%) reported mixed feeding or no breastfeeding. Overall, 94% of children had electricity in their homes. In terms of smoking exposure, 336 mothers (36%) were active smokers, while 394 (42%) were passive smokers and 216 (23%) were non-smokers. Most mothers reported having some secondary education (55%), while 32% had completed secondary education; few had primary (7%) or tertiary education (6%). Table 1 shows the general child, parental and household characteristics of children stratified by

site. Several characteristics differed across site, with participants from TC Newman being predominantly of mixed ancestry and having higher socioeconomic status compared to participants from Mbekweni. Maternal smoking was significantly more common in TC Newman compared to Mbekweni, and exclusive breastfeeding was more common in TC Newman compared to Mbekweni.

Table 1: Childhood characteristics across site.

Exposure	Total sample	Mbekweni	TC Newman	p-value
Number of children	**946**	**511**	**435**	
Child characteristics				
Sex				
Males	485(51%)	246(48%)	239(55%)	0.04
Females	461(49%)	265(52%)	196(45%)	
Ethnicity				
Mixed ancestry	437(46%)	9(2%)	428(98%)	
Black	509(54%)	502(98%)	7(2%)	<0.001
Breastfeeding				
Mixed feeding / not breastfeeding	495(52%)	298(58%)	197(45%)	<0.001
Exclusive	451(48%)	213(42%)	238(55%)	
Parental and household characteristics				
Smoking categories				
Active smoker	336(35%)	95(19%)	241(55%%)	<0.001
Passive smoker	394(42%)	243(48%)	151(35%%)	
Non-smoker	216(23%)	172(34%%)	44(10%)	
Socio-economic status				
Lowest SES	228(24%)	150(29%)	78(18%)	
Low-moderate-SES	252(27%)	143(28%)	109(25%)	<0.001
Moderate high SES	241(25%)	119(23%)	122(28%)	
High SES	225(24%)	99(19%)	126(29%)	
Education				
Tertiary	58(6%)	38(7%)	20(5%)	
Completed secondary	303(32%)	151(30%)	152(35%)	0.13
Some secondary	518(55%)	287(56%)	231(53%)	
Primary	67(7%)	35(7%)	32(7%)	
Electricity				
Yes	893(95%)	468(92%)	425(98%)	<0.001
No	50(5%)	42(8%)	8(2%)	

We further described childhood characteristics stratified by diseases status as per table 2. A total of 466 children developed pneumonia while 409 of those developed wheeze. The median age of children at LTRI episode was 10 months.

Children who acquired pneumonia were more likely to be male (58% versus 45% among those without pneumonia; p<0.001). Similarly, children with wheeze were more likely to be male (56% versus 47% among those without wheeze; p=0.006). Children who developed wheezing were also more likely to be of mixed ancestry (55% versus 40% among those without wheezing; p<0.001), but no association was observed between ethnicity and pneumonia (p=0.54). There was no significant association between breastfeeding and pneumonia (p=0.32). However, there was a statistically significant association between breastfeeding and wheezing (p=0.04), with those who developed wheeze being more likely to have received mixed feeding or no breastfeeding at 6 weeks postpartum compared to those who did not develop wheeze.

There was a statistically significant association between study site and wheezing (p<0.001), with children experiencing wheeze being more likely to have been enrolled at TC Newman (54% compared to 40% of those without wheeze). In contrast, there was no association between pneumonia and study site (p=0.53). No associations were observed between SES, the presence of electricity in homes or maternal education and either pneumonia or wheeze. Those who developed pneumonia were more likely to have parents who smoked actively compared to those who did not develop pneumonia (p=0.009). Similarly, those who developed wheezing were more likely to have mothers who smoked actively compared to those who did not develop wheezing (p=0.001).

Table 2: Characteristics of study participants by diseases status.

