GIS and Health: Enhancing Disease Surveillance and Intervention through Spatial Epidemiology

Aturinde, Augustus

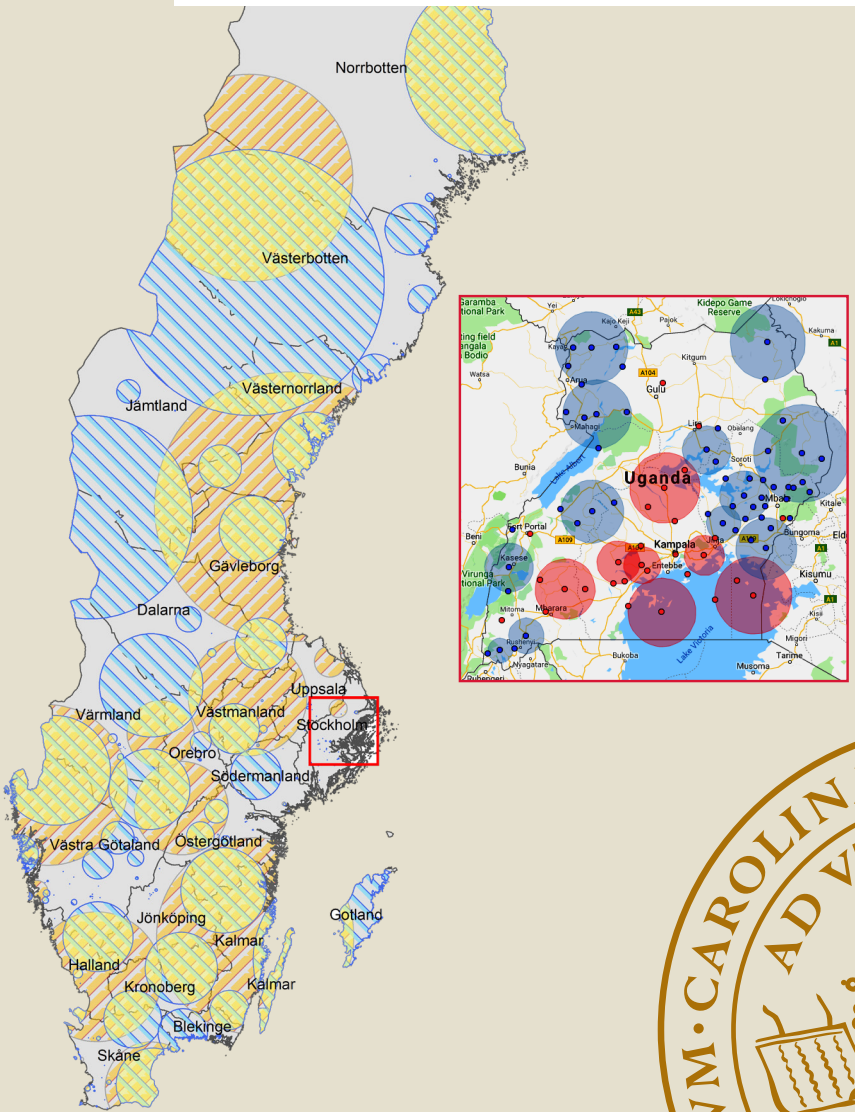
2020

Document Version:
Publisher's PDF, also known as Version of record

GIS and Health

Enhancing Disease Surveillance and Intervention through Spatial Epidemiology

AUGUSTUS ATURINDE
DEPARTMENT OF PHYSICAL GEOGRAPHY AND ECOSYSTEM SCIENCE | LUND UNIVERSITY



Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

Department of Physical Geography
and Ecosystem Science
Faculty of Science

ISBN 978-91-985015-9-9

GIS and Health

Enhancing Disease Surveillance and Intervention through Spatial Epidemiology

GIS and Health

Enhancing Disease Surveillance and Intervention through Spatial Epidemiology

Augustus Aturinde

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Science, Lund University, Sweden.
To be defended at Gotland, Geocentrum I, Sölvegatan 12, Lund, Sweden. Friday, November 27, 2020, at 9:00 am.

Faculty opponent
Associate Professor Takeshi Shirabe

KTH Royal Institute of Technology

Organization LUND UNIVERSITY Department of Physical Geography and Ecosystem Science Sölvegatan 12, SE-223 62, Lund, Sweden	**Document name:** DOCTORAL DISSERTATION **Date of issue: 2020-11-02**
Author(s) AUGUSTUS ATURINDE	**Sponsoring organization:** Swedish International Development Cooperation Agency (SIDA)

Title and subtitle: GIS and Health: Enhancing Disease Surveillance and Intervention through Spatial Epidemiology

Abstract

The success of an evidence-based intervention depends on precise and accurate evaluation of available data and information. Here, the use of robust methods for evidence evaluation is important. Epidemiology, in its conventional form, relies on statistics and mathematics to draw inferences on disease dynamics in affected populations. Interestingly, most of the data used tend to have spatial aspects to them. However, most of these statistical and mathematical methods tend to either neglect these spatial aspects or consider them as artefacts, thereby biasing the resultant estimates. Thankfully, spatial methods allow for evidence evaluation and prediction in epidemiologic data while considering their inherent spatial characteristics. This, thus, promises more precise and accurate estimates.

This thesis documents and illustrates the contribution spatial methods and spatial thinking makes to epidemiology through studies carried out in two countries with different heath-data quality realities, Uganda and Sweden. To be able to use spatial methods for epidemiology studies, proper spatial data need to be available, which is not the case in Uganda. Consequently, this study had two main aims: (1) It proposed and implemented a novel way of spatially-enabling patient registry systems in settings where the existing infrastructures do not allow for the collection of patient-level spatial details, prerequisites for fine-scale spatial analyses; (2) Where spatial data were available, spatial methods were used to study associative relationships between health outcomes and exposure factors. Spatial econometrics approaches, especially spatially autoregressive regression models were adopted. Also, consistent with location-specific epidemiologic intervention, the advantages of using spatial scan statistics, Geographically Weighted (Poisson) Regression and local entropy maps to distil model parameter estimates into their inherent spatial heterogeneities were illustrated.

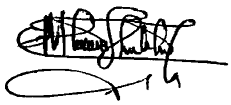
Our results illustrated that through the use of mobile and web technologies and leveraging on existing spatial data pools, systems that enable recording and storage of geospatially referenced patient records can be created. Also, spatial methods outperformed conventional statistical approaches, giving refined and more accurate parameter estimates. Finally, our study illustrates that the use of local spatial methods can inform policy and intervention better through the identification of areas with elevated disease burden or those areas worth additional scrutiny as illustrated by our study of HIV-TB coinfection areas in Uganda, the areas with high CVD-air pollution associations in Sweden, and areas with consistently high joint mortality burden for CVD and cancer among the Swedish elderly.

Overall, the incorporation of spatial approaches and spatial thinking in epidemiology cannot be overemphasized. First, by enabling the capture of fine-scale personal-level spatial data, our study promises more robust analyses and seamless data integration. Secondly, associative analyses using spatial methods showed improved results. Thirdly, identification of the areas with elevated disease burden makes identifying the primary drivers of the observed local patterns more informed and focused. Ultimately, our results inform healthcare policy and strategic intervention as the most affected areas can easily be zoned out. Therefore, by illustrating these benefits, this study contributes to epidemiology, through spatial methods, especially in the aspects of disease surveillance, informing policy, and driving possible effective intervention.

Keywords: spatial epidemiology, spatial econometrics, HIV-TB, CVD, cancer, Uganda, Sweden

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language:** ENGLISH
ISSN and key title		**ISBN: 978-91-985015-9-9**
Recipient's notes	**Number of pages** 122	Price
	Security classification	

Signature  Date: 2020-11-02

GIS and Health

Enhancing Disease Surveillance and Intervention through Spatial Epidemiology

Augustus Aturinde

LUND
UNIVERSITY

A doctoral thesis at a university in Sweden is produced either as a monograph or as a collection of papers. In the latter case, the introductory part constitutes the formal thesis, which summarizes the accompanying papers already published or manuscripts at various stages (in press, submitted or in preparation).

Cover photo by Augustus Aturinde

Faculty of Science
Department of Physical Geography and Ecosystem Science

ISBN (print): 978-91-985015-9-9
ISBN (PDF): 978-91-89187-00-9

Printed in Sweden by Media-Tryck, Lund University
Lund 2020

To our daughter,
Alexa Ankunda 'Tutu'.

Contents

List of papers

I. **Aturinde, A.**, Farnaghi, M., Pilesjö, P. & Mansourian, A. 2019. Spatial analysis of HIV-TB co-clustering in Uganda. *BMC infectious diseases,* 19, 612. DOI: https://doi.org/10.1186/s12879-019-4246-2

II. **Aturinde, A.**, Rose, N., Farnaghi, M., Maiga, G., Pilesjö, P. & Mansourian, A. 2019. Establishing spatially-enabled health registry systems using implicit spatial data pools: case study–Uganda. *BMC medical informatics and decision making,* 19, 215. DOI: https://doi.org/10.1186/s12911-019-0949-y

III. **Aturinde, A.**, Mansourian, A., Farnaghi, M., Pilesjö, P. & Sundquist, K. 2020. Spatial analysis of ambient air pollution and Cardiovascular disease (CVD) hospitalization across Sweden. *(under review for publication in GeoHealth).*

IV. **Aturinde, A.**, Mansourian, A., Farnaghi, M., Pilesjö, P. & Sundquist, K. 2020. Analysis of spatial co-occurrence between cancer and cardiovascular disease mortality and its spatial variation among the Swedish elderly (2010-2015). *(Accepted for publication in Applied Geography).*

Contributions

I. AA conceived the idea, led the study design, data preparation, implementation of the study, interpretation of the results and writing of the manuscript.

II. AA conceived the idea, led the study design, data preparation, participated in the implementation of the system, led the interpretation of the results, and writing of the manuscript.

III. AA participated in the study conceptualization and methodology, led the data preparation, implementation, analysis, interpretation of the results, and writing of the manuscript.

IV. AA conceived the idea, prepared the data, led the methodology, implementation, analysis, interpretation of the results, and writing of the manuscript.

Abstract

The success of an evidence-based intervention depends on precise and accurate evaluation of available data and information. Here, the use of robust methods for evidence evaluation is important. Epidemiology, in its conventional form, relies on statistics and mathematics to draw inferences on disease dynamics in affected populations. Interestingly, most of the data used tend to have spatial aspects to them. However, most of these statistical and mathematical methods tend to either neglect these spatial aspects or consider them as artefacts, thereby biasing the resultant estimates. Thankfully, spatial methods allow for evidence evaluation and prediction in epidemiologic data while considering their inherent spatial characteristics. This, thus, promises more precise and accurate estimates.

This thesis documents and illustrates the contribution spatial methods and spatial thinking makes to epidemiology through studies carried out in two countries with different heath-data quality realities, Uganda and Sweden. To be able to use spatial methods for epidemiology studies, proper spatial data need to be available, which is not the case in Uganda. Consequently, this study had two main aims: (1) It proposed and implemented a novel way of spatially-enabling patient registry systems in settings where the existing infrastructures do not allow for the collection of patient-level spatial details, prerequisites for fine-scale spatial analyses; (2) Where spatial data were available, spatial methods were used to study associative relationships between health outcomes and exposure factors. Spatial econometrics approaches, especially spatially autoregressive regression models were adopted. Also, consistent with location-specific epidemiologic intervention, the advantages of using spatial scan statistics, Geographically Weighted (Poisson) Regression and local entropy maps to distil model parameter estimates into their inherent spatial heterogeneities were illustrated.

Our results illustrated that through the use of mobile and web technologies and leveraging on existing spatial data pools, systems that enable recording and storage of geospatially referenced patient records can be created. Also, spatial methods outperformed conventional statistical approaches, giving refined and more accurate parameter estimates. Finally, our study illustrates that the use of local spatial methods can inform policy and intervention better through the identification of areas with

elevated disease burden or those areas worth additional scrutiny as illustrated by our study of HIV-TB coinfection areas in Uganda, the areas with high CVD-air pollution associations in Sweden, and areas with consistently high joint mortality burden for CVD and cancer among the Swedish elderly.

Overall, the incorporation of spatial approaches and spatial thinking in epidemiology cannot be overemphasized. First, by enabling the capture of fine-scale personal-level spatial data, our study promises more robust analyses and seamless data integration. Secondly, associative analyses using spatial methods showed improved results. Thirdly, identification of the areas with elevated disease burden makes identifying the primary drivers of the observed local patterns more informed and focused. Ultimately, our results inform healthcare policy and strategic intervention as the most affected areas can easily be zoned out. Therefore, by illustrating these benefits, this study contributes to epidemiology, through spatial methods, especially in the aspects of disease surveillance, informing policy and driving possible effective intervention.

Sammanfattning

Framgången av en evidensbaserad intervention beror på precis och tillförlitlig utvärdering av tillgängliga data och information. Här är användningen av robusta metoder för bevisvärdering viktig. Epidemiologi, i dess konventionella form, förlitar sig på statistik och matematik för att dra slutsatser om sjukdomars dynamik i drabbade populationer. Intressant är att de flesta data som används ofta innefattar rumsliga aspekter. Dock är det så att de flesta statistiska och matematiska metoder tenderar att antingen försumma dessa rumsliga aspekter, eller betrakta dem som artefakter och därmed öka osäkerheten i de resulterande uppskattningarna. Tack och lov möjliggör rumsliga metoder utvärdering av analys och resultat innefattande rumsliga epidemiologiska data med beaktande av deras inneboende rumsliga egenskaper. Detta kan resultera i mer precisa och exakta uppskattningar.

Denna avhandling dokumenterar och illustrerar bidraget rumsliga metoder och rumsligt tänkande gör till epidemiologi, genom studier genomförda i två länder med olika förutsättningar avseende datatillgänglighet, Uganda och Sverige. För att kunna använda rumsliga metoder för epidemiologistudier krävs korrekt rumslig information, vilket generellt inte är fallet i Uganda. Följaktligen hade denna studie två huvudmål: (1) Den föreslår och implementerar en ny modell för rumsliga patientregistreringssystem i miljöer där de befintliga infrastrukturerna inte möjliggör insamling av rumsliga detaljer på patientnivå, dvs. saknar förutsättningar för finskala rumsliga analyser; (2) Då rumsliga data finns tillgängliga, används rumsliga metoder för att studera associativa förhållanden mellan hälsoutfall och exponeringsfaktorer. Rumsliga ekonometriska tillvägagångssätt, särskilt rumsligt autoregressiva regressionsmodeller, har använts. I överensstämmelse med platsspecifik epidemiologisk intervention illustreras också fördelarna med att använda statistisk skanningsstatistik, geografiskt viktad (Poisson) regression och lokala entropikartor för att destillera parameter-uppskattningar avseende deras inneboende rumsliga heterogenitet.

Våra resultat illustrerar att genom användning av mobil- och webbteknologier, samt utnyttjande av befintliga rumsliga datapooler, kan system som möjliggör registrering och lagring av geospatialt refererade patientjournaler skapas. Dessutom överträffade

rumsliga metoder konventionella statistiska tillvägagångssätt, vilket gav förfinade och mer exakta parameteruppskattningar. Slutligen illustrerar vår studie att användningen av lokala rumsliga metoder kan informera beslutsfattare (t.ex. avseende policy och intervention) bättre genom att identifiera områden med förhöjd sjukdomsbild, eller de områden som av annan anledning är värda ytterligare granskning. Detta illustreras i våra studier av HIV-TB-infektionsområden i Uganda, områden med höga CVD-luftföroreningsföreningar i Sverige och områden med genomgående hög gemensam dödlighet för CVD och cancer bland äldre svensk befolkning.

Sammantaget kan införlivandet av rumsliga tillvägagångssätt och rumsligt tänkande i epidemiologi inte överbetonas. Först, genom att möjliggöra insamling av rumsliga data på finskalig personlig nivå, indikerar vår studie mer robusta analyser och sömlös dataintegration. För det andra visade associativa analyser med användning av rumsliga metoder förbättrade resultat. För det tredje gör identifiering av områden med förhöjd sjukdomsbild det möjligt att identifiera de primära drivkrafterna för de observerade lokala mönstren mer tillförlitligt och fokuserat. I slutändan kan våra resultat användas inom vårdpolitik och strategisk intervention eftersom de mest drabbade områdena enkelt kan identifieras och därmed regleras. Genom möjligheten att illustrera dessa fördelar ger denna studie ett bidrag till epidemiologin, genom rumsliga metoder, särskilt när det gäller övervakning av sjukdomar, information till beslutsfattare och möjligheter att driva effektiv intervention.

Acknowledgement

Significant contributions towards my PhD study have been made by many people and organizations. First, my gratitude goes to the Swedish International Development Cooperation Agency (SIDA) and the Training for Sustainable Spatially Enabled e-Services Delivery in Uganda (TSSEED) project at the College of Computing and Informatics Technology, Makerere University – Uganda in collaboration with Lund University – Sweden. I am also grateful to Kyambogo University for granting me a four-year study leave from 2016 to 2020.

Special thanks go to the team of dedicated and passionate supervisors without whom finishing this task would certainly be more uphill if not impossible. Associate Prof. Ali Mansourian, thank you for trusting in me throughout this whole journey. You welcomed me, you challenged me through our engagements, and you made your door ever open often bumping into you without appointments. You are a role model and true inspiration to any budding scientist.

Professor Petter Pilesjö, thank you for the guidance especially on aspects of clear and precise scientific communication. Your guidance often surpassed core academics and bordered on parenthood and you did this without a blink – thank you.

Dr. Mahdi Farnaghi, first, thanks for your patience. We shared more discussions than any other combination I am aware of. You always listened for the slightest of the ideas I had. We discussed, we agreed, we disagreed, but above all, you did this with one thing in mind – to better me and see me through this chapter of my life. You were a good friend and for that I am truly grateful and indebted.

Professor Kristina Sundquist, thank you for your guidance. When we first met in that meeting at the GIS centre, I didn't have the slightest idea that what I was presenting made a lot of sense. However, from that chaos, you showed that you liked my work, and throughout this process, I always looked forward to our Malmö meetings. I can only be thankful for you being on my supervisory team.

Professor Janne Rinne (my department representative), our interactions were full of life. Your approach to my issues as a student were handled with utmost clarity. For that, I am grateful.

I am grateful to the staff members and colleagues at my University, Kyambogo University. My Head of Department, Dr Ing. Ssengooba Kasule, Dean – Dr Wandera Cathy, Dr Kyakula Micheal, Dr Okia Yafesi, Dr Otukei Edward, J Rutabajuka, P Omute, S Amuko, C Tembo, L Anguyo, M Kabagambe, SP Mwesigye, T Okello, M Tusiime, S Nagujja, K Annett, J Muwanguzi, S Ainomugisha, R Basajja, N Natukunda, Snr Enyogoi (RIP), R Lubwama, G Arinaitwe, J Mubiru, I Enyogoi, D Mugerwa, I Buhamizo, Milly, I Wadembere, J Ainebyoona, T Lugolobi, Jacinta N, Mr Kirunda, and many other colleagues at the DLAS and in Management that I might not have listed here. My appreciation extends to Ms Esther Musiime and the V/C, Prof. Elly Katunguka. You made this hectic journey lighter and for that, I am truly grateful.

I am thankful to the team at Makerere University. My project PI, Associate Prof. Gilbert Maiga, Dr Nakakawa, Dr Bagarukayo, Dr Ssemaluulu, Dr Ssebugwawo, Dr Mwebaze, Associate Prof. Bainomugisha and many others at CoCIS, you made Makerere hospitable for me and for that I am grateful. Prossy, Wycliff, Pearl, Shirley at CoCIS; Nestor, Lubowa, Charles and Annet at DRGT, thank you. I am grateful to my fellow PhD students at Makerere: Ongaya, Irene, Rose, Lillian, Felix, Odongtoo, Grace, Andrew, Ismail, Sanya, Justine, Elizabeth, and many others whom we shared the struggle, I am grateful for the togetherness and atmosphere you created.

I am thankful for friends at INES. Karin, Roger, Mich, Micael, Lars H, Abdulghani, Mitch, Andreas, Jonas A, Jonas Å, Ann Å, David T, Thomas, Stefan, Lars E, Anna Maria, Maj-Lena, Paul, Anders, Irma, Per-Erik, Natalie, Yvonne, Torbern, Susanna, Britta, Eva, Rafael, and Ricardo. George, Olive, Altaaf, Pearl, Sofia, Klas, Enas, Adrian, Joel, Geerte, Tetiana, Antje, Hani, Alexandra, Tome, Fabien, Patrik, Wexin, Yanzi, Weiming, Finn, Hakim, Reza, Alex, Zhendong, Feng, Cecilia, Ehsan, Tomas, Jeppe, Pinar, Fredrik, Marcin, Per-Ola, and many others that I might not have outlined here.

I am grateful to my family. Dad, Mum, Allen, Abert, Alex, Alice, Anorld, Ann, Brian for your unending support and prayers. Mum-Robinah, uncle-Mike & family, uncle-Francis & family, uncle-Robert & family, Podi & family, Richard Kapaasi & family, Ernest & family, Amos & family, and many other families that became our own during my study. I thank my Mamba family, Leonard, Chris, Gerald, and Simon, and our family friends Prossy, Ayesiga, Oliver, Patience, Rose, Nelson, Joel & Racheal, and many more not listed.

Finally, special thanks to my dear wife, Tasha and my daughter, Alexa. Thank you for your support and love, without which my PhD experience would have been more hectic. You mean a lot to me.

1 Introduction

1.1 Background

Epidemiology entails the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems (CDC, 2012). The roots of epidemiology can be traced back to the days of the Greek physician Hippocrates of Cos (460 BC – 377 BC) who is considered the first epidemiologist (Morabia, 2004). For long, epidemiology was limited to infectious diseases through studying, documenting and analysing their spread within given populations to advise their prevention (Kuller, 1991). However, the scope of epidemiology has since expanded and currently refers to the study of any health condition that occurs in excess of normal expectancy (Gerstman, 2013). Even in this non-limiting sense, epidemiology still refers to the study of epidemics and their prevention (Kuller, 1991), and is to be differentiated from clinical medicine. The epidemiologist's primary unit of concern is, "an aggregate of human beings", as opposed to an "individual," for a clinician (Greenwood, 1935, Souris, 2019).

As a study, epidemiology is quantitative, data-driven, and relies on the systematic and unbiased collection, analysis, and interpretation of data (Dicker, 2008). Epidemiology's main objective is to uncover the relationships between the observed disease dynamics and the risk factors and to confirm that the risk factors affect the disease through some understandable mechanisms – using mathematics, statistics, and modelling (Souris, 2019). Epidemiologic data are obtained from several sources, including vital statistics data, government surveillance data and reports, health surveys, and disease registries to study factors associated with certain diseases or conditions (Torrence, 2002). However, for most of the non-communicable diseases such as heart diseases, cancer, diabetes, chronic pulmonary, and mental diseases, it is the disease registries at primary health care units (hospitals, clinics, etc.) that are often used (Boutayeb and Boutayeb, 2005). These registries capture the patient attributes like age, sex, marital status, occupation, family history of the disease, date of admission or visitation, and more importantly for our study, the absolute location of their place of work/residence, for custody purposes and/or to facilitate clinical care.

The need to record patient location details is important as places and environments influence not only the lifestyle of their occupants but also influence the diseases that affect the inhabitants. This location and season dependence of diseases is not new. In his treatise "Airs, Waters, and Places", Hippocrates (460 BC – 377 BC) expressed his conviction that man's external environment had some direct influence on his physical constitution and health, and that by studying a location's reference to the sun, the soil, the elevation, prevailing winds and the nature of the water supply, one was able to predict the character of the population and its diseases (Miller, 1962). This thinking still largely forms the basis for spatial epidemiology, a branch of epidemiology that focuses on the spatial distribution of risk factors, disease outcomes, and their spatial intersection.

Spatial epidemiology, the study of the description and analysis of spatially-indexed health data to characterize spread and possible causes (Elliott et al., 2000), principally works from the basis of three observations. First, diseases tend to vary in geographical space; second, this spatial variation is driven by the variations in the biotic and abiotic conditions that support the pathogen and its vectors and reservoirs; and third, if these biotic and abiotic conditions can be delimited on the map, then both current risk and future changes in risk should be predictable (Pavlovsky, 1966, Ostfeld et al., 2005). As such, spatial epidemiology uses the geographical distribution of disease to better understand the aetiology of disease through associations with the demographic, environmental, genetic, behavioural, socioeconomic, and infectious risk factors (Elliott and Wartenberg, 2004).

Although the importance of place in human health has long been recognised (Morabia, 2004), public health research has mostly focused on person and time, with little consideration of "the place" (Rezaeian et al., 2007). This is unfortunate as a comparison between places, times, and individuals, provides useful information for formulating and testing aetiological hypotheses (Jia, 2019). Some of the reasons for this apparent lack of interest in "the place" include lack of appropriate databases (or data not having spatial details), the complexity of spatial analysis tools, and lack of appropriate software (Hawkins, 2012, Souris, 2019). From a public health perspective, spatially-indexed epidemiologic analyses are very important in linking observed health outcomes with environmental exposures (Kirby et al., 2017). Such analyses are hence effective tools in informing healthcare policy, allocation of resources for monitoring, intervention, prevention and treatment of diseases.

1.2 Research gap

There have been various studies in relation to spatial epidemiology, targeting clinical and policy interventions. In all these studies, the existence of spatially geo-referenced health data is the required starting point. As such, to enable eventual spatial analyses on the data captured in the healthcare registries, the explicit location of the place of residence, as a patient attribute, must be captured along with other personal details. Some countries have well-developed health and population registry systems that enable capturing of this spatial data, like the existence of a personal identification number (PIN) that is linked to one's place of residence for Scandinavian countries (Brook et al., 2004). For some countries, like the USA and the United Kingdom, the PIN is not directly linked to location so registries rely on reported ZIP codes and Postcodes respectively (Elliott and Wartenberg, 2004). For most developing countries, especially in Africa, however, the lack of an addressing system means that no explicit spatial reference can be made to the location of the patients.

Underlying the PINs, ZIP codes and Postcodes are some forms of national Spatial Data Infrastructure (SDIs) that enable geocoding, and these SDIs are currently lacking in many of the resource-constrained African countries. SDIs are broadly defined as the technology, policies, standards, and human resources necessary to acquire, process, store, distribute, and improve utilization of spatial data, services, and other digital resources (Hu and Li, 2017). This, therefore, means that the lack of SDIs leads to difficulties in capturing location data generally, and patient-specific location data for our case, that would be used in both spatial epidemiologic analyses and help in the delivery of e-health services. The implication of this inability to capture fine-level spatial details is that the only spatial analyses possible are those done at coarse-level geographical aggregations. Additionally, from an analysis standpoint, most of the epidemiologic studies tend to ignore the consideration of spatial effects inherent in the morbidity and mortality data used. Failure to account for spatial effects may bias the estimates as well as affecting precision (McDonald, 2013). Accordingly, accounting for spatial dependence may improve causal inference hence policy interventions in public health problems.

This study, therefore, began by utilizing coarse-scale HIV and TB admission data and investigated their spatial co-clustering in Uganda. Here, a case was made that such coarse-scale data make targeted epidemiologic intervention difficult at best and impossible at worst, as the identification of local target foci of transmission become masked at such coarse spatial scales. We, thus, proposed, developed and implemented a digital spatially-enabled health registry system that utilized existing spatial data pools in areas without working SDIs like Uganda. The system allows for the capture

of fine-level patient spatial details at hospital consultation and/or admission. For settings with fine-scale patient spatial data already, we used spatially-explicit methods to investigate the nature of the associations between cardiovascular diseases and ambient air pollution, as well as the spatial variation of these associations across Sweden. Finally, owing to the prominence of comorbidities and their accelerated negative effects to health outcomes, fine-level spatially varying relationships between CVD and cancer were investigated in the Swedish elderly, using spatially-shared local information between the two causes of death through joint entropy analysis.

This dissertation is based on paper-compilation. As such, some repetitions especially in the general literature review, methodology and results here, and in the individual papers could not be avoided.

1.3 The aim and objectives

The aim of this study is two-folded. 1) to propose and test the possibility of using spatial data pools to create systems that enable spatial referencing of patient records, in areas where infrastructures are inexistent; (2) to use spatially-explicit methods, approaches and spatial thinking to enhance epidemiologic intervention.

Specifically, the study explored the possibilities of spatially enabling health registries and the application of spatial approaches to improve disease surveillance and disease intervention and control strategies through spatially-explicit analyses. These objectives are listed below as:

1. Adopt cluster detecting models to investigate the simultaneous spatial variation of co-infectious disease clusters from spatially aggregated data.
2. Establish a spatially-enabled patient registry system through the use of available implicit spatial data pools.
3. Adopt spatially-explicit regression models for environmental-disease surveillance.
4. Adopt joint local entropy models to investigate the spatial variation of co-morbidities and co-mortalities.

1.4 Thesis organisation

The thesis is organized into five chapters. After this introductory chapter, chapter 2 presents a review of the literature about epidemiology in general and spatial epidemiology in particular. Chapter 3 gives a detailed description of the methods and data used in the study. Chapter 4 summarizes the four resulting papers from the study. The final chapter is chapter 5 that presents the conclusions and recommendations. The resulting four papers, from which the methods, results, discussions and recommendations of this thesis are based, have been attached as a main part of the thesis.

2 Literature Review

2.1 Historical perspectives on epidemiology

The first rational explanation of disease was by the Greek physician Hippocrates of Cos (460 BC – 377 BC) who is considered as the father of medicine and the first epidemiologist (Morabia, 2004). He recognised that some forms of sickness were always present in a population, but other forms were either not usually present or, if present, exhibited seasonality in the form of being common at certain periods of the year and in certain years. Through his book “On Airs, Waters and Places”, he distinguished between "endemic" diseases, that are always present in a population and "epidemic" diseases, which can become excessively frequent and then disappear (Merrill, 2012). He, thus, was concerned about the factors responsible for local endemicity as well as reasons for epidemic prevalence (Greenwood, 1935). In this, he considered diseases as both a mass phenomenon as well as an individual occurrence and built the theory of causation based on observation of the association between disease and factors such as geography, climate, diet, and living conditions.

This association aspect of diseases and the environment was popularized by Hieronymus Fracastorius (1478 – 1553) who theorized that there exists a transference contagion, in which conveyance of a disease from an infected person to another person (hitherto uninfected) is accomplished (Duncan et al., 1988). Three types of contagion were distinguished as direct contact, germ contagion and “infection at a distance”, and these three still underlie most of the infectious disease epidemiology (Ostfeld et al., 2005). By using observation and mortality records, John Snow (1813 – 1858) was arguably the most noted epidemiologist of the nineteenth century (Howe, 1964). He identified the common of source of cholera contamination, as a water source (borehole) on Broad Street, London by plotting Cholera mortality statistics that he derived from his detailed scenario records of cholera dynamics including modes of transmission, incubation times, cause-effect association, clinical observation, scientific observation of water from different sources, as well as differences between those who got the disease and those who did not (Vinten-Johansen et al., 2003). And although John Snow was unable to identify the causing

agent of cholera, his use of statistical records enabled him to isolate contaminated water as the risk factor associated with cholera (Dicker, 2008).

Given that epidemiology is concerned with what befalls a group of human beings as opposed to individuals (Lawson et al., 2016), keeping records of morbidity and mortality in a given population is vital (Gerstman, 2013). By publishing "bills of mortality" in London weekly, John Graunt (1620 – 1674) managed to identify variations in death according to gender, residence, season and age (Rothman, 1996). Graunt's statistics were given more authority by William Farr (1807 – 1883) who organized and developed the vital statistics system as we know it and helped in the analysis of disease aetiological factors (Merrill, 2012). These aetiological factors tend to vary in both space and time.

2.2 Spatial analysis and spatial epidemiology

Epidemiology, being quantitative, begins with having recorded data. For spatial analysis to be possible, some spatial aspects of the phenomena of the population being studied must be captured. Normally, in disease-related data recording, one's residence or workplace are tagged along with the personal level details. Consequently, ZIP codes, Postcodes and Personal Identification Numbers (PINs) are used. These codes and numbers are in most cases geocoded, enabling retrieval of precise geo-locations of individual residences or workplaces. In settings where there are no geo-referenced ZIP codes, Postcodes or PINs due to lack of enabling infrastructures, fine-scale spatial analysis later alone spatial epidemiology becomes impossible. In essence, the very starting point of any form of spatial analysis on the recorded data emanates from having spatial data captured through some form of spatially enabling infrastructures.

Descriptive epidemiology focuses on the triad of person, place and time (Duncan et al., 1988). Historically, epidemiologic research focusing on "the place" has been given less attention (Kirby et al., 2017). Modern epidemiology, however, has increasingly incorporated spatial perspectives into its research design and models as the inclusion of "the place" helps in tying the observed health outcomes to the place-specific exposure factors, thus providing useful information for formulating and testing aetiological hypotheses (Jia, 2019).

Spatial epidemiology concerns "research that incorporates the spatial perspective into the design and analysis of the distribution, determinants, and outcomes of all aspects of health and well-being…" (Kirby et al., 2017). It, thus, involves the use of epidemiologic study designs that make use of spatial data or spatially derived

information. Spatial datasets provide two types of information: (1) data describing the specific locations of objects in space (and their topological relationships), and (2) data describing non-spatial attributes of the objects recorded (thematic data). For example, the spatial data set might be describing mortality count of a given disease (thematic aspect) in a given municipality (spatial aspect).

Using spatial data, we can reveal that everything is related to everything else but nearer things are more related than distant things, according to Tobler's first law of geography (Tobler, 1970). This highlights the aspect that neighbourhoods influence what is observed. Said another way, the mortality observed in one municipality is influenced by the mortality in the neighbouring municipalities. Analysis of neighbourhood process results in spatial spill-overs and spatial dependence (Anselin, 2003). More importantly, these spatial effects in the form of dependence and spatial heterogeneity result in the violation of the independent observation assumption, synonymous with conventional statistics (Yao and Stewart Fotheringham, 2016). Conventional statistics and epidemiology tend to treat these spatial effects as some form of distortion or bias.

