University of Tartu
Faculty of Science and Technology
Institute of Ecology and Earth Sciences
Department of Geography

Master's thesis in Geoinformatics for Urbanised Society (30 ECTS)

Cause Specific Death Rates and Their Geographic Aspect in Estonia

Mika Kuno

Supervisor: Evelyn Uuema, Ph.D.,

Katrin Lang, Ph.D.,

Hans Orru, Ph.D.

Tartu 2021

Abstrakt

Pealkiri: „Peamised surmapõhjused ja nende geograafiline aspekt Eestis ”

Suurem osa Eesti suremuse uuringutest on näidanud ainult staatilisi kaarte ega pole andnud dünaamilist vaadet elanikkonna suremusest ajalis-ruumilisest vaatest lähtuvalt. Suremuse hindamine dünaamilises ajaperspektiivis võiks aidata rahvatervise valdkonnas paremini otsuseid langetada. Antud uurimustöö põhieesmärk oli uurida aastatel 2005–2019 kolme peamise surmapõhjuse (vereringeelundite haigused, pahaloomulised kasvajad ning õnnetusjuhtumid, mürgistused ja traumad) ajalis-ruumilist mustrit Eestis. Suremuse geograafiliste mustrite uurimiseks rakendati ruumilise autokorrelatsiooni globaalset ja kohalikku analüüsi ning Kulldorffi skanneeringu statistikat. Statistiliselt olulist ruumilist heterogeensust täheldati surmade puhul, mis saabusid õnnetusjuhtumite, mürgistuste ja traumade tagajärjel, kuid mitte vereringeelundite haiguste ega pahaloomuliste kasvajate tõttu saabunud surmade puhul. Lisaks leiti kõigi haiguste puhul statistiliselt olulised nn kuum- ja külmpunkid. Uurimisperioodil tuvastati vähemalt kaks ruumilist suremuse kogumit erinevate surmapõhjuste ja kindlate ajaliste rühmade vahel. Selle uuringu tulemused võivad olla aluseks suremuse edasistele uuringutele Eestis.

Märksõnad: ruumianalüüs, ajalis-ruumiline analüüs, haiguste kogumid, suremus

CERCSi kood: B680 – Rahvatervishoid, epidemioloogia; S230 – Sotsiaalne geograafia

Abstract

Title: „Cause specific death rates and their geographic aspect in Estonia”

Most of the mortality research in Estonia has only shown static maps and has not provided a dynamic view of the population’s mortality status from a spatiotemporal point of view. The evaluation of mortality in a dynamic time perspective could improve decision making in public health. The main aim of the study was to investigate spatiotemporal pattern of three major causes of death (circulatory diseases, malignant neoplasms, and injury and poisoning) in Estonia from 2005 to 2019. Global and local spatial autocorrelation analyses, and Kulldorff’s scan statistic were applied to explore the geographical patterns of mortality. Overall, a statistically significant spatial heterogeneity was observed in the mortality due to injury and poisoning, but not due to circulatory diseases nor malignant neoplasms. Moreover, statistically significant clusters of hot spots and/or cold spots were identified for all diseases. There were at least two spatial clusters of mortality identified in different diseases and specific time clustering during the study period. The findings of this study can serve as a basis for further studies of mortality in Estonia.

Keywords: spatial analysis, spatiotemporal analysis, disease clusters, mortality.

CERCS code: B680 – Public health, epidemiology; S230 – Social geography

Table of contents

1. Introduction

Mortality statistics are a relevant source of information about population health — they allow to analyse current demographic conditions, suggest trends and patterns of disease risk in specific communities over time, as well as determine the prospects of potential changes in mortality conditions of the future. Geographic representation of mortality data is a valuable public health tool for formulating hypotheses about the aetiology of diseases, finding possible risk factors for disease, and identifying specific locations where changes in health policy need to be made (Gundersen, 2000).

Spatial epidemiology is an analysis of geographic variations in disease with respect to demographic, environmental, behavioural, socio-economic, genetic, and infectious risk factors (Graham et al., 2004). Geographic Information Systems (GIS) is an important asset in the field of spatial epidemiology. GIS has been widely used to analyse the spatiotemporal characteristics of both communicable diseases (K. Rahu et al., 2019) and non-communicable diseases (NCDs) (Almendra and Santana, 2020; Torres-Roman et al., 2019).

Advances in GIS and modern statistical methodology together with the availability of high-resolution, geographically referenced health databases enable to address of both tasks, i.e., inference on the geographical distribution of the disease and its prediction at new locations. For instance, global autocorrelation is used to identify the clusters of disease and local autocorrelation is used to examine regional patterns and ascertain the exact clustering location (Torres-Roman et al., 2019). The spatiotemporal scan is used to detect diseases in time and space, verify the random distribution of the disease in time and space, and ascertain the number of cases in a region (Almendra and Santana, 2020). The spatiotemporal investigation is one of the current frontiers in the evolution of GIS, which could identify key nodes in the spread of infection, defined by time spent at unique locations, can help us understand past infectious events and predict future developments (Bhunia and Shit, 2019; Lawson et al., 2016).

Numerous researchers have studied the epidemiology of different diseases in Estonia, however, several of the recent studies of cause-specific mortality have generally focused on only a single cause and/or selected causes of death, such as cancers (Ojamaa, 2019; Ojamaa et al., 2019; Pärn et al., 2019), cardiovascular diseases (Schneider et al., 2020), and neurodegenerative disorder like Parkinson's disease (Kadastik-Eerme et al., 2019). In addition to that, the studies are presenting only a static snapshot of disease distribution or analysing socio-demographic (Kaja Rahu et al., 2019), ethnicity (Reile and Leinsalu, 2017), education level, and place of residences differences (Lai and Leinsalu, 2015) – a comprehensive studies analysing the spatial and temporal distributions of mortality in

Estonia is limited. Therefore, the study aims to use a combination of GIS and statistical analysis to describe the temporal and spatial trends of mortality caused by various diseases in Estonia.

To achieve the aim of the research, the following research questions are stated about mortality in Estonia:

1. Where is the disease mortality significantly clustered?
2. How is mortality distributed spatially? What specific areas are with a high risk of disease mortality?
3. What is the spatiotemporal pattern in disease mortality in Estonia?

2. Literature Review

2.1 Spatial epidemiology

Disease is closely linked to the ideas of spatial and spatiotemporal proximity; thus, epidemiological analyses have to take both space and time into account. Spatial epidemiology is an analysis of geographic variations in a disease that incorporates the spatial perspective into the design and analysis of the distribution, determinants, and outcomes of all aspects of health (English, 1992; Kirby et al., 2017).

The beginning of spatial epidemiology dates back to the early 1800s when John Snow investigated the cholera epidemic in London by plotting the location of cholera cases using maps. In his study, Snow was able to see the spatial pattern of cholera cases in relation to potential risk factors, in this instance the locations of water pumps. He furthermore made solid use of statistics to demonstrate the connections between the quality of the source of water and cholera incidence and illustrate the spread of the disease and disease clusters around the Board Street water pump in London (Walter, 2001).

Centuries pass, technologies advance, and a growing number of tools are now available to address spatial questions. For instance, GIS can generate meaningful results by integrating data from many sources and displaying and analysing the information together.

Nowadays, statistical data analysis is the most consistent and recognised set of tools to evaluate spatial datasets. Statistical analysis with spatial or spatiotemporal datasets is called the science of spatial statistics or geostatistics. Hart first proposed the term geostatistics in 1954 to use in the mining industry (Hart, 1954). Coming to public health, nowadays, spatial statistics are commonly used to construct maps of spatial variation in disease, since it provides imperative information on how a disease is extended; which are the regions affected by the disease and forecast the next regions which have the higher prospect to be affected in order to control it (Elliot et al., 2000).

Application of spatial statistics and GIS has been widely used in the spatial epidemiological study to analyse the spatial or spatiotemporal characteristics of communicable and non-communicable diseases (Auchincloss et al., 2012; Singh et al., 2016).

There are a number of historical studies in communicable diseases. For instance, Tiwari et al. (2006) had analysed the spatial pattern of tuberculosis occurrence in 13 microscopic centres in Almora district, India. They discovered hotspots of tuberculosis in Almora district by performing spatial scan statistics created by Martin Kulldorff. The study concluded that the study has demonstrated the use of the health data, spatial scan statistic and GIS can provide necessary feedback about the statistically significant

hotspots of tuberculosis occurrence in the region to plan a more effective public health strategy. Gething et al. (2012) conducted a study to map the global endemicity of malaria. The study used data collected from 1985 to 2010 and applied a spatiotemporal Bayesian model-based geostatistical approach to estimate malaria endemicity. They identified regional differences of malaria endemicity and transmission, hotspots of the endemicity in South East Asia and small pockets of Amazonia, whereas the low endemic was found in Africa. A more recent study by Rau et al. (2020) applied Kulldorf's retrospective spatial scan statistic to cluster tick-Borne diseases in north-central Wisconsin from 2000–2016. They found that increasing non-Lyme tick-borne diseases prevalence in north-central Wisconsin varies in geographic space and time. The study concluded that public health practitioners could use their findings to increase public awareness and improve case detection.

In the area of non-communicable diseases, Erdogan (2009) performed a spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. In his case study, he employed four spatial autocorrelation techniques: two global spatial autocorrelation techniques, Moran's I and Geary's C, and two Local Indicator of the Spatial Association (LISA) techniques, local Moran's I and local Getis-Ord statistics. A similar analysis was performed by Vaz et al. (2020), where they applied local and global spatial autocorrelation to assess for self-harm by male, female in 140 Toronto neighbourhoods in the greater Toronto area, Canada. In a study about diseases of the circulatory system, Rheenen et al. (2015) integrate health care administrative datasets with GIS to identify the location of regional disparities in the spatial distribution of stroke occurrence and outcomes in Alberta, Canada. They found significant hotspots for ischemic stroke, transient ischemic attack, and in-hospital mortality, but no evidence of spatial association of stroke risk factor and ischemic stroke, transient ischemic attack, or in-hospital mortality were identified.

Moreover, GIS tools can also be used to display unmeasured characteristics like exposome and investigate relationships between disease clusters and underlying community characteristics such as socio-economic and environmental factors. For instance, Alene et al. (2017) employed global Moran's I statistic, Anselin Local Moran's I statistic, and the Getis-Ord statistic to tuberculosis mortality data of 20 districts in north-western Ethiopia to identify spatial autocorrelation of the disease over the study area and at the local scale. They found that paediatric tuberculosis was spatiotemporally clustered in north-western Ethiopia was associated with several socio-climatic factors, including urbanisation, internal migration, educational status, rainfall, and temperature. Another study by Scott et al. (2017) conducted a study about geospatial clustering of inflammatory breast cancer among US females. This study identified significant spatial clustering of county-based inflammatory breast cancer rates among

US females and their underlying social and economic environmental factors by performing global and local spatial autocorrelation analyses and statistical analysis.

Recent advancements in GIS and data availability create opportunities for researchers to improve on the traditional reporting of disease at a national or regional scale by studying variations in disease occurrence and its distribution at a local scale and allows to predict new locations of disease. Prospective space-time scan statistic developed by Martin Kulldorff can detect statistically significant active clusters of diseases for the present time period and detect emerging hotspots of disease (Chen et al., 2016; Jones et al., 2006; Kulldorff et al., 2005). In 2005, Kulldorff et al. propose a prospective space-time permutation scan statistic for the early detection of disease outbreaks with data about population-at-risk. They used data about daily analyses of hospital emergency department visits in New York City to evaluates the proposed method. The study utilised the possibility of using a new method for early disease detection surveillance systems, since the space-time permutation scan statistic has minimal assumptions about the time, geographical location, or size of the outbreak and adjusts for natural purely spatial and purely temporal variation. Jones et al. (2006) applied a prospective space-time scan statistic to Chicago's 2002 shigellosis surveillance data to evaluate active clusters of disease and detects twelve significant live clusters that all of the community clusters were located in the west-central part of the city with a temporal span of 28 days. More recently, Chen et al. (2016) designed an online analytical tool for local public health workers to prospectively detect ongoing dengue fever hotspots on a weekly basis at the village level by using a total of 57,516 confirmed cases of dengue fever in 2014 and 2015 that collected from the Taiwan Centres for Disease Control. They iteratively applied discrete Poisson model space–time scan statistics to detect the currently active cluster of dengue fever, dengue-epidemic area in each village of Tainan and Kaohsiung.

2.1.1 Statistical methods for spatial epidemiology

In spatial epidemiology, GIS and spatial statistical analysis provide unique tools to answer two fundamental questions; "Where and when diseases tend to occur?" and "Where does such pattern exist?".

The main spatial epidemiological inquiries are disease mapping, geographical correlation studies and cluster detection.

Disease mapping

Disease maps have been used since 1800s, a famous example, John Snow mapped the 1814 London cholera epidemic (Snow et al., 1936). Disease mapping provides information on disease occurrence

across a geographic space by displaying a visual summary of geographical risk and providing estimates of risk by area. Moreover, disease mapping helps to generate hypothesis before the formal examination via spatial modelling and risk assessment.

The most common approaches to visualise disease data is to use choropleth maps. A choropleth is a conventional method of summarising spatial data based on statistical, administrative, or enumeration units and can epitomise categorical and numerical data (Pfeiffer et al., 2008b). A choropleth map shows information by colouring each enumeration unit with colour, shade, dot, or hatch, providing an indication of the magnitude of the variable of interest.

From the classic study of cholera to recent analyses on emerging diseases, mapping disease pattern has been the fundamental session of spatial epidemiological analysis.

Geographical correlation studies

Geographical correlation study is to investigate geographic disparities across a population in exposure to environmental, socio-economic, and geographic variables in relation to health outcomes measured on a geographic scale – deal with the association between disease risk and the exposures (Elliott and Wartenberg, 2004). Such investigations help to study the aetiology of disease and may support supplementary research.

In recent years, there has been increasing interest in the field of exposure assessment; thus, new methods have been developed, including analytical, measurement, modelling, and statistical methods. The use of GIS together with remote sensing, personal sensors, and omics technologies is being explored to improve exposure assessment. GIS and geostatistical methods can be used to analyse the exposome of the disease. When spatial variation of a continuous attribute is often too irregular Geostatistical modelling by simple, smooth mathematical function can be applied. Kriging can be used to model the general trend of the deterministic data, the smaller local variations, and the uncorrelated noise.

Spatial clustering

Investigation of disease clusters and disease incidence/mortality near a point source is commonly known as clustering/cluster detection. Cluster methods are the most common tool for assessing spatial patterns; to determine whether the disease clusters are statistically significant and worthy of further investigation, or whether it is likely to be occurred by chance or is simply a reflection of the distribution of the population at risk. Cluster methods that evaluate whether diseases are close in space are also close in time, and vice versa is called space-time clustering.