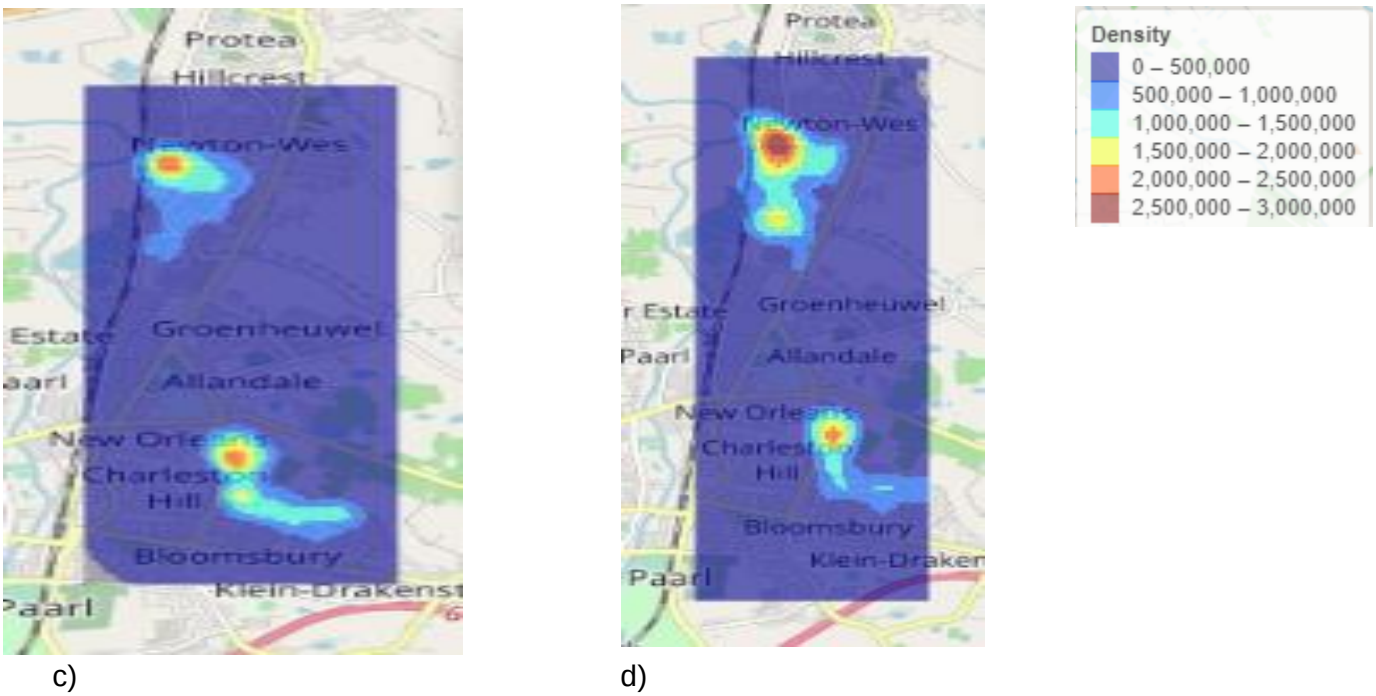
Exposure	Total sample (n=946)	Pneumonia			Wheezing		
		Pneumonia (n=466)	No pneumonia (n=480)	p-value	Wheezing (n=409)	No wheezing (n=537)	p-value
Child characteristics							
Sex							
Male	485(51%)	269(58%)	216(45%)	<0.001	232(56%)	253(47%)	0.006
Female	461(49%)	197(42%)	264(55%)		177(43%)	284(53%)	
Ethnicity							
Mixed ancestry	437(46%)	211(45%)	225(47%)	0.54	224(55%)	213(40%)	<0.001
Black	509(54%)	255(55%)	254(53%)		185(45%)	324(60%)	
Breastfeeding							
Exclusive breastfeeding	451(48%)	214(46%%)	237(49%)	0.32	179(44%)	272(51%)	0.04
Mixed feeding / not breastfeeding	495(52%)	252(54%)	243(51%)		230(55%)	265(49%)	
Parental and household characteristics							
Study site							
Mbekweni	511(54%)	257(55%)	254(53%)	0.53	188(46%)	323(60%)	<0.001
TC Newman	435(46%)	209(45%0	226(47%)		221(54%)	214(40%)	
SES-quartile							
Lowest SES	228(24%)	109(23%)	119(25%)	0.27	96(23%)	132(25%)	0.12
Low-moderate-SES	252(27%)	135(29%)	117(24%)		121(30%)	131(24%)	
Moderate high SES	241(25%)	121(26%)	120(25%)		108(26%)	133(25%)	
High SES	225(24%)	101(22%)	124(26%)		84(21%)	141(26%)	
Electricity							
Yes	893(94%)	437(94%)	456(95%)	0.19	383(94%)	510(95%)	0.12
No	50(5%)	26(6%)	24(5%)		23(6%)	27(5%)	
Smoking categories							
Active smoker	336(36%)	188(40%)	147(31%)	0.009	169(41%)	167(31%)	0.001
Passive smoker	394(42%)	185(40%)	210(44%)		166(41%)	228(42%)	
Non-smoker	216(23%)	93(20%)	123(26%)		74(18%)	142(26%)	
Education							
Tertiary	58(6%)	22(5%)	36(8%)	0.16	25(6%)	33(6%)	0.39
Completed Secondary	303(32%)	144(31%)	159(33%)		119(29%)	184(34%)	
Some secondary	518(55%)	262(56%)	256(53%)		234(57%)	284(53%)	
Primary	67(7%)	38(8%)	29(6%)		31(8%)	36(7%)	

B. 4.2 Kernel density estimation mapping:

Exploratory disease mapping was performed using kernel density estimation of point pattern data to explore locations of high incidence of pneumonia (a) and wheezing (b). The kernel density maps (Figure 2) show that cases and controls have a similar spatial concentration in two areas, namely Mbekweni and TC Newman, corresponding with the regions in which participants live. For pneumonia in particular, the spatial distribution of cases is similar to that of controls, but there is some variation in the density of wheezing cases in comparison to controls, with cases less geographically dispersed compared to controls in Mbekweni, but slightly more dispersed in TC Newman.

B. 4.2 Kernel density estimation mapping:

a) b)



c) d)

Figure 2: Kernel density estimation maps of pneumonia and wheezing.

Kernel density maps of (a) pneumonia cases, (b) pneumonia controls, (c) wheezing cases and d) wheezing controls

B.4.3 Spatial autocorrelation:

The Ripley's K function graphs show that the observed K value (black intact line) deviates significantly and is above the theoretical Poisson distribution (red dotted line) for both pneumonia (Figure 3a) and wheezing (Figure 3c), suggesting that both pneumonia and wheeze cases are spatially clustered. The K function graph depicts K- values(K(dd)) at different distances (r). The results suggest spatial autocorrelation or clustering across all distances. The K function was evaluated for 999 Monte Carlo simulations and the observed k value for both pneumonia and wheezing exhibited clustering (p=0.002). We also performed Ripley's K function for the differences between cases and controls. For pneumonia (Figure 3b), the observed K function falls within the confidence interval bounds (light grey areas), suggesting that the difference in spatial distributions between cases and controls is generally random except at very small distances. For wheezing (Figure 3d), the observed K function deviates into the darker grey bounds at both small and large distances, suggesting that the differences in spatial distribution between cases and controls is not random.