Spatial scientists and spatial epidemiologists, on the other hand, argue that these spatial effects do not constitute a bias; it is what they want to understand by evaluating its effect on the observed phenomena (Hawkins, 2012). The argument is that given the spatially structured distribution of diseases arising from aetiological processes operating in a spatially patterned environment, for example, any set of samples or representation of the disease burden (incidence, prevalence, etc.) must also contain this structure, if they are to be accurate. If spatial effects are part of the observed disease burden, and we are trying to understand the disease burden, it makes little sense to claim that spatial effects in the disease data represent some sort of bias or distortion. It, thus, follows that broad-scale epidemiologic data that do not contain spatial structure are missing key information that limits their value for understanding the disease spatial patterns being studied.

Failure to account for these spatial effects may bias the estimates and may affect precision obtained from regression models. Resultantly, accounting for spatial effects improves causal inference hence epidemiologic surveillance and policy intervention. Accounting for spatial dependence, however, calls for specialised methods of spatial statistics and spatial econometrics (Anselin, 1989) or spatial regression methods (LeSage and Pace, 2009). Additionally, for these methods to be useful in epidemiologic surveillance and targeted intervention, they must be able to distil the observed health outcomes into their local spatial heterogeneities, as well as depicting the established association relationships between the health outcomes and their independent variables at local spatial scales. The methods that deal with this kind of

local spatial autocorrelation and spatial heterogeneity form the basis for a hypothesis test for local spatial randomness, with the null hypothesis being one of spatial randomness – locally, any organisation of values in the neighbourhood is equally likely (Anselin, 2019).

Local spatial methods have gained prominence in geographical analysis in recent times. These methods are mainly concerned with local spatial heterogeneity (general considerations given in Fotheringham et al., 2002a and Lloyd, 2010) and local spatial autocorrelation generally considered under the Local Indicators of Spatial Association (LISA) framework (Anselin, 1995, Anselin and Rey, 2014). Both frameworks account for the neighbourhood through some form of spatial weights generated either through distance decay or spatial contiguity. The choice of whether to use distance decay or contiguity depends on the nature of the phenomena being studied, but tend to converge in results for most practical applications (Anselin et al., 2006).

In all, the use of these spatial methods improves the accuracy and precision of the obtained estimates. They would, therefore, improve intervention by identifying, at a local scale, which (local) risk factors are responsible for the observed health outcomes. Unfortunately, these spatial methods have not been widely applied in epidemiologic studies. This study, thus, provides numerous ways for incorporating such advanced spatial methods and spatial thinking and illustrates how doing so could improve epidemiologic surveillance through targeted intervention. Moreover, the local nature of the spatial methods adopted makes identification of areas requiring more epidemiologic intervention more straightforward – when compared with global solutions or non-spatial solutions that are more common in conventional epidemiology.

3 Data and Methods

The methods employed were primarily influenced by the nature of my study that involved working with datasets from two countries: Uganda and Sweden. The nature and spatial quality of these two groups of datasets required different approaches. For one (Uganda), the spatial scale of the available datasets was coarse while the datasets from Sweden were at fine spatial resolutions. The methods employed here also reflected this difference. Also, due to this limitation in the spatial scale of the Ugandan datasets, this inspired the proposition, design and implementation of a creative idea that included a system that allows for spatial enablement of health registry systems.

Consequently, the first group consists of the application of the different spatial methods to generate what could be interpreted as disease surveillance maps. The underlying characteristic of these approaches is that they all distil the observed or predicted disease prevalence, incidence or associations into their local spatial heterogeneities. As such, methods like spatial scan statistics, Local Indicators of Spatial Association, Geographically Weighted (Poisson) Regression and Local entropy maps, all used in this study, fall under this grouping.

Motivated by the fact that the inability to record patient spatial details limits spatial epidemiology analyses, the second group of methods is a unary category I have termed as the "development" component of the study. This is perhaps not a "method" in the strictest of the terms but a pragmatic approach used to propose, develop and implement a system that allows for spatially enabling health registry systems. It is specific to areas like Uganda where existing infrastructures do not allow for determination and recording of the precise location of the patient's residence or workplace upon admission or consultation.

3.1 Spatial statistics

3.1.1 Spatial scan statistics

Geographical disease surveillance scans for the presence of non-natural clusters of diseases in space and proceeds from the assumption that the background risk surface is flat, against which a peak (cluster) is being tested (Elliott and Wartenberg, 2004). A cluster can be defined as an unusually high concentration of disease events in a region unlikely to have happened out of chance (Turnbull et al., 1989). Spatial scan statistics is one of the methods that use point pattern to detect non-random clustering in geographical space (Kulldorff, 1997). Disease spatial cluster analysis is thus important in disease surveillance as it helps to identify areas where intervention is critical.

The earliest scan statistic was the Geographical Analytical Machine (GAM) advanced by Openshaw and colleagues (Openshaw et al., 1988). That notwithstanding, the most widely used spatial statistic is the Kulldorff spatial statistic (Sherman et al., 2014), which is both deterministic and inferential therefore allowing for identification of local clusters but also allowing for hypothesis testing and significance evaluation through the SaTScan software, and detects both circular and elliptical clusters (Chen et al., 2008, Tango and Takahashi, 2005).

As Chen et al. (2008) discussed, the SaTScan detects potential clusters by calculating the likelihood ratio (LR) given by equation (1).

$$LR_{(u)} = \left(\frac{c}{E_{[c]}}\right)^{c} \left(\frac{C-c}{C-E_{[c]}}\right)^{C-c} I\left(\frac{c}{E_{[c]}} > \frac{C-c}{C-E_{[c]}}\right) \qquad (1)$$

where C is the total number of observed cases in the study area; c is the observed number of cases within a circle; $E_{[c]}$ is the adjusted expected number within the window under the null hypothesis; $C - E_{[c]}$ is the expected number of cases outside the window, and $I\left(\frac{c}{E_{[c]}} > \frac{C-c}{C-E_{[c]}}\right)$ is the binary indicator of high-risk clusters (1) or low-risk clusters (0) or both (11). Based on the magnitude of the values of the likelihood ratio test, the set of potential clusters is then ranked and ordered. The circle with the maximum likelihood ratio among all radius sizes at all possible centroid locations is considered as the most likely cluster. The statistical significance of the clusters is determined through Monte Carlo simulations. Secondary clusters – those that have significantly large likelihood ratio but are not primary clusters can also be identified (Sherman et al., 2014).

3.1.2 Global Moran's I

The global Moran's index is used as a measure of the influence of neighbourhood values on the observed values. This neighbourhood influence is known as spatial autocorrelation and provides information about how the phenomenon under study tends to cluster in space (Cliff and Ord, 1970, Chien et al., 2015). Global Moran's I was used to estimate the degree of clustering of disease incidence rates according to equation (2).

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (2)$$

where n is the number of polygonal areas; S_0 is the sum of all weights w_{ij}, $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$; w_{ij} is the weight between observations i and j, and represents proximity between area a polygonal pair i and j; x_i is the incidence rate of a disease in the ith area; x_j is the incidence rate of a disease in the jth area; and $\bar{x}$ is the mean incidence rate of the disease under study for all the spatial polygons in the study area.

3.1.3 Local Moran's I

Whereas the global Moran's index in equation (2) shows the degree of clustering in the whole study area, it does not show variability in the clustering tendency of the phenomenon under study, across the study area. The local Moran's index, a Local Indicator of Spatial Association (LISA), was proposed by Anselin (1995) and allows for the global spatial autocorrelation to be distilled into its constituent clusters – cold spots and hotspots. The LISA of ith polygon can be calculated according to equation (3).

$$I_i = \frac{(x_i - \bar{x}) \sum_j w_{ij}(x_j - \bar{x})}{S^2} \quad (3)$$

where x_i is the incidence rate of a disease in the ith area; x_j is the incidence rate of a disease in the jth area; and $\bar{x}$ is the mean incidence rate of the disease under study for all the spatial polygons in the study area; w_{ij} is a weight parameter for a pair of polygons i and j and indicates proximity; S is the standard deviation of the disease incidence rate in the entire study area.

3.1.4 Bivariate LISA (Bi-LISA)

The bivariate local Moran's index (Bi-LISA) is an extension of the univariate local Moran's I outlined in the previous section. The Bi-LISA models the correlation between one disease prevalence rate (a) at a given location and another disease prevalence rate (b) at the neighbourhood location using equation (4).

$$I_i = \frac{(x_{ai} - \overline{x_a}) \sum_j w_{ij} (x_{\mathrm{bj}} - \overline{x_b})}{S^2} \qquad (4)$$

This approach was especially applicable for studying co-infections co-morbidities, and co-mortalities. The global and local Moran's I involved the computation of neighbourhood information captured by the spatial weight matrix. In both applications, the contiguity option of weight matrix generation was adopted.

3.2 Geographically Weighted Regression (GWR)

Due to non-stationarity of most disease variations, globally fitted spatial models (such as Ordinary Least Squares, spatial lag and spatial error models) assume stationary spatial effects, resulting in unrealistic universal relationships across the study space. Fotheringham et al. (2002b) contended that undertaking a global spatial analysis can be misleading. They thus proposed a local form of spatial modelling and analysis, termed as Geographically Weighted Regression (GWR). GWR, as shown in Figure 1, is a local form of weighted regression where the weights (W_{ij}) are calculated as an inverse function of the spatial distance (d_{ij}) between the predicted point and the data points (Fotheringham et al., 2002b). As such, near data points are given heavier weights compared to faraway points, with respect to the first law of geography: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

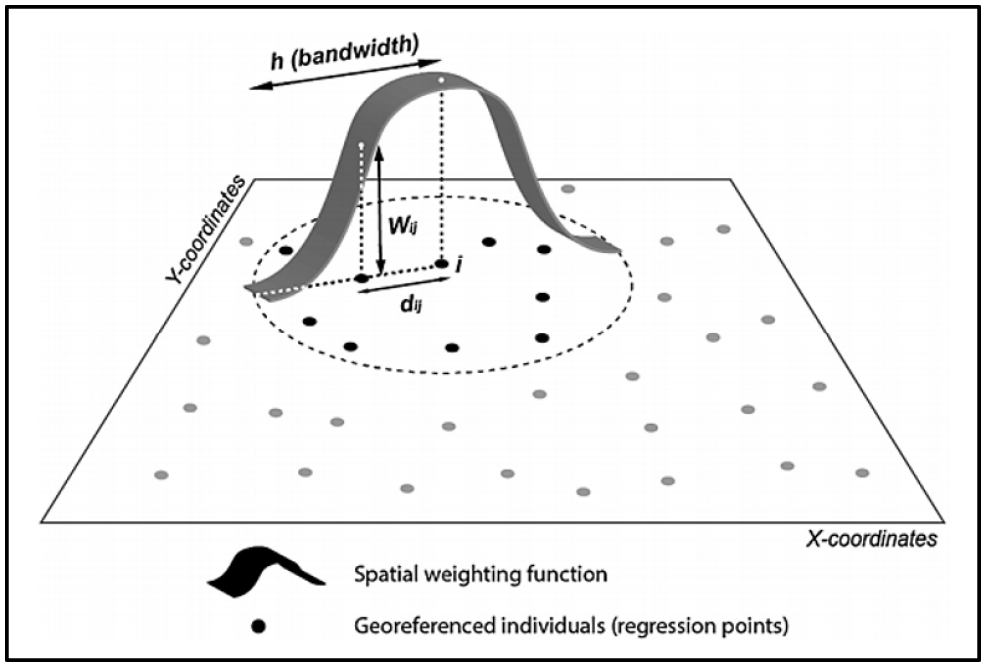


Figure 1: Schematic representation of the geographically weighted regression

For our study, we extended the Poisson variant of GWR, known as Geographically Weighted Poisson Regression (GWPR), to analyse the association between air multi-pollutants and cardiovascular disease (CVD) and the spatial variations of these associations across Sweden. The Poisson framework was used because of the count nature of the CVD records. The GWPR model can be expressed as equation (5). All analysis was done at SAMS (Small Area for Market Statistics) level, which is a census regional division, defined by Statistics Sweden (http://www.scb.se), based on homogenous types of buildings so that they approximately contain 1000 residents.

$$O_i \sim Poisson\left[N_i exp\left(\sum_k \beta_k(\boldsymbol{u}_i) x_{k,i}\right)\right] \quad (5)$$

where O_i denotes the SAMS observed CVD admission count; N_i denotes the SAMS specific underlying population; $\boldsymbol{u}_i = (u_{xi}, u_{yi})$ denotes a vector containing the two-dimensional coordinates describing the location of the particular SAMS (centroid coordinates); $x_{k,i}$ denotes the pollution variables. The regression coefficients, βs, are calculated for every SAMS (i), making them spatially varying. This makes GWPR a local spatial regression model allowing for geographically varying parameters.

3.3 Local Entropy Maps (LEM)

The concept of entropy has its roots in information theory and has been used in many application including measuring uncertainty in information theory (Gray, 2011), complexity in physics (Shannon, 1948), and diversity in ecology (Ricotta and Anand, 2006) just to mention a few. It has also, through the use of joint entropy, been used to study spatially varying multivariate relationships across space (Guo, 2010). It is this application in the spatial variability of multivariate relations that is more applicable to our study.

LEM is a non-parametric approach that proceeds from the computation of joint entropy using power-weighted minimum spanning trees (MST) as a proxy for the joint distribution of the variables (Jin and Lu, 2017). The advantage with LEM is that it does not assume a prior relationship form between the dependent and the independent variables; it also does not require specification of the underlying distribution of the data. This, therefore, makes it less restrictive in studying the nature of spatially-local relationships existing between variables (Guo, 2010).

Given that some form of assumption must be made for the data and the relationship in both LISA and spatial heterogeneity models like GW(P)R, we used a local entropy model to analyse associations without necessarily imposing assumptions on the relationship between the variables used. This promised to improve the definition of the association, especially in areas where the association is complex and not simply linear.

LEM analysis generally involves four main steps:

(1) estimation of Renyi entropy (H_λ) using the power-weighted MST length determined from the bivariate plot of the two variables, according to equation (6).

$$H_\lambda = \frac{1}{1-\lambda}\left(\log\left(M_\alpha \frac{(x_1, x_2, \ldots, x_n)}{n^\lambda}\right) - c\right) \qquad (6)$$

where x is a d-dimensional vector; $\lambda \geq 0$ is the order of the Renyi entropy; $M_\alpha(x_1, x_2, \ldots, x_n)$ is the minimum spanning tree length; n is the number of independent observations; c is a strictly positive constant that depends on the edge power, α and the dimensionality, d.

(2) evaluation for statistical significance of the obtained Renyi entropy values – converting each H_λ to p-values.

(3) processing all the p-values for the null hypothesis using several statistical tests, while controlling for the multiple testing problem.

(4) mapping and visualizing the p-values to examine for spatially varying local relationships between variables.

This particular approach was used to study spatially varying relationships between two non-communicable diseases – Cancer and CVD, among the Swedish elderly. The estimation of entropy values here also requires the definition of neighbourhood. The contiguity approach to neighbourhood specification was used.

3.4 Development – spatially enabled registry

This development is not a method if "method" is to be used in its precise terms. However, it is a pragmatic approach that was adopted to solve an existing problem. In essence, it is a combination of steps and procedures used to create a spatially enabled health registry system using existing spatial data pools.

The overall architecture is shown in Figure 2.

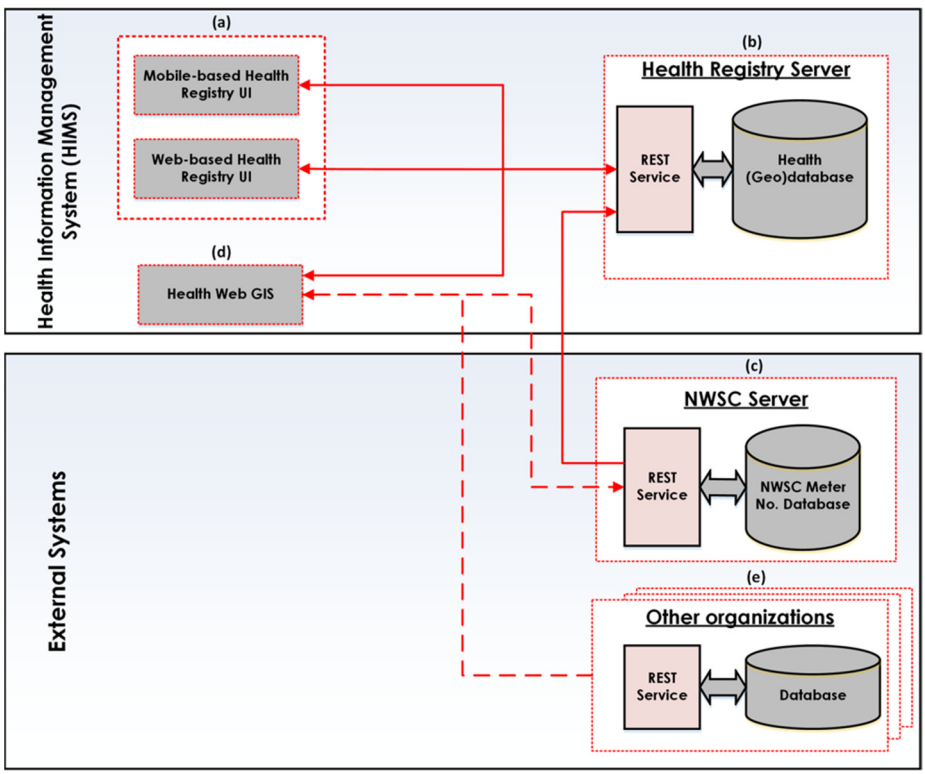


Figure 2: Schematic representation of the overall proposed spatially-enabled health registry system

The system is made up of the following components.

(a) Mobile-based health registry UI (user interface) and web-based health registry UI are used by medical personnel of healthcare centres, to register patients' admission details and patients' residential location. The geo-coordinates of patients' residence are either retrieved from the NWSC database through REST services or pined on the map using the health registry UI components.
(b) Health registry server provides the ability to save and retrieve health registry data from a (Geo)database through a REST Service.
(c) NWSC server provides the ability to access the water meter numbers and their respective geo-coordinates from the NWSC database through a REST service.
(d) Health Web GIS enables the healthcare personnel to analyse the admission data collected by the system as well as the data from other organisations that are published as REST Services
(e) These analyses can be used to answer specific spatial epidemiologic research questions.
(f) Other organizations can participate in this system by publishing their data through REST Services. Such data can then be used by the Health Web GIS component for contextual epidemiologic analysis.

The mobile-based health registry UI was developed as an android app using Java programming language. JavaScript programming languages, Cascading Style Sheets (CSS), and HyperText Markup Language (HTML) were used to develop the web-based health registry UI as well as the Health registry Web GIS. To provide mapping functionalities in the web applications, the Leaflet library (https://leafletjs.com/) was exploited.

To develop the web services, two frameworks, Service Oriented Architecture Protocol (SOAP) and REpresentational State Transfer (REST), are commonly used. However, SOAP has a heavyweight message payload thus not very favourable for resource-constrained mobile devices (Wagh and Thool, 2012). Subsequently, the REST web service framework was used in our study as its messages have a lightweight payload, hence more suitable for wireless and cellular connectivity networks synonymous with mobile devices (Wagh and Thool, 2012). The REST services were developed in Java programming language using oracle JAX-RS.

4 Results and Discussion

4.1 Introduction

In this chapter, the results obtained by applying the methods outlined in chapter 3, on the different case studies, are presented. The case studies were carried out in Uganda and Sweden, two countries with different spatial data quality realities. The study findings are majorly on (1) the proposition of an innovative way of using existing spatial data pools to create systems that enable spatial referencing of patient records in settings where existing infrastructures do not directly allow for geo-referencing of patient records – using Uganda's healthcare registry as a case study, and (2) adoption of spatially-explicit methods and approaches to enhance epidemiologic surveillance and intervention. In this regard, infectious diseases (HIV and Tuberculosis) and non-communicable diseases (Cancer and Cardiovascular disease) in Uganda and Sweden respectively were used as examples. The four accruing sub-studies, in the form of papers, are summarized next.

4.2 Summary of Paper-I

Title: Spatial analysis of HIV-TB co-clustering in Uganda

This study aimed to examine the extent to which Human Immunodeficiency Virus (HIV) and Tuberculosis (TB) spatially clustered together, in Uganda. This was motivated by the evidence available at the population level that these two diseases tend to co-exist in HIV patients, simultaneously progressing each other in co-infected patients, to the detriment of the patient's health. The World Health Organization (WHO) has since advocated for joint management of the two diseases through synchronised care and medication at a patient level. Given that epidemiologic intervention is seldom to individual patients but rather to affected communities and populations, this study geared towards establishing areas with common high (and low) prevalence rates, and more importantly spell out the likely driving factors for the observed spatial patterns.

Data from the District Health Information Software 2 system that is housed and maintained by the Ministry of Health – Uganda, was used. These were records of HIV and TB cases for the years 2015, 2016, and 2017 aggregated to the district level. The spatial methods of global and local Moran's indices, spatial scan statistics and Bivariate Local Indicators of Spatial Association were used to investigate the clustering patterns of both diseases, with the Bivariate-LISA capable of showing districts with similarly high prevalence rates in both diseases. Those were areas potentially requiring immediate coordinated attention. Highlighted too were areas that had similarly low prevalence rates, where intervention, relative to the high prevalence areas can afford to wait.

Our results showed that HIV and TB have relatively different spatial clustering patterns even when they seem globally highly correlated. They also showed that areas around the lakes, especially around Lake Victoria had persistently high joint prevalence rates, similar to some districts in Northern Uganda. The areas with persistently low joint prevalence rates were those around the Eastern districts, and around Kasese district in western Uganda. The possible reasons for these joint spatial patterns could be varied ranging from lifestyle-related factors in the lake regions, to probable influence of war in the north, to cultural practices like circumcision in the eastern and western districts with low joint prevalence rates.

Such results, depicting the spatial heterogeneity in the joint disease burden, are important as they provide actionable evidence for policy adjustment and plausible grounds for targeted intervention as the local areas affected are identified. Thus, this study through the use of spatial approaches made a significant contribution to addressing the knowledge-gaps in implementing the WHO recommendation for coordinated management of HIV and TB in the face of HIV-TB coinfection in Uganda by providing starting points for informed targeted epidemiologic intervention.

4.3 Summary of Paper-II

Title: Establishing spatially enabled health registry systems using implicit spatial data pools: case study – Uganda

This study aimed to provide a means that enables the capture of spatially high-resolution patient data upon hospital admission or consultation. The motivation was that currently, data that are captured at points of healthcare are inherently spatially aggregated to villages, parishes, counties or even districts (as was the case in Paper 1).

This inherent aggregation not only makes an intervention in case of an emergency impossible as the patient cannot be uniquely and independently identified for rescue but also makes the utility of the collected data for spatial analyses – as applied in spatial epidemiology, problematic.

This study, therefore, uses a pragmatic approach that utilizes already collected and available spatial data (called spatial data pools) from a National Water provider (NWSC) to a system that then enables the health registries to record spatially-referenced patient data upon admission or consultation. The system proposed, designed and implemented leverages on existing technology and uses interoperable web services to capture fine-level patient spatial data that is then linked with patient non-spatial information. The resultant data captured can be used in both emergency intervention as well as in fine-scale spatial analyses for epidemiologic surveillance and intervention.

This creative cost-effective solution utilizes what is already available, and is feasible for collection of spatially-indexed health records in countries with (unfortunate) data infrastructure realities similar to those of Uganda. These records can then be used in analyses for identifying spatial disease hotspots and clusters in disease incidence/prevalence rates. Additionally, the inherent integrating characteristics of spatial data can be utilized to link health outcomes with environmental exposures, improving epidemiologic provisioning, policy, and planning.

4.4 Summary of Paper III

Title: Spatial analysis of ambient air pollution and Cardiovascular disease (CVD) hospitalization across Sweden

This study aimed to analyse the association between the different breathable emission particles and the occurrence of CVD hospitalization in Sweden. Previous studies have indicated that particles in breathable air have an impact on one developing CVD or his/her CVD condition progressing. These associative studies, however, tend to do so at larger spatial scales, often using global statistics. Whereas these global summative statistics are informative, they assume homogeneity (all areas in the study region are affected the same). To aid place-specific intervention measures, local spatial analyses are required. Moreover, such kinds of studies were non-existent in Sweden.

The study uses data from the Swedish National Board of Health and Welfare (for CVD admissions), the Swedish Population Register (approximate residence of patients) and the Swedish Environmental Emissions Data (for emission data) for the

years 2005—2010. Spatial methods including global Poisson and spatially autoregressive Poisson regression models were used to analyse for global associative relationships between CVD and emission variables (Black Carbon, Carbon monoxide, Particulate matter, and Sulphur oxides) while controlling for the underlying neighbourhood deprivation through a neighbourhood deprivation index (NDI). To analyse for the more required local heterogeneities in the associations, Geographically Weighted Poisson Regression (GWPR) model was used. GWPR, being a local regression model, fits a regression at every spatial polygon (SAMS) resulting in coefficients equal in number to the number of regions in the study area. Mapping of these coefficients showed the relative variability of the association strength across Sweden.

The results from the global analyses showed that the considered air pollution variables were positively associated with CVD hospitalization across Sweden, although this was sometimes weak and unstable, mainly because CVD is multi-factorial but also possibly because of unmitigated multicollinearity existing within pollution variables. The distilled local associative heterogeneities showed more pronounced variability in the south and central parts of Sweden when compared with the northern parts. This could be driven by more anthropogenic activities being done in the south and central regions than in the northern regions of Sweden.

This study, by showing which pollutants were significantly related to CVD and where such associations were consistently persistent, contributes to the growing knowledge about CVD and its risk factors. This, therefore, provides clues on which activities could be targeted, especially those that lead to increased pollutant atmospheric loading, to reduce their influence on CVD hospitalization. Furthermore, by identifying areas of persistent high associations between air pollution and CVD identified, more focused studies could be done to learn more about the local factors responsible, for better informed future public healthcare policy and intervention.

4.5 Summary of Paper IV

Title: Analysis of spatial co-occurrence between cancer and cardiovascular disease mortality and its spatial variation among the Swedish elderly (2010-2015)

This study aimed at analysing the joint spatial distribution of cancer and CVD mortality among the Swedish elderly. This was motivated by CVD and cancer being the world's two leading causes of death, accounting for about 49% of the global deaths in 2017 (Mahase, 2019). The two diseases have also been shown to progress

each other, with most post-cancer patients dying of CVD instead. Whereas there is a seeming coincidence in morbidity, few studies have analysed for the same coincidence in mortality. This study, therefore, investigated their possible joint spatial clustering of both causes of death in the Swedish elderly with the hope that by identifying areas with consistent double burden, this result could provide much-required information for coordinated public health action aimed at addressing the double threat.

CVD and cancer mortality data for the elderly (65+) for 2010—2015 were obtained from the Swedish Healthcare Registry. Correlation analysis, global Moran's index as well as global bivariate Moran's index were used to investigate the clustering tendencies of CVD and cancer mortality at a national scale. Then spatial statistics, spatial overlay and local entropy maps were used to analyse for local joint spatial clustering of the two causes of death, resulting in variable local associations across the country.

Results from these analyses show that at the age of 65 years, males generally had higher mortality for both CVD and cancer. Beyond 87 years, however, our results show that the females overtook the males in terms of mortality. Correlation results showed that male and female mortalities were averagely positively correlated. Most importantly still, the two causes of death showed differences in spatial clustering scales. CVD clusters were almost always smaller than cancer clusters, with CVD clusters enclaving within the bigger cancer clusters. Results from local joint entropy analysis indicated that CVD and cancer were not always related across Sweden. However, whenever they were related, the relationship was mainly linear and positive.

This study contributes significantly to cancer and CVD fighting efforts in Sweden by highlighting areas where both causes of death can be considered complementary (reinforcing each other) and areas where the two should be considered as independent. This helps to tailor epidemiologic intervention and policy towards specific places, given their unique characteristics concerning the two leading causes of death. Finally, this study provides starting points for more focused studies, especially those concerned with identifying the key driving factors behind the observed associative patterns.

4.6 Synthesis of the Results

The results obtained illustrate that spatially-enable registry systems can be created using existing available spatial data pools – databases containing spatial data, but currently being used to serve other purposes. This was illustrated by designing and

implementing a health registry system that used customer spatial details captured by the National Water and Sewerage Corporation (NWSC), and retrieving these spatial details, upon admission, into the healthcare database through queries to the NWSC database. A further extension of the designed system utilized existing digital maps (Google Maps) for spatial detail retrieval, especially where one was not yet connected to the NWSC network. The patient data captured through such a system would include fine resolution location data to be used in epidemiologic interventions as well as spatial analyses.

Also, by the adopted spatial methods outperforming the conventional statistical methods, our results illustrate that spatial methods have the potential of enhancing epidemiologic interventions by providing more robust estimates than those obtained conventionally. This was illustrated, for example, by the better performance of the spatially-lagged Poisson model and the Geographically Weighed Poisson Regression model compared with the conventional Poisson model, in the Cardiovascular disease and air pollution study. Consequently, such spatial methods enhance epidemiology by providing more reliable estimates, in addition to pinpointing the areas most affected (thus desiring intervention).

Finally, by distilling the obtained associations and effects into their local spatial heterogeneities, our results illustrate how epidemiologic interventions can be more targeted, as the areas most affected are identifiable compared to when estimates are global (i.e. considering the study area as one unit). An example of this final finding was that in Uganda, whereas Tuberculosis and HIV disease rates were positively related most of the times, this correlation was not uniform across Uganda, but with some areas more pronounced than others. The same can be said for the results from the Cardiovascular disease and Cancer spatial clusters in the Swedish elderly study. Here too, the CVD-cancer obtained clusters show heterogeneities that were place-specific. Moreover, differences in cluster scaling were observed with many of the cancer clusters, where they existed, being enclaved (enclosed/enveloped) in the bigger CVD clusters. Such localized identification of most affected areas aids healthcare resource planning, appropriation and reduces epidemiologic intervention costs by providing a basis for ranking and inclusion/exclusion.

5 Conclusions and Recommendations

5.1 Conclusions

Policymakers in healthcare provisioning rely on information obtained from recorded data to make their decisions. They also rely on such data to make epidemiologic predictions of what the future holds in terms of specific disease dynamics, resource planning and intervention. It is, therefore, imperative to obtain accurate and precise estimates from the data. Given that most data about people and what affects them is spatial in nature, this study considered the application of spatial thinking and spatial methodologies to epidemiology with a particular concern for disease surveillance and epidemiologic intervention. This is based on the fact that spatial methods when compared with traditional statistical methods, give more robust estimates in the face of data with spatial characteristics – as is common with data used in epidemiology.

From the application of these spatial approaches to a number of case studies as outlined and consistent with the set-out objectives, the following conclusions can be drawn from this thesis:

(1) HIV-TB co-infection is not spatially homogeneous across Uganda. Some areas (districts) carry more of this double-burden compared to other districts. Given that the WHO recommended coordinated management of both diseases, this study pinpoints the districts where such joint intervention would be more beneficial as we race towards a TB and HIV free community.

(2) Spatially-enabled communities are achievable even in communities without conventional spatial data infrastructures. Through the use of available spatial data pools, data registry systems (patient registry system in our case) can be enabled to capture spatial details at finer spatial levels, especially when available technologies are taken advantage of.

(3) The relationship between CVD health outcomes and ambient air pollution emissions is not spatially uniform across Sweden. Whereas this is expected, it is more important to know where the relationship is more pronounced hence requiring either additional studies or immediate intervention. In our CVD-air pollution study, these areas were highlighted providing plausible evidence for healthcare policy planners in Sweden as well as providing actionable clues for possible targeted intervention.

(4) In the Swedish elderly, CVD and cancer mortalities are not always related in space. However, wherever they were related, among the numerous modes of relationships possible, this relationship was most of the times linear in nature. By showing where these relationships were always prominent and where they were not, our fourth study informs the healthcare authorities, especially those concerned with both causes of death, on areas where the two mortalities should be treated as independent entities and where they should be considered a double threat. We are convinced this information is important as the WHO, just like with HIV-TB, recommends coordinated and simultaneous management of CVD and cancer.

5.2 Recommendations

Generally, this study advocates for the utilization of spatial characteristics as a means of data integration and as a means to link the observed disease outcomes to the environmental exposure variables. It also recommends that the associative relationships established through epidemiologic studies should be broken down into their respective spatial heterogeneous constituents for only then can the intervention be focused.

Specifically, coordinated management of coinfections and comorbidities need to consider how the diseases of concern jointly cluster together in space (and time). Spatial methods provide mechanisms through which such simultaneous clustering can be investigated and evaluated, and are thus recommended for such diseases as HIV-TB, CVD-cancer, and many other diseases that seem to have complementary tendencies.

This study also recommends the use of available spatial data pools – spatial data often collected for other uses, to recreate some form of spatial references "frameworks" that can be utilized to capture and record spatially-referenced records. The advantages of spatially enabled communities are numerous. Leveraging on existing datasets and technologies, these benefits would be delivered to communities without the resources to implement conventional spatial data infrastructures.

Finally, the study of associations between health outcomes and environmental exposure should consider the spatial nature of the data – and not consider such effects as bias or artefacts. By embracing these spatial characteristics as part of the process being investigated, the resultant estimates would not only be more accurate, they would be more precise as well as indicating areas where intervention is most required; as opposed to assuming uniformity of the associations across the studied regions.

At an application level, this study mainly used deterministic models. Whereas these models were applicable in the face of the spatial nature of the data, epidemiology and specifically spatial epidemiology would benefit from future studies using probabilistic models like Bayesian ones. These models allow for more knowledge incorporation and might promise more realistic estimates. Moreover, these probabilistic models would be more practical if they account for spatial effects of spatial autocorrelation and spatial heterogeneity.

6 References

Anselin, L. 1989. What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis (89-4). *NCGIA Technical Reports.* Santa Barbara: University of California.