There are two different methods for analysing clusters; Global and Local (Besag and Newell, 1991). The term clustering applies to global methods of cluster analysis, whereas cluster detection refers to local methods of cluster analysis. Global clustering methods are used to measure the correlation among neighbouring observations, whether the pattern expressed is clustered, dispersed or random. Local cluster detection methods define the locations and the characteristics of clusters, such as their location, size, and intensity.

One approach to define clusters in data is using autocorrelation statistics. Spatial autocorrelation statistics measure the degree of spatial similarity observed among neighbouring values of an attribute over a study area. This follows Tobler's first law of geography, which states that: "*Everything is related to everything else, but near things are more related than distant things*" (Tobler, 1970).

Moran's I statistic is one of the most used ways of measuring spatial autocorrelation. Moran's I is an inferential statistics, similar to Pearson's correlation coefficient, and measures spatial autocorrelation based on feature locations and feature values (Moran, 1950).

There are many examples of the use of Moran's I in the literature. Morris and Munasinghe (1994) use Moran's I to investigate the geographic distribution of hospital admissions for pneumonia, acute respiratory infections, asthma, and chronic obstructive pulmonary disease from the United States between 1984 and 1989 using medical admissions records. The analysis demonstrates significant clustering of hospital admission rates for all four respiratory diseases among the elderly, particularly in the southeast and the northern states. Moran's I is also used by Odoi et al. (2003) to assess autocorrelation of incidence rates of human giardiasis in Ontario, Canada, and Khamis and EL-Refae (2012) used to investigate spatial autocorrelation of cases of chronic diseases in Iraq in 2007. Nyari et al. (2013) use Moran's I to analyse clustering of childhood acute leukaemia in Hungary between 1981 and 2000.

Moreover, Anselin Local Moran's I and Getis-Ord Gi* spatial statistic are used to assess the spatial clustering for high risk (hotspots) and low risk (coldspots) as well as outliers. The local Moran test measures local spatial autocorrelation and clustering within the small areas that together comprise a study area by decomposing Moran's I statistic into contributions for each area within a study area (Anselin, 1995). Getis-Ord Gi* spatial statistics identifies clusters by observing at each feature within the framework of neighbouring features. A hotspot may be exhibited if the value for a feature is high and the values for adjacent features are also high.

There are many examples of using Anselin Local Moran's I and Getis-Ord Gi* statistic to identify hotspots of disease, Tsai et al. (2009) uses these analyses to determine the location of hotspots as well

as spatial patterns of 20 leading causes of death in Taiwan in 2006 and found that the prevalence of tuberculosis in Taiwan is closely related to the location of aboriginal townships. Prasannakumar et al. (2011) The patterns of localisation and distribution of road traffic hotspots in India was assessed using Moran's I method of spatial autocorrelation, Getis-Ord Gi* statistics, demonstrating that the traffic accidents are clustered in nature. Marotta et al. (2019) examined spatial patterns of overdose deaths in New York State between 2013 and 2015. The analysis suggests that the rates of opioid overdose deaths are clustered in New York.

Scan statistics is another approach commonly used to identify clusters in data. Scan statistics or window approach is first proposed by Joseph Naus in the 1960s to applied in epidemiology, public health and astronomy (Naus, 1964). For instance, Paneth et al. (1984) used data from Monmouth Medical Center in New Jersey, the United States, in October and November 1977 and applied scan statistic to investigated a temporal cluster of left-sided congenital heart disease.

Martin Kulldorff employed a likelihood ratio test within a window placed over the study area to find the most likely cluster (Kulldorff and Nagarwalla, 1995) and extended the idea of scan statistics to multidimensional setting and varying window sizes in 1997 to apply in epidemiological analyses (Kulldorff, 1997). In Kulldorff's scan statistics, a series of circular window of varying size is constructed for each specified location, and the size of each window is set to increase continuously from zero until some fixed percentage of the total population is included. Each window absorbs the nearest neighbouring locations that fall inside it, and the maximum ratio is found after windows of different sizes have scanned the study area.

Kulldorff's spatial scan statistics has been widely used to analyse cluster of various disease, for instance, Kulldorff et al. (1997) detected clusters of high breast cancer mortality in Northeast of New York, the United States, by utilising scan statistics to demographic data and age-specific breast cancer mortality rates for 11 north-eastern states, the District of Columbia for 1988–1992. Green et al. (2003) used scan statistics to identified geographic variability of diabetes mellitus prevalence in the City of Winnipeg, Manitoba in Canada in 1998, demonstrating significant clustering and small-area variations in the prevalence of diabetes mellitus in the City of Winnipeg. Brooker et al. (2004) investigated the spatial distribution of clinical malaria during an epidemic and investigate putative risk factors, using household surveys, household geographical location from geographical positioning system and landcover types determined using high spatial resolution satellite sensor data. They identified significant spatial clusters of malaria cases using the spatial scan statistic, and the risk of malaria was higher in children with underweight, lived at lower altitudes, and low availability of drugs.

Clearly, within the epidemiological study, analysis has to take both spatial and temporal component. Kulldorff has also developed a method to detect clusters in space and time by extending the idea of a two-dimensional circular window to a three-dimensional cylinder passing through time (Kulldorff et al., 1998).

Hjalmars et al. (1999) use the space-time scan statistic to investigate clustering of childhood malignant brain tumours in Sweden between 1973-1992 and analyse all primary paediatric brain tumours derived diagnosed and reported in all relevant Swedish hospitals and physicians, neoplastic diseases registered in the Swedish Cancer Registry. They found an increase in rates of these cancers during the period 1973 to 1992, but no clustering in space or time were identified. Space-time scan statistic is also used by Ge et al. (2016) to investigate space-time clustering of tuberculosis cases in China, demonstrating an increase in rates of these cluster of tuberculosis from early spring 2010 to the end of 2011. Other recent examples of the space-time scan statistic include explorations of clustering in vector-borne disease causes severe psychological in North-east of Iran (Mollalo et al., 2015), offences and police recorded data over Stockholm, Sweden (Uittenbogaard and Ceccato, 2012), bovine spongiform encephalopathy in cattle in Hokkaido, Japan (Kadohira et al., 2008).

2.1.2 Use of spatial analysis in the epidemiological study

Researchers have implemented various methods to determine how the disease spreads and how the spatial pattern of the disease is changing. As one of the three broad categories of spatial epidemiological enquiry, spatial cluster methods are the most common tool for assessing non-random spatial patterns of disease.

The most commonly applied method for aggregated data to test for spatial autocorrelation is Moran's I (Moran 1950); the LISA and Getis Ord statistics were widely applied to identify clusters or hotspots for aggregated data (Auchincloss et al., 2012; Kulldorff et al., 2006; Singh et al., 2016). Recent research by Baptista and Queiroz (2019) uses global and local spatial autocorrelation methods to identify recent cardiovascular disease mortality patterns in Brazil. The result successfully demonstrated the dynamics of the health transition in Brazil in a recent period of time and its impacts on public health policies.

In a cancer study, Al-Ahmadi and Al-Zahrani (2013) analysed population-based records of cancers diagnosed between 1998 and 2004 to investigating non-random incidence patterns of common cancers in Saudi Arabia by employing the global Moran's I and Anselin's local Moran's I statistics. The result shows that the male lung cancer and female breast cancer exhibited positive spatial autocorrelation and small significant clusters of lung cancer, prostate cancer, and Hodgkin's disease among males in

the Eastern region and significant clusters of thyroid cancers in females in the Eastern and Riyadh regions. Shah et al. (2014) performed analyses on colorectal cancer incidence in Malaysia diagnosed in 1995 to 2011. From a total of 146 cases, hotspot analysis using Getis-Ord Gi* showed that hotspots were fell on the northeast side of Kuala Lumpur, but there was no significant global spatial autocorrelation in the study area.

Osayomi and Areola (2015) investigated spatial clustering patterns and hotspots of road traffic accidents, injuries and deaths in Nigeria. With the aid of Global Moran's I and Local Getis-Ord Gi*statistic, they found significant positive spatial autocorrelation and an accident belt in the southwestern region of Nigeria from 2002 to 2007. A more recent analysis was conducted by Santaularia et al. (2021). They employed Moran's I statistic and Anselin's LISA to assess the spatial distribution of different forms of criminal punishment with violent injuries in children in Minnesota, in the United States from 2010 to 2014, using hospital discharge data and court administrative data. The analyses demonstrated that child abuse injury, incarceration and probation are significantly geographically concentrated in similar regions, whereas violent injuries and monetary sanctions are more dispersed throughout Minnesota. Manap et al. (2021) utilised Moran's I spatial autocorrelation and the Getis–Ord Gi* statistic to predict clustering hotspots for heavy vehicle accidents in Malaysia from June 2016 to the end of May 2019. From 7276 heavy vehicles accident cases, a total of 22 heavy vehicle risk segments were identified as hotspots at significance levels from 0.10 to 0.01 with a 1355 m buffer radius.

In the epidemiological analysis, clusters of the disease may change not just in space but also in time. Space-time clustering is one of the common methods applies in broader epidemiological studies; a study conducted by Martins-Melo et al. (2016) can be an excellent example of the use of spatial and spatiotemporal clustering of disease. They studied the spatial and spatiotemporal cluster analysis to investigates nationwide trends and spatial distribution of neglected tropical disease-related mortality in Brazil from 2000 to 2011. The main spatial techniques carried out in this study were using global spatial autocorrelation to investigate the presence of disease cluster; LISA to identify significant spatial hotspots, coldspots, and outliers of the disease; and Kulldorff's space-time scan statistics to identify spatiotemporal clusters of the disease. The results demonstrated a significant spatial and spatiotemporal high-risk cluster for neglected tropical disease-related mortality in all regions. They also identified a significant association between socio-economical characteristics of the regions and the increase of mortality.

Despite the high mortality due to diseases of the circulatory system, few historical studies have been conducted. Wang et al. (2014) used spatial statistics, and spatial analysis was used to determine the

spatiotemporal distribution and identify patterns of change at the district level. Space-time scan statistics were used to identify spatiotemporal clusters of ischemic heart disease hospital admissions at the district level in Shenzhen, China, from 2003 to 2012. The results revealed that the incidence rates of ischemic heart disease increased for all districts in Shenzhen, and the highest and the lowest district predicted to have the highest incidence rates among all of the districts from 2013 to 2015. In a more recent study, space-time clustering has been successfully implemented in the 278 municipalities of Continental Portugal, where it showed how rural characteristics of municipality affects trends of cardiovascular mortality (Almendra and Santana, 2020).

Martin Kulldorff first introduced space-time scan statistics in 1997 (Kulldorff et al., 1998). The study applied space-time scan statistics to 1175 brain cancer cases in Los Alamos, New Mexico, between 1973 and 1991. They demonstrated the possibility of using space-time scan statistics for the public health official's toolbox as a screening tool for evaluating cluster alarms. After about two decades, researchers from Ecuador uses global Moran' I and space-time scan statistics to identify trends and spatial patterns of oral cancer mortality in Ecuador between 2001 and 2016 (Núñez-González et al., 2018a). They found that oral cancer mortality in Ecuador increased over the 16-year study period, predominantly in the young groups.

Núñez-González et al. (2018b) also uses spatial autocorrelation and scan statistics in suicide analysis. They analysed the spatiotemporal clusters and the spatial distribution of suicide in Ecuador, from 2011 to 2016 using all death certificates of suicide among adolescents in the 10 - 19 age groups both sex from the National Institute of Statistics and Census database in Ecuador. Suicide rates in adolescents increased over the study period; space-time scan statistics identified two significant spatial clusters for a high-risk of suicide where the primary most likely cluster included 83 cantons in Amazon and Southern Highlands regions of the country. Spatial autocorrelation through global Moran's I and LISA test showed a positive spatial autocorrelation, and a high-risk cluster was found in 14 cantons located in 6 provinces, whereas a low-rates cluster included 18 cantons located in 6 provinces. Linton et al. (2014) assessed the spatial and temporal clusters of drug activity in Baltimore, Maryland, from 2000 to 2010. A discrete Poisson model-based space-time scan statistic was used to identify statistically significant clusters of narcotic calls for service across space and time at the neighbourhood level. They found a spike of the narcotic drug after 2003 in Baltimore; significant spatial and spatiotemporal clusters in Southeast, Northeast, Northwest, and West Baltimore from 2001 to 2010 and decrease trends in East Baltimore from 2001 to 2003. The study proves the effectiveness of using space-time scan statistics to identify persistent clusters of drug activities and generates hypotheses on the association between targeted local policies and drug activity. Moreover, Song et al. (2018) used

Poisson model Kulldorff's space-time scan statistics to identify spatiotemporal clusters of a traffic accident in the United Kingdom in 2016. They were demonstrating two statistically significant clusters in the study area. In the study, they concluded that Kulldorff's space-time scan statistics is a suitable method for identifying traffic accident space-time clusters - it has the ability to pinpointing the exact location, size and period of statistically significant clusters from a vast dataset without manually designing meaningful temporal windows.

2.2 Major causes of death in Estonia

The major cause of death in Estonia are circulatory diseases, neoplasms and injuries, which constituted about 80% of overall mortality in the country in 2019 (National Institute for Health Development, 2020a). Details of clustering of causes using the International Classification of Diseases and Related Health Problems (ICD) tenth version by the World Health Organization (WHO) were described in Table 1.

Table 1. Three major causes of death in Estonia and their sub-categories

ICD-10 Code	**Disease**
C00-C97	Malignant neoplasms
C00-C75	Malignant neoplasms, stated or presumed to be primary, of specified sites, except of lymphoid, haematopoietic, and related tissue
C76-C80	Malignant neoplasms of ill-defined, secondary, and unspecified sites
C81-C96	Malignant neoplasms, stated or presumed to be primary, of lymphoid, haematopoietic, and related tissue
C-97	Malignant neoplasms of independent (primary) multiple sites
I00-I99	Diseases of the circulatory system
I00-I02	Acute rheumatic fever
I05-I09	Chronic rheumatic heart diseases
I10-I15	Hypertensive diseases
I20-I25	Ischaemic heart diseases
I26-I28	Pulmonary heart disease and diseases of pulmonary circulation
I30-I52	Other forms of heart disease
I60-I69	Cerebrovascular diseases
I70-I79	Diseases of arteries, arterioles, and capillaries
I80-I89	Diseases of veins, lymphatic vessels, and lymph nodes, not elsewhere classified

I95-I99	Other and unspecified disorders of the circulatory system
V01-Y89	Injury and poisoning
V01-X59	Accidents
X60-X84	Intentional self-harm
X85-Y09	Assault
Y10-Y34	Event of undetermined intent
Y35-Y36	Legal intervention and operations of war
Y40-Y84	Complications of medical and surgical care
Y85-Y89	Sequelae of external causes of morbidity and mortality
Y90-Y98	Supplementary factors related to causes of morbidity and mortality classified elsewhere

2.2.1 Diseases of the circulatory systems

Circulatory diseases, also often used as cardiovascular diseases (CVDs), are among the main types of NCDs, consisting of disorders or illnesses related to the circulatory system, such as the heart and blood vessels. The main causes of circulatory diseases death in the world are ischaemic heart disease, stroke and chronic obstructive pulmonary disease, responsible for approximately 16%, 11%, and 6% of total deaths in 2011 (World Health Organization, 2020).