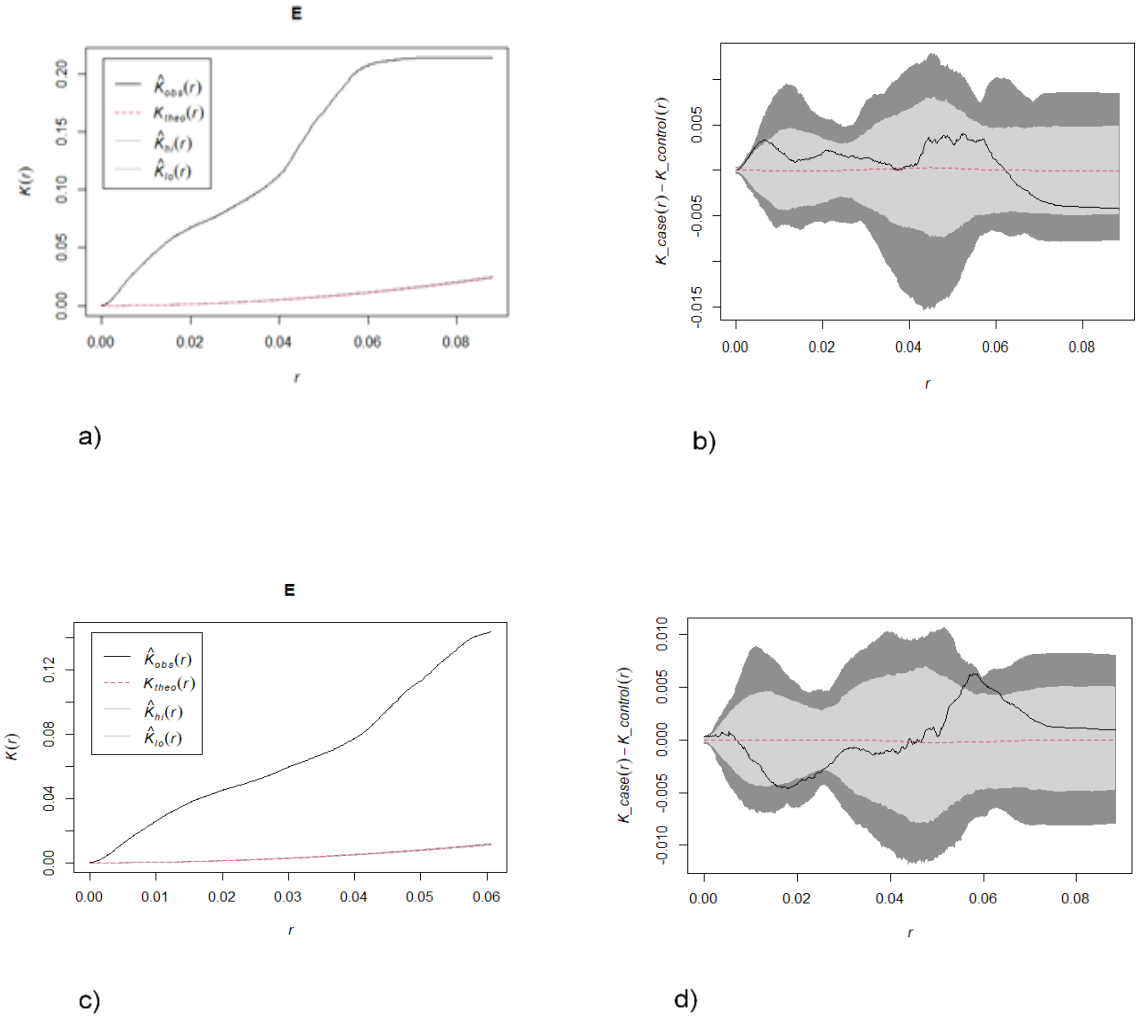


Figure 3: Ripley's K function for wheezing and pneumonia cases in children in Drakenstein using Poisson distribution. K function for pneumonia cases (a), K function for the differences between pneumonia cases and controls (b), K function for wheezing cases (c) and K function for the differences between wheezing cases and controls (d).

B.4.4 SaTScan:

Using a Bernoulli model which included both cases and controls, spatial scan statistics were performed using Kulldorff Satscan v1.9 to detect clusters for both LRTI outcomes (Table 3). For pneumonia, we observed one low-rate cluster in Mbekweni, where fewer cases than expected were observed. However, the relative risk for the observed versus expected rate for this cluster

(0.94) was close to the null and not statistically significant, suggesting that pneumonia is not clustered differently from a random distribution including 50% of the population at risk. Similarly, one most likely cluster was observed for wheezing, with the number of observed wheezing cases exceeding the number of expected cases in the TC Newman region. A relative risk of 1.38 was estimated, suggesting that wheezing cases are significantly more concentrated than expected in that region ($p<0.001$).

Table 3: Summary of spatial scan statistics from Kulldorff Satscan

LRTI outcome	Cluster detected	Relative risk (RR)	p-value	Likelihood ratio	Location
Pneumonia	Low-rate cluster	0.94	0.096	1.28	Mbekweni
Wheezing	High-rate cluster	1.38	<0.001	9.42	TC Newman

Given this observed clustering, we used the *spatstat* package in R to create a kernel density map (Figure 4) that depicts the location of the wheezing cluster detected. The cluster for wheezing was in and around the region where the study clinic TC Newman is situated.

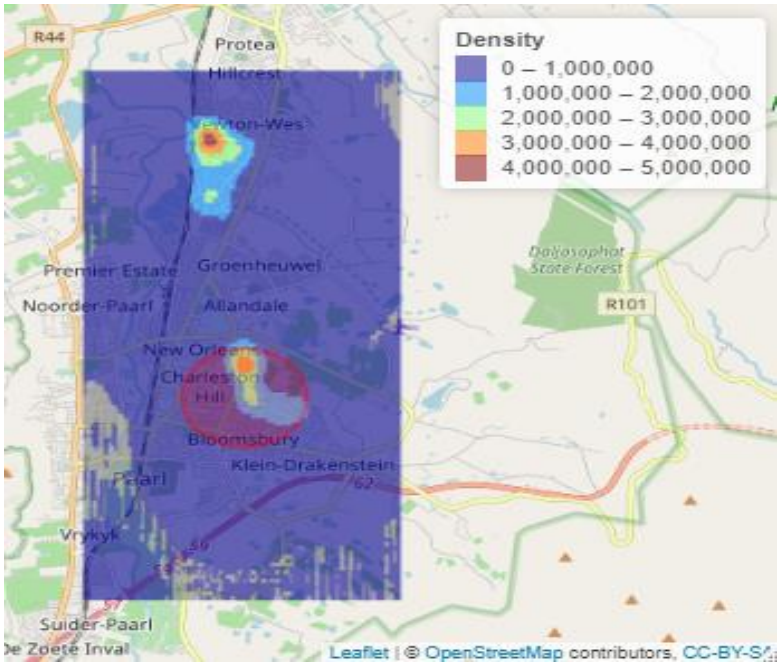


Figure 4: Kernel density map showing the location of a cluster.

A cluster is surrounded by a red circle in areas less than 1 km radius.

B.4.5 GAMs:

We built bivariate and multivariate GAMs to assess the association between covariates and pneumonia adjusting for spatial autocorrelation (Table 4A). After adjustment for SES, ethnicity, sex, education, maternal smoking and study site, we found that males were 1.67 times more likely to have pneumonia as compared to females ($p<0.001$). No significant associations were observed between pneumonia and any of ethnicity, SES, education, exposure to tobacco smoke or breastfeeding in either unadjusted or adjusted models. Children from TC Newman were less likely to have pneumonia in both bivariate (OR=0.32, p=0.003) and multivariate models (OR=0.28, p=0.05).