Anselin, L. 1995. Local indicators of spatial association—LISA. *Geographical analysis,* 27, 93-115.

Anselin, L. 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International regional science review,* 26, 153-166.

Anselin, L. 2019. A local indicator of multivariate spatial association: extending Geary's C. *Geographical Analysis,* 51, 133-150.

Anselin, L. & Rey, S. 2014. *Modern Spatial Econometrics in Practice,* Chicago, USA, Geoda Press LLC.

Anselin, L., Syabri, I. & Kho, Y. 2006. GeoDa: an introduction to spatial data analysis. *Geographical analysis,* 38, 5-22.

Boutayeb, A. & Boutayeb, S. 2005. The burden of non communicable diseases in developing countries. *International journal for equity in health,* 4, 2.

Brook, R. D., Franklin, B., Cascio, W., Hong, Y., Howard, G., Lipsett, M., Luepker, R., Mittleman, M., Samet, J. & Smith, S. C. 2004. Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association. *Circulation,* 109, 2655-2671.

CDC. 2012. *Introduction to Epidemiology* [Online]. Available: https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section1.html [Accessed January 15th 2018].

Chen, J., Roth, R. E., Naito, A. T., Lengerich, E. J. & MacEachren, A. M. 2008. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of US cervical cancer mortality. *International journal of health geographics,* 7, 57.

Chien, L.-C., Alamgir, H. & Yu, H.-L. 2015. Spatial vulnerability of fine particulate matter relative to the prevalence of diabetes in the United States. *Science of the Total Environment,* 508, 136-144.

Cliff, A. D. & Ord, K. 1970. Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography,* 46, 269-292.

Dicker, R. C. 2008. A brief review of the basic principles of epidemiology. *Field Epidemiology,* 1.

Duncan, D., Gold, R. S., Basch, C. E. & Markellis, V. C. 1988. *Epidemiology: Basis for disease prevention and health promotion*, New york: Macmillan, 1988. iv, 186 p.: illustrations; 24 cm.

Elliott, P., Wakefield, J. C., Best, N. G. & Briggs, D. J. 2000. *Spatial Epidemiology: Methods and Applications,* New York, USA, Oxford University Press, Inc.

Elliott, P. & Wartenberg, D. 2004. Spatial epidemiology: current approaches and future challenges. *Environmental health perspectives,* 112, 998.

Fotheringham, A. S., Brunsdon, C. & Charlton, M. 2002a. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons, Limited.

Fotheringham, S. A., Brundson, C. & Charlton, M. 2002b. *Geographically Weighted Regression: the analysis of spatially varying relationships,* West Sussex, John Wiley & Sons Ltd.

Gerstman, B. B. 2013. *Epidemiology kept simple: an introduction to traditional and modern epidemiology*, John Wiley & Sons.

Gray, R. M. 2011. *Entropy and information theory,* New York, Springer Science & Business Media.

Greenwood, M. 1935. *Epidemics and crowd-diseases*, Williams Norgate: London.

Guo, D. 2010. Local entropy map: A nonparametric approach to detecting spatially varying multivariate relationships. *International Journal of Geographical Information Science,* 24, 1367-1389.

Hawkins, B. A. 2012. Eight (and a half) deadly sins of spatial analysis. *Journal of Biogeography,* 39, 1-9.

Howe, G. M. 1964. A national atlas of disease mortality in the United Kingdom. *The Geographical Journal,* 130, 15-22.

Hu, Y. & Li, W. 2017. Spatial Data Infrastructures. *arXiv preprint arXiv:1707.03969*.

Jia, P. 2019. Spatial lifecourse epidemiology. *The Lancet Planetary Health,* 3, e57-e59.

Jin, H. & Lu, Y. 2017. The relationship between obesity and socioeconomic status among Texas school children and its spatial variation. *Applied Geography,* 79, 143-152.

Kirby, R. S., Delmelle, E. & Eberth, J. M. 2017. Advances in spatial epidemiology and geographic information systems. *Annals of epidemiology,* 27, 1-9.

Kulldorff, M. 1997. A spatial scan statistic. *Communications in Statistics-Theory and methods,* 26, 1481-1496.

Kuller, L. H. 1991. Epidemiology is the study of ‘‘epidemics’’ and their prevention. *American Journal of Epidemiology* 134, 1051–1056.

Lawson, A. B., Banerjee, S., Haining, R. P. & Ugarte, M. D. 2016. *Handbook of spatial epidemiology,* Boca Raton, FL., CRC Press.

LeSage, J. & Pace, R. K. 2009. *Introduction to spatial econometrics,* Boca Raton, Florida, Chapman and Hall/CRC.

Lloyd, C. D. 2010. *Local models for spatial analysis,* Boca Raton, Florida CRC press.

Mahase, E. 2019. Cancer overtakes CVD to become leading cause of death in high income countries. *BMJ: British Medical Journal (Online),* 366.

McDonald, K. 2013. Social Epidemiology and Spatial Epidemiology: An Empirical Comparison of Perspectives.

Merrill, R. M. 2012. *Introduction to epidemiology*, Jones & Bartlett Publishers.

Miller, G. 1962. Airs, Waters, and Places in History. *Journal of the history of medicine and allied sciences*, 129-140.

Morabia, A. 2004. *A history of epidemiologic methods and concepts*, Birkhäuser.

Openshaw, S., Charlton, M., Craft, A. W. & Birch, J. 1988. Investigation of leukaemia clusters by use of a geographical analysis machine. *The Lancet,* 331, 272-273.

Ostfeld, R. S., Glass, G. E. & Keesing, F. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in ecology & evolution,* 20, 328-336.

Pavlovsky, E. N. 1966. Natural Nidality of Transmissible Diseases with special reference to the Landscape Epidemiology of Zooanthroponoses. *Natural Nidality of Transmissible Diseases with special reference to the Landscape Epidemiology of Zooanthroponoses.*

Rezaeian, M., Dunn, G., St Leger, S. & Appleby, L. 2007. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *Journal of Epidemiology & Community Health,* 61, 98-102.

Ricotta, C. & Anand, M. 2006. Spatial complexity of ecological communities: Bridging the gap between probabilistic and non-probabilistic uncertainty measures. *Ecological Modelling,* 197, 59-66.

Rothman, K. J. 1996. Lessons from John Graunt. *The Lancet,* 347, 37-39.

Shannon, C. E. 1948. The mathematical theory of communication. *The Bell System Technical Journal,* 27, 379-423.

Sherman, R. L., Henry, K. A., Tannenbaum, S. L., Feaster, D. J., Kobetz, E. & Lee, D. J. 2014. Peer Reviewed: Applying Spatial Analysis Tools in Public Health: An Example Using SaTScan to Detect Geographic Targets for Colorectal Cancer Screening Interventions. *Preventing chronic disease,* 11.

Souris, M. 2019. *Epidemiology and Geography: Principles, Methods and Tools of Spatial Analysis*, John Wiley & Sons.

Tango, T. & Takahashi, K. 2005. A flexibly shaped spatial scan statistic for detecting clusters. *International journal of health geographics,* 4, 11.

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography,* 46, 234-240.

Torrence, M. E. 2002. Data sources: use in the epidemiologic study of medical devices. *Epidemiology,* 13, S10-S14.

Turnbull, B. W., Iwano, E. J., Burnett, W. S., Howe, H. L. & Clark, L. C. 1989. Monitoring for clusters of disease; Application to leukemia incidence in upstate New York. Cornell University Operations Research and Industrial Engineering.

Vinten-Johansen, P., Brody, H., Paneth, N., Rachman, S. & Rip, M. 2003. *Cholera, Chloroform, and the Science of Medicine: A Life of John Snow*, Oxford University Press, Inc.

Wagh, K. & Thool, R. 2012. A comparative study of soap vs rest web services provisioning techniques for mobile host. *Journal of Information Engineering and Applications,* 2, 12-16.

Yao, J. & Stewart Fotheringham, A. 2016. Local spatiotemporal modeling of house prices: a mixed model approach. *The Professional Geographer,* 68, 189-201.

Paper I

https://doi.org/10.1186/s12879-019-4246-2

Aturinde et al. BMC Infectious Diseases (2019) 19:612
https://doi.org/10.1186/s12879-019-4246-2

BMC Infectious Diseases

RESEARCH ARTICLE Open Access

Spatial analysis of HIV-TB co-clustering in Uganda

Augustus Aturinde[1,2,3], Mahdi Farnaghi[1], Petter Pilesjö[1,4] and Ali Mansourian[1,4*]

Abstract

Background: Tuberculosis (TB) is the leading cause of death for individuals infected with Human immunodeficiency virus (HIV). Conversely, HIV is the most important risk factor in the progression of TB from the latent to the active status. In order to manage this double epidemic situation, an integrated approach that includes HIV management in TB patients was proposed by the World Health Organization and was implemented in Uganda (one of the countries endemic with both diseases). To enable targeted intervention using the integrated approach, areas with high disease prevalence rates for TB and HIV need to be identified first. However, there is no such study in Uganda, addressing the joint spatial patterns of these two diseases.

Methods: This study uses global Moran's index, spatial scan statistics and bivariate global and local Moran's indices to investigate the geographical clustering patterns of both diseases, as individuals and as combined. The data used are TB and HIV case data for 2015, 2016 and 2017 obtained from the District Health Information Software 2 system, housed and maintained by the Ministry of Health, Uganda.

Results: Results from this analysis show that while TB and HIV diseases are highly correlated (55–76%), they exhibit relatively different spatial clustering patterns across Uganda. The joint TB/HIV prevalence shows consistent hotspot clusters around districts surrounding Lake Victoria as well as northern Uganda. These two clusters could be linked to the presence of high HIV prevalence among the fishing communities of Lake Victoria and the presence of refugees and internally displaced people camps, respectively. The consistent cold spot observed in eastern Uganda and around Kasese could be explained by low HIV prevalence in communities with circumcision tradition.

Conclusions: This study makes a significant contribution to TB/HIV public health bodies around Uganda by identifying areas with high joint disease burden, in the light of TB/HIV co-infection. It, thus, provides a valuable starting point for an informed and targeted intervention, as a positive step towards a TB and HIV-AIDS free community.

Keywords: HIV, TB, TB/HIV co-infection, Spatial co-clustering, Spatial scan statistics, Moran's I, Bivariate Moran's I, Uganda

Introduction

Tuberculosis (TB) is an airborne bacterial disease caused by *Mycobacterium tuberculosis* that most often affects the lungs. The World Health Organization (WHO) has estimated that about 10.4 million people fell ill with TB, and 1.7 million died from the disease in 2017 [1]. Human immunodeficiency virus (HIV) is one of the most important risk factors responsible for the progression of latent TB to active TB [2]. People living with HIV have a 20-fold higher risk of developing TB than those without HIV, and the risk continues to increase as the vital immunity cells (CD4) count progressively decreases [3]. HIV/TB co-infection is thus known as a 'double trouble' [4] and a public health threat especially for regions where both diseases are endemic.

Sub-Saharan Africa carries the biggest burden of both diseases, with 95% of global TB deaths and more than 70% of the global HIV burden [5]. Uganda, like the rest of Sub-Saharan countries, is plagued by the dual TB and

* Correspondence: ali.mansourian@nateko.lu.se
[1]GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University, SE-221 00 Lund, Sweden
[4]Centre for Middle Eastern Studies, Lund University, Sölvegatan 10, 223 62 Lund, Sweden
Full list of author information is available at the end of the article

HIV epidemics and is the seventh in the 22 countries with the highest TB prevalence [1]. Whereas Uganda's HIV prevalence has reduced to 6.0% in 2016 (from 7.3% in 2011, among 15–49 years old), it was still estimated that 1.3 million individuals were infected with HIV [6]. With TB/HIV co-infection at 41.5%, TB is the leading preventable cause of death among people with HIV, responsible for over 30% of HIV deaths [7].

To decrease the combined TB/HIV prevalence, the WHO formulated a framework in 2005 (modified in 2012), that aims at collaborating TB/HIV activities to manage TB in HIV patients [8]. This position is re-echoed by the WHO in its strategy to end TB in the post-2015 era of the Sustainable Development Goals (SDG) [9, 10]. The motivation for simultaneous management of TB/HIV was largely informed by the proven interactions between TB medication and HIV medication, leading to the ineffectiveness of the TB medication [3]. Additionally, both diseases complement each other with HIV quickening the progression of TB, and vice versa, for co-infected patients [11].

Due to the importance of TB and HIV co-infection, a number of scholars have endeavored to study the correlation between the two diseases. For example, while studying HIV and TB prevalence in New York, Wallace et al. [12] observed that whenever HIV infection was high in the population, there were also high numbers of patients with tuberculosis. Additionally, Corbett et al. [13], having used global TB and HIV prevalence data, concluded that both diseases exhibited similar patterns in both space and time. From a geographical perspective, Wei et al. [14] observed similar spatial clustering patterns between TB and TB/HIV co-infection in Xinjiang province, China. Similarly, Ross et al. [15] used bivariate choropleth mapping and showed that both TB and HIV were correlated and that the joint distribution for both diseases was spatially heterogeneous across Brazil. Their outputs provided an information basis for targeted intervention by the public healthcare bodies responsible for TB and HIV.

However, due to the historical lack of geographically referenced disease records, as well as lack of reliable statistics on morbidity and mortality in most African countries with high TB/HIV disease burden [6, 16], few studies have considered the simultaneous spatial patterns of these comorbidities in Africa. Luckily, with the introduction of District Health Information Software 2 (DHIS2), an open source software platform developed by Health Information System Program (HISP) to African countries, healthcare admission data for most diseases can now be recorded, hierarchically, from local to national levels [17]. For example, Gwitira et al. [5] used DHIS2 data from Zimbabwe to investigate the spatial overlaps in the distribution of HIV/AIDS and malaria. They identified 5 out of the 71 districts as clusters having high records for both HIV and malaria. These would be areas where efforts targeting minimizing both diseases would pay special attention.

In line with the WHO recommendation for collaborative management of TB and HIV, we argue that it is logical to establish the spatial joint distribution of these two co-infections in order to inform local and national intervention strategies. Whereas some studies have examined the individual spatial clustering of TB and HIV both elsewhere [18–21] and in Uganda [22], an intervention based on only one of the two complementary diseases would be ineffective.

Given that the spatial perspectives of HIV/TB co-infection are yet to be studied in Uganda, our main aim of this study, therefore, is to examine the spatial clustering of TB and HIV prevalence rates in Uganda for a three-year period (2015 to 2017) – with particular emphasis on spatial co-clustering. To the best of our knowledge, this is the first spatial study to consider co-clustering of both diseases at a national scale in Uganda. We use spatial-clustering detection and analysis techniques to identify significantly persistent clusters for TB and HIV, providing an informed basis to the Ministry of Health and partners, on the location of such co-clusters thereby potentially aiding effective joint TB/HIV intervention.

Methods

Study area

The study is carried out in Uganda, a country located within East Africa, and about 800 km from the Indian Ocean. Uganda is landlocked bordered by Kenya in the East, South Sudan in the North, Democratic Republic of Congo in the West, Tanzania in the South, and Rwanda in South West. It has a total area of 241,551 km^2, of which the land area covers 200,523 km^2. Administratively, the country is divided into one city and 122 districts (as of 2018) that are further subdivided into counties, sub-counties, parishes, and villages. Uganda's climate is equatorial with the mean temperature range of 16 °C to 30 °C, even when the Northern and Eastern regions sometimes experience relatively high temperatures exceeding 30 °C and the South Western region sometimes has temperatures below 16 °C. The relief of the study area ranges from 614 m (above mean sea level) to 5,111 m at the highest point. The 2014 national census estimated the population of Uganda to be about 35 million people.

Data

TB and HIV admission records were obtained from the DHIS2 system that is housed by the Ministry of Health of Uganda. The DHIS2 system is a community-based aggregation health information system that scales from the

lowest level to the national level [23]. The annual TB and HIV were recorded at the geocoded government healthcare facilities distributed throughout the country and aggregated to the district level. The recorded TB and HIV were all diagnosed cases, for patients that tested at centers located within a specific district. Records from 2015 to 2018 were obtained. However, at the time of acquisition (June 2018), only half of 2018 were recorded and therefore the 2018 records were excluded from the analysis.

Whereas HIV-TB coinfection records were retrievable from the DHIS2 system, they were deemed unreliable (by the staff) mainly because many health units that report to the DHIS2 do not have the capability of diagnosing both HIV and TB simultaneously. They thus report HIV and TB separately. In total, TB and HIV records were obtained for 122 districts in Uganda (based on 2018 administrative boundaries). District level population data and the district mapping shapefiles were obtained from Uganda Bureau of Statistics (https://www.ubos.org/).

TB and HIV admission counts were spatially joined to their respective district polygons for 2015, 2016 and 2017. The TB and HIV disease prevalence was calculated by dividing the total number of each disease cases in each district by the total human population in the district to obtain the population-adjusted district level prevalence rates. For all the years, the population used was that from the 2014 Uganda national census, and the resultant trends are visualized through Fig. 1. As can be observed, the prevalence rates for both TB and HIV, for any given year, are not uniform across Uganda.

Statistical analysis

To understand the characteristics of the TB and HIV data, global pattern analysis was conducted. This involved computing for Spearman's correlation – an overall measure of the linear relationship between TB and HIV district-recorded prevalence rates. The influence of neighborhood prevalence rates on the district-observed prevalence rates was also investigated. This spatial tendency is known as spatial autocorrelation

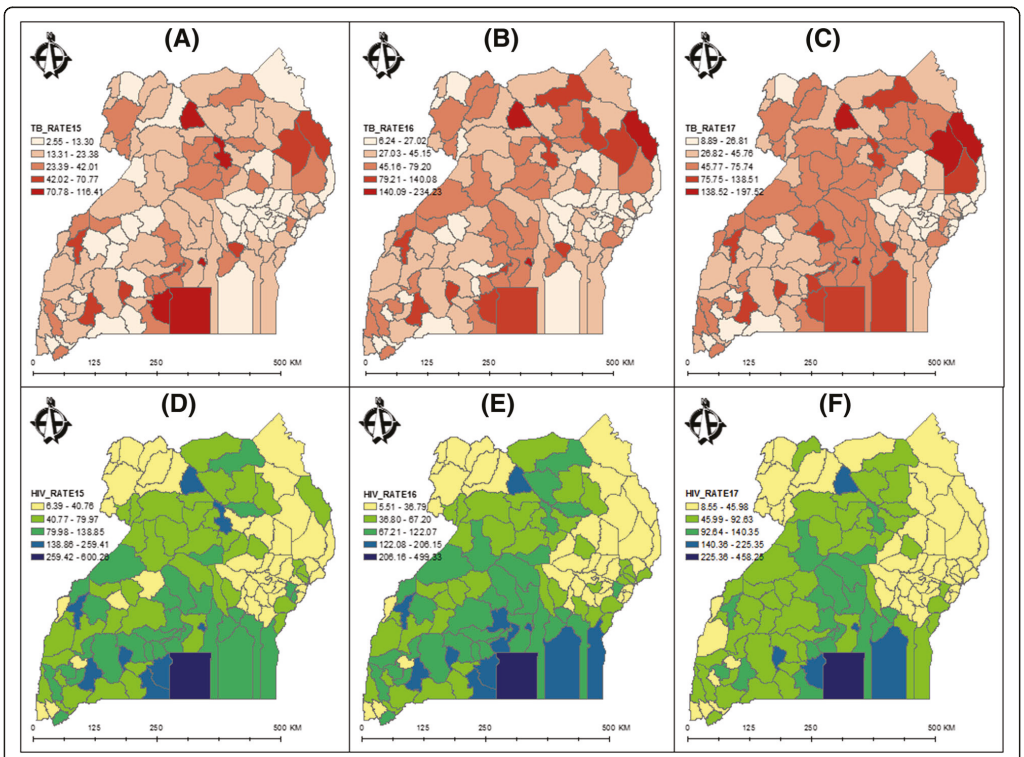


Fig. 1 TB and HIV prevalence rates per 10,000 people in Uganda from 2015 to 2017 (**a**, **b** and **c** for TB; **d**, **e** and **f** for HIV)

and was globally investigated by using global Moran's Index and bivariate global Moran's Index that identified whether the data were spatially autocorrelated or not. Then, spatial scan statistics (SaTScan) was used to extract the local spatial clusters and mark the areas of high risk to inform prevention intervention.

The global analyses were concerned with summarizing the trends within the data, when viewed at a Uganda national level, for the years 2015, 2016 and 2017. SaTScan was concerned with identifying the location and the shape of significant local clusters (hotspots and cold spots) in the study area.

Given that reliable records for HIV-TB coinfection were not available (as discussed in section 2.2), bivariate local Moran's Index was used to investigate the simultaneous occurrence and hence co-clustering in both TB and HIV. It reports areas with hotspots (High-High), cold spots (Low-Low) and discordant (High-Low or Low-High) clusters. To ensure the robustness of the obtained clusters, 9,999 randomizations were allowed for this analysis.

Global pattern analysis

Spearman's correlation analysis was used as a statistical measure for the strength of the linear relationship between TB and HIV district-paired data. The global Moran's I was used to examine the spatial auto-correlation in the TB and HIV prevalence rates. Generally, spatial autocorrelation can be understood as the measure of the influence that the neighborhood values have on the observed values [24–26]. It stems from Tobler's first law of Geography: "everything is related to everything else, but near things are more related than distant things" [27]. This required computation of contiguity information through the generation of the spatial weight matrix. Rooks contiguity was used in this study [28]. Moran's I relates the average TB or HIV prevalence rate within each neighborhood (spatial lag) and the standardized TB or HIV prevalence rate. The global Moran's I and bivariate global Moran's I were performed using GeoDa software [29].

Spatial scan statistics

District-specific TB and HIV clusters were detected by applying Kulldorff's spatial scan statistics [30]. The same technique has been widely used in many applications [14, 18–20, 22, 31]. Spatial scan statistics have reasonable sensitivity and specificity [32]. This enhances their efficiency and accuracy when compared with other cluster detection methods, such as Bayesian disease mapping [5]. The Spatial clusters were detected based on the Poisson probability model, with the underlying assumption that the observed TB and HIV cases in each district, when adjusted for the population at risk, result from a random process [32].

The basic idea behind SaTScan is to impose circular windows of various sizes across the study area, and at each location, defined by the district centroid location in this study; a comparison is made between the disease rate within the window and that outside of it. Under the discrete Poisson assumption, SaTScan [33] detects potential clusters by calculating the likelihood ratio (LR) given by eq. (1).

$$LR_{(u)} = \left(\frac{c}{E_{[c]}}\right)^{c}\left(\frac{C-c}{C-E_{[c]}}\right)^{C-c} I\left(\frac{c}{E_{[c]}} > \frac{C-c}{C-E_{[c]}}\right) \qquad (1)$$

where C is the total number of TB or HIV cases in the study area; c is the observed number of TB or HIV cases within a circle; $E_{[c]}$ is the adjusted expected number within the window under the null hypothesis; $C - E_{[c]}$ is the expected number of TB or HIV cases outside the window, and $I\left(\frac{c}{E_{[c]}} > \frac{C-c}{C-E_{[c]}}\right)$ is the binary indicator of high-risk clusters (1) or low-risk clusters (0) or both (11). Based on the magnitude of the values of the likelihood ratio test, the set of potential clusters are then ranked and ordered. The circle with the maximum likelihood ratio among all radius sizes at all possible centroid locations is considered as the most likely cluster. The statistical significance of the clusters is determined through Monte Carlo simulations (999 simulations).

Within the SaTScan software, the "spatial" option to 2015, 2016, and 2017 TB and HIV case data, both High and Low rates (Hotspots and Coldspots) were analyzed. The user-defined maximum radius of the circular spatial window was varied, starting at 5% and incremented by 5% until it reached 50%. The obtained results were not affected by the choice of the radius selected. The default value of 50% of the population at risk, as advised by Kulldorff [30] was thus maintained.

Bivariate local Moran's I

The bivariate local Moran's I, also called BiLISA, is an extension of the univariate local Moran's I to model the correlation between one variable (e.g. TB) at a location, and a different variable (e.g. HIV) at the neighboring locations. The bivariate Moran's I (for TB) of the i th district can be calculated as eq. (2).

$$I_i = \frac{(x_{TB} - \bar{x}_{HIV}) \sum_j w_{ij}(x_{TB} - \bar{x}_{HIV})}{S^2} \qquad (2)$$

where x_i = the TB prevalence rate for the i th district; $\bar{x}$ = the mean HIV prevalence rate for all districts in the study area; x_j = the TB prevalence rate for the j th district; w_{ij} = a weight parameter for the pair of districts i and j that represent proximity; S = the standard deviation of the TB prevalence rates in the entire study area. The

same was done for HIV, with TB and HIV switching positions.

Results

Global pattern analysis

Table 1 represents these global summary statistics for the study period.

It can be observed that for both diseases, the Moran's I is significantly positive (at 95% confidence interval) – disqualifying the null hypothesis that observations are spatially independent (Moran's I of zero). The positive Moran's I values in Table 1 show that neighboring districts tend to have similar prevalence rates for both TB and HIV. Also, for the whole study period, HIV was consistently more spatially correlated than TB. The significantly positive bivariate Moran's I showed that overall, the observed TB rates were positively influenced by the HIV rates in the neighborhood and vice versa. The computed Spearman's correlation showed that the two diseases were highly correlated through the correlation varied with time. The correlation was highest for 2015 (76%), lowest for 2016 (55%) and moderately high (60%) for 2017.

Clustering analysis

To distill out areas with probable clusters of TB and HIV, spatial scan statistics (discrete Poisson) were employed and the result is shown in Fig. 2.

It can be observed in Fig. 2 that TB high clusters were largely around Lake Victoria and in the central north and one consistent high cluster in the northeast. There is a noticeable reduction in the number of big high clusters from 2015 (six), to 2016 (four), and 2017 (three). The TB low clusters were concentrated in the West and the East (with the central axis dominated with high clusters). On the other hand, HIV high clusters were consistently concentrated in the south, around Lake Victoria and the central parts of Uganda, throughout the study period. The low clusters were generally concentrated in the east, northeast, northwest, and southwest.

Table 1 Moran's I and Correlation for TB and HIV (2015–2017)

	2015	2016	2017
Moran's I			
TB	0.118	0.069	0.129
HIV	1.239	0.327	0.377
Bivariate Global Moran's I			
TB/HIV	0.112	0.074	0.110
Spearman's Correlation			
TB/HIV	0.759	0.548	0.602

Co-clustering analysis

To this end, the concentration has been on the spatial global trends or local clustering patterns in the individual disease prevalence rates. To investigate the simultaneous variation of TB and HIV prevalence in Uganda, the study area was segmented into 9 regions ("bins") based on the study area coordinates, and the linear relationships between the prevalence rates regenerated. Given that relatively similar clustering trends were observed throughout the study period, it was considered that any single year would be representative of the study period. Figure 3 shows the resultant relationships after regionalization, for 2015.

Figure 3 shows the spatial variation of TB-HIV relationship across Uganda for 2015 (the pattern is observed for 2016 and 2017). It illustrates that across Uganda, TB generally had a positive association with HIV and this relationship varies significantly across the geographical space. For example, across the diagonal (plots g, e, and c), the gradient is consistently around 0.2 and significant (at 95% confidence interval) for plots g and e, and not significant for plot c. The region with the highest spatial relationship between the two diseases is the middle upper-most region (b). The lowest right region (plot i) has a negative relationship between TB and HIV, though it is not statistically significant.

To model the simultaneous occurrence and hence co-clustering of both diseases in space, the bivariate local Moran's I was used to show areas where similar disease rates were clustered; characterizing the resultant clusters into High-High, Low-Low, Low-High and High-Low clusters as shown by Fig. 4.

Figure 4 illustrates that generally, there are two High-High TB/HIV occurrence and co-clusters: one around Lake Victoria consisting districts of Kalangala, Mpigi, Kyotera, Kalungu, Masaka, and Mukono, and the other in the north-central districts of Pader and Omoro, in 2015. There is a Low-Low TB/HIV occurrence and co-cluster in the east consisting districts of Butebo, Kaliro, Pallisa, Kumi, Bukwo, and Kibuku, and another central-west co-cluster in Kyegegwa district. Six districts appear as discordant clusters with Lwengo, Wakiso, and Kotido appearing as Low-High, while Rukungiri, Kabale, and Mbale appear as High-Low, for 2015.

For 2016 and 2017, the trends in TB/HIV occurrence and co-clusters are more or less the same as for 2015 with generally a High-High TB/HIV occurrence and co-cluster around the Lake Victoria region and north-central, and a Low-Low occurrence and co-cluster in the east that progressively increase in size with time. For 2016, Koboko in the northwest appears as a cold cluster, though it again became insignificant in 2017. Apart from Mbale and Kotido discordant clusters that are consistent throughout the study period, other discordant clusters (Rukungiri, Kabale, Lamwo, Omoro) are temporally unstable. Also, the districts

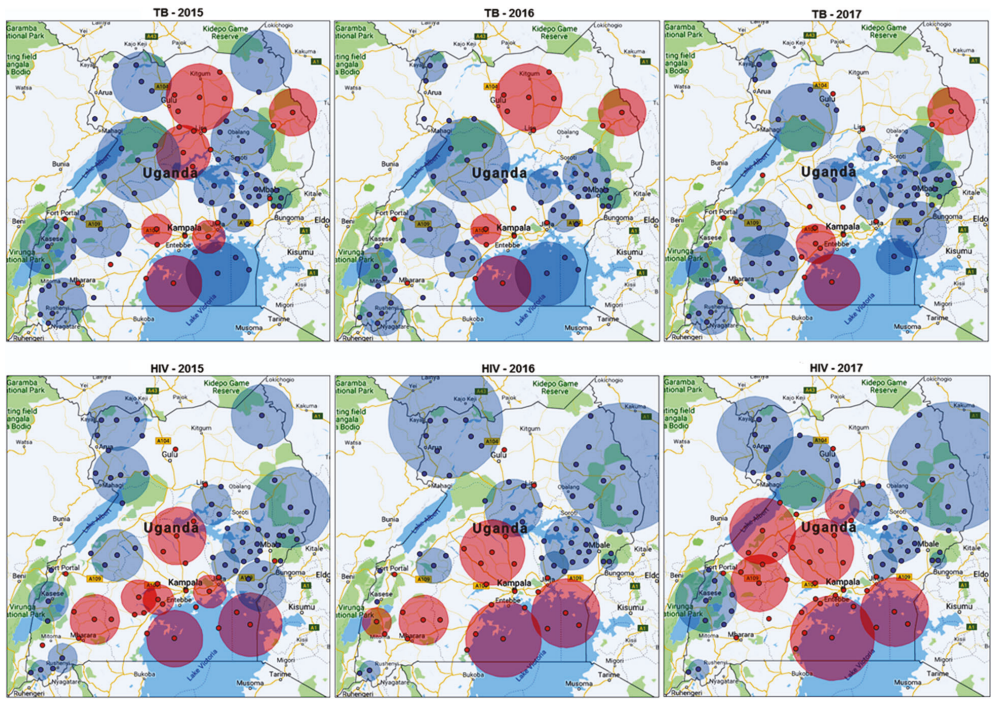


Fig. 2 TB and HIV High (RED) and Low (BLUE) clusters across Uganda (2015–2017)

of Kalangala, Masaka, and Kyotera, consisting of the lower south High-High co-cluster, are unstable throughout the study period.

Discussion

Epidemiological intervention based on the homogeneity of disease patterns often results in non-optimised utilization of the available resources where resources are dedicated to areas that do not require them, at the expense of the areas that require them more [34]. Through the use of spatial methods, health outcomes data can be distilled into their spatial heterogeneity, providing a basis for the explanation of the observed heterogeneity on the basis of existing local risk factors [35].

Our analysis shows that TB and HIV prevalence is geographically heterogeneous. This spatial variability is consistent with the results from the 2016 Uganda Population HIV Impact Assessment (UPHIA) which indicated that the magnitude of HIV prevalence varied considerably across Uganda from a low of 2.8% in West-Nile to 7.7% in the southwestern region [7]. Similarly, our results were consistent with those from the first nationwide community-based TB prevalence survey in 2014/15. Here, it was established that TB was about 1.3 times more prevalent among the urban population than rural residents; approximately three times more prevalent among men than women; nearly three times more prevalent among HIV-negative than HIV-positive individuals; and that TB hotspots exist in both urban and rural areas [36].

These two national surveys for HIV and TB confirm that both epidemics significantly vary across the Ugandan geographic space. However, they do not explicitly identify where the disease clusters are located, making targeted intervention difficult if not impossible. In our study, we identified the clusters exhibited by each disease, as well as the combined occurrence and clustering of both diseases. We also found that the two diseases were highly correlated, hence qualifying the need to manage both diseases simultaneously [9, 37]. Our analysis found a 76, 55, and 60% correlation between TB and HIV for 2015, 2016 and 2017, respectively. This was consistent with results by Dye [16] who observed up to 50% correlation between the two diseases in South Africa, Zambia, and Zimbabwe.