In Estonia, although the mortality due to circulatory diseases has decreased, it still remains the main cause of death in Estonia with the share of CVDs from total mortality in 2019 being 50 % (National Institute for Health Development, 2020b)

2.2.2 Malignant neoplasm

A neoplasm, also often called a tumour, is an abnormal mass of tissue growth resulting from the excessive and uncontrolled proliferation of cells. When the tumours are abnormal and grow uncontrollably, invade and destroy nearby tissues, spread to other tissues within the body or metastasise. They are concluded as a malignant neoplasm, are collectively known as cancers. Malignant neoplasms is the second leading cause of death worldwide. According to GLOBACAN, in 2020, around 19.3 million cancer cases were newly diagnosed and 9.9 million deaths globally, 4.3 million new cancer cases and 1.9 million death in Europe. The cancer burden and mortality have steadily increased and estimated to rise by 62.5% and reach 16 million in 2040 (International Agency for Research on Cancer, 2020).

In Estonia, the share of cancer from total mortality in 2019 was 24 %, accounting for around 7.8 thousand new cancer cases diagnosed and 4009 deaths and expected to reach nearly 5 thousand death in 2040 (International Agency for Research on Cancer, 2020; National Institute for Health Development, 2020a).

2.2.3 Injury and poisoning

Injury is an act of violence against others or oneself, road traffic crashes, burns, drowning, falls, and poisonings. Injuries are among the most prominent public health concerns globally, and three leading causes of injury are road traffic injuries, suicide, and falls. According to WHO, deaths caused by injuries are predicted to rise; road traffic injuries are predicted to become the 7th leading cause of death; falls rising to become the 17th leading cause of death; and suicide remaining in the top 20 by 2030 (World Health Organization, 2014).

In Estonia, although the mortality of injuries has decreased, injuries are still the third major causes of death, accounting for 849 Estonian inhabitants died in 2017 (National Institute for Health Development, 2020b; Panov, 2018). The main causes of injury deaths are alcohol and narcotic intoxication and suicide; intoxication was common in men, whereas most deaths in women were suicide-related cases (Panov, 2018).

3. Data and Methodology

3.1 Study area

Estonia is situated in north-eastern Europe, bordering the Gulf of Finland and the Baltic Sea, between Latvia and Russia, with an area of 45,227 km^2. Estonia is administratively divided into 15 counties (first-level administrative division) and as of December 2020 had a population of 1,328,976 people (As of December 2020). The study was conducted based on the county level (Figure 1).

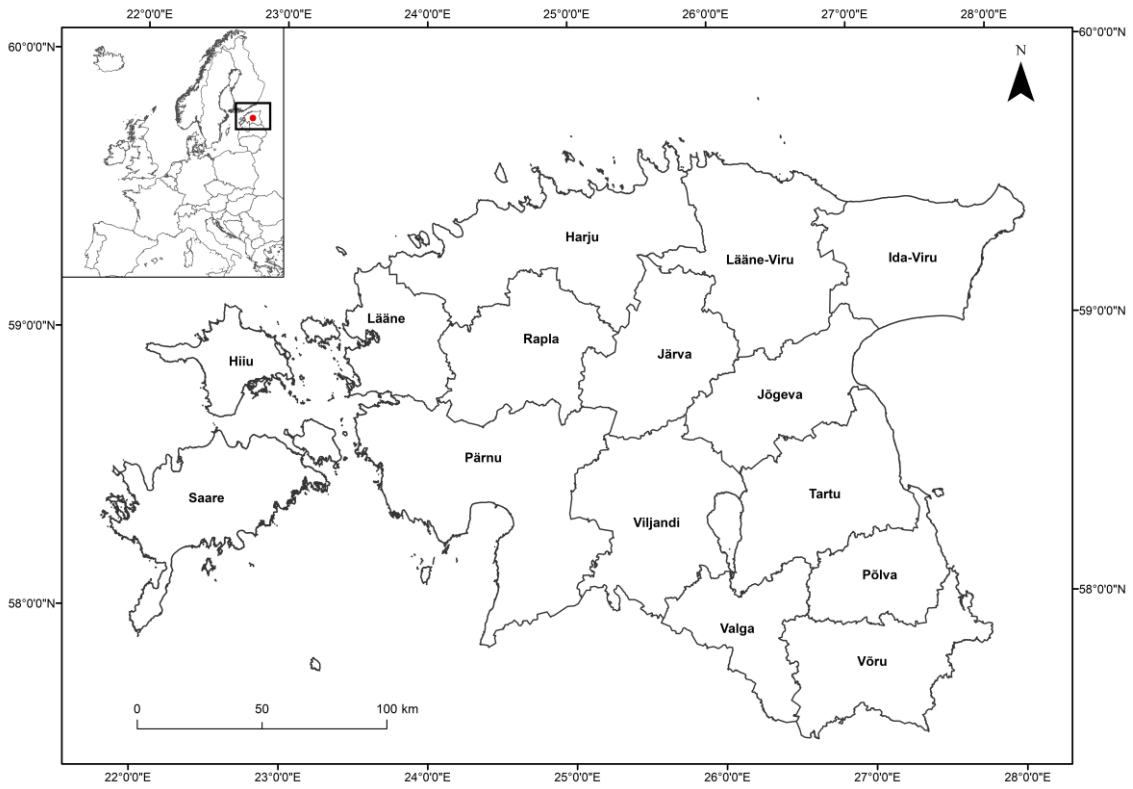

Figure 1. Study area map

3.2 Data

Mortality data of three major causes of death in Estonia was acquired from the Health Statistics and Health Research Database administrated by National Institute for Health Development (accessed 11 November 2020). These data are publicly available free of charge at the website of the Health Statistics and Health Research Database.

The Health Statistics and Health Research Database provides mortality data that have standardised using 'European standard population', which is published in 2013 (European Commission. Eurostat., 2013). This standard population is given in annexe 1. The data were stratified by the main cause of death and covered 15 years, 2005 to 2019 and included all deaths in Estonia. Causes of death are coded according to the ICD tenth revision by WHO. Mortality due to Malignant neoplasms (C00-C97),

Diseases of the circulatory system (I00-I99), Injury and poisoning (V01-Y89) were included in the study.

Spatial data, county boundary map, was acquired from Estonia Land Board, and population dataset of 2005–2019 for each county was obtained from Statistic Estonia.

The analyses were based exclusively on publicly available secondary data for which there was no possibility of identifying individuals.

3.3 Methodology

A simplified flowchart of this study can be found in Figure 2. The input datasets underwent pre-processing, then followed by spatial analysis. In order to examine spatiotemporal trends, three different analyses were conducted: a general mortality trend analysis, a spatial cluster analysis, and a spatial-temporal cluster analysis (Table 2).

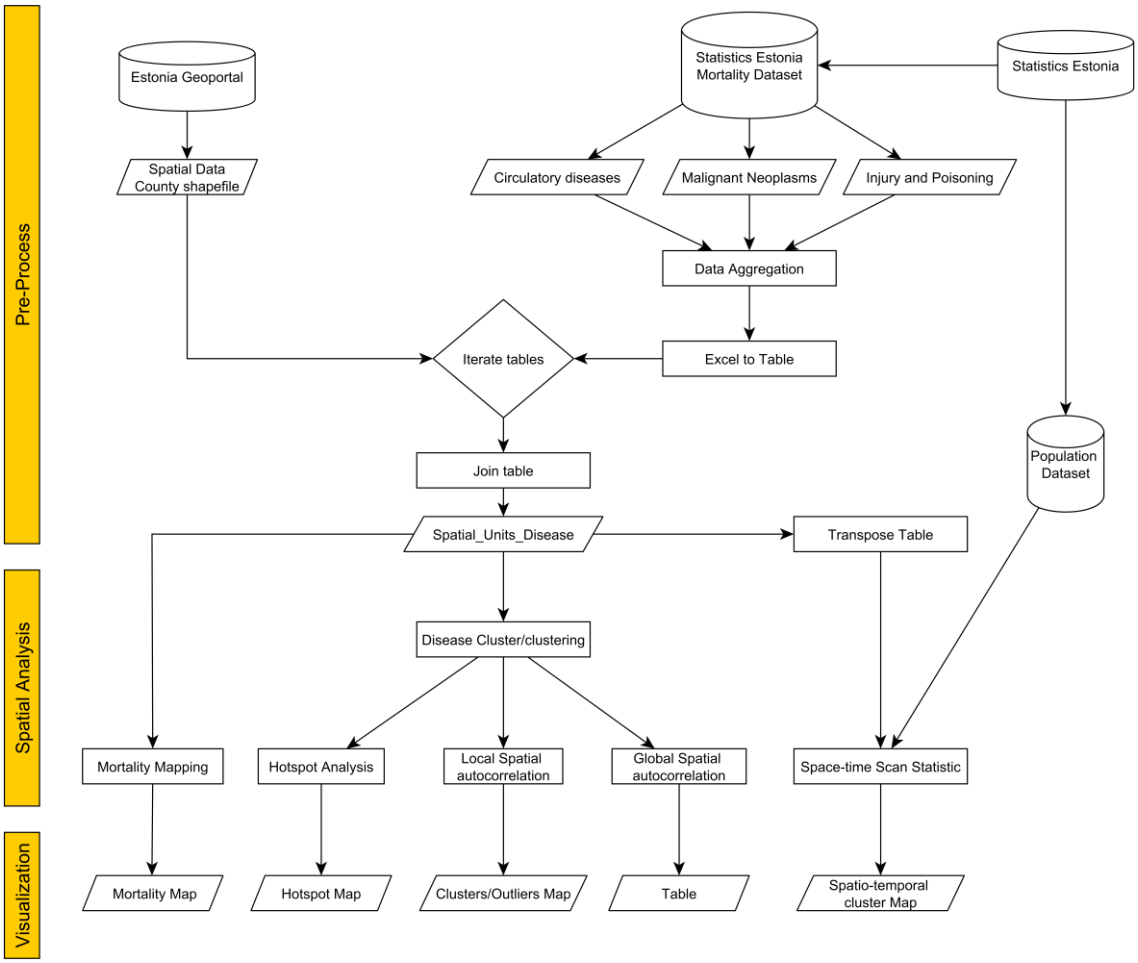

Figure 2. Overview of the workflow

Table 2. Overview of analytical tools

Research objective	Type of analysis	Software
Where is the disease mortality significantly clustered?	Global spatial autocorrelation	ArcGIS
How is mortality distributed spatially? What specific areas are with a high risk of disease mortality?	Local spatial autocorrelation (Cluster and Outlier analysis) Hotspot analysis (Getis-Ord Gi* test)	ArcGIS
What is the spatiotemporal pattern in disease mortality in Estonia?	Space-time model and Spatial Variation in Temporal Trends Model	ArcGIS, SaTScan

3.3.2 Mortality trends analysis

The mortality data were aggregated over a period of three years in order to reduce the year by year fluctuations due to small mortality numbers in small counties for statistical processing. Obtained mortality dataset was manipulated before joining spatial data and mortality data, then joined spatial data and mortality data for further analysis.

Standardised mortality rates were divided into five-classes and displayed accordingly by different shades of colours. The mortality maps of 3-yearly intervals were mapped using ArcMap v10.6.1. to describe and compare the geographic pattern of disease mortality in Estonia over the study period.

3.3.3 Global Index of Spatial autocorrelation – Grobal Moran's I

Global indexes of spatial autocorrelation were used to assess the similarity, or spatial dependence, across counties with respect to mortality rates. In other words, are counties with similar mortality rates located close together or are mortality rates randomly distributed across Estonia?

The most applied method to test for global spatial autocorrelation is Moran's I (Pfeiffer et al., 2008a; Singh et al., 2016); it quantifies the similarity of an outcome variable among areas that are defined as spatially related (Moran, 1950).

$$I = \frac{N}{\Sigma_i \Sigma_j w_{ij}} \times \frac{\Sigma_i \Sigma_j w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\Sigma_i (y_i - \bar{y})^2}$$

where:

N = units

y_j = attribute value for each unit *i*

w_{ij} = wight for unit *i* and *j*

The expected value for Moran's I is

$$E\ (I) = \frac{-w_i}{n-1}$$

with w_i as the sum of the row elements.

Like classical correlation, Moran's I lies between +1 and -1, where zero indicates the null hypothesis of no clustering — the spatial allocation of the observations is random. When the Moran's I is positive, it means clustering of areas of similar attribute values, whereas a negative coefficient indicates that neighbouring areas tend to have dissimilar attribute values.

In order to decide the significant difference between any given pattern and a random pattern of the result, z-statistic is delivered from the mean and variance of I index, then compared with p-value or the critical value found in the normal table. High values of the Moran's I and corresponding z-scores greater than 1.96 indicate that there is statistically significant clustering across Estonia ($p < 0.05$), whereas low Moran's I and z-scores less than −1.96 indicate that there is statistically significant regularity.

In this study, the global Moran's I tool in ArcGIS was used which returns the z-score and *p*-value.

3.3.3. Local indicators of spatial association (LISA) – Getis-Ord Gi* and Anselin Local Moran's I

Since Global tests of autocorrelation such as Moran's I do not indicate where clusters of high or low mortality rates might occur; thus, two LISAs to examine clusters of counties with high or low mortality of the disease; the Getis-Ord Gi* statistic and Anselin's LISA.

Getis-Ord Gi* statistics given (Getis and Ord, 1992)

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}{}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}}$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight feature i and j, *n* is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} {x_j}^2}{n} - (\bar{X})^2}$$

The Getis-Ord Gi* statistic generates a z-score and corresponding *p*-value for each county. For statistically significant positive z scores ($p < 0.05$), the larger the z score (where z-scores greater than 1.96) is, the more intense the clustering of high values (a significant "hotspot") and z-scores lower than −1.96 indicate a significant "coldspot" or the more intense the clustering of low values ($p < 0.05$).