Similarly, we built GAMs for wheezing (Table 4B). The odds of wheezing were found to be significantly higher in males (OR=1.39, p=0.02) as compared to females, with a similar association observed in the multivariate model (OR=1.40, p=0.02). Children with mixed ancestry were 3.23 times as likely to have wheeze (p<0.001) as compared to Black children and the covariate adjusted model showed a 41% increase in the odds of wheezing in children with mixed ancestry (p<0.001). Compared to children in the highest SES category, children in all other SES categories were more likely to have wheeze in bivariate analysis, although these associations were slightly attenuated in the multivariate models. Those who developed wheeze were 14 times as likely to reside in TC Newman compared to Mbekweni (p<0.001), although this association was not significant in the multivariate model. No association was observed between wheeze and either of caregiver smoking or education. Finally, children who were mixed feeding or not breastfeeding at 6 weeks of age were more likely to have wheeze compared to children who were exclusively breastfed in both bivariate and multivariate analyses (p=0.01 in both).

Table 4: Generalised additive models of associations for (A) pneumonia and (B) wheezing.

Covariate	Bivariate Odds ratio	p-value	Multivariate Odds ratio	p-value
(A) Pneumonia				
Sex: Male (versus female)	1.63	<0.001	1.67	<0.001
Ethnicity: Mixed ancestry (versus Black)	0.91	0.77	0.70	0.28
SES (versus Highest)				
Moderate-high	1.24	0.24	1.10	0.64
Moderate-low	1.42	0.06	1.11	0.64
Lowest	1.10	0.62	0.74	0.24
Caregiver smoking categories (versus active smoker)				
Passive smoker	0.99	0.99	0.93	0.86
Non smoker	0.79	0.58	0.76	0.53
Mixed feeding / not breastfeeding (versus Exclusive breastfeeding)	1.09	0.48	1.07	0.62
Study site: TC Newman (versus Mbekweni)	0.32	0.003	0.28	0.05
Caregiver education (versus Tertiary education)				
Primary completed	2.18	0.04	2.14	0.07
Some secondary	1.66	0.08	1.42	0.28
Secondary completed	1.43	0.24	1.31	0.39
(B) Wheezing				
Sex: Male (versus female)	1.39	0.02	1.40	0.02
Ethnicity: Mixed ancestry (versus Black)	3.23	<0.001	1.41	0.08
SES (versus Highest)				
Moderate-high	1.42	0.07	1.41	0.07
Moderate-low	1.70	0.01	1.60	0.03
Lowest	1.43	0.07	1.18	0.40
Caregiver smoking categories (versus active)				
Passive smoker	1.11	0.80	1.14	0.80
Non smoker	0.82	0.63	0.88	0.70
Breastfeeding practices (Mixed versus exclusive breastfeeding)	1.43	0.01	1.46	0.01
Study site: TC Newman (versus Mbekweni)	14.01	<0.001	9.40	0.10
Caregiver education (versus Tertiary education)				
Primary	1.04	0.91	0.78	0.60
Some Secondary	0.98	0.96	0.68	0.30
Completed Secondary	0.74	0.32	0.61	0.10

GAM produced covariate adjusted smooth maps using the modgam package, with the maps showing locations with increased or decreased risk of LRTIs. Red areas show locations with a high odds of being a case while blue show locations with a decreased odds of be ing a case. Areas that were identified as statistically significant were denoted with black contour lines or bands on the maps. Figure 4a shows a crude heatmap for pneumonia. This map is flat with areas of high odds of being case (global p=0.03). After adjustment for covariates (Figure 4b), the map had more pronounced hot and cold spots (global p<0.03), suggesting the presence of confounding. Bivariate smooth maps were created adjusting for one covariate at a time and the apparent differences were due to confounding by SES. Conversely both crude (global p=0.028) and adjusted maps (global p<0.001) were flat for wheezing. The crude map (Figure 4c) does not show any hotspots while the adjusted Figure 4d shows a few hotspots by contour lines. Adjusting for potential spatial confounders made no appreciable differences in the maps suggesting that there is no spatial confounding.