Even with such high correlation, TB and HIV show relatively different spatial clustering patterns across

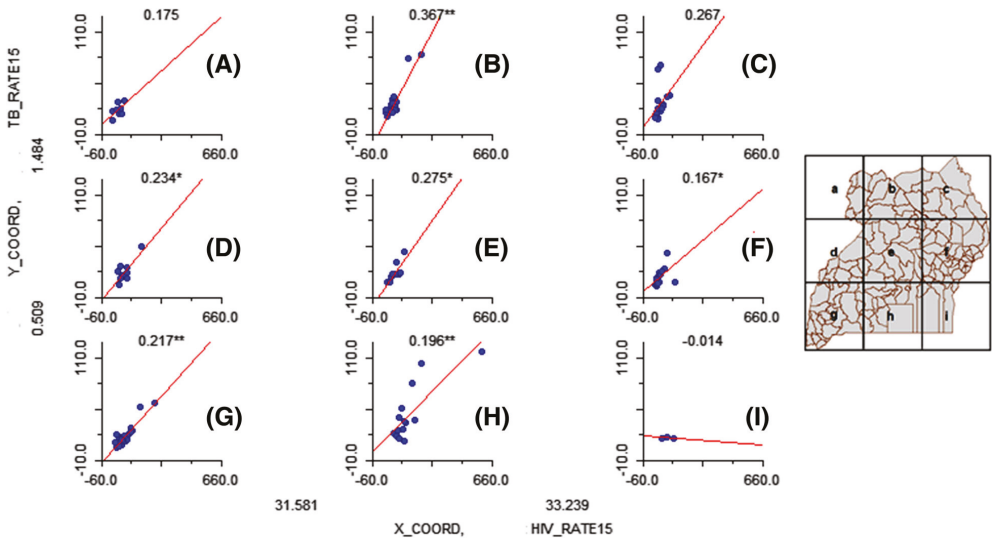


Fig. 3 Spatial variation of TB condition on HIV across Uganda, 2015. Letters **a**-**i** represent the regions from which the variations are derived as shown by the map on the right

Uganda, as observed in the location of clusters in Fig. 2. For example, there were consistent TB hotspots in the greater northern and north-eastern parts of Uganda. This trend was not the same for HIV whose clusters were persistently concentrated around central and southern parts of Uganda, especially around districts in or surrounding Lake Victoria. Also, persistent cold spot clusters for both HIV and TB were observed in the eastern, north-western, and the very south-western (around Kabale) districts of Uganda. These low prevalence rates, especially for HIV, were consistent with projections by the United Nations Programme on HIV/AIDS [38].

Given that in Uganda HIV is more studied than TB, and considering the contribution of HIV in TB progression within TB/HIV co-infected persons [11], the observed TB/HIV geographical clustering trends can easily be explained from an HIV than from a TB standpoint. In Uganda, HIV was first discovered in a rural fishing community of Rakai district, in 1982 and some of the patients surveyed then had TB [39]. Since then, HIV has spread to almost all parts of the country, with some areas more affected than others, so that a more recent study by Bbosa et al. [40] found that these fishing communities are no longer sources but sinks of HIV infection. Even still, this geographical variability in HIV, which is the main risk factor for the progression of latent TB to active TB [4, 9, 14], can be explained by the variability in the underlying socio-economic, behavioral, and cultural factors [41]. Apart from HIV, other population-level risk factors for TB include poor living and working conditions, malnutrition, smoking, diabetes, alcohol abuse, poverty, contact with persons with active TB (health workers, family members), overcrowding and indoor air pollution [42–44].

The most pronounced TB/HIV hotspot co-cluster observed in this study consisted of districts around Lake Victoria; it is thus worth discussing the most likely risk factors around the lake regions. Uganda's fishing communities have been listed among the most-at-risk population with the highest prevalence rate of 15–40% compared to 7% in the general population [36]. In an exclusive study about HIV infections in the fishing communities of Lake Victoria, Opio et al. [45] found the HIV prevalence to be 22%. They also found that these communities were underserved with HIV prevention, care, and support services when compared with other communities. Also, previous studies have shown that fishing communities have fatalistic attitudes, with some viewing HIV infection as less risky than drowning while fishing [46]. Moreover, Ondondo et al. [41], while studying the fishing communities on the Kenyan side of Lake Victoria concluded that the high HIV prevalence (23.3%) could be explained by high-risk unsafe sex practiced within fishing communities. We thus think that the TB/HIV hotspot around Lake Victoria is driven by the high HIV prevalence rates among the fishing communities,

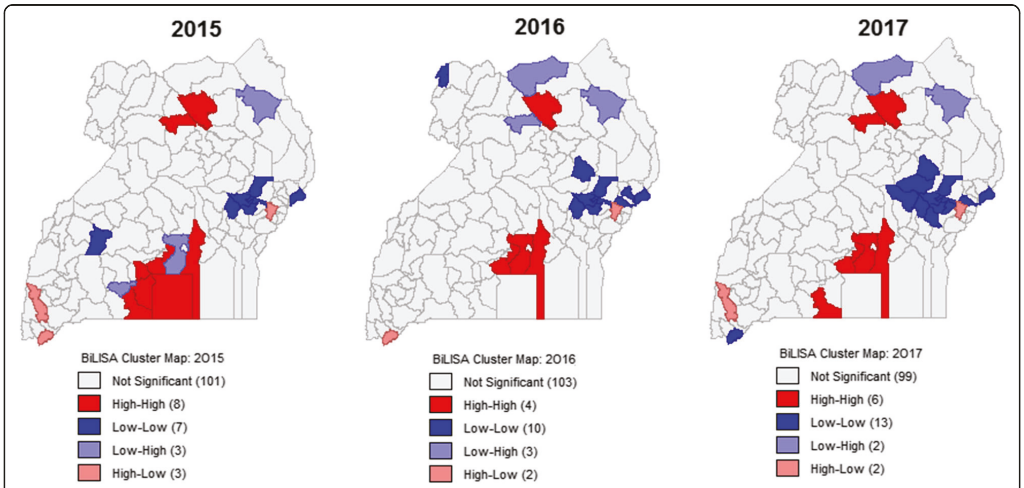


Fig. 4 Spatial Co-clustering of TB and HIV across Uganda (2015–2017)

explained by confounding overlap of lack of TB/HIV support, behavioral and high-risk sex life.

This study observed another TB/HIV hotspot co-cluster in northern Uganda (Pader and Omoro). Its presence could be attributed to the presence of refugees, mainly from South Sudan, and to people that were initially internally displaced into camps, during the Lord's Resistance Army (LRA) war that happened in northern Uganda until 2008. Refugee camps and congested places have been shown to increase TB prevalence [6] and the HIV/AIDS is also known to progress in such settings [47] even when this complex relationship is not well documented [48]. What is not contested, however, is that living in such camps reduces the communities' resilience to such epidemics [48].

The contribution of HIV to the observed TB/HIV co-clustering notwithstanding, one cannot rule out the contribution of other known TB risk factors. These factors were discussed by Narasimhan et al. [49] and were characterized into personal factors, including age, gender, proximity to active TB, malnutrition, diabetes, and environmental factors, including overcrowding, smoking, occupational risk, dangerous alcohol consumption, indoor air pollution.

Finally, this study observed consistent TB/HIV cold spots, especially in eastern Uganda. These were areas, around Mbale district (discordant cluster), that had low prevalence rates for both TB and HIV – consistent with district estimates by UNAIDS [38], especially for HIV. This eastern cold spot could be linked to the traditional practice of male circumcision among the people in those communities – Bagisu and Sebei [50]. Also, from the HIV clusters observed in Fig. 2, Kasese district (inhabited mainly by Bakonjo) has a consistent cold cluster. Male circumcision has for long been associated with reduced risk in acquiring HIV infection. The World Health Organization, based on male circumcision studies from Kisumu in Kenya [51], Rakai district in Uganda [52], and an earlier study from South Africa [53] that had realised 53, 51, and 60% reduction in HIV acquisition risk, respectively, recommended safe male circumcision as an additional measure to reduce HIV acquisition in men [54]. Relatedly, Opio et al. [45] observed higher prevalence rates in uncircumcised men (27%) compared to their circumcised counterparts (11%). We thus submit that the observed TB/HIV cold spot clusters could be attributed mainly to low HIV prevalence rates, which are in turn mediated through culturally practiced male circumcision practices.

Whereas this study achieved its set objective of analyzing the areas in Uganda with elevated prevalence rates for HIV and TB, there were some limitations, especially regarding data availability. Data were available at the district level – which is a larger aggregate level. These results could be more informative had the analysis been done on a finer geographical level (like parish or village level). Also, data about other risk factors for both TB and HIV was not available – this data would have been used to do a more informative spatial regression analysis. These aspects shall be considered in future studies.

Conclusions

Given that for most HIV patients, TB is responsible for more than half the mortalities, and given that HIV

increases the chances of developing active TB by up to 20-folds, scientific evaluation of places where these two diseases are persistently prevalent is not only important but essential for effective management of both diseases. Our study analyzed for joint spatial clustering of TB and HIV. To the best of our knowledge, this is the first spatial study to consider both diseases at a national scale in Uganda, using DHIS2 data. By identifying areas where both diseases co-cluster for the period 2015 to 2017, this study provides valuable information to healthcare policy concerned with these two complementary and endemic diseases in Uganda.

Our analysis identifies the middle-south regions around Lake Victoria (Kalangala, Masaka, Rakai, Mukono, Wakiso, and Mpigi) and some districts in northern Uganda (Pader and Omoro) to be of special interest, as they constitute hotspots. The districts of Kabale and Mbale constitute discordant districts (areas of relatively high prevalence rates in the neighborhood of low prevalence rates, and vice versa) while other eastern districts are significantly cold spots. By aligning healthcare policy and intervention efforts with this obtained spatial heterogeneity in both disease prevalence rates, our study provides an informed starting point towards simultaneous management of TB and HIV.

Abbreviations

DHIS2: District Health Information Software 2; HISP: Health Information System Program; HIV: Human immunodeficiency virus; LRA: Lord's Resistance Army; SDG: Sustainable Development Goals; TB: Tuberculosis; WHO: World Health Organization

Acknowledgments

The authors acknowledge the staff at the Ministry of Health, Uganda especially Mr. Stephen Akena Abwoye for providing the data used in this study. We also acknowledge Drs. Rahman Sanya and Dr. Ernest Mwebaze from the Department of Information Technology, and Mr. Amanya Geoffrey from the Infectious Disease Institute, Makerere University, for their willingness to participate in the initial discussions about the study.

Authors' contributions

AA participated in the study design, implementation of the system, and writing of the manuscript; MF participated in the implementation of the study, review, and editing; PP participated in designing the study and supervision and AM participated the study design, coordination, supervision, and critical commenting the paper. All authors read and approved the final manuscript.

Funding

This research is financed by the Swedish International Development Cooperation Agency (SIDA) under award number 2015–003424 to Ali Mansourian. The authors would like to appreciate SIDA for the funding that enabled the successful implementation of this study.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The need for Ethical approval was waived by the Lund University Ethical Review Board, since the data used is publicly available from the DHIS2 system - Ministry of Health Uganda (https://hmis2.health.go.ug/).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University, SE-221 00 Lund, Sweden. [2]College of Computing and Information Science, Makerere University, Kampala, Uganda. [3]Department of Lands and Architectural Studies, Kyambogo University, Kampala, Uganda. [4]Centre for Middle Eastern Studies, Lund University, Sölvegatan 10, 223 62 Lund, Sweden.

Received: 5 April 2019 Accepted: 30 June 2019
Published online: 12 July 2019

References

1. WHO. Tuberculosis 2018 [cited 2018 20th July]. Available from: http://www.who.int/en/news-room/fact-sheets/detail/tuberculosis.
2. Davy-Mendez T, Shiau R, Okada RC, Moss NJ, Huang S, Murgai N, et al. Combining surveillance systems to investigate local trends in tuberculosis-HIV co-infection. AIDS Care. 2019:1–8. https://doi.org/10.1080/09540121.2019.1576845.
3. Padmapriyadarsini C, Narendran G, Swaminathan S. Diagnosis & treatment of tuberculosis in HIV co-infected patients. Indian J Med Res. 2011;134(6):850.
4. McShane H. Co-infection with HIV and TB: double trouble. Int J STD AIDS. 2005;16(2):95–101.
5. Gwitira I, Murwira A, Mberikunashe J, Masocha M. Spatial overlaps in the distribution of HIV/AIDS and malaria in Zimbabwe. BMC Infect Dis. 2018;18(1):598.
6. MoH Uganda. NATIONAL TUBERCULOSIS AND LEPROSY CONTROL PROGRAMME: revised National Strategic Plan 2015/16–2019/20. Kampala, Uganda. Uganda: Ministry of Health; 2017.
7. Uganda AIDS Commission. Uganda HIV/AIDS country progress report July 2016-June 2017. Kampala: Uganda AIDS Commission, Ministry of Health, Republic of Uganda; 2017.
8. Hermans SM, Castelnuovo B, Katabira C, Mbidde P, Lange JM, Hoepelman AI, et al. Integration of HIV and TB services results in improved TB treatment outcomes and earlier, prioritized ART initiation in a large urban HIV clinic in Uganda. J Acquir Immune Defic Syndr. 2012;60(2):e29.
9. Uplekar M, Weil D, Lonnroth K, Jaramillo E, Lienhardt C, Dias HM, et al. WHO's new end TB strategy. Lancet. 2015;385(9979):1799–801.
10. Lönnroth K, Raviglione M. The WHO's new end TB strategy in the post-2015 era of the sustainable development goals. Trans R Soc Trop Med Hyg. 2016;110(3):148–50.
11. Shankar EM, Vignesh R, Ellegård R, Barathan M, Chong YK, Bador MK, et al. HIV–mycobacterium tuberculosis co-infection: a 'danger-couple model'of disease pathogenesis. Pathog Dis. 2014;70(2):110–8.
12. Wallace R, Wallace D, Andrews H, Fullilove R, Fullilove M. The spatiotemporal dynamics of AIDS and TB in the New York metropolitan region from a sociogeographic perspective: understanding the linkages of central city and suburbs. Environ Plan A. 1995;27(7):1085–108.
13. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, et al. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch Intern Med. 2003;163(9):1009–21.
14. Wei W, Wei-Sheng Z, Ahan A, Ci Y, Wei-Wen Z, Ming-Qin C. The characteristics of TB epidemic and TB/HIV co-infection epidemic: a 2007–2013 retrospective study in Urumqi, Xinjiang Province, China. PLoS One. 2016;11(10):e0164947.
15. Ross JM, Henry NJ, Dwyer-Lindgren LA, de Paula Lobo A, de Souza FM, Biehl MH, et al. Progress toward eliminating TB and HIV deaths in Brazil, 2001–2015: a spatial assessment. BMC Med. 2018;16(1):144.
16. Dye C. Global epidemiology of tuberculosis. Lancet. 2006;367(9514):938–40.
17. Karuri J, Waiganjo P, Daniel O, Manya A. DHIS2: the tool to improve health data demand and use in Kenya. J Health Inform Dev Ctries. 2014;8(1)38–60.
18. Tiwari N, Adhikari C, Tewari A, Kandpal V. Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic. Int J Health Geogr. 2006;5(1):33.
19. Smith CM, Maguire H, Anderson C, Macdonald N, Hayward AC. Multiple large clusters of tuberculosis in London: a cross-sectional analysis of molecular and spatial data. ERJ Open Res. 2017;3(1):00098–2016.

20. Smith C, Lessells R, Grant A, Herbst K, Tanser F. Spatial clustering of drug-resistant tuberculosis in Hlabisa subdistrict, KwaZulu-Natal, 2011–2015. Int J Tuberc Lung Dis. 2018;22(3):287–93.
21. Chamie G, Kato-Maeda M, Emperador DM, Wandera B, Mugagga O, Crandall J, et al. Spatial overlap links seemingly unconnected genotype-matched TB cases in rural Uganda. PLoS One. 2018;13(2):e0192666.
22. Chimoyi LA, Musenge E. Spatial analysis of factors associated with HIV infection among young people in Uganda, 2011. BMC Public Health. 2014;14(1):555.
23. Kiberu VM, Matovu JK, Makumbi F, Kyozira C, Mukooyo E, Wanyenze RK. Strengthening district-based health reporting through the district health management information software system: the Ugandan experience. BMC Med Inform Decis Mak. 2014;14(1):40.
24. Getis A. Spatial statistics. Geogr Inform Syst. 1999;1:239–51.
25. Anselin L. Spatial econometrics: methods and models. Dordrecht: Springer-Science+Business Media; 1988.
26. Cliff AD, Ord K. Spatial autocorrelation: a review of existing and new measures with applications. Econ Geogr. 1970;46(sup1):269–92.
27. Tobler WR. A computer movie simulating urban growth in the Detroit region. Econ Geogr. 1970;46(sup1):234–40.
28. Anselin L, Rey S. Modern spatial econometrics in practice. Chicago: Geoda Press LLC; 2014.
29. Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. Geogr Anal. 2006;38(1):5–22.
30. Kulldorff M. A spatial scan statistic. Commun Stat Theory Methods. 1997; 26(6):1481–96.
31. Rajabi M, Mansourian A, Pilesjö P, Åström DO, Cederin K, Sundquist K. Exploring spatial patterns of cardiovascular disease in Sweden between 2000 and 2010. Scand J Public Health. 2018;46:647–58.
32. Wand H, Ramjee G. Targeting the hotspots: investigating spatial and demographic variations in HIV infection in small communities in South Africa. J Int AIDS Soc. 2010;13(1):41.
33. Chen J, Roth RE, Naito AT, Lengerich EJ, MacEachren AM. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of US cervical cancer mortality. Int J Health Geogr. 2008;7(1):57.
34. Meyer-Rath G, McGillen JB, Cuadros DF, Hallett TB, Bhatt S, Wabiri N, et al. Targeting the right interventions to the right people and places: the role of geospatial analysis in HIV program planning. AIDS. 2018;32(8):957.
35. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. Environ Health Perspect. 2004;112(9):998.
36. MoH Uganda. The Uganda National Tuberculosis Prevalence Survey, 2014–2015: survey report. 2017.
37. WHO. Global Tuberculosis Report 2018. Geneva: World Health Organization; 2018.
38. UNAIDS. Uganda developing subnational estimates of HIV prevalence and the number of people living with HIV. 2014.
39. Serwadda D, Sewankambo N, Carswell J, Bayley A, Tedder R, Weiss R, et al. Slim disease: a new disease in Uganda and its association with HTLV-III infection. Lancet. 1985;326(8460):849–52.
40. Bbosa N, Ssemwanga D, Nsubuga RN, Salazar-Gonzalez JF, Salazar MG, Nanyonjo M, et al. Phylogeography of HIV-1 suggests that Ugandan fishing communities are a sink for, not a source of, virus from general populations. Sci Rep. 2019;9(1):1051.
41. Ondondo RO, Ng'ang'a ZW, Mpoke S, Kiptoo M, Bukusi EA. Prevalence and incidence of HIV infection among fishermen along Lake Victoria Beaches in Kisumu County, Kenya; 2014.
42. Lönnroth K, Jaramillo E, Williams BG, Dye C, Raviglione M. Drivers of tuberculosis epidemics: the role of risk factors and social determinants. Soc Sci Med. 2009;68(12):2240–6.
43. Sharma D, Sharma J, Deo N, Bisht D. Prevalence and risk factors of tuberculosis in developing countries through health care workers. Microb Pathog. 2018;124:279–83.
44. Getahun H, Matteelli A, Abubakar I, Aziz MA, Baddeley A, Barreira D, et al. Management of latent mycobacterium tuberculosis infection: WHO guidelines for low tuberculosis burden countries. Eur Respir J. 2015;46(6):1563–76.
45. Opio A, Muyonga M, Mulumba N. HIV infection in fishing communities of Lake Victoria Basin of Uganda–a cross-sectional sero-behavioral survey. PLoS One. 2013;8(8):e70770.
46. Akumu J, Odongkara K, Masette M, Khaidhiwa M. Prevalence and impacts of HIV/AIDS and other diseases, indigenous knowledge and nutritional status of fisher communities of Lake Albert. Jinja: NaFIRRI National Agricultural Research Organisation; 2006.
47. Serbessa MK, Mariam DH, Kassa A, Alwan F, Kloos H. HIV/AIDS among pastoralists and refugees in north-East Africa: a neglected problem. Afr J AIDS Res. 2016;15(1):45–54.
48. Spiegel PB. HIV/AIDS among conflict-affected and displaced populations: dispelling myths and taking action. Disasters. 2004;28(3):322–39.
49. Narasimhan P, Wood J, MacIntyre CR, Mathai D. Risk factors for tuberculosis. Pulm Med. 2013;2013:1–11.
50. Sarvestani AS, Bufumbo L, Geiger JD, Sienko KH. Traditional male circumcision in Uganda: a qualitative focus group discussion analysis. PLoS One. 2012;7(10):e45316.
51. Bailey RC, Moses S, Parker CB, Agot K, Maclean I, Krieger JN, et al. Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. Lancet. 2007;369(9562):643–56.
52. Gray RH, Kigozi G, Serwadda D. Male circumcision for HIV prevention in young men in Rakai, Uganda: a randomized trial. Lancet. 2007;396:675–66.
53. Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 trial. PLoS Med. 2005;2(11):e298.
54. WHO. New data on male circumcision and HIV prevention: policy and programme implications: WHO. Geneva: World Health Organization; 2007. p. 9241595981.

Publisher's Note

Paper II

https://doi.org/10.1186/s12911-019-0949-y

Aturinde et al. BMC Medical Informatics and Decision Making (2019) 19:215
https://doi.org/10.1186/s12911-019-0949-y

BMC Medical Informatics and Decision Making

RESEARCH ARTICLE Open Access

Establishing spatially-enabled health registry systems using implicit spatial data pools: case study – Uganda

Augustus Aturinde[1], Nakasi Rose[2], Mahdi Farnaghi[1], Gilbert Maiga[2], Petter Pilesjö[1] and Ali Mansourian[1*]

Abstract

Background: Spatial epidemiological analyses primarily depend on spatially-indexed medical records. Some countries have devised ways of capturing patient-specific spatial details using ZIP codes, postcodes or personal numbers, which are geocoded. However, for most resource-constrained African countries, the absence of a means to capture patient resident location as well as inexistence of spatial data infrastructures makes capturing of patient-level spatial data unattainable.

Methods: This paper proposes and demonstrates a creative low-cost solution to address the issue. The solution is based on using interoperable web services to capture fine-scale locational information from existing "spatial data pools" and link them to the patients' information.

Results: Based on a case study in Uganda, the paper presents the idea and develops a prototype for a spatially-enabled health registry system that allows for fine-level spatial epidemiological analyses.

Conclusion: It has been shown and discussed that the proposed solution is feasible for implementation and the collected spatially-indexed data can be used in spatial epidemiological analyses to identify hotspot areas with elevated disease incidence rates, link health outcomes to environmental exposures, and generally improve healthcare planning and provisioning.

Keywords: Spatially-enabled health registry, SDI, RESTful web services, Spatial epidemiology, Mobile-GIS, Uganda

Background

The central paradigm of epidemiology is that disease patterns in populations can be systematically analyzed to understand causes and possible control of diseases. This involves comparisons of differences and similarities in disease patterns over time and between places, to gain new insights about the disease [1]. Given that epidemiology is concerned with disease patterns in human populations as opposed to individuals [2], and that these populations tend to inhabit space in non-homogeneous ways, the resulting disease patterns are often non-homogeneous and spatially-dependent [3].

Spatial heterogeneity in both disease risk and disease incidence at fine-spatial scales is well documented and is driven by genetic, social and environmental factors that subsequently affect exposure and response to infection [4]. As such, most health-related issues such as outbreaks and other epidemiological threats are better understood from a spatial-temporal perspective [5]. This then necessitates the recording of fine-scale spatial details of patients along with other personal data upon hospital admission.

Different countries have devised different mechanisms to enable the capture of fine-scale spatial details of persons. These include the use of postcodes for the UK, ZIP codes for the USA, and personal numbers for Scandinavian countries [4], to mention but a few. These codes and numbers are geocoded and therefore enable the capture of spatial positions of patients at a high spatial resolution level, upon hospital admission. Subsequently, these spatial positions are used in epidemiological analyses to identify where disease incidents are common

* Correspondence: ali.mansourian@nateko.lu.se
[1]GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, 223 62 Lund, Sweden
Full list of author information is available at the end of the article

and identify possible local risk factors involved to help in intervention, prevention, and control.

For most developing and resource-constrained countries, mainly in Africa, however, there are no such systems for home and personal addressing [6–8]. The lack of such ways to enable capture of high resolution personal spatial details forces the healthcare personnel to inevitably aggregate patient data, upon admission, to coarser level administrative units. The implication of this coarse-level aggregation is that fine-scale heterogeneity and the role of local contextual factors may be masked. The masking of such heterogeneity derails intervention and control programmes, especially for infectious diseases, as identification of target foci of transmission and local risk factors are obscured at larger scales [9].

Collection of spatial data is an expensive and time-consuming venture. For sustainability, there is a need for approaches that allow for reusability and sharing of already collected spatial data. Reusability and sharing of spatial data are at the core of spatial data infrastructures (SDI). However, given that traditional establishment of SDIs requires a top-down approach with financing originating from national governments [10], their establishment in resource-constrained countries, such as Uganda, have so far not been very feasible. However, SDIs offer advantages that cannot be ignored.

This study pragmatically addresses this challenge by making use of existing 'spatial data pools'. Spatial data pools in this study are defined as spatial data that have already been collected by different organizations and for other purposes but can be reused for another purpose – for spatial location of patients in our case. By linking this data to patient information, the patient information is georeferenced. In this paper, we address this idea through developing a prototype system using lightweight web services technologies and mobile-based Geographical Information System (GIS) to create a patient registry system that enables recording of such fine-scale patient spatial data, upon hospital/healthcare centre admissions.

The advantage of our system is that it can be integrated with existing health registry (information) systems such as those present in Uganda to make them spatially enabled. It also addresses the challenge of access to desktop computers at healthcare registries by enabling health workers to use their mobile devices to register patient records that can then be visualized and analyzed on the web, reducing the use of paper-based databases – a situation too common in developing countries, like Uganda.

We thus contend that with the proper use of existing spatial data resources, African societies can be spatially enabled by incorporating spatial data in different national information systems in health, tax, police, etcetera that are developing fast.

Related studies

A variety of approaches have been used, in different studies, to collect disease/health-related data with fine spatial resolution. However, these approaches (as reviewed in this session) do not provide a solution for sustainable and continued data collection, on extensive scales like at countrywide levels.

Karas [6] proposed the use of Global Positioning System (GPS) technology to capture a patient's homestead location in areas where the patient's address may be indicated as: "after crossing the river, climb the third hill on the left". He thus advocated for a latitude and longitude file system, especially for rural African hospital, in tracking infection outbreaks and enabling spatial-epidemiological analyses. Using the GPS/GIS approach, Tanser and Wilkinson [11] quantified the improvement in access to Tuberculosis care in Hlabisa, South Africa, when the hospital, clinic, community health workers (CHW), and patient locations are known. They found that by using key locations, the mean distance from patient homestead to point of care (hospital, clinic or CHW) reduced from 29.6 km to just 1.9 km. Similarly, Dwolatzky et al. [12] while studying patient adherence to TB medication in Johannesburg, South Africa, used handheld computing devices (personal digital assistants – PDA) with GPS capabilities to trace the location of patients. By comparing the time taken while using the device, and while not, they found that using PDA/GPS devices reduced the locating time by up to 50%.

Whereas these studies showed considerable success in the use of GPS and PDAs, it must be appreciated that they were used in small towns, where the mapping of individual patients is possible. Scaling up of this approach to a regional or national level would be too expensive to sustain. Consequently, the conference of the African Federation of Emergency Medicine recommended for the use of existing mobile technology to optimally solve patient location problems in Africa [7].

The use of mobile technology in the provision of medical care is not new. Working from the knowledge that Dementia patients are at a higher risk of wandering and getting lost due to a decline in cognitive functioning [13], Huang et al. [14] implemented a pilot program that sends the GPS coordinates of the patient, using a passer-by phone, to service centre personnel, using near-field communication tags embedded in the patient's wristband. In a similar approach, Mendoza et al. [15] proposed tracking and locating of patients with Alzheimer's disease in a nursing home by the use of a wearable tracking device. The device continuously transmits the patient location and sends notification messages to a monitoring database whenever the patient wanders beyond the designated limits. Whereas these are good approaches, they are very case-specific and work best for small special groups of

patients. Also, the need for programmable devices makes costs unbearable and makes the approach impractical especially when large numbers of patients, in a resource-constrained setting, spread all over a large area are involved.

Fornace et al. [8] used android tablet-based applications to geo-locate malaria patients in rural Philippines and Indonesia. Their study provided a way of obtaining high-resolution spatial data in resource-constrained societies with poor internet connectivity, even though their approach is affected by the same scalability issues identified earlier. Additionally, their approach has an inherent requirement for retrospective collection of patient spatial data after patients had been discharged, making it laborious and prone to missing some people. Finally, in the face of another disease, there would be a need for another fieldwork to collect patient location data hence no reusability of the already recorded spatial data.

The challenges in these previous studies can be summarised as (1) retrospective collection of patient location after hospital discharge is both laborious and may miss out some patients; (2) it is challenging to scale up retrospective collection of patient location details when a large area and a large number of patients is concerned; and (3) the patient registry system at the healthcare units is not improved. These three challenges to health informatics, i.e. completeness of records, scalability, and improvement of the existing registry system, are addressed in this study.

Methods

Currently, there are three main health registry systems implemented by the Ministry of Health (MoH) in Uganda.

- OpenMRS – globally adopted open source electronic health registry, used mainly for HIV/AIDS reporting in Uganda [16].
- mTrac – an SMS-based health system originally used to report real-time stocks status of malaria drugs and vaccine at health facilities, but was modified to handle reporting of disease admissions too [17].
- eHMIS-DHIS2 – a community-based aggregation system that scales from the lowest level to the national level [18].

eHMIS-DHIS2 (hereinafter called DHIS2) is currently the official health reporting system in Uganda. The system provides monthly summaries of health status at the district level based on the data from both mTrac and OpenMRS platforms. These monthly aggregates are then transmitted to the national level for archiving, summarising and maintenance. The inherent aggregate architecture in both mTrac and DHIS2 does not allow them to capture personal-level spatial details. Also, whereas OpenMRS records personal details, it does not capture the patient's residential location along with other patient details, upon hospital admission or consultation.

While there is no addressing system in Uganda to collect or geo-reference patient residential locations, the National Water and Sewerage Corporation (NWSC), a central body responsible for water distribution in Uganda, has a database including the geo-coordinates of water meters which are identified through unique water meter numbers. This database is updated regularly with the development of the water network in Uganda. Also, all connected households have access to their water meter numbers, which is written on their monthly water invoices.

The suggested idea by this study is that by asking the patients to report their household NWSC meter numbers along with other personal details upon hospital admission, the spatial locations of their residences could be uniquely identified. Subsequently, they can be used not only in e-health services delivery but also in epidemiological analyses, intervention, and control. This way, the georeferenced meter numbers act as our spatial data pool. Enablement of this capability requires a linkage between the NWSC database and the health registry database in an interoperable way. Also, tools/systems are needed for the digital collection and registry of patients' information.

The overall architecture of the system

Figure 1 shows the overall architecture of a system for a spatially-enabled health registry and how it enables eventual spatial epidemiological analyses. The system is made up of the following components.

(a). *Mobile-based health registry UI* (user interface) and *web-based health registry UI* are used by medical personnel of healthcare centres, to register patients' admission details and patients' residential location. The geo-coordinates of patients' residence are either retrieved from the NWSC database through REST services or pined on the map using the health registry UI components.
(b). *Health registry server* provides the ability to save in and retrieve health registry data from a (Geo) database through a REST Service.
(c). *NWSC server* provides the ability to access the water meter numbers and their respective geo-coordinates from the NWSC database through a REST service.
(d). *Health Web GIS* enables the healthcare personnel to analyse the admission data collected by the system as well as the data from other organisations

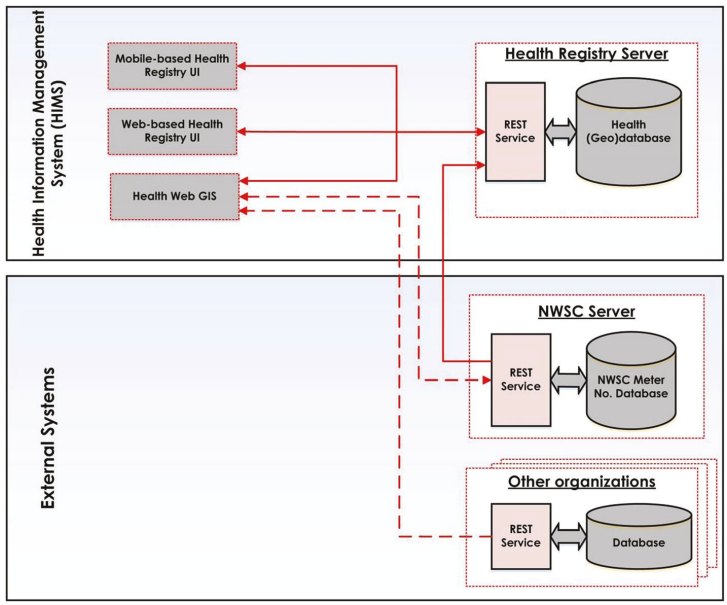


Fig. 1 Schematic representation of the proposed spatially-enabled health registry system

that are published as REST Services (e). These analyses can be used to answer specific spatial epidemiological research questions.

(e). Other organizations can participate in this system by publishing their data through REST Services. Such data can then be used by the Health Web GIS component for contextual epidemiological analysis.

At any hospital or healthcare centre within the country, a mobile application providing the *mobile-based health registry UI* (a) is installed on handheld devices of medical personnel responsible for patient data recording. Alternatively, *web-based health registry UI* (a) on their computers can be used. The medical staffs, including nurses and doctors, use these UIs for registering patient's information including disease diagnoses and their residential locations.

To record patient residential locations, patients are asked to specify their water meter numbers (provided and maintained by NWSC). The UI components also have a background map, with satellite imagery, by which the approximate location of the patients' houses can be navigated and marked for those without water meter numbers (possibly not yet connected to the water network) or are unable to access them for some other reason. The registered information is sent to and stored in the health registry (geo)database. In the health registry (geo) database (b), a patient's information is geo-coded by using geo-coordinates received from the NWSC database (c) (with meter number as the unique identity).

A Health Web GIS system (d), with spatial analysis functions including those needed for spatial epidemiology, has access to the health registry database (b) and other databases (e) (e.g. environmental data from related organisations). So, spatial epidemiology analysis is possible to detect disease hotspots, outbreaks, monitor the progress of diseases in space and time, prepare prevention plans, etc.