Additionally, Anselin's LISA was used to detects local spatial autocorrelation or "hotspot" and spatial outliers, by decomposing Moran's I statistic into contributions for each area within a study region (Anselin, 1995). These indicators detect clusters of either similar or dissimilar disease frequency values around a given observation.

$$I(i) = z_i \, \Sigma \, w_{ij} z_j$$

where:

z_i = Standardised scores of attribute values for unit *i*

z_j = Standardised scores of attribute values for unit j

w_{ij}= weight matrix w_{ij}

When the I(i) is positive, it means that a grouping of similar values (higher or lower than average) and a negative value indicates a combination of dissimilar values (e.g. low value surrounded by high values). Based on these local Moran's Index and associated p-values, each county can then be classified by cluster/outlier type (COType). Counties fall into one of 5 groups (ESRI, 2019):

- Not part of a cluster;
- Hotspot: High mortality rates and surrounded by a cluster of high mortality rates (HH = "high-high");
- Coldspot: Low mortality rates and surrounded by a cluster of low mortality rates (LL = "low-low");
- Spatial outlier: High mortality rates surrounded primarily by low mortality rates (HL = "high-low");
- Spatial outlier: Low mortality rates is surrounded primarily by high mortality rates (LH = "low-high").

In this study, Hotspot analysis (Getis-Ord Gi*) and Cluster and Outlier Analysis (Anselin Local Moran's I) tool in ArcMap v10.6.1. was used for each year and significant clusters were then visualised.

3.3.4. Space-time scan statistic

In epidemiology, space-time clustering is used because geographical clusters of disease change over time in each of the census districts. Space-time scan statistic detects diseases in time and space, verify the random distribution of the disease in time and space, and ascertain the number of cases in a region (Singh et al., 2016). The method was developed by Kulldorff in 1997, based on a scan statistic, extending a two-dimensional circular window to a three-dimensional cylinder passing through time (Kulldorff, 1997). The cylinder base is geographic, and the radius of the window varies from zero to a user-specified upper limit. The height of the cylinder corresponds to time within the study period.

When the scan statistic is used to evaluate the spatial variation in temporal trends, the scanning window is purely spatial in nature and the temporal trend is calculated both inside and outside the scanning window, for each possible geographical position with each possible window size (Kulldorff, 1997). The null hypothesis is that the geographical trends in different windows are the same, while the alternative is that the trends are different. Based on these hypotheses, the mean relative risk (RR) and Log-Likelihood Ratio (LLR), based on the number of observed cases versus the number of expected cases both inside and outside the window, were calculated. A lower LLR value indicates that the difference in trends is occurred by chance. The window with the largest likelihood statistic is considered as the most likely cluster which the trend inside the window is least likely to be the same as the trends outside the window, while others are considered as secondary clusters.

The likelihood function for a specific window is proportional to (Kulldorff, 1997):

$$\left(\frac{c}{E[c]}\right)^{c}\left(\frac{C-c}{C-E[c]}\right)^{C-c} I()$$

Where:

C = Total number of cases,

C = Observed number of cases within the window and

$E[c]$ = Expected number of cases within the window under the null hypothesis

In this study, first, retrospective Kulldorff's space-time scan statistics with the discrete Poisson model was used to identify high or low-risk spatiotemporal clusters during a specific time frame. Next, the

spatial variation with temporal trends using a discrete Poisson model was applied to determine if there was a significant change over the study period.

For all disease conditions, areas were scanned simultaneously for statistically significant high and low rates. For all disease condition, the circular scanning window was set to default, a maximum size of 50% of the population and a size considered sufficiently large to ensure that small and large clusters could be included. The size of the temporal window was also set to the default (50 %). Additionally, a one-calendar time period of aggregation over the fifteen-year time frame of the study was used to assess any long-term patterns in the clustering of disease. A p-value was generated using 999 Monte Carlo simulations.

Space-time cluster analyses were conducted using SaTScan™ software, which was developed under the joint auspices of (i) Martin Kulldorff, (ii) the National Cancer Institute, and (iii) Farzad Mostashari of the New York City Department of Health and Mental Hygiene. Statistically significant space-time clusters of mortality rates with $p < 0.05$ were mapped using ArcMap v10.6.1.

4. Results

This chapter presents the results by following the methodological sequence outlined in the previous chapter for each disease.

4.1 Trends of Mortality

In general, trends of mortality for diseases of the circulatory system and injury and poisoning were declined. For malignant neoplasm, some counties showed an increasing trend. The mortality for death causes varied to a big extent (Table 3).

Table 3. Summary statistics of standardised mortality rate (per 100,000 inhabitants) for three major causes of death in Estonia

		2005 2007	2008 2010	2011 2013	2014 2016	2017 2019
Diseases of the circulatory system	Min	826.88	761.23	678.38	602.54	526.99
	Max	1143.93	1077.39	903.31	830.06	768.49
	Mean	973.03 ± 93.55	900.08 ± 88.04	764.83 ± 66.65	700.64 ± 58.02	635.16 ± 67.47
Malignant neoplasm	Min	272.37	247.74	283.17	279.18	270.45
	Max	326.73	357.74	322.19	363.63	396.24
	Mean	297.66 ± 15.88	292.70 ± 24.20	300.54 ± 12.47	308.78 ± 23.22	306.04 ± 29.84
Injury and poisoning	Min	104.57	74.44	66.18	52.34	52.19
	Max	163.12	132.62	118.05	102.01	89.26
	Mean	129.37 ± 18.97	98.69 ± 18.98	91.50 ± 16.14	73.36 ± 14.54	68.47 ± 10.96

Figures 3–5 show disease maps of standardised mortality rates per 100,000 inhabitants of three disease for each county in Estonia. Each county is coloured according to the category into which its corresponding attribute value falls; the county with dark red colour has the higher mortality rates, whereas county with light red colour has the lower mortality rates.

For death caused by diseases of the circulatory system, the highest mortality (1143.93 per 100,000 population at risk) was reported in Põlva county in 2005–2007, while the lowest occurred in Harju county (526.29 per 100,000 population at risk) in 2017–2019.

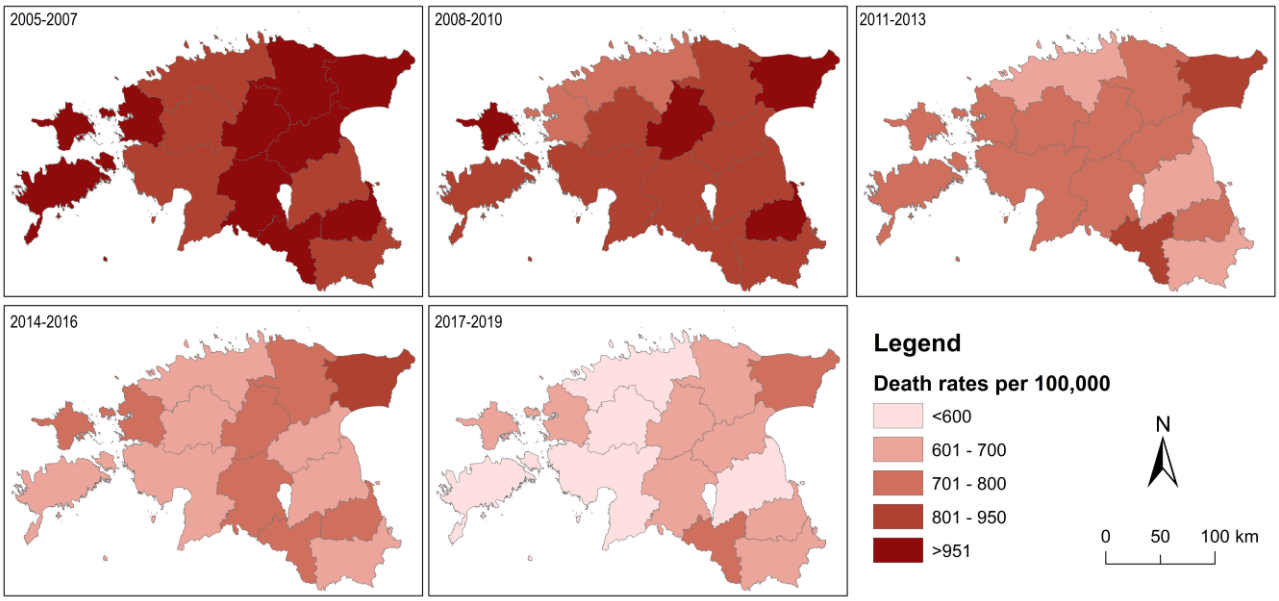

Figure 3. Standardised mortality rates of diseases caused by diseases of the circulatory system for 3-yearly intervals from 2005–2019

For malignant neoplasm, both the highest and the lowest occurred mortality were observed in 2017–2019, in Hiiu county (396.24 per 100,000 population at risk) and Harju county (270.45 per 100,000 population at risk), respectively.

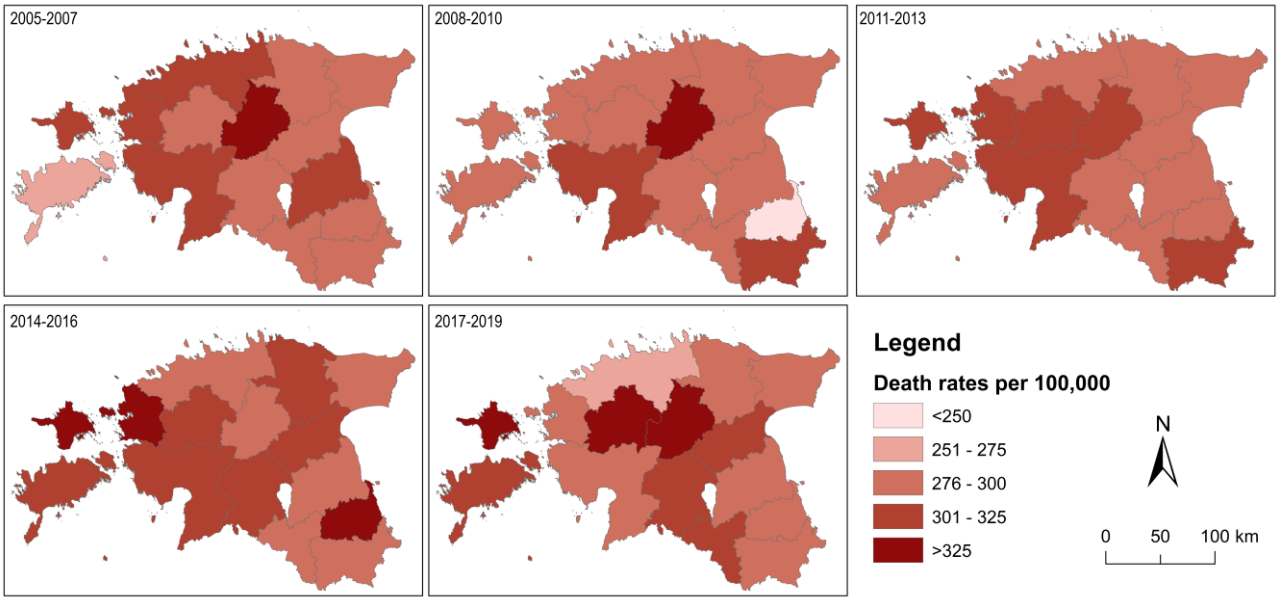

Figure 4. Standardised mortality rates of malignant neoplasm for 3-yearly intervals from 2005–2019

For injury and poisoning, the highest mortality (163.12 per 100,000 population at risk) was reported in Ida-Viru county in 2005–2007, while the lowest occurred in Harju county (52.19 per 100,000 population at risk) in 2017–2019.

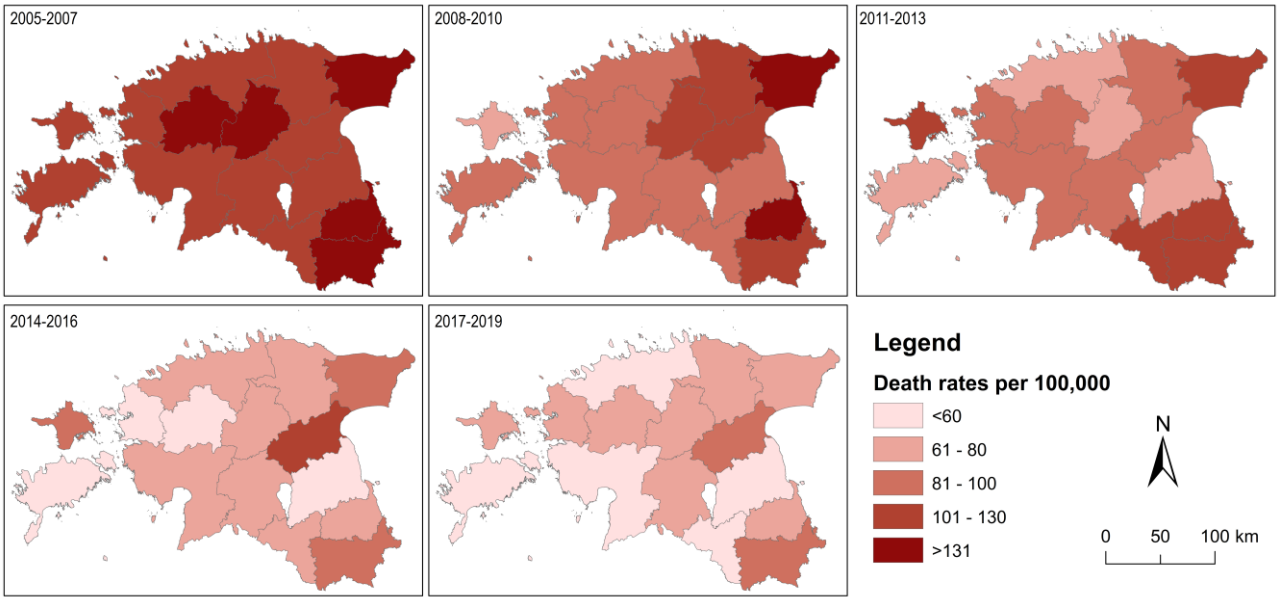

Figure 5. Standardised mortality rates of injury and poisoning for 3-yearly intervals from 2005–2019

4.2 Results for spatial Clustering

Results of the Global Moran's Index and local indicator of the spatial association are reported here.

4.2.1. Clustering of Mortality rates – Moran's I

Moran Indexes of diseases of the circulatory system and malignant neoplasms were statistically not significant at a level of 0.05, meaning the null hypothesis of the observed pattern of values being randomly distributed could be accepted. In other words, across Estonia, the distribution of disease was found to be randomly clustered.

On the other hand, for injury and poisoning, the Global Moran's I and the corresponding z-score for the years 2008–2010 suggest that there was significant spatial autocorrelation of county-level mortality rates (Moran's Index =0. 33, z-score =2.94, p <0.05). The Moran's Indexes of all 15 counties in these years were positives, indicating that, across Estonia, counties with similar mortality rates tend to locate near each other.