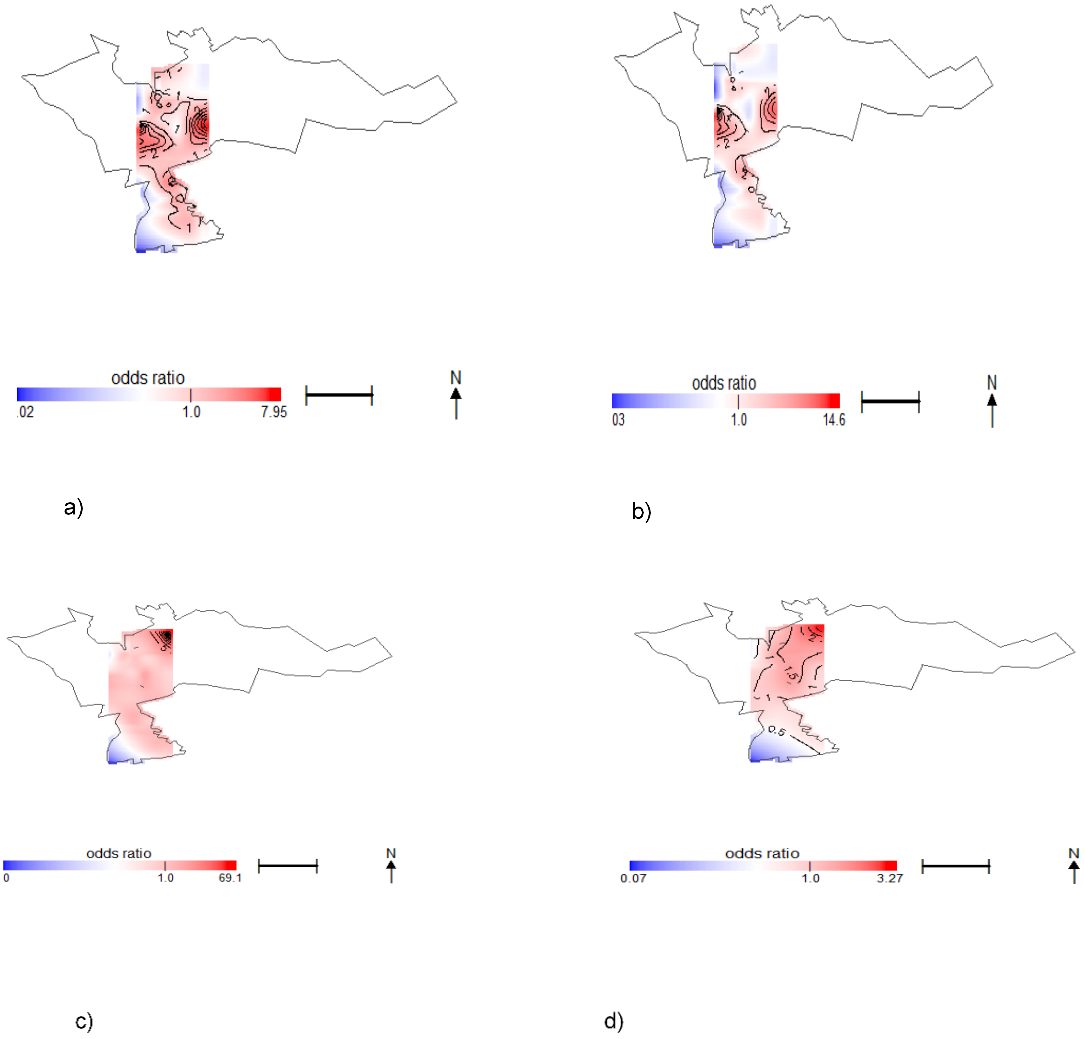


Figure 4: Covariate adjusted maps a) Crude pneumonia (top left), b) covariate adjusted pneumonia (top right), c) crude wheezing (bottom left) and d) covariate adjusted wheezing (bottom right).

B.5 Discussion:

This study examined the spatial epidemiology and clustering of childhood pneumonia and wheeze as well as individual-, household- and community-level risk factors for these outcomes in two communities in Drakenstein, South Africa. In bivariate analysis that did not take into account spatial clustering, we observed that male children were more likely to have both pneumonia and wheeze compared to females. We also found that children with mixed ancestry were more likely to have wheezing than black African children, consistent with the finding of higher levels of wheeze in the community of TC Newman. Children who experienced wheeze were more likely to be mixed feeding or not breastfeeding at 6 weeks of age compared to those who did not experience wheeze. Finally, household smoking was associated with both pneumonia and wheeze. Exploratory disease mapping suggested that the spatial distribution of pneumonia cases is similar to that of controls, but that there is some variation in the density of wheezing cases in comparison to controls. Spatial analysis suggested that both pneumonia and wheeze are spatially clustered, and that the differences in the spatial distribution between wheezing cases and controls is not random. Alongside these results, we observed one most likely cluster for wheezing, situated in the TC Newman area. Finally, GAMs suggested a statistically significant relationship between location and both pneumonia and wheeze, with cases of pneumonia less likely in Mbekweni and cases of wheeze more likely in TC Newman. Similar to analyses that did not take spatial clustering into account, male sex was associated with a higher odd of both pneumonia and wheeze; and wheeze was associated with having mixed ancestry as well as receiving mixed feeding or no breastfeeding.

The finding that pneumonia and wheeze are more common among males compared to females is consistent with other ''non-spatial'' studies (Falagas et al., 2007) and corresponds with a generally high incidence of LRTIs of any aetiology in boys (Teck et al., 2019). In both spatial and non-spatial analyses, children with mixed ancestry were more likely to experience wheeze in this study, consistent with the higher levels of wheeze observed in TC Newman.

Population characteristics such as socio-economic and cultural differences in the study locations may explain this association. For example, data from this setting suggests that smoking is more common in mixed ancestry households (Vanker et al., 2016; Vanker et al.,2017). Our findings show that there was no association between SES and either of pneumonia or wheeze, despite LRTIs mainly affecting impoverished areas with high social vulnerability (Pina et al.,2020, Ramesh et al.,2012). Many studies have demonstrated the association between SES and LRTIs through potential pathways which include increased exposure to infectious agents due to household crowding, lack of immunization, poor health seeking behavior, malnutrition and poor parental education levels (Beamer et al., 2016; Almasi et al., 2021). Additional research should explore these pathways in order to target interventions. Similarly, breastfeeding practices were not associated with pneumonia and wheezing in the spatially adjusted model, contrary to what is known in the literature about breastfeeding and LRTIs (Jackson et al., 2013). Lastly, our GAM findings suggest that smoking was not associated with either pneumonia or wheeze. The discrepancies in the associations observed between GAM and Chi-squared tests may be attributed to adjusting for site location and geographic clustering in GAM. Our GAM findings contradict what is known in literature about the relationship between smoking and pneumonia and wheeze (Ramesh et al., 2012, Le Roux et al., 2019). There was a positive association between study location (TC Newman) and wheeze, consistent with the cluster observed in TC Newman. This finding suggests that the incidence of wheeze in Drakenstein may be spatially dependent. This warrants further investigation of risk factors that may be associated with the elevated incidence of wheeze in that area.

While associations between LTRIs and risk factors have been well documented in the DCHS (Vanker et al., 2016; MacGinty et al., 2018), it is important to consider adjusting for spatial confounding given that risk factors tend to cluster in space, and spatial models may unmask associations at a particular location. This study thus addresses an important gap in the literature by investigating the spatial clustering of childhood pneumonia and wheeze and

factors associated with each in this cohort. Simply examining kernel density maps would have made it difficult to identify random patterns and clusters and areas with aggregated cases of wheeze and pneumonia (Mollalo et al., 2015). Therefore, point pattern and clustering analyses were necessary and pointed to a cluster of wheeze cases in TC Newman, where the number of observed cases exceeded the number of expected cases. This could be attributed to cultural and socio-demographic differences between the two study locations, given that possible risk factors for wheeze vary by location. Many known risk factors for LRTIs are spatially diverse and therefore spatial patterns will vary across space. A recent non-spatial analysis conducted in the two study areas revealed that smoking was more prevalent in TC Newman than Mbekweni and therefore smoking may be associated with wheezing in that location (MacGinty et al.,2018). The same study showed that children from TC Newman had elevated odds of developing wheeze. This finding suggests that the cluster detected in TC Newman is not explainable by chance. For pneumonia, a secondary cluster detected was not statistically significant.