Development and implementation

The *mobile-based health registry UI* was developed as an android app by Java programming language. JavaScript programming languages, Cascading Style Sheets (CSS), and HyperText Markup Language (HTML) were used to develop the *web-based health registry UI* as well as the *Health registry Web GIS*. In order to provide mapping functionalities in the web applications, the Leaflet library (https://leafletjs.com/) was exploited. To develop the web services, two frameworks: Service Oriented Architecture Protocol (SOAP) and REpresentational State Transfer (REST) are commonly used. However, SOAP has a heavyweight message payload thus not very favourable for resource-constrained mobile devices [19]. Subsequently, the REST web service framework was

used in our study as its messages have a lightweight payload hence more suitable for wireless and cellular connectivity networks synonymous with mobile devices [19]. The REST services were developed in Java programming language using oracle JAX-RS.

The system was implemented as a prototype and presented to the Ministry of Health officials. The idea was welcomed as a prospective candidate to complement patient registry systems in Uganda since it covers the gap in existing systems which do not record the absolute spatial location of patients' homes. Discussions about the adoption of this system are ongoing and decisions will have to be made at higher levels of the Ministry of Health.

Results

Below we describe a scenario where a patient visits a healthcare unit and how the system is used to capture the patient details, including the patient's home location.

When a patient visits a healthcare centre, the healthcare provider (doctor, nurse, etc.) collects and registers the patient's personal information and the disease diagnosis through the *mobile-based* or *web-based health registry UIs*. To avoid duplication of patient records, the National Identity Number (NIN) is required for every patient. As mentioned before, the water meter number of the patient is also asked to geocode a patient's residence location. Figure 2 shows the interface of the mobile app, with its corresponding geo-coordinates as synched from the NWSC database.

If a patient has no meter number, a background map with satellite imagery is launched (Fig. 3) with navigation capabilities to navigate to the patient's home. The same high resolution background map would be used in instances where the patient forgets his meter number or is unable to provide it for some other reason. Upon identifying the home, a single click on the home retrieves its coordinates, prompting the user to save the coordinates against the patient's record. All the information gets stored in the health registry database upon saving.

The coordinates are therefore either retrieved from the geocoded meter numbers or from the background satellite imagery map. All the information gets stored in the health registry database.

Through the Health Web GIS interface, the location of patients and diseases can be retrieved and used for analysis. Using this interface, one may find the spread of a particular disease by viewing and analysing the recorded data or may integrate the patient and disease data to other data to perform more specialized spatial analyses. Below we illustrate the possible uses of the system using an example.

Figure 4 illustrates a snippet of what can be done with the recorded patient information. In Fig. 4 (a), the incidence data are plotted to show where the incidences are spatially located. By applying cluster analysis, one can study where the incidences could be high and where they are low. In (B), we use point density to highlight areas with more admissions. This results in the visualisation of potential clusters on the map. To investigate which of these potential clusters constitute a hotspot or cold spot, hotspot analysis may be applied. Fig. 4 (c) shows the outputs of hotspot analysis. It shows that among the identified density clusters, only a few are statistically significant (arising not by chance). There is a very pronounced significant hotspot as shown by the GiZscore, on the far right of Fig. 4 (c) and some noticeable ones to the left of this far right hotspot.

This fine detail spatial analysis could not be possible with the current health registry systems. In terms of

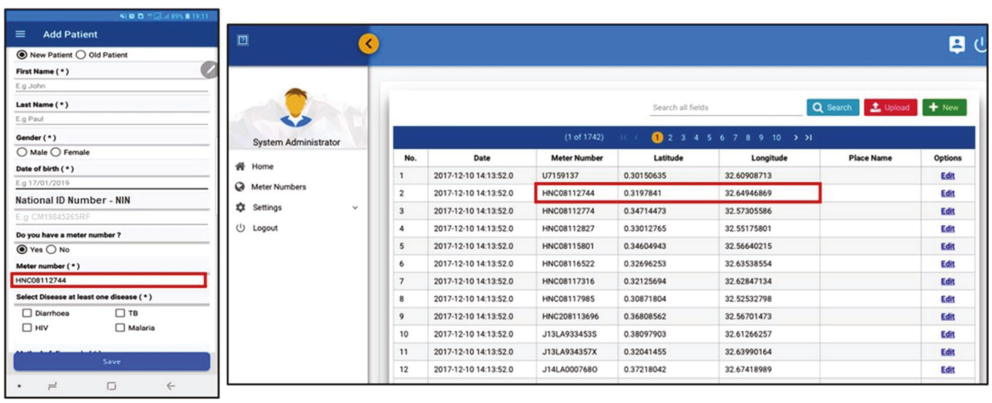


Fig. 2 Mobile App UI for a patient with a METER NUMBER and their geo-coordinates from the NWSC database

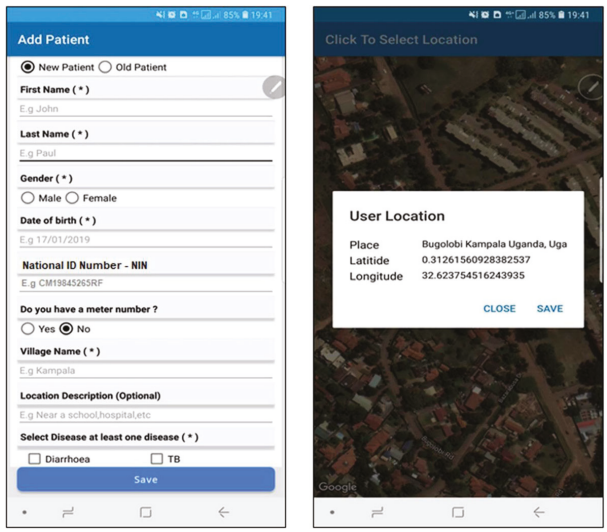


Fig. 3 Patient without a meter number and geo-coordinates collected using a background high resolution map

directing intervention, the healthcare professionals and policymakers are better informed with this kind of tool, not only in disease surveillance but also in associative analyses to evaluate the environmental factors at play in specific disease scenarios.

Discussion

Williamson et al. [20] described a spatially enabled society as an emerging cultural and governance revolution offered by spatial information technology and spatially equipped citizenry that change the way economies, people and environment are organised and managed for the better. Thus, spatial enablement, itself is a consequence of the realisation of SDI promises through developing spatial information products, smarter delivery of services, improved risk management and better macroeconomic decision making.

The establishment of SDIs is still largely dependent on national governments' initiatives [21] especially through the establishment of complete national land cadastre

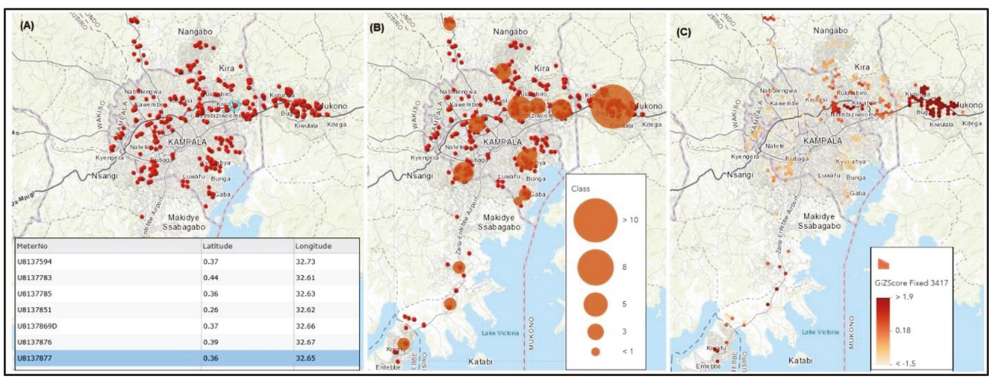


Fig. 4 Some of the possible analyses with geocoded patient records: Incidence (**a**), Density Clustering (**b**), and Hotspot Analysis (**c**). (Map generated from sample collected data and produced by the authors. The background map is from ArcGIS-Online)

systems [20, 22], even when synergistic two-sided involvement of public institutions and entrepreneurs in SDIs is currently being encouraged [23]. The overarching benefit of such SDI establishments lies in the better national management of both spatial information and information organised according to location. However, due to the lack of a means to capture spatial data as well as the inexistence of SDIs in most of the resource-constrained African countries, SDIs and hence spatial enablement of society have not been achievable in most parts of Africa. For example, a country like Uganda does not even have a street addressing system and has an incomplete cadastre.

Our study shows that by using creative approaches, operational systems can be developed to achieve a spatially enabled society in many aspects, as we have done in this study using the case of healthcare registry. Wallace [24] illustrated that spatial enablement has two stages: the first stage involves utilising of imagery to answer the basic question of "where am I?" while the second stage involves linking all data with a geocoded reference to allow for spatial analysis and spatial decision making.

In this study, we utilized spatially referenced water utility data and RESTful web services to implement a prototype of a spatially enabled health registry system in Uganda. Our system achieves the enablement stages outlined by Wallace [24] by (1) using publically available google maps satellite imagery services to identify patient residencies in instances where the patients do not have the meter numbers, and (2) using geo-referenced water utility meter numbers from NWSC to enable geocoding of patient-level records upon hospital admission, using RESTful web services. We used RESTful web services for they have a lightweight message load and is more adapted for mobile network connectivity [19, 25]. This kind of system can be generalised to other (East African) countries that have such spatially referenced utility data like Kenya [26], Tanzania and Rwanda whose utility sector is relatively similar to the Ugandan sector.

Our system, to the best of our knowledge, is the first spatially enabled health registry system in Uganda. Being a digital system, our system like other existing health information registry systems (OpenMRS, mTrac, DHIS2) helps to get rid of paper databases that are still common in Uganda. Additionally, our system is easy to integrate with existing systems, especially the OpenMRS [16] that inherently records individual patient-level details, as opposed to aggregate systems like mTrac [17] or DHIS2 [18]. Finally and more importantly, our system allows for the collection of spatially referenced medical records that can be used in spatial epidemiological analyses and health planning as we illustrate in Fig. 4. The proposed solution for spatial-enablement of health registry is not expensive to implement, especially in comparison with a real implementation of an SDI.

In terms of enabling the capturing of geographical coordinates of patient homes, our study compares to systems by Tanser and Wilkinson [11], Dwolatzky et al. [12] and Fornace et al. [8]. However, unlike these previous studies, our study achieves this but also has capabilities of enabling existing registry systems at the healthcare units. It, subsequently, eliminates the need to capture the patient home location retrospectively after being discharged from the hospital.

The advantages of such a system that enables the recording of the spatial location of patients' homes are that it makes tracking of infectious diseases, identification of health trends, identification of disease clusters and linking of environmental exposure to health outcomes, possible. It also improves service delivery as medication can be delivered on doorsteps, especially necessary for diseases (mainly terminal illnesses) that are better managed from home.

Whereas our study achieves what it set out to achieve, we acknowledge some challenges that were either encountered in the study or challenges that could influence the adoption of such a system. For example, there were some difficulties in convincing NWSC to share their meter number data, as the solution was to help the healthcare industry, not the water industry. We solved this issue by convincing NWSC on how their help can contribute to increasing the health and the quality of life in Uganda. The system also requires some preliminary training of medical personnel to use the system especially map reading and navigation when the patient does not have a meter number (possibly not yet connected to the national water network). The time required to navigate to patients' homes might slow down the registration process and decrease efficiency, especially for novice users. However, from a cost-benefit perspective, the benefits of collecting positional information of patients are more than the cost/time spent. Additionally, we are convinced that this system is extendable to most developing countries whose spatial data infrastructure situation is similar to Uganda's (incomplete digital cadastre, no geocoded street names and no postcodes). However, the lack of any geocoded data could be a limitation to applying the suggested idea in those countries. Finally, the patients in such a system cooperatively give their home location details, when there are enforcing laws, like laws in many European countries. Privacy and confidentiality laws are needed in Uganda to hinder the distribution of individual patient's records (for example only aggregated data may be analysed and distributed [4]) and also to satisfy ethical constraints.

Conclusions

Recording of patient residential locations normally requires the use of established country-specific spatial data infrastructures that are inexistent in most developing countries. By using geo-coded data already collected for utility services provision, RESTful web services and mobile technology, our study provides a valuable possible improvement to existing electronic health registry systems that enable them to be spatially-enabled hence increasing their return on investment. The return on investment is in the form of extra capabilities that spatially-enabled health registry systems have over currently existing ones such as identification of areas with elevated disease incidence rates, identification disease trends across space and time, aiding targeted intervention as well as linking environmental exposures to health outcomes. Finally, by explicitly capturing patients' residential locations, such services as location-aware emergency and prescription delivery can be enabled thereby improving general healthcare planning and provisioning.

Abbreviations

AIDS: Acquired Immune Deficiency Syndrome; CHW: Community Health Workers; CSS: Cascading Style Sheets; GIS: Geographical Information Systems; GPS: Global Positioning System; HIV: Human Immunodeficiency Virus; HTML: HyperText Markup Language; NWSC: National Water and Sewerage Corporation; REST: REpresentative State Transfer; SDI: Spatial Data Infrastructure; SOAP: Service Oriented Architecture Protocol; UI: User Interface

Acknowledgements

The authors would like to appreciate SIDA for the funding that enabled the successful implementation of this study. Additiionally, the authors acknowledge National Water and Sewerage Corporation, Uganda especially Mr. Kanyesigye Christopher, for granting them the data used. They also acknowledge the staff at the Ministry of Health, Uganda especially Dr. Eddie Mukoyo for granting them access to the ministry's existing health registry systems.

Authors' contributions

AA participated in the study design, implementation of the system, and writing of the manuscript; MF participated in the design of the system and reviewing the manuscript; NR participated in the implementation of the system; GM participated in designing the system, supervision for implementation of the system, data collection and critical commenting the paper; PP participated in designing the study and supervision; AM participated the design, supervision, and coordination of the study as well as critical reading and commenting the paper. All authors read and approved the final manuscript.

Funding

The first three authors of this study are financed by the Swedish International Development Cooperation Agency (SIDA) under award number 2015–003424 to Ali Mansourian.
Role of the Funder: The funding organization had no direct role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; and decision to submit the manuscript for publication.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to integrity and legal reasons but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

[1]GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, 223 62 Lund, Sweden. [2]College of Computing and Information Science, Makerere University, Kampala, Uganda.

Received: 22 February 2019 Accepted: 24 October 2019
Published online: 08 November 2019

References

1. Bhopal RS. Concepts of epidemiology: integrating the ideas, theories, principles, and methods of epidemiology: Oxford University press; 2016.
2. Kirby RS, Delmelle E, Eberth JM. Advances in spatial epidemiology and geographic information systems. Ann Epidemiol. 2017;27(1):1–9.
3. Lawson AB, Banerjee S, Haining RP, Ugarte MD. In: Fitzmaurice G, editor. Handbook of spatial epidemiology. Boca Raton: CRC Press; 2016.
4. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. Environ Health Perspect. 2004;112(9):998.
5. Granell C, Fernández ÓB, Díaz L. Geospatial information infrastructures to address spatial needs in health: collaboration, challenges and opportunities. Futur Gener Comput Syst. 2014;31:213–22.
6. Karas JA. Tracing patients in rural Africa. Lancet. 1996;348(9041):1598.
7. Stein C, Mould-Millman N-K, De Vries S, Wallis L. Access to out-of-hospital emergency care in Africa: consensus conference recommendations. Afr J Emerg Med. 2016;6(3):158–61.
8. Fornace KM, Surendra H, Abidin TR, Reyes R, Macalinao ML, Stresman G, et al. Use of mobile technology-based participatory mapping approaches to geolocate health facility attendees for disease surveillance in low resource settings. Int J Health Geogr. 2018;17(1):21.
9. Bannister-Tyrrell M, Verdonck K, Hausmann-Muela S, Gryseels C, Ribera JM, Grietens KP. Defining micro-epidemiology for malaria elimination: systematic review and meta-analysis. Malar J. 2017;16(1):164.
10. Sjoukema J-W, Bregt A, Crompvoets J. Evolving spatial data infrastructures and the role of adaptive governance. ISPRS Int J Geo Inf. 2017;6(8):254.
11. Tanser F, Wilkinson D. Spatial implications of the tuberculosis DOTS strategy in rural South Africa: a novel application of geographical information system and global positioning system technologies. Tropical Med Int Health. 1999; 4(10):634–8.
12. Dwolatzky B, Trengove E, Struthers H, McIntyre JA, Martinson NA. Linking the global positioning system (GPS) to a personal digital assistant (PDA) to support tuberculosis control in South Africa: a pilot study. Int J Health Geogr. 2006;5(1):34.
13. Juzwishin D, Mueen M, Cruz AM, Ruptash T, Barnard S, Sebastianski M, et al. Characteristics of a successful collaboration in evaluation of a health care innovation: lessons learned from GPS locator technology for dementia clients. Innovation and Entrepreneurship in Health. 2017;4:1–8.
14. Huang JC-S, Lin Y-T, Yu JK-L, Liu K, Kuo Y-H, editors. A wearable NFC wristband to locate dementia patients through a participatory sensing system. 2015 International Conference on Healthcare Informatics. Dallas: IEEE; 2015.
15. Mendoza MB, Bergado CA, De Castro JLB, Siasat RGT, editors. Tracking system for patients with Alzheimer's disease in a nursing home. TENCON 2017-2017 IEEE Region 10 Conference. Penang: IEEE; 2017.
16. Tierney WM, Achieng M, Baker E, Bell A, Biondich PG, Braitstein P, et al. editors. Experience implementing electronic health records in three East African countries. Cape Town: MedInfo; 2010.
17. Abandu J, Kivunike FN. Immunisation-notification adoption model: strategies for implementing mobile electronic notification of mothers in Uganda. Int J Telemed Clin Pract. 2017;2(2):121–39.
18. Kiberu VM, Matovu JK, Makumbi F, Kyozira C, Mukooyo E, Wanyenze RK. Strengthening district-based health reporting through the district health management information software system: the Ugandan experience. BMC Med Inform Decis Mak. 2014;14(1):40.
19. Wagh K, Thool R. A comparative study of soap vs rest web services provisioning techniques for mobile host. J Inf Eng Appl. 2012;2(5):12–6.

20. Williamson I, Rajabifard A, Wallace J, Bennett R. Spatially enabled society; 2011.
21. Acharya PS, Sarda NL. National Data Registry (NDR) for the National Spatial Data Infrastructure (NSDI). In: Sarda NL, Acharya PS, Sen S, editors. Geospatial Infrastructure, Applications and Technologies: India Case Studies. Singapore: Springer Singapore; 2018. p. 17-29.
22. Rajabifard A, Binns A, Williamson I. SDI design to facilitate spatially enabled society. In: Towards a Spatially Enabled Society, Melbourne: The University of Melbourne; 2007. p. 219–32.
23. Jabbour C, Rey-Valette H, Maurel P, Salles J-M. Spatial data infrastructure management: a two-sided market approach for strategic reflections. Int J Inf Manag. 2019;45:69–82.
24. Wallace J. Spatially enabling mortgage markets in Australia. In: Towards a spatially enabled societyards a spatially enabled society; 2007. p. 119–38.
25. Jabbar S, Khan M, Silva BN, Han K. A REST-based industrial web of things' framework for smart warehousing. J Supercomput. 2018;74(9):4419–33.
26. Gakii Gatua J, Ikiara M, Kabubo-Mariara J, Mwaura M, Whittington D. Water and sanitation service delivery, pricing, and the poor: an empirical estimate of subsidy incidence in Nairobi. Kenya Water Resour Res. 2016;52(6):4845–62.

Publisher's Note

Paper III

Spatial analysis of ambient air pollution and Cardiovascular disease (CVD) hospitalization across Sweden

Augustus Aturinde[1,2,3], Mahdi Farnaghi[1], Ali Mansourian[1*], Petter Pilesjö[1], and Kristina Sundquist[4]

[1] GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University, SE-221 00 Lund, Sweden; augustus.aturinde@nateko.lu.se; mahdi.farnaghi@nateko.lu.se; ali.mansourian@nateko.lu.se; petter.pilesjo@gis.lu.se

[2] College of Computing and Information Science, Makerere University, Kampala, Uganda

[3] Department of Lands and Architectural Studies, Kyambogo University, Kampala, Uganda

[4] Center for Primary Health Care Research, Department of Clinical Sciences, Malmö, Lund University, Sweden; kristina.sundquist@med.lu.se

*Corresponding author: Ali Mansourian (ali.mansourian@nateko.lu.se); Tel.: +46 46 222 1733

Key Points:

- There are significant place-specific associations between air pollutants and CVD admissions across Sweden.
- The southern parts of Sweden show more spatial variability of these place-specific associations than other parts.
- More epidemiologic emphasis should be placed on local impacts of air pollution on CVD outcomes in Sweden.

Abstract

The associations of multiple pollutants and Cardiovascular disease (CVD) morbidity, and the spatial variations of these associations have not been nationally studied in Sweden. The main aim of this study was, thus, to spatially analyse the associations between ambient air pollution (Black Carbon, Carbon monoxide, Particulate matter (both <10μm and <2.5μm in diameter) and Sulphur oxides considered) and CVD admissions while controlling for neighbourhood deprivation across Sweden from 2005 to 2010. Annual emission estimates across Sweden along with admission records for coronary heart disease, Ischemic stroke, atherosclerotic and aortic disease were obtained and aggregated at Small Areas for Market Statistics level. Global associations were analysed using global Poisson regression and spatially autoregressive Poisson regression models. Spatial non-stationarity of the associations was analysed using Geographically Weighted Poisson Regression. Generally, weak but significant associations were observed between most of the air pollutants and CVD admissions. These associations were non-homogeneous, with more variability in the southern parts of Sweden. Our study demonstrates significant spatially varying associations between ambient air pollution and CVD admissions across Sweden and provides an empirical basis for developing healthcare policies and intervention strategies with more emphasis on local impacts of ambient air pollution on CVD outcomes in Sweden.

1. Introduction

The association between short- and long-term exposure to ambient air pollution as pollution within the outdoor breathable air, especially particulate matter (PM), and Cardiovascular diseases (CVD) morbidity and mortality have been investigated in many studies (Atkinson et al. 2013; Beelen et al. 2008; Bell et al. 2008; Brook et al. 2004; Dominici et al. 2006; Grahame and Schlesinger 2010; Le Tertre et al. 2002; Lim et al. 2014; Luo et al. 2016; Meister et al. 2012; Pope et al. 2004; Qiu et al. 2017; Stockfelt et al. 2017; Sun et al. 2010; Sunyer et al. 2003; Zhang et al. 2014). The World Health Organisation (WHO) defines Cardiovascular diseases (CVD) as a group of disorders of the heart and blood vessels and includes coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic heart disease, congenital heart disease, deep vein thrombosis and pulmonary embolism (WHO 2019) and was the leading cause of death with over 17.9 million premature deaths in 2016 (Hadley et al. 2018). In a relatively recent study of Global Burden of Diseases, it was estimated that about 4.7 million deaths in 2015 were attributable to ambient particle mass with a diameter less than 2.5 μm (PM2.5) (Cohen et al. 2017), mainly through CVD (Thurston and Newman 2018).

To manage the effects of multiple air pollutants on CVD health outcomes, healthcare policies and intervention efforts need to be informed of where in space these associative effects are particularly more pronounced in order to devise place-specific intervention approaches, with a full view of the environmental, demographic and social-economic conditions prevailing in the specific places. This would subsequently enable effective pollutant-specific measures to be designed and implemented for different areas.

In Sweden, whereas some scholars have studied air pollution and CVD, they have only considered selected cities. For example, Le Tertre et al. (2002) and Sunyer et al. (2003) observed significant short-term effects of air pollution on CVD admissions in eight European cities,

including Stockholm. Stockfelt et al. (2017) observed long-term effects of total and source-specific particulate matter (both <10μm in diameter (PM10) and PM2.5) air pollution on CVD incidence in Gothenburg city, calling for efforts to reduce air pollution if its negative health effects are to be minimised. Additionally, Segersson et al. (2017) studied the health impacts of source-specific air pollution (PM10, PM2.5 and Black Carbon (BC)) in Stockholm, Gothenburg and Umea cities. They concluded that the majority of the observed premature deaths were related to local emissions and that road traffic and residential wood combustion had the largest impact.

To study the effect of these local emissions on CVD hospitalization requires the use of local spatial regression models. However, such local studies have not yet been done in Sweden. So, whereas the effect of different air pollutants might be known for some selected cities (mainly Stockholm, Umea and Gothenburg), the multi-pollutant associations with CVD across the whole of Sweden remain to be studied. Moreover, single-city analyses have been shown to be prone to publication bias (Chen et al. 2017), where authors choose to publish only cities with positive associations.

The main objective of this study, therefore, was to analyse multi-pollutant (PM10, PM2.5, BC, Sulphur oxides (SOx), and Carbon monoxide (CO)) associations with CVD and their spatial variation across Sweden for the years 2005 to 2010, using Geographically Weighted Poisson Regression (GWPR). This was done while accounting for underlying neighbourhood deprivation, using the computed Neighbourhood Deprivation Index (NDI) from the four socioeconomic factors low education, unemployment, low income, and recipient of social welfare (Winkleby et al. 2007), an index that is independently associated with CVD (Lawlor et al. 2005; Li et al. 2019; Sundquist et al. 2004a). The advantage with the GW(P)R framework lies in its robustness to the effects of multicollinearity (Fotheringham and Oshan 2016), a condition common with multi-pollutant data (Stockfelt et al. 2017). Our goal was thus to identify how the strength of the association between CVD hospitalization and each of the ambient air pollution variables varies across Sweden while accounting for underlying socioeconomic factors through NDI. Areas of particularly high associations provide opportunities for further research to pinpoint the possible causality factors as well as aiding targeted sensitization, intervention and control measures.

2. Literature Review

The pathophysiological pathways of CVD as triggered by particulate matter (PM) air pollution were investigated by Pope et al. (2004), identifying pulmonary and systemic inflammation, accelerated atherosclerosis, and altered cardiac autonomic function as possible mechanisms. These mechanisms are re-echoed by Vidale and Campana (2018).

Whereas most studies have almost exclusively concentrated on PM air pollution, especially PM10 and PM2.5 (Bell et al. 2008; Cohen et al. 2017; Dominici et al. 2006; Lim et al. 2014; Meister et al. 2012; Pope et al. 2004; Qiu et al. 2017; Segersson et al. 2017; Stockfelt et al. 2017; Zhang et al. 2014), others have also studied other gaseous air pollutants like nitrogen oxides (NOx), CO, sulphur dioxide (SO2), ozone (O3) and BC (Atkinson et al. 2013; Beelen et al. 2008; Grahame and Schlesinger 2010; Le Tertre et al. 2002; Sun et al. 2010; Sunyer et al. 2003) with an assumption

that they trigger the same pathways in CVD as triggered by particulate matter. For most of these studies, their motivating question could be generally summarized as: "in my study area, is ambient air pollution significantly associated with CVD morbidity and/or mortality; if so, to what extent?" Consequently, most of these studies have been limited to the boundaries of single cities. Other studies have considered multiple cities but only for comparison reasons (Segersson et al. 2017; Zhang et al. 2014) to evaluate where the effects of air pollution are contributing more to CVD health outcomes. However, such studies ignore the fact that even within cities, the association between air pollution and CVD can be and is often heterogeneous (Luo et al. 2016).

In a recent paper detailing clinical handling of the CVD-air pollution challenge, Hadley et al. (2018) highlighted the importance of geospatial maps in identifying areas of elevated CVD risk from ambient air pollution to aid targeted intervention at individual and population level. Their recommendations call for localised spatial regression models to help distil the heterogeneity within the relative risk, but also to link the different risk factors to CVD outcomes. Among other risk factors, neighbourhood deprivation has previously been shown to independently predict heart disease morbidity (Sundquist et al. 2004a). Using a Neighbourhood Deprivation Index (NDI), Winkleby et al. (2007) found that age-adjusted Coronary Heart Disease (CHD) incidence and case fatality from CHD was about twice as high for persons in high versus low deprivation neighbourhoods in Sweden. Similarly, having accounted for age and other individual-level factors, Lawlor et al. (2005) found that the odds for CHD were 27% higher for women in British wards with higher deprivation scores than the median score. More recently, Li et al. (2019), showed after adjusting for potential confounders a significant and still retained association between neighbourhood deprivation and heart failure among patients with diabetes mellitus in Sweden.

To address spatial heterogeneity, different studies have used different methods that are subsequently discussed. For example, Alexeeff et al. (2018) used Cox proportional hazard regression to study the association between the incidence of CVD and long-term exposure to transport-related air pollution (TRAP), including nitrogen dioxide (NO2), nitric oxide (NO), and BC, in Oakland California. Their results show that street-level variation in TRAP exposure within urban neighbourhoods significantly contributes to differences in risk of CVD events. However, as with most studies using Cox proportional hazard regression (Jerrett et al. 2017; Qiu et al. 2017; Stockfelt et al. 2017), the heterogeneity in the association was not explicitly accounted for. Failure to account for spatial heterogeneity and spatial autocorrelation, a phenomenon where similar values tend to be near each other, has been shown to lead to underestimation of the uncertainty associated with the effects of air pollution on health outcomes (Burnett et al. 2001).

Luo et al. (2016) used a mixed Cox proportional hazard model to analyse the spatial heterogeneity of the effects of NO2 on Cardiovascular mortality in the 16 districts of Beijing. They applied conditional logistic regression to evaluate the district-specific effects of NO2 on Cardiovascular mortality. Their results showed independent and spatially varied effects of NO2 on CVD mortalities, providing actionable evidence of districts with higher risk. They, however, also did not explicitly handle the spatial effects of spatial autocorrelation and spatial heterogeneity within the NO2 and CVD data. Blangiardo et al. (2016) used a two-stage Bayesian model, first to

estimate the concentration of NO2 from sparse monitoring stations to spatial units (used by the Clinical Commissioning Group - CCG) across England, and secondly to investigate the effect of NO2 on prescription rates of chronic respiratory diseases using integrated nested Laplace approximations. However, given the nature of the prescription data used (aggregated at CCG level), they could not make inference at the individual level or link the data with hospital admissions. Additionally, the use of Bayesian-based methods is extremely computer intensive resulting in lengthy processing times for large datasets.

Regarding methods that explicitly address spatial heterogeneity, most of them fall within the category of Geographically Weighted Regression (GWR), with slight modifications to account for the nature of the data being modelled (Gomes et al. 2017). For example, whereas both GWR and Geographically Weighted Poisson Regression (GWPR) can be used for modelling of spatially heterogeneous processes, allowing for relationships between a response and a set of covariates to vary across geography (Fotheringham et al. 2015; Nakaya 2015), they are different modelling frameworks – GWR assumes Gaussian outcomes (Fotheringham et al. 2002) and GWPR assumes Poisson counts (Nakaya et al. 2005). Poisson counts are more appropriate for modelling small area disease rates, especially where the local expected number is low (Nakaya et al. 2005), as was with our case. For data with overdispersion, Geographically Weighted Negative Binomial Regression (GWNBR) model is sometimes preferred (da Silva and Rodrigues 2014). These models have been used in many studies and compare differently.

By using scan statistics and GWR, Lim et al. (2014) investigated the correlation between PM10 and CVD mortality (daily counts of death from 2008 to 2010) in Seoul metropolitan area, South Korea. They concluded that CVD mortality was related to the concentration of PM10 and that this relationship was heterogeneous across their study area. Since count data was used in their study, we argue that GWPR would have been a more appropriate model. By comparing the Root Means Square Error (RMSE) from GWPR and global negative binomial (GNB) models, Li et al. (2013) found that GWPR performed better than the GNB model since it had a lower RMSE. Gomes et al. (2017) also studied the performances of GNB, GWPR and GWNBR models. They concluded that GWPR and GWNBR models performed better than the GNB model. They also asserted that GWNBR had performed better than GWPR, judging by the Akaike Information Criterion (AIC) metric. However, in their study, GWPR outperformed GWNBR when judged using the RMSE metric as used by Li et al. (2013). Additionally, the use of GWNBR resulted in a wider bandwidth, hence banding effects were observed in the obtained coefficient maps (more homogeneous). Given that our primary concern was spatial heterogeneity of the associations, between the two (GWNBR and GWPR), GWPR was more tailored for our specific problem.

GWPR has been used in analysing local variations in associations between health outcomes (disease counts, incidence rates, mortality risks, etc.) and a set of environmental and socio-economic characteristics (Alves et al. 2016; Feuillet et al. 2015; Nakaya et al. 2005). Specific to CVD, GWPR was used by Chen et al. (2010) to examine the non-stationary effects of extreme cold on mortality in Taiwan. By studying these non-homogeneous spatial patterns between disease

outcomes and a set of variables, these studies provide actionable tools in managing diseases and increase our understanding of how geography influences these associations.

In Sweden, however, such local spatial regression analyses for CVD and ambient air pollution at a countrywide level have hitherto not been studied. Moreover, CVD is the highest cause of death in Sweden with about 91,000 deaths in 2015 (Brooke et al. 2017). The prevalence of CVD in 2015 was 1,942,532 cases in 2015; approximately 20% of the 9.747 million Swedish population in 2015 (Wilkins et al. 2017). Whereas some studies have been done on CVD and air pollution in Sweden (Le Tertre et al. 2002; Segersson et al. 2017; Stockfelt et al. 2017; Sunyer et al. 2003), they only considered selected cities (Stockholm, Gothenburg, and Umea), and so the multi-pollutant effect of air pollution on CVD and the spatial variation of such effects across Sweden remains to be studied. Our aim was to adopt a Poisson modelling framework to analyse the association between PM10, PM2.5, BC, CO, and SOx and CVD hospitalization while accounting for neighbourhood deprivation, and the spatial variation of this association across the whole Sweden.

3. Materials and Methods

3.1. Data Acquisition

3.1.1. Cardiovascular data

The CVD data used in this study are based on Swedish hospital records of CVD admission between January 1st, 2005 and December 31st, 2010. According to the World Health Organization's International Classification of Diseases (ICD-10), the following CVDs were considered: Coronary heart disease (CHD) codes including I20, I21, I22, I23, I24, I25; Ischemic stroke codes including I63 (excluding I63.6), I65, I66, I67.2, I67.8, G45 and G46 (G46 was only included when it was in combination with another diagnosis), and atherosclerotic and aortic disease codes including I70, I71, I72, I73 (excluding I73.0, I73.1), I74 and I77.1.