4.2.2 Local autocorrelation – Getis-Ord Gi* and Anselin Local Moran's I

The result of global autocorrelation does not tell if the clustering of high value or low value were attributes to the Moran's index. Here, local autocorrelation localises specific clusters and determine the magnitude of spatial autocorrelation at a local level.

Significant clusters of counties with high (hotspots) and low (coldspots) mortality of three major causes of death in Estonia, as assessed by the Getis-Ord Gi* tool for the years 2005–2007 and 2017–2019, can be seen in Figure 6, 8, and 10. Anselin Local Moran's I confirmed some of the significant hot and coldspots identified by the Getis Ord G * tool. Additionally, several counties were categorised as spatial outliers (Figure 7,9,11).

Disease of Circulatory system

There was no hot and coldspot with a significance above 95% CI. However, a significant coldspot with 90% CI was found in Lääne county in 2017–2019 and Anselin Local Moran's I confirmed its significance (LL cluster). Tartu identified as LH outliers in 2005–2007 and 2011–2019, this indicates that the death caused by diseases of the circulatory system in Tartu was low, but the surrounding county has a high mortality.

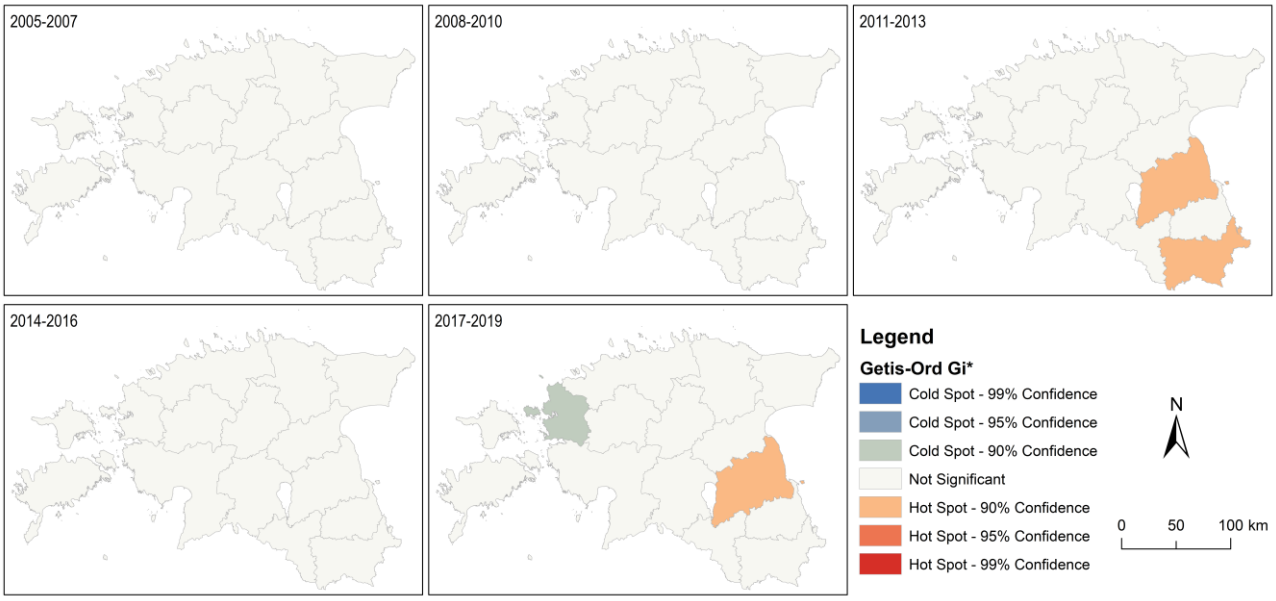

Figure 6. Hotspot analysis (Getis–Ord Gi*) of mortality caused by diseases of the circulatory system in Estonia during 2005–2019

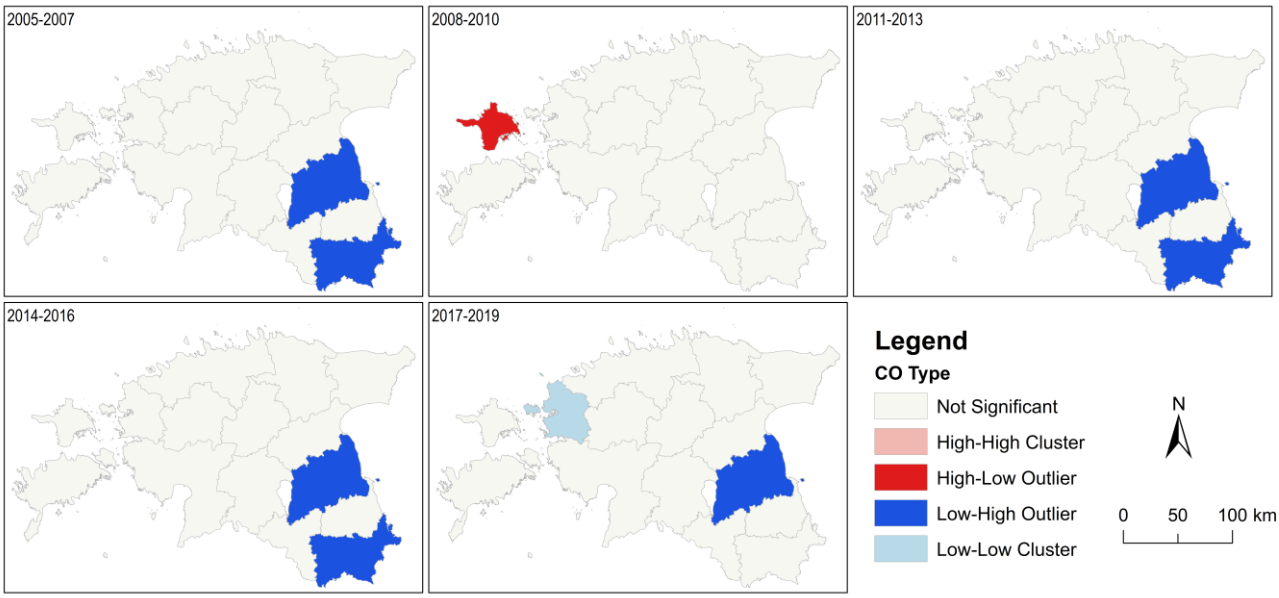

Figure 7. Cluster and Outlier analysis map (Anseline Local Moran's I) of mortality caused by diseases of the circulatory system in Estonia during 2005–2019

Malignant neoplasm

Significant coldspots (95% confidence and LL cluster) were seen in Võru county in 2005–2007 and in Tartu county in the year 2008–2010 and 2011–2013 and the significance of the clusters were also confirmed by Anselin Local Moran's I test. There was no significant hotspot observed in this disease.

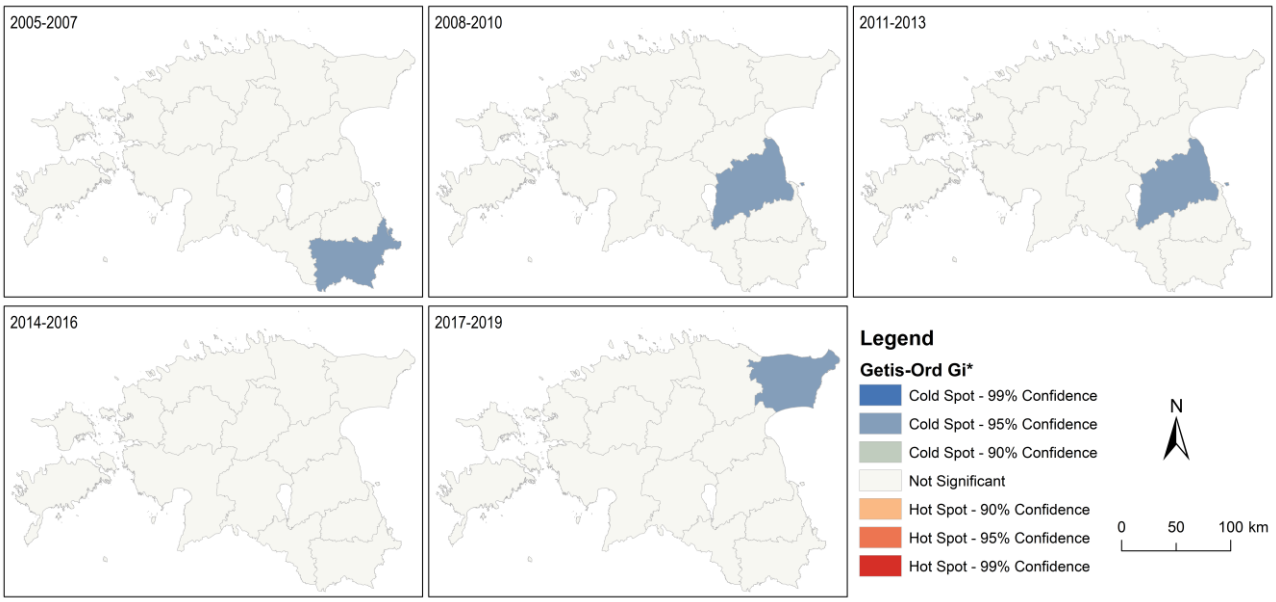

Figure 8. Hotspot analysis (Getis–Ord Gi*) of mortality by malignant neoplasm in Estonia during 2005–2019

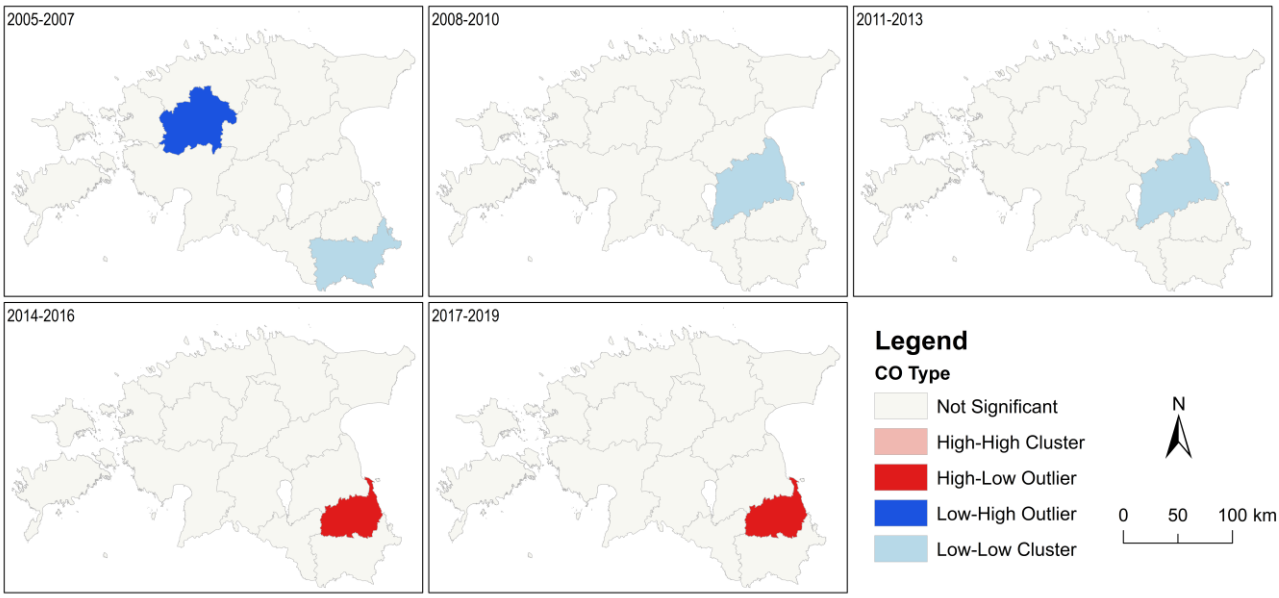

Figure 9. Cluster and Outlier analysis map (Anseline Local Moran's I) of mortality by malignant neoplasm in Estonia during 2005–2019

Injury and poisoning

Five counties in the West part of Estonia (Lääne county, Pärnu county, Hiiu county, Saare County, Rapla county) were part of coldspot, however, a significant coldspot (95%) was only seen in Lääne county in the year 2008–2010, 2017–2019. Additionally, an extreme significant hotspot (99% confidence and HH cluster) was appeared in Ida-Viru county in 2008-2010 and a significant hotspot (95% confidence and HH cluster) was found in Põlva county in the year 2011–2013. These two significant hotspots were also observed in Anselin Local Moran's I test. Interestingly, Pärnu identified as a coldspot (LL cluster) in 2008–2010 but changed to HL outliers in 2014–2016, suggesting that, in 2014–2016, either morality in Pärnu increased or mortality of neighbouring countries decreased.

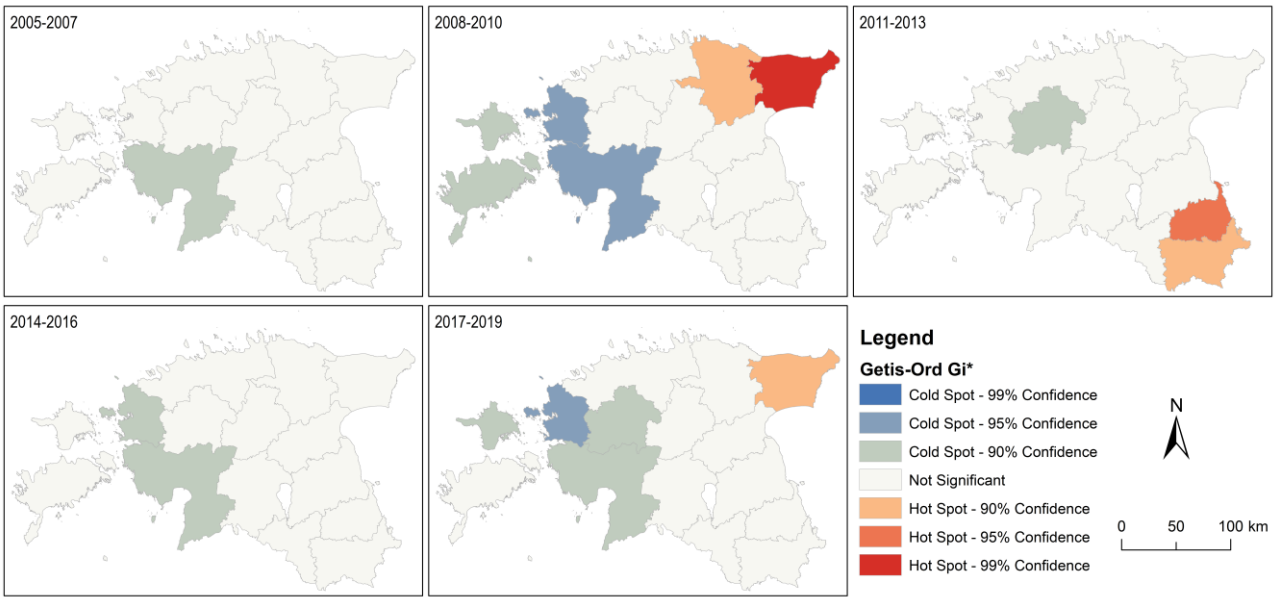

Figure 10. Hotspot analysis (Getis–Ord Gi*) of mortality caused by injury and poisoning in Estonia during 2005–2019