The measures of effect were estimated using GAM which provides a framework to test hypotheses (Yazdy et al., 2015; Hoffman et al., 2010) and generate covariate adjusted maps to depict high risk locations while adjusting for confounding. Using GAM, we generated crude maps and covariate adjusted maps to illustrate areas of increased and decreased LRTIs while adjusting for confounding. The covariate adjusted smooth maps showed that there was potential for confounding for pneumonia because the covariate adjusted map is more pronounced while no confounding was detected in the wheezing map because the map is relatively flat. The variation observed in the maps depends on the span size which is the value that determines the amount of data the smoother will use in the smoothing process. We used sp = 0.15 for all maps and it has been published that the smaller the span size the higher the variability that will be depicted in a map and increased random pattern detection (Yazdy et al.,

2015; Hoffman et al., 2010). Consequently, there is a compromise between variability and bias (Yazdy et al., 2015; Hoffman et al., 2010).

Several limitations of these findings should be considered. This is a secondary analysis that uses data which was not previously collected for spatial analysis, which posed challenges for analysis. The study areas were very small compared to other published studies that explore spatial clustering and the clusters detected may not mimic clustering when a larger population is sampled. For GAM maps, the visualization of hotspots is partially dependent on span size, with more variability visualized and consequently more bias with smaller span sizes (Hoffman et al., 2010). We are not able to rule out the possibility of residual confounding. We suggest that future spatial work should collect environmental factors and be included in the analysis using different approaches and techniques to overcome disadvantages of modelling in R to strengthen the fidelity of the results. Furthermore, wheezing was both diagnosed and self-reported which may have resulted in misclassification and there is potential for information bias as parents may have not recalled all wheezing episodes. Finally, these data arise from two communities in South Africa and the findings may not be generalizable to other areas as risk factors are geographically dependent and vary spatially.

Conclusions:

Our results suggest that while pneumonia is spatially dispersed across the study areas, cases of wheeze tend to cluster geographically, with a high-rate cluster identified in the TC Newman area. These results further reiterate that "space" is a critical epidemiological concept that influences disease patterns and distribution,and suggest that focused public health interventions are needed to reduce the incidence of LRTIs in the area of TC Newman specifically. In particular, coordinated health system strengthening and mobilization of resources to address social and health inequalities may be needed, as well as more research on non-pharmaceutical interventions such as health education, health promotion and advocacy. To our knowledge, this is the first study to explore spatial variation of LRTIs in a

low-income setting in South Africa. Future spatial studies are needed to better understand factors influencing diseases and ultimately improve health outcomes.

Abbreviations

LRTIs – Lower respiratory tract infections

KDE – Kernel density estimation

DCHS - Drakenstein Child Health Study

GIS -geographic information system

GPS – Global positioning system

GAM – Generalized additive model

SES – socio-economic status

LMIC – Low-middle income countries

RR – Risk ratio

OR – Odds ratio

ArcGIS – A geographic information system for working with maps and geographic information.

SaTScan – A software that analyses spatial data and detects space-time disease clusters

R – A statistical software

B.6 References

Adane MM, Alene GD, Mereta ST, Wanyonyi KL. Prevalence and risk factors of acute lower respiratory infection among children living in biomass fuel using households: a community-based cross-sectional study in Northwest Ethiopia. BMC public health. 2020 Dec;20(1):1-3.

Adeloye D, Chan KY, Rudan I, Campbell H. An estimate of asthma prevalence in Africa: a systematic analysis. Croatian medical journal. 2013 Dec 15;54(6):519-31.

Beamer PI, Lothrop N, Lu Z, Ascher R, Ernst K, Stern DA, Billheimer D, Wright AL, Martinez FD. Spatial clusters of child lower respiratory illnesses associated with community-level risk factors. Pediatric pulmonology. 2016 Jun;51(6):633-42.

Boloursaz, M.R., Lotfian, F., Aghahosseini, F., Cheraghvandi, A., Khalilzadeh, S., Farjah, A. and Boloursaz, M. Epidemiology of lower respiratory tract infections in children. *Journal of Comprehensive Pediatrics*. 2013 July 1;*4*(2): 93-8.

Carlos HA, Shi X, Sargent J, Tanski S, Berke EM. Density estimation and adaptive bandwidths: a primer for public health practitioners. International journal of health geographics. 2010 Dec;9(1):1-8.

Feldman, C. and Shaddock, E. Epidemiology of lower respiratory tract infections in adults. *Expert Review of Respiratory Medicine*. 2019 May 5;*13*(1):63-77.

Gehring U, Cyrys J, Sedlmeir G, Brunekreef B, Bellander T, Fischer P, Bauer CP, Reinhardt D, Wichmann HE, Heinrich J. Traffic-related air pollution and respiratory health during the first 2 yrs of life. European respiratory journal. 2002 Apr 1;19(4):690-8.

Hinman SE, Blackburn JK, Curtis A. Spatial and temporal structure of typhoid outbreaks in Washington, DC, 1906–1909: evaluating local clustering with the G i* statistic. International Journal of Health Geographics. 2006 Dec;5(1):1-7.

Hoffman K, Webster TF, Weinberg JM, Aschengrau A, Janulewicz PA, White RF, Vieira VM. Spatial analysis of learning and developmental disorders in upper Cape Cod, Massachusetts using generalized additive models. International Journal of Health Geographics. 2010 Dec;9(1):1-1.

Hossain MZ, Bambrick H, Wraith D, Tong S, Khan AF, Hore SK, Hu W. Sociodemographic, climatic variability and lower respiratory tract infections: a systematic literature review. International journal of biometeorology. 2019 Feb 15;63(2):209-19.

Jackson S, Mathews KH, Pulanić D, Falconer R, Rudan I, Campbell H, Nair H. Risk factors for severe acute lower respiratory infections in children–a systematic review and meta-analysis. Croatian medical journal. 2013 Apr 15;54(2):110-21.

Jackson S, Mathews KH, Pulanić D, Falconer R, Rudan I, Campbell H, Nair H. Risk factors for severe acute lower respiratory infections in children–a systematic review and meta-analysis. Croatian medical journal. 2013 Apr 15;54(2):110-21.