Hospital admissions including date of admission were obtained from the Swedish National Board of Health and Welfare and comprised 538,573 hospital admissions across Sweden for the years 2005 to 2010 as shown in Figure 1. From National Population Registers, the approximate location of each patient within 100m was also obtained, providing a basis for spatial aggregations.

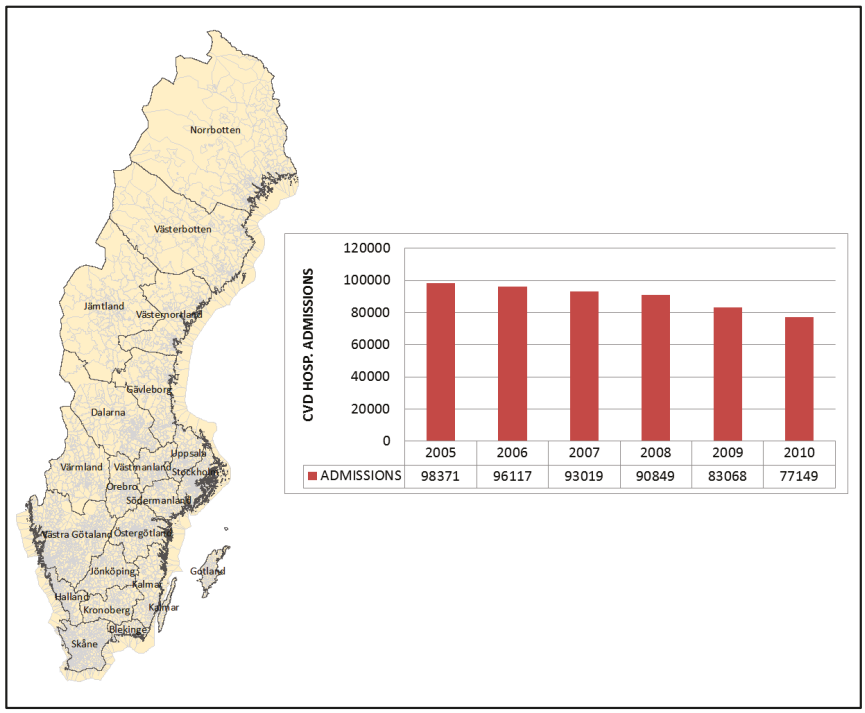


Figure 1. Sweden Counties (with SAMS) and Trends of CVD Hospital Admissions in Sweden from 2005 to 2010

3.1.2. Emission data

The emission data (hereinafter used interchangeably with "air pollution") used in this research was based on the Swedish Environmental Emissions Data (SMED) and consists of particulate matter (PM10 and PM2.5), Black Carbon, Sulphur Oxides and Carbon monoxide emission records across Sweden, in a 1 km by 1 km grid resolution for the period 2005 to 2010. The details for the calculation of the 1 km by 1 km grid estimates and the validation of these estimates were discussed by Gidhagen et al. (2009). The SMED consortium uses this very emission inventories for the report of greenhouse gases to the European Commission, under the Climate Convention obligation (Gidhagen et al. 2009; Gidhagen et al. 2013). The data was generated by SMED as annual averages from eight sectors of the power supply, industrial processes, product usage, transportation, work machines, agriculture, waste and sewerage, and international aviation and shipping (SMED 2016). The same data has previously been used, at the urban level, by Stockfelt et al. (2017) in their study of air pollution and CVD in the city of Gothenburg, Sweden.

3.1.3. Neighbourhood Deprivation Index

NDI is a summary measure used to characterise neighbourhood-level deprivation. Deprivation indicators that have been used in previous studies were identified to characterise neighbourhoods; principal component analysis was then used to generate the SAMS specific z-score (first principal component) indicative of NDI. Four variables were selected for persons aged 25 – 64 years. These four were low educational status (<10 years of formal education); unemployment (not employed,

excluding full-time students, those completing compulsory military service, and early retirees); social welfare recipient (receiving social welfare support); and low income (income from all sources, defined as less than 50% of individual median income.

3.1.4. Scale of modelling

All modelling and analysis were carried out at SAMS (Small Area for Market Statistics) level, which is a census regional division, defined by Statistics Sweden (http://www.scb.se), based on homogenous types of buildings so that they approximately contain 1000 residents. The admissions at the individual level and emission values for each pollutant were aggregated to these SAMS blocks. For SAMS whose underlying population was less than 50 persons, they were excluded from the analysis as their inclusion would lead to unstable statistical estimates (Sundquist et al. 2004b; Sundquist and Yang 2007). This reduced the original number of SAMS from the original 9194 to 8419.

3.2. Study methodology

We used the Poisson framework to model the associative relationships between ambient air pollution and CVD admissions in Sweden while accounting for neighbourhood deprivation. Poisson was chosen because the observed CVD admissions were recorded as counts and also given that the local expected number was low.

Global Poisson model (GPM) was applied first to recognize the relations of individual pollutants with CVD, in addition to understanding the significance of these relations at the global level. The existence of spatial correlation in data results in biased estimates of the global models (Anselin 1988; Anselin and Rey 1991). Given that the GPM does not account for spatial effects in the observed CVD, a spatial lag term was introduced to address the influence of neighbourhood values on the observed CVD values to the GPM, leading to the spatially auto-regression Poisson (SAR-Poisson) model. This was important as CVD cases in a region are also influenced by the underlying socio-economic, demographic and environmental factors (Poulter 1999), which are seldom random in space. However, SAR-Poisson, being a global model, does not handle local spatial heterogeneity in the obtained associations.

Regression models that allow for geographical weighting are better suited for handling spatial heterogeneity (Nakaya 2015). We thus employed Geographically Weighted Poisson Regression (GWPR) model that allows for the establishment of coefficient terms and all other regression parameters for each spatial unit (8419 units for our case).

3.3. Statistical methods

We investigated the associations between annual ambient air pollution exposures (PM10, PM2.5, BC, SOx and CO) from the eight sectors (power supply, industrial processes, product usage, transportation, work machines, agriculture, waste and sewerage, and international aviation and shipping), NDI and CVD admission count. Let y_i be the CVD admission count for a particular SAMS (i). Denote the five ambient air pollution determinants, and NDI as $x_{k,i}$, $k = 1, 2, \ldots, 6$. The SAMS specific population is denoted by N_i. The conventional GPM can then be specified by equation (1).

$$y_i \sim Poisson\left[N_i exp\left(\sum_k \beta_k x_{k,i}\right)\right] \quad (1)$$

To model for the possible existence of spatial dependence in the observed CVD admissions, SAR-Poisson model was derived from the original GPM in equation (1) by adding a spatial lag term, $\rho(W_i y_i)$, shown in equation (2).

$$y_i \sim Poisson\left[N_i exp\left(\rho(W_i y_i) + \sum_k \beta_k x_{k,i}\right)\right] \quad (2)$$

where $W_i y_i$ is the spatially lagged dependent variable for the weights matrix W_i, and ρ is a spatially lagged coefficient. The weight matrix was defined by considering n-nearest neighbours.

GWPR extends this traditional model by allowing for all parameters to vary with geographical location, defined by SAMS in our study. This introduces a location parameter, $\boldsymbol{u}_i = (u_{xi}, u_{yi})$, a vector containing the two-dimensional coordinates describing the location of the particular SAMS (centroid coordinates). The Poisson model in equation (1) can be rewritten as equation (3).

$$O_i \sim Poisson\left[N_i exp\left(\sum_k \beta_k(\boldsymbol{u}_i) x_{k,i}\right)\right] \quad (3)$$

The regression coefficients βs in equation (3) are calculated for each and every SAMS (i), making them spatially varying. This makes GWPR a local spatial regression model allowing for geographically varying parameters. In our study, there were a total of 8419 estimated coefficients corresponding to the 8419 SAMS used.

The geographical weighting in GWPR is such that a kernel window is placed around every SAMS, and the βs are computed using all the data contained within the kernel window, allowing for neighbourhood data to contribute to the value of β at that specific SAMS. A bi-square adaptive weighting kernel, defined by equation (4), was used for our study.

$$w_{ik} = \left[1 - \left(\frac{d_{ik}}{h}\right)^2\right]^2 \text{ if } d_{ik} \le h;\ otherwise\ 0 \quad (4)$$

where d_{ik} is the distance between SAMS (i) and the nearby SAMS (k). Observations closer to SAMS (i) would carry more weights and have greater impacts on parameter estimates than those far away, in accordance with the first law of geography (Tobler 1970). h is a constant bandwidth defining the neighbourhood, and for GWPR, it denotes a value that yields the lowest Akaike Information Criterion (AIC), a metric that deals with the trade-off between the goodness of fit of the model and model simplicity, defined by equation (5), through the bandwidth selection procedure (Nakaya et al. 2005).

$$AIC(G) = D(G) + 2K(G) \quad (5)$$

where D and K denote the deviance and the effective number of parameters in the model with bandwidth G, respectively.

4. Results

To examine the possible determinants the spatial variation of CVD across Sweden and over time, each ambient air pollution variable and NDI were regressed against the SAMS-specific CVD outcome. Table 1 shows the model estimates for the GPM model combining all the independent

variables. Generally, the associations were both positive and negative except for SOx that was positive throughout.

Table 1. Global Poisson model Estimates (2005 to 2010)

	Estimate (2005)	Estimate (2006)	Estimate (2007)	Estimate (2008)	Estimate (2009)	Estimate (2010)
(Intercept)	-0.207*** [1]	-0.2364***	-0.271***	-0.3117***	-0.3911***	-0.4653**
NDI	-0.058***	0.05548***	0.05565***	0.05709***	-0.0618***	-0.06027***
BC	-5.275E-07**	-4.846E-07**	-8.8E-07***	-1.1E-06***	-1.8E-06***	-1.6E-06***
CO	-7.064E-09***	-7.53E-09***	-4.1E-09**	3.56E-09***	3.85E-09*	3.98E-09.
PM10	2.295E-07***	1.559E-07***	2.23E-07***	1.3E-07***	5.28E-08	-3.7E-08
PM25	-2.798E-07***	-1.173E-07**	-2.6E-07***	-1.8E-07***	-1E-07.	3.49E-08
SOx	3.19E-08***	1.531E-08*	3.12E-08***	3.82E-08***	4.92E-08***	3.51E-08***

[1] Significant codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 2 shows the SAR-Poisson model estimates. All the variables generally remained significant throughout the years, consistent with the results from the GPM. However, the introduction of the lag term created some changes in the nature of the associations observed in the overall GPM. For example, PM10 and BC become consistently negative while PM2.5 becomes consistently positive, just as SOx and the lag term. CO and the NDI term retain their mixed associations.

Table 2. Spatial Autoregressive Global Poisson model Estimate (2005 to 2010)

	Estimate (2005)	Estimate (2006)	Estimate (2007)	Estimate (2008)	Estimate (2009)	Estimate (2010)
(Intercept)	-0.3904***	-0.4181***	-0.4546***	-0.4975***	-0.5718***	-0.6523***
lag	0.01426***	0.01464***	0.01518***	0.01585***	0.01656***	0.01834***
NDI	-0.04152***	0.03937***	0.03987***	0.04134***	-0.04672***	-0.04522***
BC	-3.926E-07**	-2.47E-07.	-6.3E-07***	-8.7E-07***	-1.2E-06***	-1.3E-06***
CO	-5.843E-09***	-3.807E-09**	-3.6E-09*	3.17E-09***	2.02E-09	4.67E-09*
PM10	-3.321E-07***	-2.619E-07***	-3.3E-07***	-4.2E-07***	-5E-07***	-5.8E-07***
PM25	4.747E-07***	3.837E-07***	4.92E-07***	5.76E-07***	6.76E-07***	7.67E-07***
SOx	1.156E-08	7.031E-09	8.4E-09	2.2E-08*	2.25E-08*	1.52E-08.

This unstable nature of the associations could be possibly due to multicollinearity existing within the air pollution and NDI variables. Indeed, by computing for the Variation Inflation Factor (VIF) statistic for the five variables, values ranging from 2 to 20 were obtained. Ideally, these values should be less than 5; values between 5 and 10 indicate moderate multicollinearity while values above 10 indicate extreme multicollinearity (Alves et al. 2016). It thus showed that we were dealing with a substantial amount of multicollinearity.

Being a local regression model, GPWR accounts for spatial heterogeneity and is robust against multicollinearity (Fotheringham and Oshan 2016). Figure 2 shows the performance of the three models: the GPM, the SAR-Poisson model and the GWPR model. It shows that GWPR has

consistently lower AIC values, followed by SAR-Poisson, and GPM has the highest AIC values throughout the study period.

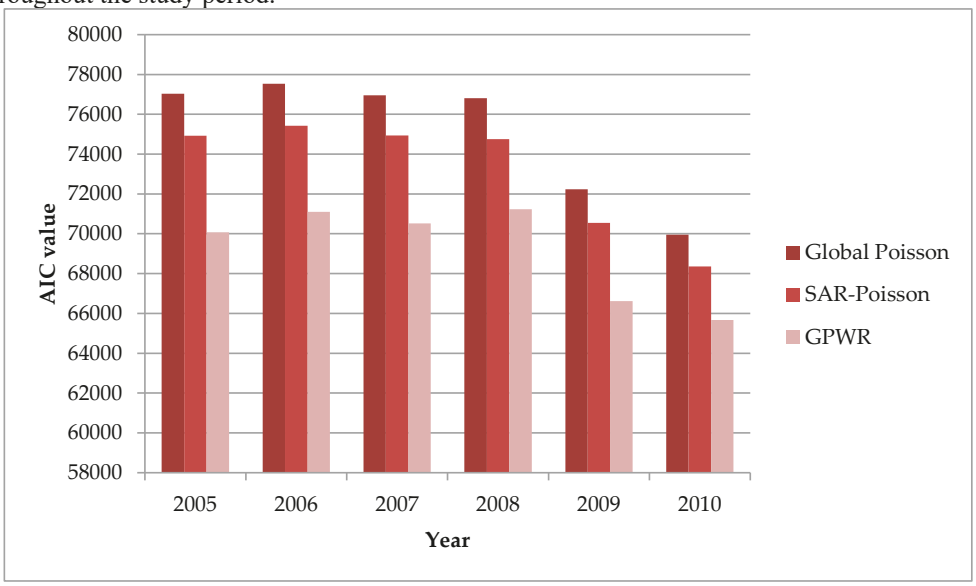


Figure 2. Model Performance by Akaike Information Criterion (AIC) for the 3 models

Table 3 shows the summary of parameter estimates as obtained from the GWPR model. They are described by the minimum, lower quartile, median, upper quartile, and the maximum. Given that the parameter values were not standardised in the global models, the intercept is the only comparable parameter between the global and these local estimates. We note that the median intercept coefficient estimates for both models (GWPR and overall GPM) were relatively similar, for all the years. Additionally, some parameter estimates range from negative to positive over the study area, exhibiting a wider dynamic range compared to the averaged values reported by GPM.

Table 3. Summary statistics for varying (Local) coefficients from the GWPR model

Coefficients	Minimum of Coefficients	Lower Quartile of Coefficients	Median of Coefficients	Upper Quartile of Coefficients	Maximum of Coefficients
2005 Intercept	0.898	1.316	1.438	1.572	1.841
NDI_2005	-0.750	-0.513	-0.429	-0.283	-0.093
BC	-2.565	-0.213	-0.016	0.217	1.490
CO	-1.874	-0.123	0.074	0.371	1.629
PM10	-6.988	-0.976	-0.181	0.508	3.759
PM25	-2.857	-0.339	0.152	0.784	4.847
SOx	-5.605	-0.156	-0.039	0.038	6.241
2006 Intercept	0.873	1.302	1.403	1.551	1.793
NDI_2006	0.090	0.268	0.440	0.509	0.783
BC	-3.038	-0.313	-0.057	0.174	1.558
CO	-1.888	-0.043	0.144	0.443	1.543

PM10	-4.268	-0.893	-0.109	0.446	4.085
PM25	-1.559	-0.210	0.075	0.577	2.569
SOx	-5.365	-0.151	-0.027	0.090	6.384
2007 Intercept	0.813	1.263	1.371	1.531	1.759
NDI_2007	0.092	0.257	0.419	0.499	0.776
BC	-2.534	-0.209	-0.001	0.231	1.870
CO	-2.005	-0.078	0.113	0.366	1.321
PM10	-5.239	-0.895	-0.227	0.375	4.561
PM25	-2.553	-0.314	0.158	0.539	4.691
SOx	-5.113	-0.126	-0.021	0.062	8.773
2008 Intercept	0.715	1.243	1.357	1.492	1.733
NDI_2008	0.097	0.257	0.407	0.486	0.782
BC	-3.846	-0.128	0.100	0.333	1.404
CO	-0.499	-0.143	-0.032	0.023	0.228
PM10	-5.158	-0.691	-0.185	0.509	7.446
PM25	-3.548	-0.443	0.151	0.678	5.132
SOx	-5.250	-0.196	-0.048	0.031	9.307
2009 Intercept	0.654	1.131	1.250	1.414	1.738
NDI_2009	-0.784	-0.463	-0.395	-0.272	-0.098
BC	-3.313	-0.280	-0.051	0.189	1.976
CO	-2.149	-0.002	0.188	0.408	1.575
PM10	-5.582	-1.368	-0.351	0.391	8.087
PM25	-2.951	-0.228	0.314	0.879	5.191
SOx	-5.787	-0.128	-0.018	0.128	10.021
2010 Intercept	0.520	1.036	1.185	1.338	1.764
NDI_2010	-0.842	-0.450	-0.400	-0.288	-0.100
BC	-2.314	-0.283	-0.092	0.061	2.092
CO	-2.116	-0.006	0.228	0.544	1.563
PM10	-7.036	-1.370	-0.480	0.392	4.762
PM25	-5.386	-0.188	0.401	0.973	6.651
SOx	-4.454	-0.190	-0.034	0.122	12.116

The spatial variation of the associations between CVD and the five air pollution variables (BC, CO, PM10, PM2.5, and SOx) were visualised by maps. The mean spatial variations of the coefficients of all the five air pollutants are given by Figure 3. This was obtained by averaging SAMS-specific coefficients for individual pollutants, over the six-year period. Labels (a), (b), (c), (d), and (e) in Figure 3 was used to distinguish the spatial coefficient variations for BC, CO, SOx, PM10, and PM2.5, respectively. By averaging, areas of particularly persistent high associations were highlighted. For example, BC (a) shows a moderately strong association in Gotland (an island in the southeast), across mid-lower-western regions and across mid-upper-western regions of Sweden. Weaker associations for BC are observed mainly in the northern parts of Sweden. PM10 (d) is the most pronounced with strongest associations in the mid-lower parts of Sweden and persistently moderate to strong associations in the North. CO (b) and SOx (c) show generally moderate associations while PM2.5 (e) shows generally low associations with CVD, across

Sweden. The upper northern part particularly shows lower associations with PM2.5 over the six-year study period.

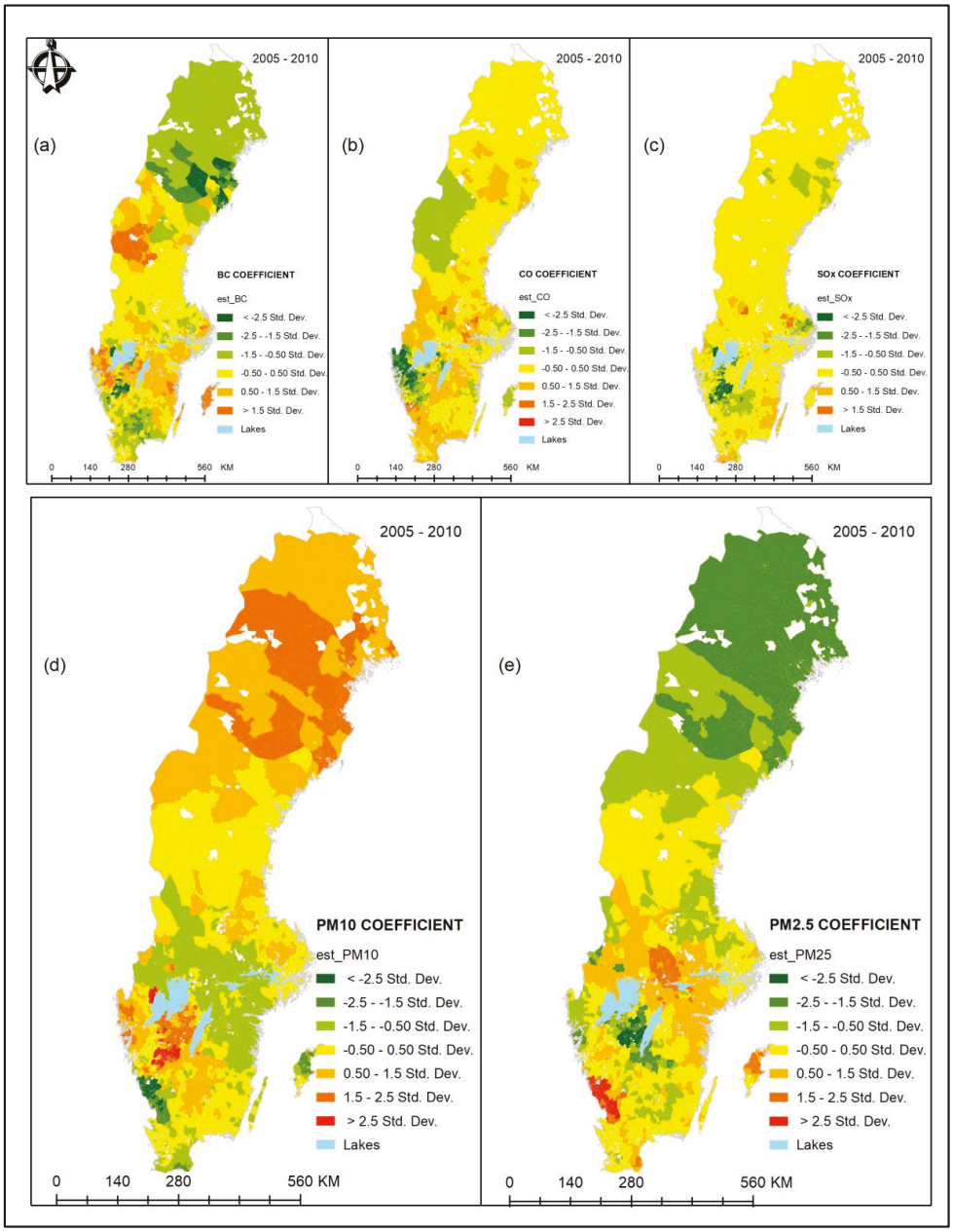


Figure 3. Combined Spatial Variations (2005 -2010) in BC (a) CO (b), SOx (c), PM10 (d) and PM2.5 (e)

5. Discussion

This study employed a spatial perspective with emphasis on spatial non-stationarity to assess the association between ambient air pollution variables (BC, CO, PM10, PM2.5, and SOx) and CVD admissions while accounting for neighbourhood deprivation in Sweden from 2005 to 2010. Neighbourhood deprivation in the form of the established index, NDI, was used to account for the underlying socio-economic variables mainly because it has been shown to independently predict heart disease morbidity. The results of our analysis also showed that NDI has a consistently significant association with CVD in Sweden, as indicated by the results of the global Poisson models. This was consistent with the results obtained by previous studies in Britain (Lawlor et al. 2005) and Sweden (Kawakami et al. 2011; Sundquist et al. 2004b; Winkleby et al. 2007).

The GWPR model as an effective tool to evaluate these non-stationary relations was used to compute the spatially-varying regression parameter estimates across Sweden. GWPR assumes a Poisson distribution to model count-based outcomes and is hence statistically more appropriate than conventional regression models based on Gaussian distribution like conditional and simultaneous autoregressive models (Chen et al. 2010). Through the Akaike Information Criterion (AIC) statistic, Figure 2 shows that GWPR was the best model fit against global Poisson models as measured by the lowest AIC value. It was followed by SAR-Poisson, and global Poisson model was the worst of the models. To examine how GWPR successfully captured the spatially non-stationary variations in the coefficient parameters, Table 3 was examined. Here, some estimated coefficients range from negative to positive over the study area. This indicates how GPWR was able to capture the spatial non-stationarity, and how the global models (in Table 1 and Table 2) can be misleading by assuming constant association coefficients across the study area.

We argue that traffic-related PM10 could be responsible for the persistent strong associations with CVD in the middle-south and southwest of Sweden (Figure 3). This position is consistent with the results of Segersson et al. (2017) who contended that PM10 and BC are primarily produced by road traffic through both wear particles and exhaust. However, the moderately stronger associations of PM10 in the northern part of Sweden were unclear to attribute to any specific source. PM2.5, CO, and BC are known to be mainly produced by residential wood combustion and road traffic sources (Segersson et al. 2017; Stockfelt et al. 2017). We thus speculate that residential wood heating, fuel burning, and road traffic could be largely responsible for the observed spatial patterns between BC, PM10, PM2.5, and CO air pollution and CVD in Figure 3. On the other hand, SOx is known to be a shipping pollutant due to the high sulphur content of marine fuels (Nikopoulou et al. 2013). Therefore, the patterns observed in SOx associations could be attributed to the numerous navigational routes along the coastline (especially the southern half of Sweden).

It should be emphasized that our interpretation of place-specific association obtained in this study is more general as pollutants may exhibit associations in places away from their sources. This noncommittal interpretation was called for by Meister et al. (2012) who cautioned about the interpreting place-specific associations of PM2.5 and health outcomes as large portions of it in cities tend to be transported over long distances.

Our findings, having considered spatial heterogeneity, were consistent with conclusions from previous studies (Alves et al. 2016; Feuillet et al. 2015; Gomes et al. 2017; Li et al. 2013; Lim et al. 2014; Nakaya 2015; Nakaya et al. 2005) regarding the heterogeneity of relations. From global models, we observed generally weak but highly significant associations between air pollution variables and CVD, evidenced by relatively small coefficient values. These weak and mixed associations were also observed by Stockfelt et al. (2017) and Taj et al. (2017) in their city-specific

studies in Sweden. This could be generally explained by the fact that CVD is multi-factorial (Poulter 1999) and influenced by many other lifestyle and socio-economic determinants, like smoking, hypertension, lack of exercises, to mention but a few, in addition to ambient air pollution. However, they could also be due to the combined effects of overdispersion and multicollinearity in the data (Gomes et al. 2017).

Overdispersion issues are well handled by Global Negative Binomial models (GNB), and would as such be a better alternative (Alves et al. 2016; Li et al. 2013). However, given the spatial nature of the data as evidenced by the significant lag term in the SAR-Poisson model, we reasoned that the probabilistic mechanisms used by global GNB to handle such overdispersion would overlook its specific local-scale causes, which was also mentioned by Alves et al. (2016). Moreover, our tests showed that the global GNB results were not very different from the GPM results. Additionally, the unobserved heterogeneity as computed from the density of variance of GNB random effects (Rodríguez 2019) had the quartile ranges [0.185 (Q1); 0.581 (Q2); and 1.373 (Q3)], meaning that CVD admissions at the lower quantile of the unobserved heterogeneity were 81% lower than expected, CVD admissions at the median were 8% higher and those at the upper quantile were 37% higher than expected. The observed overdispersion was therefore in part due to heterogeneity, which is better handled by local spatial models.

While for the local model, GWNBR is known to better handle overdispersion than GWPR (da Silva and Rodrigues 2014), applying GWNBR on a section of our dataset resulted in banding effects, characterised by homogeneous regions in the resultant coefficient maps. This could be attributed to the bandwidth selection procedure converging at wider bandwidths for GWNBR which shows that GWNBR was not able to handle spatial heterogeneity. This was consistent with results obtained by Gomes et al. (2017). It was thus a split-decision between better handling of either overdispersion (GWNBR) or heterogeneity (GWPR). Since spatial heterogeneity was our primary goal, GWPR was selected for our analysis.

Multicollinearity between air pollution variables has been highlighted by Stockfelt et al. (2017) as the limitation responsible for few multi-pollution studies like our own. However, in an elaborate study of GWR, Fotheringham and Oshan (2016) illustrated that GWR is robust even under extreme multicollinearity and produces reliable results.

Finally, whereas this study achieved what it set out to do, the authors are aware that this being an ecological study, there is a need to acknowledge ecological bias. Given that all data (CVD and air pollution) had to be aggregated to SAMS level (underlying population was available at this level), the obtained associations cannot reflect the would-be associations at the individual level.

6. Conclusions

The primary contribution of this study is the global as well as local analyses of the association between several established air pollutants and CVD in Sweden, on a nationwide basis while accounting for socio-economic factors through an established neighbourhood deprivation index. It has successfully demonstrated that multi-pollutant associations with CVD are not homogenous across Sweden and is, to the best of our knowledge, the first nationwide study that spatially analyses multi-pollutant data and CVD with a particular focus on spatial non-stationarity. In this six-year study of CVD admission counts and ambient air pollution, we found generally weak but statistically significant global associations between main particulate matter pollutants and CVD admissions. More importantly, using GWPR, we found these associations to be non-homogeneous but varied across space. Generally, more dynamism in the observed patterns was associated with southern parts of Sweden than with the northern regions. These results are, despite certain

limitations, useful because they indicate that health policies targeting air pollution and CVD preventive and management efforts in Sweden may be defined at local levels rather than at a global (national – in this case) level. Moreover, with areas of persistent high associations between air pollution and CVD identified, more focused studies could be conducted in these areas to learn more about the drivers of such associations so as to better inform future healthcare policy and intervention efforts.

Conflict of Interest

The authors declare no conflict of interest.

Funding Information

This study was part of a project funded by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number R01HL116381 to Kristina Sundquist. The first and the second authors of this study, Augustus Aturinde and Mahdi Farnaghi, are financed by the Swedish International Development Cooperation Agency (SIDA) under award number 2015-003424 to Ali Mansourian.

Availability of data and material

The datasets generated and/or analysed during the current study are not publicly available due to integrity and legal reasons but are available from the corresponding author on reasonable request.