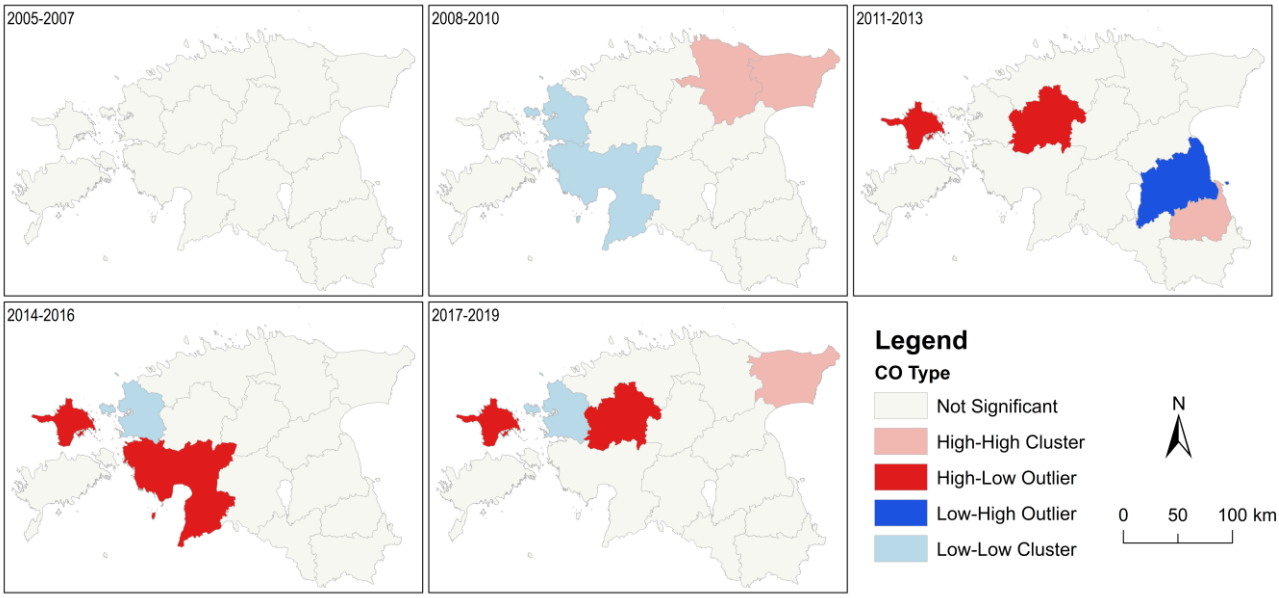

Figure 11. Cluster and Outlier analysis map (Anseline Local Moran's I) of mortality caused by injury and poisoning in Estonia during 2005–2019

4.3 Results for space-time clustering

4.3.1 Space-time clustering

Disease of Circulatory system

As shown in Table 7, the most likely cluster was located in Harju county from 2011–2017 (RR=0.08, LLR=9424.204), which there is a lower cluster of death caused by diseases of the circulatory system occurred in the county. Secondary clusters were identified in Valga, Rapla, Järva, Lääne-viru, Jõgeva, Tartu, Pärnu, Viljandi, Põlva, Võru (RR=6.35, LLR=9091.02, p<0.001) in 2007 and Saare, Hiiu, Lääne (RR=6.3, LLR=7682.083, p<0.001), from 2007–2013.

Table 4. Spatiotemporal clusters of deaths caused by diseases of the circulatory system in Estonia during 2005–2019

	County	Time Frame	Observed	Expected	RR	LLR	p-value
Most likely cluster	Harju	2011 - 2017	1280	12393.28	0.08	9424.20	<0.001
Secondary cluster	Valga, Rapla, Järva, Lääne-viru, Jõgeva, Tartu, Pärnu, Viljandi, Põlva, Võru	2007	9647	1759.8	6.35	9091.02	<0.001
	Saare, Hiiu, Lääne	2007-2013	8114	1455.11	6.3	7682.08	<0.001

Malignant neoplasms

The space-time scan statistic identified three statistically significant clusters of mortality due to Malignant neoplasm that occurred during 2005–2019 (Table 8). The most likely lower cluster occurred in Harju county during 2011–2017 (RR =0.09, LLR=3376.72, p <0.001), which suggests that lower mortality occurred during these years as compared to other years. The second cluster occurred in counties in west Estonia (Saare, Hiiu, Lääne, Rapla counties) during 2013–2019 (RR = 3.40, p <0.001) and in southern Estonia in the year 2019 (RR =6.63, LLR=2678.04, p <0.001), suggesting that higher mortality occurred during this year as compared to other years.

Table 5. Spatiotemporal clusters of deaths caused by malignant neoplasm in Estonia during 2005-2019

	County	Time Frame	Observed	Expected	RR	LLR	p-value
Most likely cluster	Harju	2011-2017	564	4691.54	0.09	3376.72	<0.001
Secondary cluster	Saare, Hiiu, Lääne, Rapla	2013-2019	3857	806.24	5.56	3211.29	<0.001
	Valga, Järva, Lääne-viru, Ida-Viru, Jõgeva, Tartu, Viljandi, Põlva, Võru	2019	2689	451.48	6.63	2678.04	<0.001

Injury and poisoning

As shown in Table 9, the most likely cluster was located in Valga, Rapla, Järva, Lääne-viru, Jõgeva, Tartu, Pärnu, Viljandi, Põlva, Võru counties in 2007 (RR=7.69, LLR=1422.54, p<0.001), which there is a higher cluster of death caused by injury and poisoning occurred in these counties. Secondary cluster was identified in Harju (RR=0.08, LLR=1119.26, p<0.001), from 2011–2017.

Table 6. Spatiotemporal clusters of deaths caused by injury and poisoning in Estonia during 2005-2019

	County	Time Frame	Observed	Expected	RR	LLR	p-value
Most likely cluster	Valga, Rapla, Järva, Lääne-viru, Jõgeva, Tartu, Pärnu, Viljandi, Põlva, Võru	2007	1306	203.38	7.69	1422.54	<0.001
Secondary cluster	Harju	2011 - 2017	136	1432.28	0.08	1119.25	<0.001

4.3.2 Spatial variation in temporal trends

Figure 12 shows the result of spatial variation in temporal trend analysis. Detailed information of the clusters can be found in Annex 3.

Disease of Circulatory system

A total of four clusters (one most likely cluster and three secondary clusters) were identified throughout the study period. All significant clusters were an increasing pattern in both inside and outside the window. Figure 12 shows the first two significant clusters; the most likely cluster of lower change rate was located in Ida-Viru county (LLR=155.52, $p<0.001$ from 2005–2019. Secondary clusters were found in Hiiu and Saare (LLR=57.58, $p<0.001$).

Malignant neoplasm

Figure 12 showed the spatial variation over time of mortality caused by Malignant neoplasms in Estonia. A total of two significant clusters (most likely cluster of the lower rate and secondary cluster of higher rates) were identified. All significant clusters were in an increasing pattern in both inside and outside the window.

The most likely cluster was found in Ida-Viru (LLR=42.03, RR=0.55, $p<0.001$), a slower increase in mortality was detected in this county, the 15% annual increase rate among which was significantly slower than the observed in other counties (7.68% annually). Secondary higher rate clusters were detected in Valga, Järva, Pärnu, Jõgeva, Tartu, Viljandi, Põlva, Võru (LLR=18.09, RR=2.16, $p<0.001$).

Injury and poisoning

As shown in Figure 12, a total of two significant clusters (1 most likely cluster of the higher rate and one secondary cluster of lower rates) were identified. All significant clusters were in an increasing pattern both inside and outside the window. The most likely cluster of higher change rate was located in Ida-Viru county (LLR=16.10, $p<0.00$). The mean annual change in rate for this county reached 5.43%, which was almost four times that in other counties (1.35%). Lower rate cluster was found in Harju, Lääne, Rapla, Järva, Jõgeva (LLR=7.61, $p<0.006$).

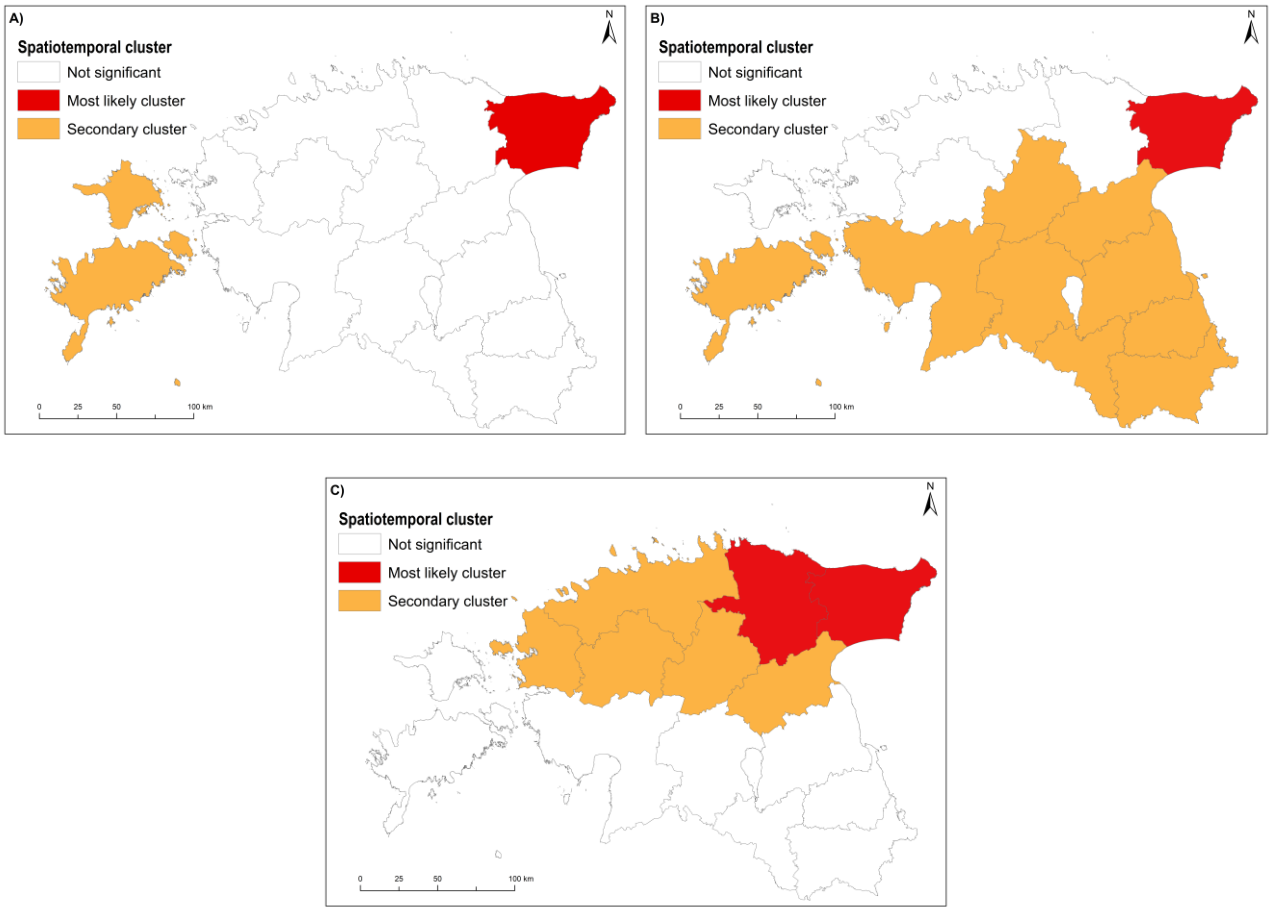

Figure 12. Spatiotemporal clusters of death caused by A) diseases of the circulatory system; B) malignant neoplasms; and C) injury and poisoning

5. Discussion

5.1 Spatiotemporal patterns in mortality rate

This study describes the geographic and temporal distribution of three major causes of death in Estonia over the period of 2005 to 2019. The methods adopted in this study combine spatial clustering and space-time clustering, allowing the identification of counties characterised by higher and/or lower disease mortality and of counties with temporal trends different from the rest of the country. As the major cause of mortality in Estonia (Statistika andmebaas, 2020), mortality due to diseases of the circulatory systems, malignant neoplasms, and injuries have been conducted frequently, however, existing studies relevant to the diseases were mostly either conducted on a regional level and/or have ignored the spatial patterns. To the best of author's knowledge, this is the first study assessing the presence of clusters of three major cause of death in Estonia using geospatial analysis at the county level in Estonia.

Moran's Index of mortality caused by diseases of the circulatory system and malignant neoplasms were not statistically significant, suggesting that the general pattern of the mortality is randomly distributed across Estonia. However, several significant local clusters were captured by LISA and Hotspot analysis. On the other hand, results of the global and local Moran's I confirmed that mortality rates of injury and poisoning exhibit spatial dependence, and several hotspots and coldspots were identified across Estonia. Furthermore, positive spatial autocorrelations were identified in the patterns of mortality due to injury and poisoning in 2008–2010, indicating that the mortality rates were not spatially homogeneous and were also different in the study area.

Most of the occurred clusters of Hotspot analysis and LISA were also detected as part of scan statistic's most likely or secondary likely clusters, which highlights the reliability of the resulted maps. This study found at least two space-time clusters during the study period in different counties of the country. The space-time clustering might provide information about potential underlying risk factors the shifted through time.

The spatiotemporal pattern of mortality of three major causes of death in Estonia, diseases of the circulatory system, malignant neoplasms and injury and poisoning, has been changing between 2005 and 2019, where a general increasing trend was identified with an annual average increase of 3.82%, 8.30%, and 2.07%, respectively, with significant temporal and spatial variation among the counties. The difference in the spatial and temporal pattern of mortality might be due to the geographic disparities or inequality in socio-economic status (Lai and Leinsalu, 2015).

For diseases of the circulatory system mortality, there were no significant clusters captured by both Getis-Ord Gi* test and LISA. However, significant increase and decrease trends were captured during the study period.

Previous studies have discussed that ethnic status affects mortality of the disease, where the Russians have a higher mortality rate in cardiovascular diseases (Kaldmäe et al., 2017; Leinsalu et al., 2004) and mortality was declined among Russian ethnicity and the lower educated county on Ida-Viru county (Lai and Leinsalu, 2015).