Knobbe RB, Diallo A, Fall A, Gueye AD, Dieng A, van Immerzeel TD, Ba A, Diop A, Diop A, Niang M, Boye CS. Pathogens causing respiratory tract infections in children less than 5 years of age in Senegal. Microbiology insights. 2019 Dec;12:1178636119890885.

Le Roux DM, Myer L, Nicol MP, Zar HJ. Incidence and severity of childhood pneumonia in the first year of life in a South African birth cohort: the Drakenstein Child Health Study. The Lancet Global Health. 2015 Feb 1;3(2):e95-103.

Le Roux DM, Nicol MP, Myer L, Vanker A, Stadler JA, von Delft E, Zar HJ. Lower respiratory tract infections in children in a well-vaccinated South African birth cohort: spectrum of disease and risk factors. Clinical Infectious Diseases. 2019 Oct 15;69(9):1588-96.

Mollalo A, Alimohammadi A, Shirzadi MR, Malek MR. Geographic information system-based analysis of the spatial and spatio-temporal distribution of zoonotic cutaneous

leishmaniasis in Golestan Province, north-east of Iran. Zoonoses and public health. 2015 Feb;62(1):18-28.

Mollalo A, Vahedi B, Bhattarai S, Hopkins LC, Banik S, Vahedi B. Predicting the hotspots of age-adjusted mortality rates of lower respiratory infection across the continental United States: Integration of GIS, spatial statistics and machine learning algorithms. International journal of medical informatics. 2020 Oct 1;142:104248.

Pina JC, Alves LS, Arroyo LH, Arcêncio RA, Gondim EC, de Carvalho Furtado MC, de Mello DF. Using geo-spatial analysis for assessing the risk of hospital admissions due to community-acquired pneumonia in under-5 children and its association with socially vulnerable areas (Brazil). BMC pediatrics. 2020 Dec;20(1):1-4.

Rahman MM, Shahidullah M. Risk factors for acute respiratory infections among the slum infants of Dhaka city. Bangladesh Medical Research Council Bulletin. 2001 Aug 1;27(2):55-62.

Ramesh Bhat Y, Manjunath N, Sanjay D, Dhanya Y. Association of indoor air pollution with acute lower respiratory tract infections in children under 5 years of age. Paediatrics and international child health. 2012 Aug 1;32(3):132-5.

Sanchez J, Stryhn H, Flensburg M, Ersbøll AK, Dohoo I. Temporal and spatial analysis of the 1999 outbreak of acute clinical infectious bursal disease in broiler flocks in Denmark. Preventive veterinary medicine. 2005 Oct 12;71(3-4):209-23.

Seidu, A.A., Dickson, K.S., Ahinkorah, B.O., Amu, H., Darteh, E.K.M. and Kumi-Kyereme, A., Prevalence and determinants of acute lower respiratory infections among children under-five years in sub–Saharan Africa: evidence from demographic and health surveys. *SSM-population health*. 2019 March 21;*8*: 100443.

Vadrevu KP, Badarinath KV. Spatial pattern analysis of fire events in Central India–A case study. Geocarto International. 2009 Apr 1;24(2):115-31.

Yazdy MM, Werler MM, Anderka M, Langlois PH, Vieira VM. Spatial analysis of gastroschisis in Massachusetts and Texas. Annals of epidemiology. 2015 Jan 1;25(1):7-14.

PART C: APPENDIX

C.1. Human research ethics committee letter of approval

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room G50- Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6492
Email: hrec-enquiries@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

07 December 2020

HREC REF: 554/2020

Dr J Ncayiyana
Division of Epidemiology & Biostatistics
Room 537, 5th Floor, Falmouth Building-FHS
Email: - jabulani.ncayiyana@uct.ac.za
Student: namhlabhenxa@gmail.com

Dear Dr Ncayiyana

PROJECT TITLE: SPATIAL PATTERNS AND CORELATES OF LOWER RESPIRATORY TRACT ILLNESSES IN CHILDREN FROM DRAKENSTEIN, WESTERN CAPE-MASTER'S CANDIDATE-MISS NAMHLA BHENXA-SUB-STUDY LINKED TO 401/2009

Thank you for your response letter, addressing the issues raised by the Faculty of Health Sciences Human Research Ethics Committee (HREC).

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study.

This approval is subject to strict adherence to the HREC recommendations regarding research involving human participants during COVID -19, dated 17 March 2020 & 06 July 2020.

Approval is granted for one year until the 30 December 2021.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

The HREC acknowledge that the student: - Miss Namhla Bhenxa will also be involved in this study.

Please quote the HREC REF in all your correspondence.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval, where necessary, before the research may occur.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE

Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938
NHREC-registration number: REC-210208-007

This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use: Good Clinical Practice (ICH GCP), South African Good Clinical Practice Guidelines (DoH 2006), based on the Association of the British Pharmaceutical Industry Guidelines (ABPI), and Declaration of Helsinki (2013) guidelines. The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

C.2 SaTScan Results

```
                                SaTScan v9.6
                                ____________


Program run on: Sun Feb 14 21:00:19 2021

Retrospective Space-Time analysis
scanning for clusters with high or low rates
using the Bernoulli model.

_____________________________________________________________________________

SUMMARY OF DATA

Study period.....................: 2012/1/1 to 2018/12/31
Number of locations..............: 2
Total population.................: 966
Total number of cases............: 466
Percent cases in area............: 48.2

_____________________________________________________________________________

CLUSTERS DETECTED

1.Location IDs included.: location3
  Coordinates / radius..: (33.730900 S, 18.970800 E) / 0 km
  Time frame............: 2012/1/1 to 2012/12/31
  Population............: 459
  Number of cases.......: 209
  Expected cases........: 221.42
  Observed / expected...: 0.94
  Relative risk.........: 0.90
  Percent cases in area.: 45.5
  Log likelihood ratio..: 1.283455
  Monte Carlo rank......: 57/523
  P-value...............: 0.096

NOTE: The sequential Monte Carlo procedure was used to terminate the calculations after 522
```