References

Alexeeff SE et al. (2018) High-resolution mapping of traffic related air pollution with Google street view cars and incidence of cardiovascular events within neighborhoods in Oakland, CA Environmental Health 17:38

Alves AT, Nobre FF, Waller LA (2016) Exploring spatial patterns in the associations between local AIDS incidence and socioeconomic and demographic variables in the state of Rio de Janeiro, Brazil Spatial and spatio-temporal epidemiology 17:85-93

Anselin L (1988) Spatial Econometrics: Methods and Models. Kluwer Academic Publishers

Anselin L, Rey S (1991) Properties of tests for spatial dependence in linear regression models Geographical analysis 23:112-131

Atkinson RW, Carey IM, Kent AJ, van Staa TP, Anderson HR, Cook DG (2013) Long-term exposure to outdoor air pollution and incidence of cardiovascular diseases Epidemiology 24:44-53

Beelen R et al. (2008) Long-term effects of traffic-related air pollution on mortality in a Dutch cohort (NLCS-AIR study) Environmental health perspectives 116:196

Bell ML, Ebisu K, Peng RD, Walker J, Samet JM, Zeger SL, Dominici F (2008) Seasonal and regional short-term effects of fine particles on hospital admissions in 202 US counties, 1999–2005 American journal of epidemiology 168:1301-1310

Blangiardo M, Finazzi F, Cameletti M (2016) Two-stage Bayesian model to evaluate the effect of air pollution on chronic respiratory diseases using drug prescriptions Spatial and spatio-temporal epidemiology 18:1-12

Brook RD et al. (2004) Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association Circulation 109:2655-2671

Brooke HL et al. (2017) The Swedish cause of death register European Journal of Epidemiology 32:765-773

Burnett R, Ma R, Jerrett M, Goldberg MS, Cakmak S, Pope 3rd C, Krewski D (2001) The spatial association between community air pollution and mortality: a new method of analyzing correlated geographic cohort data Environmental health perspectives 109:375

Chen R et al. (2017) Fine particulate air pollution and daily mortality. a nationwide analysis in 272 chinese cities American journal of respiratory and critical care medicine 196:73-81

Chen VY-J, Wu P-C, Yang T-C, Su H-J (2010) Examining non-stationary effects of social determinants on cardiovascular mortality after cold surges in Taiwan Science of the Total Environment 408:2042-2049

Cohen AJ et al. (2017) Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015 The Lancet 389:1907-1918

da Silva AR, Rodrigues TCV (2014) Geographically weighted negative binomial regression—incorporating overdispersion Statistics and Computing 24:769-783

Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, Samet JM (2006) Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases Jama 295:1127-1134

Feuillet T et al. (2015) Spatial heterogeneity of the relationships between environmental characteristics and active commuting: towards a locally varying social ecological model International journal of health geographics 14:12

Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. John Wiley & Sons, Limited,

Fotheringham AS, Crespo R, Yao J (2015) Geographical and temporal weighted regression (GTWR) Geographical Analysis 47:431-452

Fotheringham AS, Oshan TM (2016) Geographically weighted regression and multicollinearity: dispelling the myth Journal of Geographical Systems 18:303-329

Gidhagen L, Johansson H, Omstedt G (2009) SIMAIR—evaluation tool for meeting the EU directive on air pollution limits Atmospheric Environment 43:1029-1036

Gidhagen L, Omstedt G, Pershagen G, Willers S, Bellander T (2013) High-resolution modeling of residential outdoor particulate levels in Sweden Journal of Exposure Science and Environmental Epidemiology 23:306

Gomes MJTL, Cunto F, da Silva AR (2017) Geographically weighted negative binomial regression applied to zonal level safety performance models Accident Analysis & Prevention 106:254-261

Grahame TJ, Schlesinger RB (2010) Cardiovascular health and particulate vehicular emissions: a critical evaluation of the evidence Air Quality, Atmosphere & Health 3:3-27

Hadley MB, Baumgartner J, Vedanthan R (2018) Developing a clinical approach to air pollution and cardiovascular health Circulation 137:725-742

Jerrett M et al. (2017) Comparing the health effects of ambient particulate matter estimated using ground-based versus remote sensing exposure estimates Environmental health perspectives 125:552

Kawakami N, Li X, Sundquist K (2011) Health-promoting and health-damaging neighbourhood resources and coronary heart disease: a follow-up study of 2 165 000 people J Epidemiol Community Health 65:866-872

Lawlor DA, Davey Smith G, Patel R, Ebrahim S (2005) Life-course socioeconomic position, area deprivation, and coronary heart disease: findings from the British Women's Heart and Health Study American journal of public health 95:91-97

Le Tertre A et al. (2002) Short-term effects of particulate air pollution on cardiovascular diseases in eight European cities Journal of Epidemiology & Community Health 56:773-779

Li X, Sundquist J, Forsberg P-O, Sundquist K (2019) Association Between Neighborhood Deprivation and Heart Failure Among Patients With Diabetes Mellitus: A 10-Year Follow-Up Study in Sweden Journal of cardiac failure

Li Z, Wang W, Liu P, Bigham JM, Ragland DR (2013) Using geographically weighted Poisson regression for county-level crash modeling in California Safety science 58:89-97

Lim Y-R, Bae H-J, Lim Y-H, Yu S, Kim G-B, Cho Y-S (2014) Spatial analysis of PM10 and cardiovascular mortality in the Seoul metropolitan area Environmental health and toxicology 29

Luo K et al. (2016) Acute effects of nitrogen dioxide on cardiovascular mortality in Beijing: an exploration of spatial heterogeneity and the district-specific predictors Scientific reports 6:38328

Meister K, Johansson C, Forsberg B (2012) Estimated short-term effects of coarse particles on daily mortality in Stockholm, Sweden Environmental health perspectives 120:431

Nakaya T (2015) Geographically Weighted Generalised Linear Modelling. In: Brunsdon C, Singleton A (eds) Geocomputation: A Practical Primer. Sage,

Nakaya T, Fotheringham AS, Brunsdon C, Charlton M (2005) Geographically weighted Poisson regression for disease association mapping Statistics in medicine 24:2695-2717

Nikopoulou Z, Cullinane K, Jensen A (2013) The role of a cap-and-trade market in reducing NO x and SO x emissions: Prospects and benefits for ships within the Northern European ECA Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment 227:136-154

Pope CA, Burnett RT, Thurston GD, Thun MJ, Calle EE, Krewski D, Godleski JJ (2004) Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease Circulation 109:71-77

Poulter N (1999) Coronary heart disease is a multifactorial disease American Journal of Hypertension 12:92S-95S

Qiu H, Sun S, Tsang H, Wong C-M, Lee RS-y, Schooling CM, Tian L (2017) Fine particulate matter exposure and incidence of stroke A cohort study in Hong Kong Neurology 88:1709-1717

Rodríguez G (2019) Models for Over-Dispersed Count Data vol 2019. Princeton University, Princeton University

Segersson D, Eneroth K, Gidhagen L, Johansson C, Omstedt G, Nylén AE, Forsberg B (2017) Health Impact of PM10, PM2. 5 and Black Carbon Exposure Due to Different Source Sectors in Stockholm, Gothenburg and Umea, Sweden International Journal of Environmental Research and Public Health 14:742

SMED (2016) Description of Methods and Quality of Spatially Distributed Emissions to Air During 2016. Contract no 309 1235. (in Swedish, original title: Metod-och kvalitetsbeskrivning för geografiskt fördelade emissioner till luft under 2016). SMED,

Stockfelt L et al. (2017) Long-term effects of total and source-specific particulate air pollution on incident cardiovascular disease in Gothenburg, Sweden Environmental Research 158:61-71

Sun Q, Hong X, Wold LE (2010) Cardiovascular effects of ambient particulate air pollution exposure Circulation 121:2755-2765

Sundquist K, Malmström M, Johansson S (2004a) Neighbourhood deprivation and incidence of coronary heart disease: a multilevel study of 2.6 million women and men in Sweden Journal of Epidemiology & Community Health 58:71-77

Sundquist K, Winkleby M, Ahlén H, Johansson S-E (2004b) Neighborhood socioeconomic environment and incidence of coronary heart disease: a follow-up study of 25,319 women and men in Sweden American journal of epidemiology 159:655-662

Sundquist K, Yang M (2007) Linking social capital and self-rated health: a multilevel analysis of 11,175 men and women in Sweden Health & place 13:324-334

Sunyer J et al. (2003) The association of daily sulfur dioxide air pollution levels with hospital admissions for cardiovascular diseases in Europe (The Aphea-II study) European heart journal 24:752-760

Taj T, Malmqvist E, Stroh E, Oudin Åström D, Jakobsson K, Oudin A (2017) Short-Term Associations between Air Pollution Concentrations and Respiratory Health—Comparing Primary Health Care Visits, Hospital Admissions, and Emergency Department Visits in a Multi-Municipality Study International Journal of Environmental Research and Public Health 14:587

Thurston GD, Newman JD (2018) Walking to a pathway for cardiovascular effects of air pollution The Lancet 391:291-292

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region Economic geography 46:234-240

Vidale S, Campana C (2018) Ambient air pollution and cardiovascular diseases: From bench to bedside European journal of preventive cardiology 25:818-825

WHO (2019) Cardiovascular disease. World Health Organisation. http://www.who.int/cardiovascular_diseases/en/. Accessed June 30th 2019

Wilkins E, Wilson L, Wickramasinghe K, Bhatnagar P, Rayner M, Townsend N (2017) European Cardiovascular Disease Statistics 2017. European Heart Network, Brussels

Winkleby M, Sundquist K, Cubbin C (2007) Inequities in CHD incidence and case fatality by neighborhood deprivation American journal of preventive medicine 32:97-106

Zhang L-w et al. (2014) Long-term exposure to high particulate matter pollution and cardiovascular mortality: a 12-year cohort study in four cities in northern China Environment international 62:41-47

Paper IV

Analysis of spatial co-occurrence between cancer and cardiovascular disease mortality and its spatial variation among the Swedish elderly (2010-2015)

Augustus Aturinde[1,2,3*], Ali Mansourian[1], Mahdi Farnaghi[1], Petter Pilesjö[1], Kristina Sundquist[4] and Gilbert Maiga[2]

[1] GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University, SE-221 00 Lund, Sweden; augustus.aturinde@nateko.lu.se; ali.mansourian@nateko.lu.se; mahdi.farnaghi@nateko.lu.se; petter.pilesjo@gis.lu.se

[2] College of Computing and Information Science, Makerere University, Kampala, Uganda; gilmaiga@gmail.com

[3] Department of Lands and Architectural Studies, Kyambogo University, Kampala, Uganda

[4] Center for Primary Health Care Research, Department of Clinical Sciences, Malmö, Lund University, Sweden; kristina.sundquist@med.lu.se

* Correspondence: augustus.aturinde@nateko.lu.se; Tel.: +46 46 222 1687

Abstract

CVD and cancer are the two leading causes of death worldwide. Improvement in cancer early detection and treatment has resulted in an increased number of cancer survivors. However, many of the survivors tend to develop CVD often leading to their demise. Conversely, people with pre-existing CVD conditions, especially the elderly, have increased chances of developing cancer and dying from the same. The World Health Organization, consequently, recommends joint management of both diseases. However, in Sweden, as with many other countries, few studies have explored the nature of the associations between the two disease mortalities and their spatial variation at a population level.

This study uses correlation, global Moran's index and global bivariate Moran's index to investigate national trends of cancer and CVD crude mortality rates in the Swedish elderly. Spatial scan statistics, spatial overlay and local entropy maps were used to analyse for spatial co-occurrence, local joint spatial clustering and associations in the 2010—2015 cancer and CVD crude mortality rates for the Swedish elderly (65+ years). Mortality data were obtained from the Swedish Healthcare Registry.

Our results showed that throughout the years of study, the correlation between cancer and CVD crude mortality rates was averagely positive. Spatial correlation analysis (univariate and bivariate) showed that the contribution of the neighbourhood mortality rates to the observed mortality rates was weak, though significant. From cluster analysis, the cancer and CVD crude mortality rates showed differences in clustering spatial scales with CVD clustering at a smaller scale. Finally, local entropy maps showed that cancer and CVD crude mortality rates were not always related across Sweden, but whenever they were, the relationship was mainly positive and linear.

This study contributes to cancer and CVD public health efforts in Sweden by identifying areas where the two causes of death spatially co-occur, and where the two exhibit no spatial overlap. This provides a valuable starting ground for more focused studies to identify local drivers and/or informs coordinated targeted intervention in both causes of death.

Keywords: Cancer, CVD, spatial variation, Local Entropy Map (LEM), spatial scan statistics, Swedish elderly

1 Background

Globally, and especially in the industrialized economies, more people are likely to die of cancer and cardiovascular diseases than of other diseases. Cardiovascular diseases (CVD), a group of disorders affecting the heart and blood vessels, are the leading cause of death, accounting for about 31% of all global deaths and with 18 million people estimated to have died from CVD in 2016 (WHO 2019b). Cancer, a generic term for a large group of diseases that lead to the growth of abnormal cells that outgrow their boundaries and invade other adjoining body parts, is the second leading cause of death and is estimated to have accounted for 9.6 million global deaths in 2018 (WHO 2019a).

Improvements in early detection and treatment of cancer have led to an increase in the number of cancer survivors (Miller et al. 2019). However, there are concerns that most cancer survivors tend to die from cardiovascular diseases (Armenian et al. 2016). Research has indicated that cardiotoxic therapies like anthracyclines and chest radiation (Rugbjerg et al. 2014) interact with the heart structures in such a way that predisposes one to CVD. These pathways were comprehensively discussed by Moslehi (2016) in their study on cardiovascular effects of targeted cancer therapies. Additionally, cancer survivors tend to be more likely to have CVD known risk factors like diabetes, hypertension, obesity, etc. (Armenian et al. 2016). In spite of this seemingly direct link between the two diseases, many studies have tended to study them independently. Indeed, a recent commentary article by Blaes and Shenoy (2019) asks the question as to whether it is time to include cancer in cardiovascular risk prediction tools.

Whereas the answer to this inclusion/exclusion question is wide, we contend that by analysing for the simultaneous spatial occurrence of clusters in these two diseases (cancer and CVD), the results would provide a plausible forward step towards simultaneous management of the two diseases as advised by the WHO. To this end, previous studies have not addressed the possible spatial co-occurrence of clusters for cancer and CVD. Moreover, a spatial co-occurrence study would establish areas where both disease morbidities and /or mortalities jointly cluster, thereby supporting surveillance and informing targeted intervention. Perhaps more importantly, such a study would form a basis for more focused studies to establish the driving factors behind the simultaneous clustering, in addition to providing plausible clues as to whether cancer ought to be considered a risk factor in CVD prediction and/or vice versa.

This paper addresses this gap by investigating the simultaneous spatial clustering and local association of Cancer and CVD in the Swedish elderly (65+ years). Mortality data from the Swedish Healthcare Registry were used to identify people, sixty-five and above years of age, that died of either cancer or CVD. Correlation analysis through Pearson's r, global Moran's and bivariate Moran's indices were used to examine the correlation tendencies for crude CVD and cancer mortality rates. Then scan statistics and local entropy map was used to identify areas where in space the two mortality rates co-clustered and related together through the years 2010 to 2015.

2 Related studies

The relationship between cancer and CVD has been previously investigated (Miller et al. 2019, Armenian et al. 2016, Rugbjerg et al. 2014, Strongman et al. 2019). These and many more studies seem to agree that chemotherapy and radiotherapy, characteristic of cancer treatment, pre-exposes the cancer survivors to CVD (Moslehi 2016), often leading to their death (Blaes and Shenoy 2019). For example, in a retrospective cohort study of cancer survivors and controls (aged 40 years or older) from Southern California, USA, Armenian et al. (2016) observed that cancer survivors with CVD had a 3.8 fold risk of all-cause mortality when compared with the survivors without CVD.

Similarly, in a population-based cohort study of women diagnosed with early-stage breast cancer (EBC) in Ontario, Canada, Abdel-Qadir et al. (2017) analysed competing risk of death from CVD and from EBC (n = 98,999). The authors found that CVD was the leading competing risk within the study cohort.

Relatedly, having noticed that evidence was scarce on the specific CVDs in survivors of a wide array of cancers, Strongman et al. (2019) used large-scale electronic health records from linked databases in the UK to establish these risk factors. Using a record of 10,825 cancer survivors and 523,541 controls (aged 18 years or older), they compared a range of CVD outcomes using crude and adjusted Cox models. They observed an increased risk of venous thromboembolism in patients of 18 out of the 20 site-specific cancers; heart failure increase in 10 out of the 20 cancers. Also, they observed elevated risks of arrhythmia, coronary heart disease, stroke, pericarditis, and valvular heart disease for multiple cancer sites. This reassures the thinking that cardiotoxicity associated with cancer treatment increases the risk of cancer survivors developing cardiovascular diseases, and most probably dying from the same.

These studies and several other similar ones (Mahase 2019, Blaes and Shenoy 2019, Giza et al. 2017) highlight the urgent need for simultaneous management of both cancer (e.g. breast, lung, kidney, ovarian, etc.) and CVD. From a patient-management perspective, Strongman et al. (2019) elaborated the challenge in managing post-cancer CVD as long-term follow-up is being done mostly by cancer specialists who in turn put less focus on CVD sequelae. This position was re-echoed by Moslehi (2016) who called for closer collaboration between cardiologists and oncologists in order to manage this double threat. From a surveillance point of view, we argue that analysing for spatial co-occurrence of clusters in both CVD and cancer mortality rates would highlight areas of elevated joint-mortality across the region of study, providing workable evidence on where areas of most need are located. Subsequently, this would better inform coordinated oncology and cardiology intervention efforts for better management of both diseases, hence reduced mortalities.

The use of spatial methods to identify and/or qualify risk factors of disease mortality is not new in epidemiology, and certainly not new to cancer and CVD. The advantage of using the spatial approach lies in the ability to tie environmental exposure to the observed health outcomes

(Elliott and Wartenberg 2004). For example, in what is arguably the pioneering work on spatial clustering, Openshaw et al. (1988) used the 'geographical analysis machine' to identify five clusters of child leukaemia in the North of England. Their analysis gave a light on the possibility of other environmental factors, like air pollution other than residual radioactive radiation, being at play for the observed leukaemia clustering. Since then, considerable research using cluster-hunting spatial methods, especially spatial scan statistics, has been done for many infectious diseases and mortalities including HIV (Cuadros, Awad and Abu-Raddad 2013, Wand and Ramjee 2010), TB (Tiwari et al. 2006, Smith et al. 2017, Smith et al. 2018), measles (Tang et al. 2017), scarlet fever (Zhang et al. 2017), foot and mouth (Deng et al. 2013), and malaria (Gwitira et al. 2018) to mention but a few. For non-communicable diseases, CVD clustering has been investigated by Rajabi et al. (2018), colorectal cancer by Sherman et al. (2014), and cervical cancer by Chen et al. (2008), to list a few.

Whereas some spatial co-clustering efforts have been done for HIV/TB (Aturinde et al. 2019), and HIV/malaria (Gwitira et al. 2018), no previous studies investigating the simultaneous spatial clustering for non-communicable diseases like CVD and cancer were available in the consulted literature. Co-clustering analyses support surveillance and intervention efforts by identifying areas most affected by comorbidities and mortalities, thereby providing grounds for more focussed studies to identify the local driving factors. This study, therefore, aims at investigating the spatial co-occurrence of CVD and cancer mortality through analysing the simultaneous spatial clustering patterns and associations of cancer and CVD in the Swedish elderly (65+ years) using mortality data from the Swedish Healthcare Registry. To this end, cluster-searching methods, especially spatial scan statistics, and local entropy maps were used.

3 Methods

The motivation behind bivariate relationship analysis lies in the need to illustrate that the event of the first phenomenon is somehow nearer of farther to the event of the second phenomenon than would be expected at random (Souris and Bichaud 2011). This is especially important in epidemiology as it helps in linking the two health outcomes through possible common underlying environmental or socioeconomic risk factors. In our case, the need was to illustrate that the event of population-adjusted CVD mortality was nearer to the event of Cancer mortality in the Swedish elderly across Sweden for the period 2010 to 2015.

First, preliminary analysis of the non-spatial relationship between the two mortalities rates was done using Pearson's correlation. Given the possibility of having spatial effects in the CVD and cancer mortality rate, Moran's indices for both were therefore calculated. However, given that the objective of the study was to investigate the clustering and simultaneous co-occurrence of CVD and Cancer, spatial scan statistics (through SaTScan) and local entropy analyses were respectively applied.

3.1 Data

3.1.1 Cardiovascular and Cancer data

The CVD data used in this study are based on Swedish CVD mortality records for people 65+ years old between January 1st, 2010 and December 31st, 2015. According to the World Health Organization's International Classification of Diseases (ICD-10), the following CVDs were considered: Coronary heart disease (CHD) codes including I20, I21, I22, I23, I24, I25; Ischemic stroke codes including I63 (excluding I63.6), I65, I66, I67.2, I67.8, G45 and G46 (G46 was only included when it was in combination with another diagnosis), and atherosclerotic and aortic disease codes including I70, I71, I72, I73 (excluding I73.0, I73.1), I74 and I77.1.

For Cancer, the following ICD-10 codes were included: Neoplasms (C00-C97) and Neoplasm of uncertain or unknown behaviour of other and unspecified sites (D48*).

Mortality data including date of death were obtained from the Swedish Healthcare Registry for the years 2010 to 2015. National Population Registers provided the approximate location of each deceased, providing a basis for spatial aggregations to Small Area for Market Statistics (SAMS) level. Statistics Sweden (https://www.scb.se/) defines SAMS as a census regional division based on homogenous types of buildings so that they approximately contain 1000 residents. Additionally, SAMS polygons with population were obtained from the National Population Registers. Some SAMS whose underlying population was less than 50 persons were excluded from the analysis as their inclusion would lead to unstable statistical estimates (Sundquist and Yang 2007, Sundquist et al. 2004). The exclusion reduced the number of SAMS from 9194 to 8419. Crude mortality rates were thus calculated from the aggregated mortality records and at the SAMS-level population.

3.2 Statistical analysis

3.2.1 Correlation

The relationship between cancer and CVD mortality was investigated, preliminarily, by using Pearson's correlation, with the underlying assumption being that the observed mortality due to the two disease mortalities is random across Sweden. As with most health-related data, however, this assumption of independence of occurrence tends to be violated by data having spatial effects – meaning that nearby observations tend to have similar values than far observations, with respect to Tobler's first law of geography (Tobler 1970). To evaluate the contribution of the neighbourhood to the observed mortality in both causes of death, global Moran's Index (Cliff and Ord 1970) was calculated. Further, to evaluate the neighbourhood contribution of cancer to the observed CVD and vice versa, global Bivariate Moran's I (Anselin, Syabri and Kho 2006, Anselin and Rey 2014) was calculated.

3.2.2 Spatial scan statistics

SAMS-specific CVD and cancer mortality spatial clusters were detected using Kulldorff's spatial scan statistics (Kulldorff 1997). This technique has previously been used in clustering analyses

of CVD (Rajabi et al. 2018), tuberculosis (Smith et al. 2018, Smith et al. 2017), HIV (Gwitira et al. 2018), HIV-TB coinfection (Aturinde et al. 2019), to mention but a few. CVD and cancer clusters were detected using the Poisson probability model, with the assumption that the observed CVD and cancer mortality, condition on the population at risk, is random (Kulldorff 1997). The technique works by imposing circular windows of varying sizes across the study area (different locations defined by SAMS centroids). Then the mortality rate within the window was compared with the mortality rate outside the window. The potential clusters (assuming discrete Poisson) are then detected through evaluation of the calculated likelihood ratio (LR) given by equation (1).

$$\mathrm{LR}_i = \left(\frac{z}{E_{[z]}}\right)^{z} \left(\frac{Z-z}{Z-E_{[z]}}\right)^{Z-z} I\left(\frac{z}{E_{[z]}} > \frac{Z-z}{Z-E_{[z]}}\right) \quad (1)$$

where Z is the total number of CVD or Cancer mortality rate in Sweden; z is the observed mortality rate within the window; $E_{[z]}$ is the adjusted expected mortality rate within the window – under the null hypothesis; $I()$ is the binary indicator for high-risk, low-risk or both (evaluating 1, 0, or 11 respectively). The obtained LR values were then ranked and ordered with the window with maximum LR values among all possible radius options and in all possible centroid locations considered the most likely cluster. For this analysis, the statistical significance of the clusters was established through 999 Monte Carlo simulations. SaTScan v9.6 software was used with the "spatial" option and the default user-defined maximum radius of the circular window maintained at 50% of the population at risk (sensitivity analysis of this radius showed no effect on the obtained results).

3.2.3 Local entropy map (LEM)

Entropy has its roots in thermodynamics and is a fundamental concept in information theory (Naimi et al. 2019). Entropy-based approaches have been used as a measure of complexity in physics (Shannon 1948), diversity in ecology (Ricotta and Anand 2006), and uncertainty in information theory (Gray 2011). In geography, joint entropy has been used to study spatially varying multivariate relationships (Guo 2010). It is this aspect of its application in spatially varying relationships that is utilized in this study.

LEM is a non-parametric approach that allows for exploration of spatially varying relationships within variables (CVD and cancer crude mortality rates) observed at functional units (SAMS in our case) by computing for joint entropy using power-weighted minimum spanning trees as a proxy for the joint distribution of the variables (Jin and Lu 2017). Other spatial methods like Local Indicator of Spatial Association (LISA) and Geographically Weighted Regression (GWR) have been previously used to study spatially clustering patterns of phenomena. However, given their assumption of a prior relationship form (mainly linear), and the requirement for one to know the underlying distribution, a less constraining approach in terms of prior relationship and the prior distribution of the data, was preferred. To detect simultaneous local relationships, LEM analysis generally involves four main steps (Guo 2010).

Firstly, the normalized power-weighted Minimum Spanning Tree (MST) length is determined from the bivariate plot of the two variables, and used to estimate the Renyi entropy value (H_λ) for the multivariate/bivariate dataset in each region using equation (2).

$$H_\lambda = \frac{1}{1-\lambda}\left(\log\left(M_\alpha \frac{(x_1, x_2, \dots, x_n)}{n^\lambda}\right) - c\right) \quad (2)$$

where x is a d-dimensional vector; $\lambda \geq 0$ is the order of the Renyi entropy; $M_\alpha(x_1, x_2, \dots, x_n)$ is the minimum spanning tree length; n is the number of independent observations; c is a strictly positive constant that depends on the edge power, α and the dimensionality, d. Secondly, the obtained Renyi entropy values are evaluated for their statistical significance using a permutation-based approach. This converts each local H_λ to a p-value. Thirdly, all p-values (one for each region) are then processed with several statistical tests for the null hypothesis – independence (while controlling for the multiple-testing problem). The False Discovery Rate (FDR) method (Benjamini and Hochberg 1995) was used to control for the multiple-testing problem. It involves ranking the p-values in ascending order (e.g. $p_1 < p_2 \dots < p_n$); finding the first $p_i < \alpha_s * \frac{i}{n}$ and assigning $\alpha_s * \frac{i}{n}$ as its critical value. Fourthly and finally, the p-values are mapped and visualised for the examination of spatially varying local relationships between variables.

In our analysis, LEMs were used to explore SAMS-specific spatial relationships between CVD and cancer mortality rates in the Swedish elderly. The estimation of entropy values from the length of the minimum spanning tree requires the definition of neighbourhood, k and the edge power, α. The influence of k on the resulting entropy map is such that when a larger k is used, more robust entropy values are obtained since more data points are used in representing the relationship. However, this is achieved at the expense of local heterogeneity (with smaller k value). On the other hand, the influence of α is such that decreasing α value reduces the distance between long and short edges, thereby resulting in more robustness against noise and outliers. In his sensitivity analysis, Guo (2010) obtained better results with k in the range ($35-50$) and α in the range ($0.25-0.5$). We conducted a sensitivity analysis of these two parameters (k and α) on our CVD and cancer mortality data and found that the combination $k = 50$ and $\alpha = 0.5$ resulted in fewer type 1 (false discovery) errors. These two parameter values were therefore adopted this for our analysis.

The algorithm for the generation of LEM is provided by ESRI (2019). It basically relies on the comparison of the computed statistical metrics in the form of corrected Akaike Information Criterion (AICc) and adjusted R^2 to segment the local relationships into the six classes of Positive Linear, Negative Linear, Convex, Concave, Undefined complex and Non-significant. Figure 1 schematically summarizes the used algorithm (ESRI 2019).

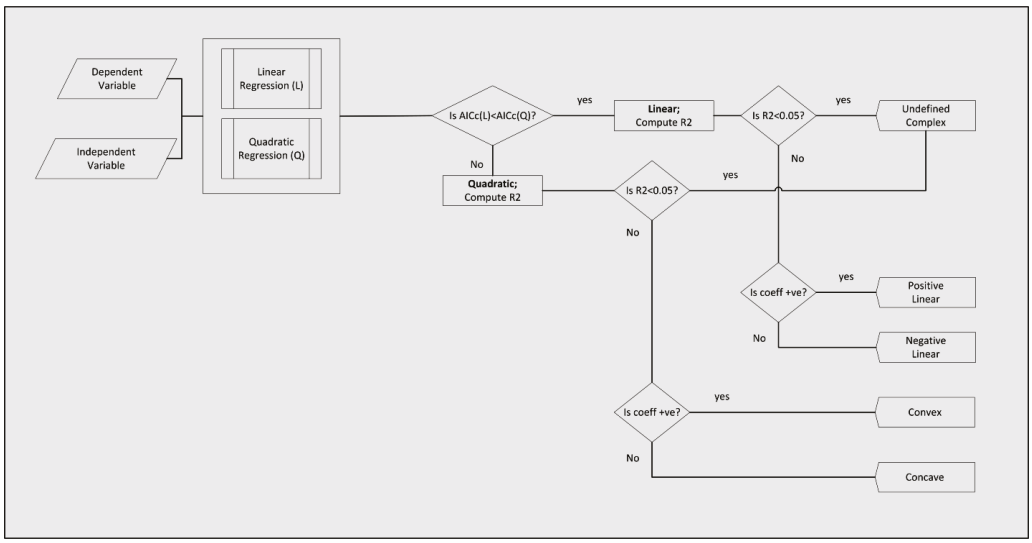
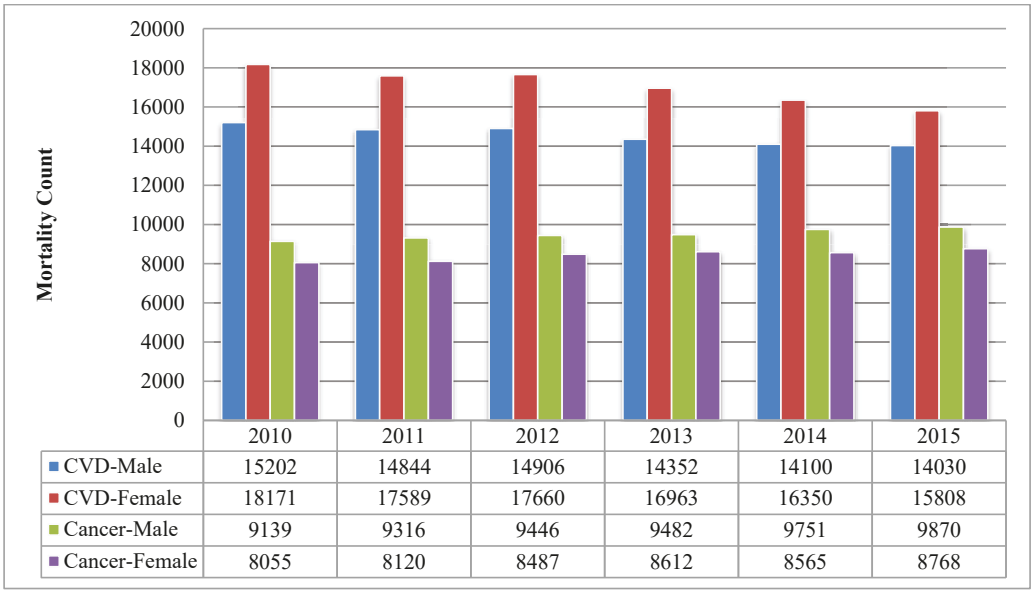


Figure 1: Schematic representation of the Local Entropy Map algorithm

4 Results

4.1 Trends in mortality of CVD and cancer by sex and age

To explore trends within the mortality data, Figure 2 and Figure 3 were used for visualization. It should be noted that these visualizations depict the country-wide mortality before aggregation to SAMS. Figure 2 shows the total mortality per year, stratified by sex, while Figure 3 shows the trends in sex-stratified mortality by age.



	2010	2011	2012	2013	2014	2015
CVD-Male	15202	14844	14906	14352	14100	14030
CVD-Female	18171	17589	17660	16963	16350	15808
Cancer-Male	9139	9316	9446	9482	9751	9870
Cancer-Female	8055	8120	8487	8612	8565	8768

Figure 2: Swedish elderly – total CVD and cancer mortality by sex (2010—2015)

Generally, CVD total mortality in both males and females decreases for the period 2010 to 2015 by 1172 and 2363, respectively. Cancer mortality in both males and females slightly increases by 730 and 713, respectively during the study period.

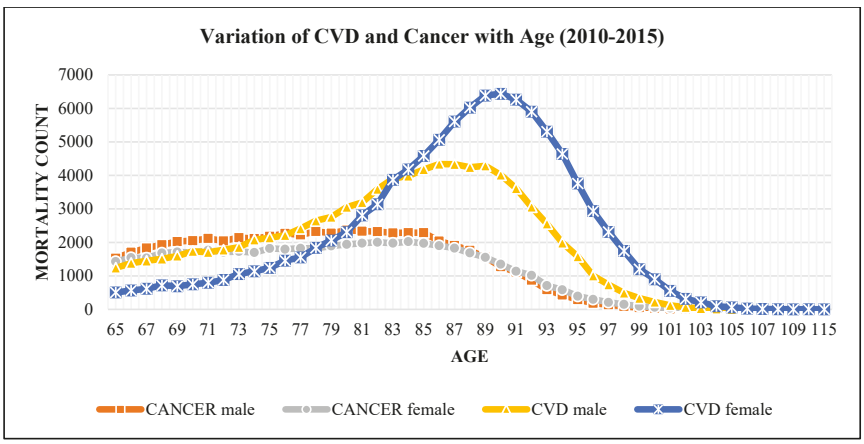


Figure 3: Variation of CVD and cancer mortality in the Swedish elderly by age (2010-2015)

Figure 3 shows that cancer mortality curves for both males and females were close to each other, with the male mortality slightly higher than the female until around 87 years where the order reverses and female mortality becomes slightly higher. Conversely, the CVD mortality curves showed more variability with male mortality starting at a higher value compared to the female curve. The male CVD mortality curve peaks at about 87 years, before starting to drop.

The female CVD curve increases almost exponentially from 65 years to 91 years, before starting to drop almost exponentially too. Finally, it can be observed that for both CVD and cancer mortality, the females tended to outlive their male counterparts.

4.2 Correlation analysis

To distil the possible relationships at a national scale, the results from Pearson's correlation and Moran's indices ($p<0.05$ for Pearson's and 9,999 randomizations for Moran's) are presented in Table 1.

Table 1: Correlation analysis of CVD and cancer crude mortality rates among the Swedish elderly (2010-2015)

	2010	**2011**	**2012**	**2013**	**2014**	**2015**
Pearson's r Correlation						
Cancer & CVD [M]	0.504	0.457	0.445	0.458	0.489	0.418
Cancer & CVD [F]	0.490	0.445	0.387	0.475	0.508	0.470
Cancer & CVD [Total]	0.636	0.625	0.540	0.597	0.635	0.577
Moran's I						
CVD [M]	0.143	0.158	0.156	0.173	0.145	0.177
CVD [F]	0.090	0.082	0.093	0.079	0.100	0.072
Cancer [M]	0.135	0.103	0.126	0.116	0.096	0.093
Cancer [F]	0.106	0.100	0.079	0.093	0.099	0.094
CVD [Total]	0.133	0.135	0.143	0.151	0.146	0.139
Cancer [Total]	0.164	0.134	0.152	0.144	0.137	0.135
Bivariate Moran's I						
Cancer & CVD [M]	0.130	0.117	0.128	0.129	0.118	0.135
Cancer & CVD [F]	0.105	0.087	0.093	0.089	0.098	0.088
Cancer & CVD [Total]	0.145	0.137	0.134	0.137	0.141	0.136

Table 1 shows that the correlation between CVD and cancer mortality rates was averagely positive – 51% (mean = 50.9, s.d = 0.08), in general, for the six-year study period. The correlation coefficients ($p<0.05$) showed no apparent temporal trends. When spatial effects were considered in the univariate setting (Moran's I analyses), the results showed that the index was small and positive – indicating that the influence of neighbourhood mortality rates on the SAMS-observed mortatilty rates was weak though significant (mean = 0.123, s.d = 0.03). In the bivariate Moran's I, also weak but significant (mean = 0.120, s.d = 0.02) influence of neighbourhood CVD mortality rates on the SAMS-level observed cancer mortality rates, and vice versa, was observed. In both spatial cases, there were no apparent temporal trends in the Moran's indices.

To ensure the robustness of the obtained indices (for both univariate and bivariate), 9,999 randomizations were allowed used for this analysis.

4.3 Cluster detection

The analyses in Table 1 being global in nature, do not show where, in Sweden, these relationships were significant and where they were not. To investigate for simultaneous clustering of CVD and cancer crude mortality rates, the mortality-clusters analysed using spatial scan statistics (SaTScan – discrete Poisson) were spatially overlaid to identify overlapping areas. The results of this overlay are shown in Figure 4 (2010) and Figure 5 (2015).