Spatial clusters of mortality due to diseases of the circulatory system have been identified in different countries (Almendra and Santana, 2020; Baptista and Queiroz, 2019; Rajabi et al., 2018) but none of these studies has been conducted in Estonia where the mortality rates are higher. In Portugal, Almendra and Santana found space-time clustering of CVD mortality between 1991 and 2017 (Almendra and Santana, 2020); they found that clusters with rural characteristics tend to present higher CVD mortality risk. In Brazil, Baptista and Queiroz assessed clustering during 1996 and 2015 and found spatial clustering in small census areas suggesting changes in the CVD mortality might be affected by socio-economic conditions, access to health care and social norms (Baptista and Queiroz, 2019). In Sweden, Rajabi et al. studied clustering of cardiovascular disease during 2000 and 2010 and found evidence of spatial clustering that tended to be hotspots for long periods, where the hotspots were found in municipalities with low population density and high age (Rajabi et al., 2018).

Similar results are also found in the present study, where the higher trends were identified in two isolated counties in western Estonia, Hiiu and Saare county. These areas are characterised by the lower population density, the lowest proportion of young people, and a low average monthly gross salary. On the other hand, a decreasing trend has identified in Ida-Viru county, where one of the major cities in Estonia is located there. The observed mortality trends seem to be affected by rural characteristics and the ethnic composition of the clusters. Noticeably, the decreasing trend of Ida-Viru county, which is caused by either mortality in the county is increasing less than in other counties or it has a rate that is decreasing more than outside the cluster.

For malignant neoplasms, the coldspots were found in different parts of Estonia in different years, with more consistency in the eastern part of the country with high population density. Tartu County is a densely populated area and approximately 11% of the total population of Estonia lives there. The coldspot also represents areas with a different population structure: a younger, probably healthier, and more mobile population. The spatiotemporal analysis identified that the Ida-Viru has a low trend - it has a rate that is increasing less than other counties. Interestingly, two counties in south-eastern Estonia, Tartu and Võru county, were identified as the coldspots in 2008–2013 and 2005–2007, respectively.

However, spatiotemporal analysis shows that these counties are a member of the cluster that has a high trend, in other words, the mortality rate in Tartu and Võru had either increasing more than outside the cluster or decreasing less than outside the cluster.

Several studies have analysed socio-economic and racial/ethnic disparities in cancer death hotspots (Baburin et al., 2016; Bermudi et al., 2020; Gross, 2007; O'Connor et al., 2018; Singh and Jemal, 2017). Singh and Jemal conducted a census-based study of all-cancers mortality in the United States from 1950 to 2014; they found that the lower education and income groups had higher mortality rates of all cancers (Singh and Jemal, 2017). O'Connor et al. conducted a cross-sectional study of 3135 in the US counties and found that cancer mortality is higher in low-income counties in the United States (O'Connor et al., 2018). For site-specific cancer, lung cancer mortality was significantly associated with lower high school graduation rates (Gross, 2007); breast cancer mortality is high among elderly women (Baburin et al., 2016); cervical cancer mortality remained with the highest rates in lower socio-economic status (Bermudi et al., 2020).

Nevertheless, in contrast to the previous studies, the present study did not find any significant cluster of high-risk mortality. Lack of clustering of malignant neoplasms should be interpreted with care. In most divergence of results in epidemiology, lack of consistency leads to low credibility for associations between potential etiologic factors and diseases (Syriopoulou et al., 2021; Vandenbroucke et al., 2016). Additionally, it should be noted that the causes of death considered here are lump up of several specific-cancer types and some type of cancer has clear aetiology, such as potential environmental factors, that leads to uneven distribution across the country. Further disaggregation to investigate the contribution of each cancer type across space and time is necessary.

For injury and poisoning, hotspots were found in northeast counties. These areas are characterised by relatively high population density and low life expectancy. On the other hand, the coldspot was found in the northwest county with low population density, high age, and low average monthly gross salary. Most of the occurred clusters of Hotspot analysis and LISA were also detected as part of scan statistic's most likely or secondary likely clusters, which highlights the reliability of the resulted maps. The most significant spatiotemporal cluster for high rate was detected in Ida-Viru county, whereas Lääne county identified as a lower rate with lower relative risk. The probable reasons why these particular locations are typical hotspots for injury and poisoning might be; first, these areas consist mostly industrial enterprises and work is more intensive, which are the higher incidence of occupational accidents and fatal accidents (Ministry of Social Affairs, 2018). Furthermore, although the mortality of manufacturing areas is now in decline, mortality rates for transport accidents, suicide, homicide and

alcohol-related causes remained higher among area with a high proportion of Russian and the lower educated residents (Lai and Leinsalu, 2015).

Noticeably, Harju county identified has a low rate in spatiotemporal analysis for all three major causes of death in Estonia, during 2011–2017. Recent studies suggesting that this increase less than other counties in Estonia is attributed to higher education level, higher life expectancy and high gross wages and employment rate (Lai and Leinsalu, 2015).

5.2 Limitation of the study

Even though some analyses produced a statistically significant result, there were several limitations in the study.

Different spatial analysis was applied to investigate the spatial pattern of the major cause of death in Estonia. Unlike spatial scan statistic, spatial autocorrelation analysis can detect any cluster shape but it works properly when the input features class is more than 30 features (ESRI, 2019). Since the analyses were carried at the county level, there are only 15 features for the input feature class for spatial analyses by ArcGIS. Thus, further work on a smaller scale, i.e., municipality level could be done. However, in several municipalities the subsequent limitation is very small number of death cases and high variation between years.

It is also worth mentioning that SatScan has limitations on detecting very large cluster and tends to include neighbouring areas that do not have the risks due to its fixed scan window (Coleman et al., 2009). The results from the method are sensitive to the maximum cluster size parameter; too large of a maximum-size can hide smaller core clusters, and too small of a maximum-size can miss significant, regional-level clusters (Chen et al., 2008). This study used the default maximum-size setting of 50% to generates sensitivities (Kulldorff, 1997) by preventing granular clusters with lower rates. However, there is no defined method to select a proper scanning window and timeframe. Therefore, a sophisticated method should be applied to find the proper window size and shape in future studies of spatial scan statistic.

Since this study examines three different causes of death, the cluster detection analyses does not take into account all possible confounders. There might be a possibility that the identified clusters are affected by endogenous or exogenous factors such as socio-economic status, low-income housing affordability, and location of treatment and support centres among others, depending on the cause of death. The spatial dependencies of mortality rates with socio-economic or environmental conditions

could be examined in the important area using other study designs to assess association or causation, for future studies.

6. Conclusion

This study utilised GIS and statistical analysis to capture geographic patterns of the mortality of the three major cause of death in Estonia. Four approaches were used; Global Moran's I was used to explore the structure of patterns; Getis-Ord Gi* analysis and Anselin's Local Moran's I were used to find local disease clusters and Kulldorff's scan statistics were applied to investigate the spatial and spatiotemporal patterns of mortality.

This study has shown that there is a significant positive spatial autocorrelation for injury and poisoning in 2008-2010; an extreme significant hotspot was appeared in Ida-Viru county and a significant coldspot was seen in Lääne county. Only local spatial clustering was identified for mortality due to diseases of the circulatory system and malignant neoplasms; a significant coldspot was located in Lääne county in 2017–2019 and in Tartu in the year 2008–2010 and 2011–2013, respectively.

In Kulldorff's scan statistic of mortality, research questions as the finding of the spatial and spatiotemporal patterns of mortality were answered. The result of Kulldorff's space-time cluster analysis in this study indicates that a lower cluster for the major cause of death in Estonia was identified in Harju county from 2011–2017. Spatial variation in temporal trends analysis determined a slower increase in mortality caused by diseases of the circulatory system and malignant neoplasms was detected in Ida-Viru county in 2005-2019, while a higher change rate was found in Ida-Viru county where the mean annual change in rate for this county reaches 5.43%.

Recommendations

The data used in this study is at the county levels. For spatial clustering analysis using ArcGIS, ESRI recommends to use more than 30 input features for a more reliable result (ESRI, 2019). As found in previous studies, this study has found significant spatial clusters of disease even the input features is less than 30. Further significance will be investigated if the analyses are conducted on a smaller scale, i.e., municipalities level.

This study uses a default setting of the maximum cluster size at 50% of the population, but it is important to remember that there is no defined method to select a proper scanning window and timeframe for Kulldorf's scan statistics. Further research is needed to determine the appropriate window size and shape for analysis in Estonia.

Lastly, the findings of this study can serve as a basis for further studies of mortality in Estonia. For instance, the spatial dependencies of mortality rates with underlying socio-economic or environmental conditions could be examined to assess association or causation. Especially for mortality due to

malignant neoplasm, spatiotemporal analysis using specific cancer type disaggregated data could be conducted to investigate the contribution of each cancer type across space and time. Moreover, the results could be used to improve the unequal spatial accessibility of health services in areas that are more rural and isolated since rurality may contribute to the spatial disparity of mortality in Estonia.

Kokkuvõte

Lõputöö pealkiri: „Peamised surmapõhjused ja nende geograafiline aspekt Eestis ”

Autor: Mika Kuno

Kokkuvõte

Suremusstatistika on asjakohane teabeallikas elanikkonna tervise kohta. See võimaldab analüüsida praeguseid demograafilisi tingimusi, ennustada aja jooksul surmapõhjuste riski trende ja mustreid kindlates kogukondades ning määrata kindlaks suremapõhjuste võimalikke muutuseid tulevikus. GIS on oluline vahend haiguste geograafiliste variatsioonide uurimiseks. GIS-i ja ruumistatistika rakendamine võimaldab analüüsida haiguste ja neile järgneda võiva suremuse ajalis-ruumilisi omadusi.

Suurem osa Eesti suremusuuringutest näitab ainult staatilisi kaarte ega anna dünaamilist ülevaadet elanikkonna suremusest ajalis-ruumilisest vaatest lähtuvalt. Pealegi on kõige hiljutisemad suremapõhjuste uuringud keskendunud üldjuhul ainult ühele surma põhjusele ning on analüüsitud põhinedes erinevustele sotsiaaldemograafias, rahvuses, haridustasemes ja elukohas. Seetõttu kombineeriti antud uuringus GIS-i ja statistilist analüüsi, et uurida aastatel 2005–2019 kolme peamise surmapõhjuse ajalis-ruumilisi mustreid Eestis. Pahaloomuliste kasvajate, vereringesüsteemi haiguste ning vigastuste ja mürgituste poolt põhjustatud suremuse andmed aastate 2005–2019 Eesti kohta saadi Statistikaameti hallatavast Eesti Riikliku Tervise Arengu Instituudi surmapõhjuste registrist. Haiguste suremuse ajalis-ruumiliste mustrite uurimiseks rakendati leviala analüüsi, ruumilise autokorrelatsiooni analüüse ja Kulldorffi skanneeringu statistikat.

Leiud näitasid statistiliselt olulist ruumilist heterogeensust õnnetusjuhtumite, mürgistuste ja traumade põhjustatud suremuses kogu Eestis, kuid vereringeelundite haiguste ega pahaloomuliste kasvajate puhul seda ei leitud. Ruumilise klasteranalüüsi tulemused näitasid Ida-Viru maakonnas õnnetusjuhtumite, mürgistuste ja traumade suremuse märkimisväärset leviala aastatel 2008–2010, samas kui Lääne maakonnas oli see ebaoluline. Vereringesüsteemi haiguste ja pahaloomuliste kasvajate põhjustatud suremuse levialas ei olnud spetsiifilist mustrit; samas nende vähene levik leiti Lääne maakonnas aastatel 2017–2019 ja Tartu maakonnas aastatel 2008–2013.

Aegruumilise klasteranalüüsi põhjal tuvastati uurimisperioodil vähemalt kaks ruumilist suremuse kogumit erinevate haiguste ja spetsiifiliste ajaliste rühmade vahel. Üks neist tuvastati Harju maakonnas aastatel 2011–2017, mis viitas sellele, et selles maakonnas vähenes suremus teiste maakondadega võrreldes rohkem.

Selles uuringus kirjeldatakse GIS-i ja ruumistatistikat kasutades kolme peamise surmapõhjuse geograafilist ja ajalist jaotust Eestis ajavahemikul 2005–2019. Autorile teadaolevalt on see esimene uuring Eestis, milles hinnatakse kolme peamise surmapõhjuse kogumite olemasolu, kasutades ruumianalüüsi maakondlikul tasandil. Selle uuringu tulemused võivad olla aluseks suremuse edasistele uuringutele Eestis.

Acknowledgements

Foremost, I am extremely grateful to my supervisors Evelyn Uuemaa, Ph.D., Katrin Lang, Ph.D., and Hans Orru, Ph.D., for their guidance and continuous support throughout all stages of this thesis. Without their persistent help, I would not have finished my work. I would like to extend my sincerest thanks to everyone who supported and helped me to complete the graduate study.

References

Al-Ahmadi, K., Al-Zahrani, A., 2013. Spatial Autocorrelation of Cancer Incidence in Saudi Arabia. Int. J. Environ. Res. Public. Health 10, 7207–7228. https://doi.org/10.3390/ijerph10127207

Alene, K.A., Viney, K., McBryde, E.S., Clements, A.C.A., 2017. Spatiotemporal transmission and socio-climatic factors related to paediatric tuberculosis in north-western Ethiopia. Geospatial Health 12. https://doi.org/10.4081/gh.2017.575

Almendra, R., Santana, P., 2020. Mortality from cardiovascular diseases in the municipalities of mainland Portugal: spatiotemporal evolution between 1991 and 2017. Geogr. Environ. Sustain. 13, 128–133. https://doi.org/10.24057/2071-9388-2020-06

Anselin, L., 1995. Local Indicators of Spatial Association—LISA. Geogr. Anal. 27, 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Auchincloss, A.H., Gebreab, S.Y., Mair, C., Diez Roux, A.V., 2012. A Review of Spatial Methods in Epidemiology, 2000–2010. Annu. Rev. Public Health 33, 107–122. https://doi.org/10.1146/annurev-publhealth-031811-124655

Baburin, A., Aareleid, T., Rahu, M., Reedik, L., Innos, K., 2016. Recent changes in breast cancer incidence and mortality in Estonia: Transition to the west. Acta Oncol. 55, 728–733. https://doi.org/10.3109/0284186X.2015.1125014

Baptista, E.A., Queiroz, B.L., 2019. Spatial Analysis of Mortality by Cardiovascular Disease in the Adult Population: A Study for Brazilian Micro-Regions Between 1996 and 2015. Spat. Demogr. 7, 83–101. https://doi.org/10.1007/s40980-019-00050-6

Bermudi, P.M.M., Pellini, A.C.G., Rebolledo, E.A.S., Diniz, C.S.G., Aguiar, B.S. de, Ribeiro, A.G., Failla, M.A., Baquero, O.S., Chiaravalloti-Neto, F., 2020. Spatial pattern of mortality from breast and cervical cancer in the city of São Paulo. Rev. Saúde Pública 54. https://doi.org/10.11606/s1518-8787.2020054002447

Besag, J., Newell, J., 1991. The Detection of Clusters in Rare Diseases. J. R. Stat. Soc. Ser. A Stat. Soc. 154, 143–155. https://doi.org/10.2307/2982708

Bhunia, G.S., Shit, P.K., 2019. Geospatial Analysis of Public Health. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-01680-7

Brooker, S., Clarke, S., Njagi, J.K., Polack, S., Mugo, B., Estambale, B., Muchiri, E., Magnussen, P., Cox, J., 2004. Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya. Trop. Med. Int. Health 9, 757–766. https://doi.org/10.1111/j.1365-3156.2004.01272.x

Chen, C.-C., Teng, Y.-C., Lin, B.-C., Fan, I.-C., Chan, T.-C., 2016. Online platform for applying space–time scan statistics for prospectively detecting emerging hot spots of dengue fever. Int. J. Health Geogr. 15, 43. https://doi.org/10.1186/s12942-016-0072-6

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 U.S. cervical cancer mortality. Int. J. Health Geogr. 7, 57. https://doi.org/10.1186/1476-072X-7-57

Coleman, Marlize, Coleman, Michael, Mabuza, A.M., Kok, G., Coetzee, M., Durrheim, D.N., 2009. Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. Malar. J. 8, 68. https://doi.org/10.1186/1475-2875-8-68

Elliot, P., Wakefield, J.C., Best, N.G., Briggs, D.J., 2000. Spatial epidemiology: methods and applications. Spat. Epidemiol. Methods Appl.