Figure 1: A snippet of detailed pneumonia SaTScan results as discussed in main document.

```
                                    SaTScan v9.6
                                    ____________


Program run on: Mon Jan 11 16:50:41 2021

Retrospective Space-Time analysis
scanning for clusters with high or low rates
using the Bernoulli model.

-------------------------------------------------------------------------------

SUMMARY OF DATA

Study period.....................: 2012/1/1 to 2018/12/31
Number of locations..............: 2
Total population.................: 946
Total number of cases............: 409
Percent cases in area............: 43.2

-------------------------------------------------------------------------------

CLUSTERS DETECTED

1.Location IDs included.: location3
  Coordinates / radius..: (33.746460 S, 18.923960 E) / 0 km
  Time frame............: 2012/1/1 to 2012/12/31
  Population............: 435
  Number of cases.......: 221
  Expected cases........: 188.07
  Observed / expected...: 1.18
  Relative risk.........: 1.38
  Percent cases in area.: 50.8
  Log likelihood ratio..: 9.416224
  Monte Carlo rank......: 1/1000
  P-value...............: 0.000000017

-------------------------------------------------------------------------------
```

Figure 1: A snippet of detailed wheezing SaTScan results as discussed in main document.

C.3 BMC Journal article instructions

General guidelines for BMC Infectious diseases journal

The title page should:

- present a title that includes, if appropriate, the study design e.g.:
    - "A versus B in the treatment of C: a randomized controlled trial", "X is a risk factor for Y: a case control study", "What is the impact of factor X on subject Y: A systematic review".
    - or for non-clinical or non-research studies a description of what the article reports
- list the full names and institutional addresses for all authors.
    - if a collaboration group should be listed as an author, please list the Group name as an author. If you would like the names of the individual members of the Group to be searchable through their individual PubMed records, please include this information in the "Acknowledgements" section in accordance with the instructions below.
- indicate the corresponding author.

The Abstract should not exceed 350 words with minimal the use of abbreviations and do not cite references in the abstract. Reports of randomized controlled trials should follow the **CONSORT** extension for abstracts. The abstract must include the following separate sections:

- **Background:** the context and purpose of the study.
- **Methods:** how the study was performed and statistical tests used.
- **Results:** the main findings.
- **Conclusions:** brief summary and potential implications
- **Keywords**

Three to ten keywords representing the main content of the article.

The Background section should explain the background to the study, its aims, a summary of the existing literature and why this study was necessary or its contribution to the field.

The methods section should include:

- the aim, design and setting of the study
- the characteristics of participants or description of materials
- a clear description of all processes, interventions and comparisons. Generic drug names should generally be used. When proprietary brands are used in research, include the brand names in parentheses
- the type of statistical analysis used, including a power calculation if appropriate

Results should include the findings of the study including, if appropriate, results of statistical analysis which must be included either in the text or as tables and figures.

Discussion should discuss the implications of the findings in context of existing research and highlight limitations of the study.

Conculsions should state clearly the main conclusions and provide an explanation of the importance and relevance of the study reported.

If abbreviations are used in the text they should be defined in the text at first use, and a list of abbreviations should be provided.

All manuscripts must contain the following sections under the heading 'Declarations':

- Ethics approval and consent to participate
- Consent for publication
- Availability of data and materials
- Competing interests
- Funding
- Authors' contributions
- Acknowledgements
- Authors' information (optional)

Please see below for details on the information to be included in these sections.

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

Manuscripts reporting studies involving human participants, human data or human tissue must:

- include a statement on ethics approval and consent (even where the need for approval was waived)
- include the name of the ethics committee that approved the study and the committee's reference number if appropriate.

Studies involving animals must include a statement on ethics approval and for experimental studies involving client-owned animals, authors must also include a statement on informed consent from the client or owner.

If your manuscript does not report on or involve the use of any animal or human data or tissue, please state "Not applicable" in this section.

If your manuscript contains any individual person's data in any form (including any individual details, images or videos), consent for publication must be obtained from that person, or in the case of children, their parent or legal guardian. All presentations of case reports must have consent for publication.

If your manuscript does not contain data from any individual person, please state "Not applicable" in this section.

All manuscripts must include an 'Availability of data and materials' statement. Data availability statements should include information on where data supporting the results reported in the article can be found including, where applicable, hyperlinks to publicly archived datasets analysed or generated during the study. By data we mean the minimal dataset that would be necessary to interpret, replicate and build upon the findings reported in the article. We recognise it is not always possible to share research data publicly, for instance when individual privacy could be compromised, and in such instances data availability should still be stated in the manuscript along with any conditions for access.

All financial and non-financial competing interests must be declared in this section.

Please use the authors initials to refer to each authors' competing interests in this section.

If you do not have any competing interests, please state "The authors declare that they have no competing interests" in this section.

All sources of funding for the research reported should be declared. The role of the funding body in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript should be declared.

The individual contributions of authors to the manuscript should be specified in this section.

Please acknowledge anyone who contributed towards the article who does not meet the criteria for authorship including anyone who provided professional writing services or materials. Authors should obtain permission to acknowledge from all those mentioned in the

If you do not have anyone to acknowledge, please write "Not applicable" in this section.

Group authorship (for manuscripts involving a collaboration group): if you would like the names of the individual members of a collaboration Group to be searchable through their individual PubMed records, please ensure that the title of the collaboration Group is included on the title page and in the submission system and include collaborating author names as the last paragraph of the “Acknowledgements” section. Please add authors in the format First Name, Middle initial(s) (optional), Last Name. You can add institution or country information for each author if you wish, but this should be consistent across all authors.

Please note that individual names may not be present in the PubMed record at the time a published article is initially included in PubMed as it takes PubMed additional time to code this information.

Authors information is optional.

You may choose to use this section to include any relevant information about the author(s) that may aid the reader's interpretation of the article, and understand the standpoint of the author(s). This may include details about the authors' qualifications, current positions they hold at institutions or societies, or any other relevant background information. Please refer to

authors using their initials. Note this section should not be used to describe any competing interests.

Footnotes can be used to give additional information, which may include the citation of a reference included in the reference list. They should not consist solely of a reference citation, and they should never include the bibliographic details of a reference. They should also not contain any figures or tables. Footnotes to the text are numbered consecutively; those to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data). Footnotes to the title or the authors of the article are not given reference symbols. Always use footnotes instead of endnotes.

The preferred reference style is Vancouver.