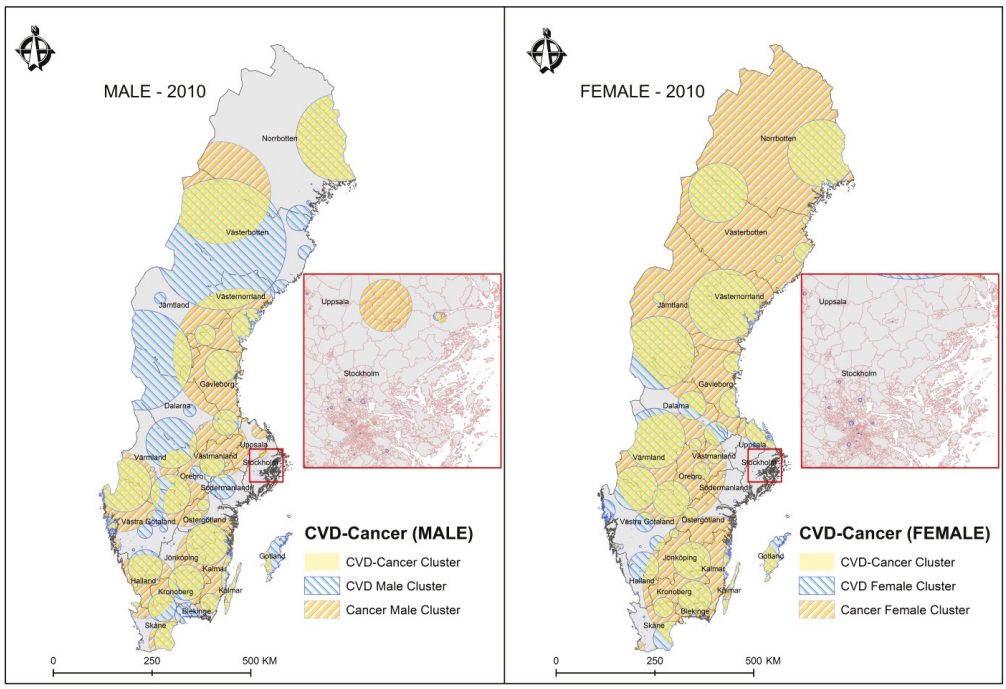


Figure 4: Spatial clustering of CVD and cancer crude mortality rate across Sweden, 2010 (Inset: Stockholm)

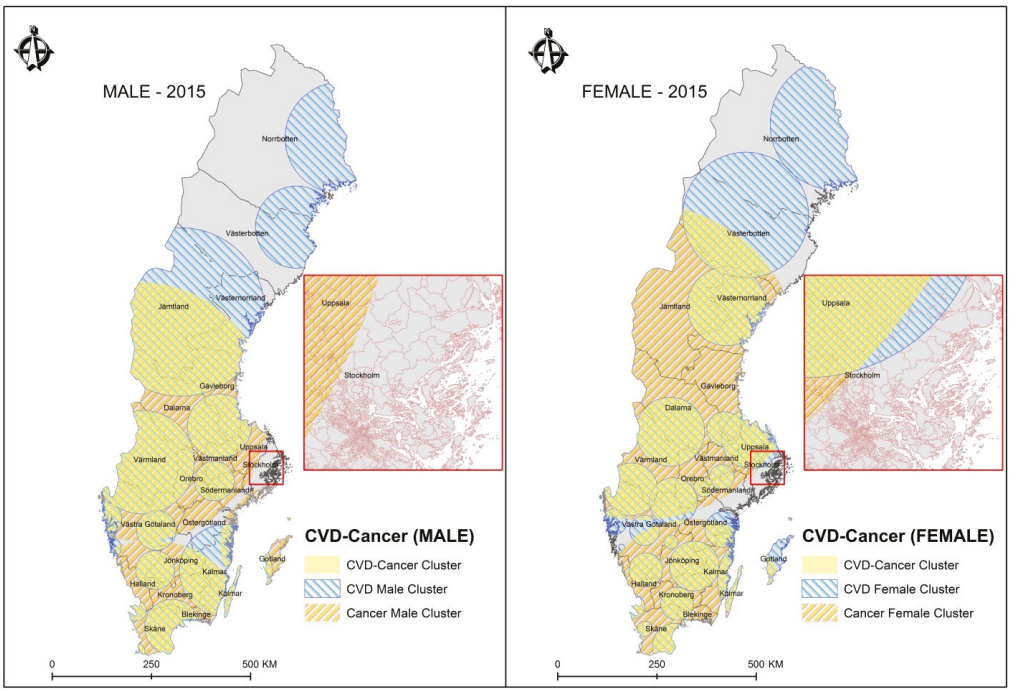


Figure 5: Spatial clustering of CVD and cancer crude mortality rate across Sweden, 2015 (Inset: Stockholm)

Figure 4 and Figure 5 show that CVD and cancer clusters overlap most of the time. It also shows that most of the CVD mortality rate clusters tend to fall within the cancer mortality rate clusters as cancer clusters appear to be larger than CVD clusters. This was common in the central and southern parts of Sweden and points to different spatial clustering scales for the two mortalities. For example, the percentage overlap (in terms of area) between cancer and CVD for 2010 shows that for males, there was a 57.1% overlap while for women, there was a 90.4% overlap. This was calculated based on the common areas when compared with the area of CVD clusters. We noted minimum clustering around Stockholm area, most common clustering was around the areas in the south, mid-central, upper left and right for men while the south, mid-central and the greater part of the north had common CVD and cancer clustering.

4.4 Local Entropy map analysis

Figures 4 and Figure 5 are to the effect that there exists spatial co-occurrence between crude mortality rate clusters of CVD and cancer, and the two mortality rates have somewhat different clustering scales. However, they do not show the nature of the relationship between these two causes of death. To investigate the nature of their simultaneous relationships, local entropy maps were generated and are shown in Figure 6. Whereas we analysed for local entropy using male/female stratification, we obtained unstable estimates throughout the years of analysis (when compared with the results from spatial scan statistics). We thus adopted the use of total

crude SAMS mortality rates. For brevity and given that the patterns are more or less similar, only the years 2010 and 2015 were visualized. The remaining years were summarized in Table 2.

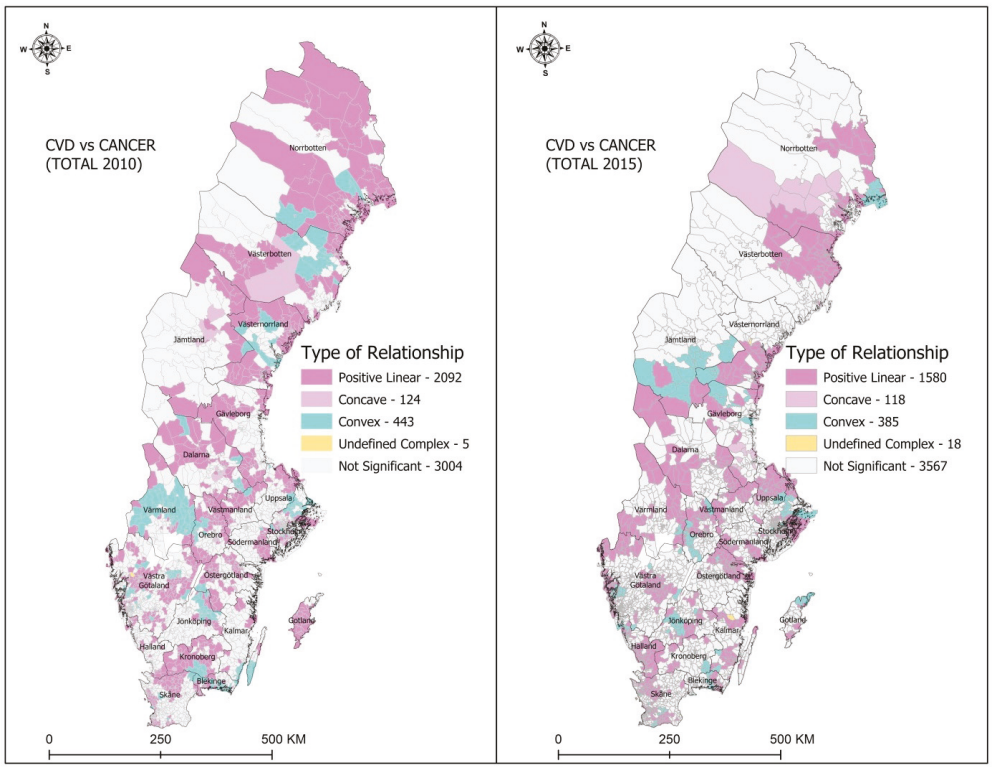


Figure 6: Local entropy map – relationship types between CVD and cancer crude mortality rate across Sweden (2010 and 2015)

Figure 6 shows local entropy maps for the years 2010 and 2015, using total mortality rates for CVD and cancer. These results largely reflect what Figure 4 and Figure 5 show, only that mortalities for males and females were combined to get the total in Figure 6. For example, we noticed that for both there were many associations and combined clusters in the mid and southern parts of Sweden, as well as the right northern parts. Visual inspection and interpretation of the legend show that more than half of Sweden indicated no significant relationship between the two mortalities. Also, in areas where there was a significant relationship, the majority of these relationships were positive linear. That said, other complex relationships like concave, convex, and undefined complex existed. These other non-linear relationships were scattered around the study area, with the convex one being the more pronounced. A convex relationship indicated that the dependent variable (CVD) changed by a convex curve as the explanatory variable (cancer) increased, resulting in an upward-arcing

curve (concave – results in a downward-arcing curve). The percentages for each relationship from 2010 to 2015 are provided by Table 2 (k = 50, α = 0.5).

Table 2: Percentage of the different relationships from LEM of Swedish elderly (2010—2015)

	2010	2011	2012	2013	2014	2015
Positive Linear (%)	37	34	31	27	26	28
Concave (%)	2	1	5	4	2	2
Convex (%)	8	5	6	3	5	7
Undefined Complex (%)	0	0	0	0	0	0
Not Significant (%)	53	59	58	66	67	63
Total	100	100	100	100	100	100

Table 2 shows that consistently, for more than half of Sweden the relationship between the two mortality rates was not significant. It indicates that whenever there was a significant relationship, this relationship was positively linear in 78% (±4%) of the occasions; it was convex in 14% (±3%); concave in 7% (±3%); and undefined complex in 1% (±0.4%) of the occasions. Therefore, and generally, it shows that for the areas where the relationship was established, it tended to be in the order positive linear, convex, concave, and lastly undefined complex

5 Discussion

5.1 Trends in mortality by sexes and age

Our results showed that for both cancer and CVD, males had higher mortality compared to their female counterparts. This is consistent with Sweden-specific CVD admission results by Rajabi et al. (2018) who observed that men had significantly higher admission rates in the time period 2000—2011. It can only be inferred that these higher rates account for the slightly higher mortality rates in males in our results. Similarly, in their study of all incidence primary cancer cases in the nationwide Swedish cancer registry for the period 1970—2014, Radkiewicz et al. (2017) observed significantly higher mortality in men for 27 out of 39 cancer sites when compared to women who had only 2 out of 39 sites. Unlike other studies, we observed that the trends reversed and female mortality superseded the men's for both CVD and cancer, past the age of 83 years and 87 years, respectively. The non-observance of the reversed trends with advanced age in previous studies could be due to the use of general (all-age) population compared to our study that is concerned with the elderly (65+ years).

5.2 Correlation and spatial patterns of crude mortality rates

Our exploratory study adopted several approaches to capture both the general global trends as well as the local patterns in the CVD and cancer data in terms of local clusters and local associations. Correlation analysis allowed for quantitative evaluation of the strength of the relationship between CVD and cancer mortality rates by reducing the information contained in n SAMS observations to a single number, falling into a normed interval (Filzmoser and Hron 2009). Spatial scan statistics, known for reasonable specificity and sensitivity (Gwitira et al.

2018), scaled-down the global patterns into local spatial clusters (Lawson et al. 2016) whose overlay showed areas of potential joint CVD and cancer mortality. More importantly, the local relationships within the joint mortality were distilled by entropy through LEM. The advantage of the entropy approach lies in its non-parametric nature – no prior specification of the underlying data model is required (Guo 2010), and no assumptions made on the type of relationship existing between the variables, consequently making it more applicable in such exploratory studies (Jin and Lu 2017). Moreover, to the best of our knowledge, this is the first study considering joint clustering and association of CVD and cancer in Swedish elderly therefore assumptions on the relationship nature would miss out on other non-conformal relationships established in this study.

Our results show that CVD and cancer mortality rates in the general Swedish elderly were averagely positively correlated (51%). This was further indicated by the small but significant global bivariate Moran's I which measured the influence of neighbourhood mortality of each disease to the observed mortality rate in the other (e.g. neighbourhood influence of cancer mortality rate on the observed CVD mortality rate at each SAMS). Additionally, our clustering results showed that CVD and Cancer mortality rate among the elderly Swedish population did not always cluster together in space with about 60% of the country, including areas known for high population density like Stockholm, experiencing no significant relationships for the study period 2010—2015. This could mean that for a significant portion of Sweden, different mechanisms exist between CVD and cancer mortality rates, pointing to the need to still consider and treat these diseases as different. This, perhaps, justifies why for long these two diseases have been treated as separate disease entities (Giza et al. 2017).

For the remaining 40% of Sweden, however, there were significant relationships between CVD and cancer mortality rates in the Swedish elderly. Our results from local entropy analysis showed that CVD and cancer mortality rates were positively related to each other. This positive relationship has been widely studied, and could be explained by the two causes of death having mechanistic overlaps (Weaver et al. 2013). These underlying mechanisms could be in the form of shared CVD risk factors like cigarette smoking, obesity, hypertension, diabetes, hyperlipidaemia, and physical inactivity (Giza et al. 2017, Koene et al. 2016) , becoming more pronounced in the old age (65+) of our current mortality study. Aside from the common risk factors, this positive relationship could be linked to cardiotoxic effects of cancer treatment in the form of radiotherapy, chemotherapy, and hormonal therapies (Moslehi 2016, Blaes and Shenoy 2019), as well as the inherent physical inactivity and weight gain are known to exacerbate CVD (Koene et al. 2016). However, given the exploratory nature of our study, our aim was to establish spatial co-occurrence, and not to establishing the underlying mechanisms between these two causes of death. What is new to existing knowledge is that the relationships between these two mortality rates are not always positive and linear, but complex too. These convex, concave and undefined complex relationships as we observed, were also observed by Guo (2010) and Jin and Lu (2017) in similar studies.

Unique to this study was that CVD and cancer mortality rates had different spatial clustering scales, with CVD clustering at a smaller scale. By spatially overlaying the clusters obtained from scan statistics, the differences in the clustering scales became apparent. Spatial scan statistics are known to have reasonable sensitivity and specificity when compared to other cluster detecting methods (Wand and Ramjee 2010, Gwitira et al. 2018). This difference in clustering, in essence, indicates that at a population scale, CVD mortality clusters exist as enclaves within cancer mortality cluster territories. This is consistent with the fact that different cancers have differing effects on CVD. For example, in a population-based study of CVD risk factors in survivors of 20 cancers, Strongman et al. (2019) found that whereas there was an increased CVD risk for breast and lung cancers, malignant melanoma, and non-Hodgkin lymphoma, they found a reduced CVD risk for prostate cancer and no association for other nine cancer, including bladder cancer. In relation to our result, it shows that CVD mortality enclaving in the background of bigger cancer mortality clusters could be highlighting areas with elevated cancer-induced or cancer-promoted CVD mortalities. These areas of joint mortality rate burden would benefit from coordinated joint management of both cancer and CVD, while still diseases, as proposed by the WHO and an increasing number of scholars (Koene et al. 2016, Coviello 2018, Moslehi 2016, Strongman et al. 2019).

5.3 Limitations

Like many empirical studies, the analysis in this study has some limitations. Firstly, the scale of analysis was limited to the SAMS level as dictated by the population data. SAMS is the official level of population-level statistics dissemination in Sweden. As such, the results obtained were scale-dependent and would slightly change if the analysis was to be done at another scale. This scale-dependence problem is better known as the Modifiable Area Unit Problem (MAUP) (Nelson and Brewer 2017). Secondly, being an ecological study, no personal level data was used and therefore the observed patterns could not be distilled to their potential risk factors. Also, the observed patterns at the ecological scale could be different if observed at individual scales. However, given that the study set out to analyse the spatial clustering patterns and relationships between CVD and cancer mortality rates, we are convinced that this study achieved its set objectives.

6 Conclusions

The relationships between cancer and CVD crude mortality rates in the elderly and their spatial variation across Sweden were explored in both males and females for the period 2010—2015. Our clustering and overlay analysis showed that the two causes of death have different spatial clustering scales with CVD's smaller mortality clusters existing as enclaves within the bigger cancer mortality clusters. Additionally, joint entropy analysis showed that cancer and CVD mortality rates were not always related pointing to the need for public health planners to consider cancer and CVD as separate entities in some areas. However, where they were significantly related, the relationship tended to be positive linear, pointing to the need for public

health planners to study and consider common mechanisms driving the double mortality rates in those areas. By spatially designating areas with elevated CVD and cancer joint mortality rates, our study provides a stepping stone to the CVD and cancer healthcare community by informing the authorities of areas where simultaneous and better management of both diseases could reduce eventual mortalities, especially among the elderly in Sweden. Finally, our study provides an initial stage upon which more focused epidemiologic studies can be made to establish the underlying mechanisms and possible place-specific risk factors behind areas with elevated CVD-cancer mortality rates within the Swedish elderly.

Acknowledgement

We would wish to appreciate the Swedish International Development Cooperation Agency (SIDA), the Swedish Research Council and the Swedish Heart-Lung foundation for providing the funding for this research.

References

Abdel-Qadir, H., P. C. Austin, D. S. Lee, E. Amir, J. V. Tu, P. Thavendiranathan, K. Fung & G. M. Anderson (2017) A population-based study of cardiovascular mortality following early-stage breast cancer. *JAMA cardiology,* 2**,** 88-93.

Anselin, L. & S. Rey. 2014. *Modern Spatial Econometrics in Practice*. Chicago, USA: Geoda Press LLC.

Anselin, L., I. Syabri & Y. Kho (2006) GeoDa: an introduction to spatial data analysis. *Geographical analysis,* 38**,** 5-22.

Armenian, S. H., L. Xu, B. Ky, C. Sun, L. T. Farol, S. K. Pal, P. S. Douglas, S. Bhatia & C. Chao (2016) Cardiovascular disease among survivors of adult-onset cancer: a community-based retrospective cohort study. *Journal of Clinical Oncology,* 34**,** 1122-1130.

Aturinde, A., M. Farnaghi, P. Pilesjö & A. Mansourian (2019) Spatial analysis of HIV-TB co-clustering in Uganda. *BMC infectious diseases,* 19**,** 612.

Benjamini, Y. & Y. Hochberg (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological),* 57**,** 289-300.

Blaes, A. H. & C. Shenoy (2019) Is it time to include cancer in cardiovascular risk prediction tools? *The Lancet,* 394**,** 986-988.

Chen, J., R. E. Roth, A. T. Naito, E. J. Lengerich & A. M. MacEachren (2008) Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of US cervical cancer mortality. *International journal of health geographics,* 7**,** 57.

Cliff, A. D. & K. Ord (1970) Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography,* 46**,** 269-292.

Coviello, J. S. (2018) Cardiovascular and Cancer Risk: The Role of Cardio-oncology. *Journal of the advanced practitioner in oncology,* 9**,** 160.

Cuadros, D. F., S. F. Awad & L. J. Abu-Raddad (2013) Mapping HIV clustering: a strategy for identifying populations at high risk ofHIV infection in sub-Saharan Africa. *International journal of health geographics,* 12**,** 28.

Deng, T., Y. Huang, S. Yu, J. Gu, C. Huang, G. Xiao & Y. Hao (2013) Spatial-temporal clusters and risk factors of hand, foot, and mouth disease at the district level in Guangdong Province, China. *PloS one,* 8**,** e56943.

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

How Local Bivariate Relationships works - ArcGIS Pro. 2019 Esri Inc.

Filzmoser, P. & K. Hron (2009) Correlation analysis for compositional data. *Mathematical Geosciences,* 41**,** 905.

Giza, D. E., G. Iliescu, S. Hassan, K. Marmagkiolis & C. Iliescu (2017) Cancer as a risk factor for cardiovascular disease. *Current oncology reports,* 19**,** 39.

Gray, R. M. 2011. *Entropy and information theory*. New York: Springer Science & Business Media.

Guo, D. (2010) Local entropy map: A nonparametric approach to detecting spatially varying multivariate relationships. *International Journal of Geographical Information Science,* 24**,** 1367-1389.

Gwitira, I., A. Murwira, J. Mberikunashe & M. Masocha (2018) Spatial overlaps in the distribution of HIV/AIDS and malaria in Zimbabwe. *BMC infectious diseases,* 18**,** 598.

Jin, H. & Y. Lu (2017) The relationship between obesity and socioeconomic status among Texas school children and its spatial variation. *Applied geography,* 79**,** 143-152.

Koene, R. J., A. E. Prizment, A. Blaes & S. H. Konety (2016) Shared risk factors in cardiovascular disease and cancer. *Circulation,* 133**,** 1104-1114.

Kulldorff, M. (1997) A spatial scan statistic. *Communications in Statistics-Theory and methods,* 26**,** 1481-1496.

Lawson, A. B., S. Banerjee, R. P. Haining & M. D. Ugarte. 2016. *Handbook of spatial epidemiology*. Boca Raton, FL.: CRC Press.

Mahase, E. (2019) Cancer overtakes CVD to become leading cause of death in high income countries. *BMJ: British Medical Journal (Online),* 366.

Miller, K. D., L. Nogueira, A. B. Mariotto, J. H. Rowland, K. R. Yabroff, C. M. Alfano, A. Jemal, J. L. Kramer & R. L. Siegel (2019) Cancer treatment and survivorship statistics, 2019. *CA: A Cancer Journal for Clinicians,* 69**,** 363-385.

Moslehi, J. J. (2016) Cardiovascular toxic effects of targeted cancer therapies. *New England Journal of Medicine,* 375**,** 1457-1467.

Naimi, B., N. A. Hamm, T. A. Groen, A. K. Skidmore, A. G. Toxopeus & S. Alibakhshi (2019) ELSA: Entropy-based local indicator of spatial association. *Spatial statistics,* 29**,** 66-88.

Nelson, J. K. & C. A. Brewer (2017) Evaluating data stability in aggregation structures across spatial scales: revisiting the modifiable areal unit problem. *Cartography and Geographic Information Science,* 44**,** 35-50.

Openshaw, S., M. Charlton, A. W. Craft & J. Birch (1988) Investigation of leukaemia clusters by use of a geographical analysis machine. *The Lancet,* 331**,** 272-273.

Radkiewicz, C., A. L. Johansson, P. W. Dickman, M. Lambe & G. Edgren (2017) Sex differences in cancer risk and survival: A Swedish cohort study. *European Journal of Cancer,* 84**,** 130-140.

Rajabi, M., A. Mansourian, P. Pilesjö, D. O. Åström, K. Cederin & K. Sundquist (2018) Exploring spatial patterns of cardiovascular disease in Sweden between 2000 and 2010. *Scandinavian journal of public health***,** 1403494818780845.

Ricotta, C. & M. Anand (2006) Spatial complexity of ecological communities: Bridging the gap between probabilistic and non-probabilistic uncertainty measures. *Ecological Modelling,* 197**,** 59-66.

Rugbjerg, K., L. Mellemkjær, J. D. Boice, L. Køber, M. Ewertz & J. H. Olsen (2014) Cardiovascular disease in survivors of adolescent and young adult cancer: a Danish cohort study, 1943–2009. *JNCI: Journal of the National Cancer Institute,* 106.

Shannon, C. E. (1948) The mathematical theory of communication. *The Bell System Technical Journal,* 27**,** 379-423.

Sherman, R. L., K. A. Henry, S. L. Tannenbaum, D. J. Feaster, E. Kobetz & D. J. Lee (2014) Peer Reviewed: Applying Spatial Analysis Tools in Public Health: An Example Using SaTScan to Detect Geographic Targets for Colorectal Cancer Screening Interventions. *Preventing chronic disease,* 11.

Smith, C., R. Lessells, A. Grant, K. Herbst & F. Tanser (2018) Spatial clustering of drug-resistant tuberculosis in Hlabisa subdistrict, KwaZulu-Natal, 2011–2015. *The international journal of tuberculosis and lung disease,* 22**,** 287-293.

Smith, C. M., H. Maguire, C. Anderson, N. Macdonald & A. C. Hayward (2017) Multiple large clusters of tuberculosis in London: a cross-sectional analysis of molecular and spatial data. *ERJ Open Research,* 3**,** 00098-2016.

Souris, M. & L. Bichaud (2011) Statistical methods for bivariate spatial analysis in marked points. Examples in spatial epidemiology. *Spatial and spatio-temporal epidemiology,* 2**,** 227-234.

Strongman, H., S. Gadd, A. Matthews, K. E. Mansfield, S. Stanway, A. R. Lyon, I. dos-Santos-Silva, L. Smeeth & K. Bhaskaran (2019) Medium and long-term risks of specific cardiovascular diseases in survivors of 20 adult cancers: a population-based cohort study using multiple linked UK electronic health records databases. *The Lancet,* 394**,** 1041-1054.

Sundquist, K., M. Winkleby, H. Ahlén & S.-E. Johansson (2004) Neighborhood socioeconomic environment and incidence of coronary heart disease: a follow-up study of 25,319 women and men in Sweden. *American journal of epidemiology,* 159**,** 655-662.

Sundquist, K. & M. Yang (2007) Linking social capital and self-rated health: a multilevel analysis of 11,175 men and women in Sweden. *Health & place,* 13**,** 324-334.

Tang, X., A. Geater, E. McNeil, Q. Deng, A. Dong & G. Zhong (2017) Spatial, temporal and spatio-temporal clusters of measles incidence at the county level in Guangxi, China during 2004–2014: flexibly shaped scan statistics. *BMC infectious diseases,* 17**,** 243.

Tiwari, N., C. Adhikari, A. Tewari & V. Kandpal (2006) Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic. *International journal of health geographics,* 5**,** 33.

Tobler, W. R. (1970) A computer movie simulating urban growth in the Detroit region. *Economic geography,* 46**,** 234-240.

Wand, H. & G. Ramjee (2010) Targeting the hotspots: investigating spatial and demographic variations in HIV infection in small communities in South Africa. *Journal of the International AIDS Society,* 13**,** 41.

Weaver, K. E., R. E. Foraker, C. M. Alfano, J. H. Rowland, N. K. Arora, K. M. Bellizzi, A. S. Hamilton, I. Oakley-Girvan, G. Keel & N. M. Aziz (2013) Cardiovascular risk factors among long-term survivors of breast, prostate, colorectal, and gynecologic cancers: a gap in survivorship care? *Journal of Cancer Survivorship,* 7**,** 253-261.

WHO. 2019a. Cancer. World Health Organisation.

---. 2019b. Cardiovascular disease. World Health Organisation.

Zhang, Q., W. Liu, W. Ma, Y. Shi, Y. Wu, Y. Li, S. Liang, Y. Zhu & M. Zhou (2017) Spatiotemporal epidemiology of scarlet fever in Jiangsu Province, China, 2005–2015. *BMC infectious diseases,* 17**,** 596.

Avhandlingar från Institutionen för naturgeografi och ekosystemanalys (INES), Lunds universitet

---

Dissertations from Department of Physical Geography and Ecosystem Science, University of Lund

---

Martin Sjöström, 2012: Satellite remote sensing of primary production in semi-arid Africa.

Zhenlin Yang, 2012: Small-scale climate variability and its ecosystem impacts in the sub-Arctic.

Ara Toomanian, 2012: Methods to improve and evaluate spatial data infrastructures.

Michal Heliasz, 2012: Spatial and temporal dynamics of subarctic birch forest carbon exchange.

Abdulghani Hasan, 2012: Spatially distributed hydrological modelling: wetness derived from digital elevation models to estimate peatland carbon.

Julia Bosiö, 2013: A green future with thawing permafrost mires?: a study of climate-vegetation interactions in European subarctic peatlands. (Lic.)

Anders Ahlström, 2013: Terrestrial ecosystem interactions with global climate and socio-economics.

Kerstin Baumanns, 2013: Drivers of global land use change: are increasing demands for food and bioenergy offset by technological change and yield increase? (Lic.)

Yengoh Genesis Tambang, 2013: Explaining agricultural yield gaps in Cameroon.

Jörgen Olofsson, 2013: The Earth: climate and anthropogenic interactions in a long time perspective.

David Wårlind, 2013: The role of carbon-nitrogen interactions for terrestrial ecosystem dynamics under global change: a modelling perspective.

Elin Sundqvist, 2014: Methane exchange in a boreal forest: the role of soils, vegetation and forest management.

Julie Mari Falk, 2014: Plant-soil-herbivore interactions in a high Arctic wetland: feedbacks to the carbon cycle.

Finn Hedefalk, 2014: Life histories across space and time: methods for including geographic factors on the micro-level in longitudinal demographic research. (Lic.)

Sadegh Jamali, 2014: Analyzing vegetation trends with sensor data from earth observation satellites.

Cecilia Olsson, 2014: Tree phenology modelling in the boreal and temperate climate zones : timing of spring and autumn events.

Jing Tang, 2014: Linking distributed hydrological processes with ecosystem vegetation dynamics and carbon cycling: modelling studies in a subarctic catchment of northern Sweden.

Wenxin Zhang, 2015: The role of biogeophysical feedbacks and their impacts in the arctic and boreal climate system.

Lina Eklund, 2015: "No Friends but the Mountains": understanding population mobility and land dynamics in Iraqi Kurdistan.

Stefan Olin, 2015: Ecosystems in the Anthropocene: the role of cropland management for carbon and nitrogen cycle processes.

Thomas Möckel, 2015: Hyperspectral and multispectral remote sensing for mapping grassland vegetation.

Hongxiao Jin, 2015: Remote sensing phenology at European northern latitudes: from ground spectral towers to satellites.

Bakhtiyor Pulatov, 2015: Potential impact of climate change on European agriculture: a case study of potato and Colorado potato beetle.

Christian Stiegler, 2016: Surface energy exchange and land-atmosphere interactions of Arctic and subarctic tundra ecosystems under climate change.

Per-Ola Olsson, 2016: Monitoring insect defoliation in forests with time-series of satellite based remote sensing data: near real-time methods and impact on the carbon balance.

Jonas Dalmayne, 2016: Monitoring biodiversity in cultural landscapes: development of remote sensing- and GIS-based methods.

Balathandayuthabani Panneer Selvam, 2016: Reactive dissolved organic carbon dynamics in a changing environment: experimental evidence from soil and water.

Kerstin Engström, 2016: Pathways to future cropland: assessing uncertainties in socio-economic processes by applying a global land-use model.

Finn Hedefalk, 2016: Life paths through space and time: adding the micro-level geographic context to longitudinal historical demographic research.

Ehsan Abdolmajidi, 2016: Modeling and improving Spatial Data Infrastructure (SDI).

Giuliana Zanchi, 2016: Modelling nutrient transport from forest ecosystems to surface waters.

Florian Sallaba, 2016: Biophysical and human controls of land productivity under global change: development and demonstration of parsimonious modelling techniques.

Norbert Pirk, 2017: Tundra meets atmosphere: seasonal dynamics of trace gas exchange in the High Arctic.

Minchao Wu, 2017: Land-atmosphere interactions and regional Earth system dynamics due to natural and anthropogenic vegetation changes.

Niklas Boke-Olén, 2017: Global savannah phenology: integrating earth observation, ecosystem modelling, and PhenoCams.

Abdulhakim M. Abdi, 2017: Primary production in African drylands: quantifying supply and demand using earth observation and socio-ecological data.

Nitin Chaudhary, 2017: Peatland dynamics in response to past and potential future climate change.

Ylva van Meeningen, 2017: Is genetic diversity more important for terpene emissions than latitudinal adaptation?: using genetically identical trees to better understand emission fluctuations across a European latitudinal gradient.

Patrik Vestin, 2017: Effects of forest management on greenhouse gas fluxes in a boreal forest.

Mohammadreza Rajabi, 2017: Spatial modeling and simulation for disease surveillance.

Jan Blanke, 2018: European ecosystems on a changing planet: integrating climate change and land-use intensity data.

Min Wang, 2018: Characteristics of BVOC emissions from a Swedish boreal forest: using chambers to capture biogenic volatile organic compounds (BVOCs) from trees and forest floor.

Wilhelm Dubber, 2018: Natural and social dimensions of forest carbon accounting.

Emma Johansson, 2018: Large-Scale Land Acquisitions as a Driver of Socio-Environmental Change: From the Pixel to the Globe.

Helen Eriksson, 2018: Harmonisation of geographic data: between geographic levels, hierarcic structures and over time. (Lic.)

Zhendong Wu, 2018: Modelling the terrestrial carbon cycle: drivers, benchmarks, and model-data fusion.

Zhanzhang Cai, 2019: Vegetation observation in the big data era: Sentinel-2 data for mapping the seasonality of land vegetation.

Fabien Rizinjirabake, 2019: Dissolved organic carbon in tropical watersheds: linking field observation and ecohydrological modelling.

Jeppe Ågård Kristensen, 2019: Biogeochemistry in Subarctic birch forests: perspectives on insect herbivory.

Yanzi Yan, 2019: The role of hydrological cycle in forest ecosystems: flow path, nutrient cycling and water-carbon interaction.

George Oriangi, 2019: Urban resilience to climate change shocks and stresses in Mbale municipality in Eastern Uganda.

Alex Paulo Lubida, 2019: Investigating spatial data infrastructure planning in Tanzania using system modelling and social concepts.

Weiming Huang, 2020: Geospatial data and knowledge on the Web: knowledge-based geospatial data integration and visualisation with Semantic Web technologies.

Oskar Löfgren, 2020: Remote sensing of grassland communities: integrated effects of soil nutrients and habitat age.

Altaaf Mechiche-Alami, 2020: Food security in a changing climate: the role of cropland intensification and land acquisitions across Africa.

Helen Eriksson, 2020: Harmonisation of 3D geodata – a prerequisite for a digital information flow for applications in the planning and building sector.

Augustus Aturinde, 2020: GIS and Health: Enhancing Disease Surveillance and Intervention through Spatial Epidemiology.