Elliott, P., Wartenberg, D., 2004. Spatial Epidemiology: Current Approaches and Future Challenges. Environ. Health Perspect. 112, 998–1006. https://doi.org/10.1289/ehp.6735

English, D.R., 1992. Geographical epidemiology and ecological studies. Geogr. Environ. Epidemiol. Methods Small-Area Stud. 3–13.

Erdogan, S., 2009. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. J. Safety Res. 40, 341–351. https://doi.org/10.1016/j.jsr.2009.07.006

ESRI, 2019. Cluster and Outlier Analysis (Anselin Local Moran's I)—Help | ArcGIS Desktop [WWW Document]. URL https://desktop.arcgis.com/en/arcmap/10.6/tools/spatial-statistics-toolbox/cluster-and-outlier-analysis-anselin-local-moran-s.htm (accessed 3.8.21).

ESRI Inc., 2018. ArcMap, ArcGIS. ESRI Inc.

European Commission. Eurostat., 2013. Revision of the European Standard Population: report of Eurostat's task force : 2013 edition. Publications Office, LU.

Ge, E., Zhang, X., Wang, X., Wei, X., 2016. Spatial and temporal analysis of tuberculosis in Zhejiang Province, China, 2009-2012. Infect. Dis. Poverty 5, 1–10. https://doi.org/10.1186/s40249-016-0104-2

Gething, P.W., Elyazar, I.R.F., Moyes, C.L., Smith, D.L., Battle, K.E., Guerra, C.A., Patil, A.P., Tatem, A.J., Howes, R.E., Myers, M.F., George, D.B., Horby, P., Wertheim, H.F.L., Price, R.N., Müeller, I., Baird, J.K., Hay, S.I., 2012. A Long Neglected World Malaria Map: Plasmodium vivax Endemicity in 2010. PLoS Negl. Trop. Dis. 6, e1814. https://doi.org/10.1371/journal.pntd.0001814

Getis, A., Ord, J.K., 1992. The Analysis of Spatial Association by Use of Distance Statistics. Geogr. Anal. 24, 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Graham, A.J., Atkinson, P.M., Danson, F.M., 2004. Spatial analysis for epidemiology. Acta Trop. 91, 219–225. https://doi.org/10.1016/j.actatropica.2004.05.001

Green, C., Hoppa, R.D., Young, T.K., Blanchard, J.F., 2003. Geographic analysis of diabetes prevalence in an urban area. Soc. Sci. Med. 57, 551–560. https://doi.org/10.1016/S0277-9536(02)00380-5

Gross, D., 2007. The Association of Lung Cancer Mortality with Income and Education in Kentucky Counties 2.

Gundersen, L., 2000. Mapping it out: using atlases to detect patterns in health care, disease, and mortality. Ann. Intern. Med. 133, 161–164. https://doi.org/10.7326/0003-4819-133-2-200007180-00102

Hart, J.F., 1954. Central Tendency in Areal Distributions. Econ. Geogr. 30, 48–59. https://doi.org/10.2307/142234

Hjalmars, U., Kulldorff, M., Wahlqvist, Y., Lannering, B., 1999. Increased incidence rates but no space–time clustering of childhood astrocytoma in Sweden, 1973–1992. Cancer 85, 2077–2090. https://doi.org/10.1002/(SICI)1097-0142(19990501)85:9<2077::AID-CNCR27>3.0.CO;2-4

International Agency for Research on Cancer, 2020. Global Cancer Observatory [WWW Document]. URL https://gco.iarc.fr/ (accessed 1.10.21).

Jones, R.C., Liberatore, M., Fernandez, J.R., Gerber, S.I., 2006. Use of a Prospective Space-Time Scan Statistic to Prioritize Shigellosis Case Investigations in an Urban Jurisdiction. Public Health Rep. 121, 133–139. https://doi.org/10.1177/003335490612100206

Kadastik-Eerme, L., Taba, N., Asser, T., Taba, P., 2019. Incidence and Mortality of Parkinson's Disease in Estonia. Neuroepidemiology 53, 63–72. https://doi.org/10.1159/000499756

Kadohira, M., Stevenson, M.A., Kanayama, T., Morris, R.S., 2008. Epidemiology of bovine spongiform encephalopathy in cattle in Hokkaido, Japan, between September 2001 and December 2006. Vet. Rec. 163, 709–713. https://doi.org/10.1136/vr.163.24.709

Kaldmäe, M., Zemtsovskaja, G., Abina, J., Land, T., Viigimaa, M., 2017. Prevalence of cardiovascular disease risk factors in Tallinn, Estonia. Medicina (Mex.) 53, 268–276. https://doi.org/10.1016/j.medici.2017.07.002

Khamis, F.G., EL-Refae, G.A., 2012. Investigating spatial clustering of chronic diseases at governorate level in Iraq-2007. J. Appl. Sci. 12, 1564–1571. https://doi.org/10.3923/jas.2012.1564.1571

Kirby, R.S., Delmelle, E., Eberth, J.M., 2017. Advances in spatial epidemiology and geographic information systems. Ann. Epidemiol., GIS and Spatial Methods in Epidemiology Symposium 27, 1–9. https://doi.org/10.1016/j.annepidem.2016.12.001

Kulldorff, M., 1997. A spatial scan statistic. Commun. Stat. - Theory Methods 26, 1481–1496. https://doi.org/10.1080/03610929708831995

Kulldorff, M., Athas, W.F., Feuer, E.J., Miller, B.A., Key, C.R., 1998. Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos, New Mexico. Am. J. Public Health 88, 1377–1380.

Kulldorff, M., Feuer, E.J., Miller, B.A., Freedma, L.S., 1997. Breast Cancer Clusters in the Northeast United States: A Geographic Analysis. Am. J. Epidemiol. 146, 161–170. https://doi.org/10.1093/oxfordjournals.aje.a009247

Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R., Mostashari, F., 2005. A space-time permutation scan statistic for disease outbreak detection. PLoS Med. 2, e59. https://doi.org/10.1371/journal.pmed.0020059

Kulldorff, M., Nagarwalla, N., 1995. Spatial disease clusters: Detection and inference. Stat. Med. 14, 799–810. https://doi.org/10.1002/sim.4780140809

Kulldorff, M., Song, C., Gregorio, D., Samociuk, H., DeChello, L., 2006. Cancer Map Patterns: Are They Random or Not? Am. J. Prev. Med. 30, S37–S49. https://doi.org/10.1016/j.amepre.2005.09.009

Lai, T., Leinsalu, M., 2015. Trends and inequalities in mortality of noncommunicable diseases: Case study for Estonia. World Health Organ. 65.

Lawson, A.B., Banerjee, S., Haining, R.P., Ugarte, M.D. (Eds.), 2016. Handbook of Spatial Epidemiology, Handbook of Spatial Epidemiology. CRC Press. https://doi.org/10.1201/b19470

Leinsalu, M., Vågerö, D., Kunst, A.E., 2004. Increasing ethnic differences in mortality in Estonia after the collapse of the Soviet Union. J. Epidemiol. Community Health 58, 583–589. https://doi.org/10.1136/jech.2003.013755

Liisi Panov, 2018. Injuries in Estonia 2017. Natl. Inst. Health Dev.

Linton, S.L., Jennings, J.M., Latkin, C.A., Gomez, M.B., Mehta, S.H., 2014. Application of Space-Time Scan Statistics to Describe Geographic and Temporal Clustering of Visible Drug Activity. J. Urban Health 91, 940–956. https://doi.org/10.1007/s11524-014-9890-7

Maa-amet, 2020. Administrative and Settlement Division.

Manap, N., Borhan, M.N., Yazid, M.R.M., Hambali, M.K.A., Rohan, A., 2021. Identification of Hotspot Segments with a Risk of Heavy-Vehicle Accidents Based on Spatial Analysis at Controlled-Access Highway. Sustainability 13, 1487. https://doi.org/10.3390/su13031487

Marotta, P.L., Hunt, T., Gilbert, L., Wu, E., Goddard-Eckrich, D., El-Bassel, N., 2019. Assessing Spatial Relationships between Prescription Drugs, Race, and Overdose in New York State from 2013 to 2015. J. Psychoactive Drugs 51, 360–370. https://doi.org/10.1080/02791072.2019.1599472

Martins-Melo, F.R., Ramos, A.N., Alencar, C.H., Heukelbach, J., 2016. Trends and spatial patterns of mortality related to neglected tropical diseases in Brazil. Parasite Epidemiol. Control 1, 56–65. https://doi.org/10.1016/j.parepi.2016.03.002

Ministry of Social Affairs, 2018. "National Health Plan (NHP) 2009–2020“ Performance Report for 2013–2014.

Mollalo, A., Alimohammadi, A., Shirzadi, M.R., Malek, M.R., 2015. Geographic Information System-Based Analysis of the Spatial and Spatio-Temporal Distribution of Zoonotic Cutaneous Leishmaniasis in Golestan Province, North-East of Iran. Zoonoses Public Health 62, 18–28. https://doi.org/10.1111/zph.12109

Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37, 17–23. https://doi.org/10.2307/2332142

Morris, R.D., Munasinghe, R.L., 1994. Geographic Variability in Hospital Admission Rates for Respiratory Disease Among the Elderly in the United States. Chest 106, 1172–1181. https://doi.org/10.1378/chest.106.4.1172

National Institute for Health Development, 2020a. SD21: Deaths by cause, sex and age group [WWW Document]. Tervisestatistika Ja Terviseuuringute Andm. URL

http://statistika.tai.ee/pxweb/en/Andmebaas/Andmebaas__01Rahvastik__04Surmad/SD21.px/ (accessed 12.28.20).

National Institute for Health Development, 2020b. SD23: Standardised mortality rates per 100 000 inhabitants by cause of death, sex and county.

Naus, J.I., 1964. Clustering of Random Points in Line and Plane (Ph.D.). ProQuest Diss. Theses. Harvard University, United States -- Massachusetts.

Núñez-González, S., Delgado-Ron, J.A., Gault, C., Simancas-Racines, D., 2018a. Trends and Spatial Patterns of Oral Cancer Mortality in Ecuador, 2001–2016. Int. J. Dent. 2018, e6086595. https://doi.org/10.1155/2018/6086595

Núñez-González, S., Lara-Vinueza, A.G., Gault, C., Delgado-Ron, J.A., 2018b. Trends and Spatial Patterns of Suicide Among Adolescent in Ecuador, 1997-2016. Clin. Pract. Epidemiol. Ment. Health 14. https://doi.org/10.2174/1745017901814010283

Nyari, T.A., Ottóffy, G., Bartyik, K., Thurzó, L., Solymosi, N., Cserni, G., Parker, L., McNally, R.J.Q., 2013. Spatial clustering of childhood acute lymphoblastic leukaemia in Hungary. Pathol. Oncol. Res. 19, 297–302. https://doi.org/10.1007/s12253-012-9582-0

O'Connor, J.M., Sedghi, T., Dhodapkar, M., Kane, M.J., Gross, C.P., 2018. Factors Associated With Cancer Disparities Among Low-, Medium-, and High-Income US Counties. JAMA Netw. Open 1. https://doi.org/10.1001/jamanetworkopen.2018.3146

Odoi, A., Martin, S.W., Michel, P., Holt, J., Middleton, D., Wilson, J., 2003. Geographical and temporal distribution of human giardiasis in Ontorio, Canada. Int. J. Health Geogr. 2. https://doi.org/10.1186/1476-072X-2-5

Ojamaa, K., 2019. Epidemiology of gynecological cancer in Estonia.

Ojamaa, K., Veerus, P., Baburin, A., Everaus, H., Innos, K., 2019. Increasing incidence and survival of corpus uteri cancer in Estonia over the past two decades. Cancer Epidemiol. 62, 101566. https://doi.org/10.1016/j.canep.2019.101566

Osayomi, T., Areola, A.A., 2015. Geospatial Analysis of Road Traffic Accidents, Injuries and Deaths in Nigeria. Indones. J. Geogr. 47, 88–98. https://doi.org/10.22146/ijg.6749

Paneth, N., Kiely, M., Hegyi, T., Hiatt, I.M., 1984. Investigation of a temporal cluster of left sided congenital heart disease. J. Epidemiol. Community Health 38, 340–344. https://doi.org/10.1136/jech.38.4.340

Panov, L., 2018. Injuries in Estonia 2017. Natl. Inst. Health Dev.

Pärn, M.-L., Innos, K., Baburin, A., Kiivet, R.-A., Jaal, J., 2019. Gastric cancer trends in Estonia 1995–2014 by age, subsite, morphology and stage. Acta Oncol. 58, 283–289. https://doi.org/10.1080/0284186X.2018.1546058

Pfeiffer, D.U., Robinson, T.P., Stevenson, M., Stevens, K.B., Rogers, D.J., Clements, A.C.A., 2008a. Spatial clustering of disease and global estimates of spatial clustering, in: Spatial Clustering of Disease and Global Estimates of Spatial Clustering. Oxford University Press, pp. 32–44.

Pfeiffer, D.U., Timothy, R., Mark, S., Kim, S., David, R., Archie, C. (Eds.), 2008b. Spatial analysis in epidemiology. Oxford University Press, Oxford ; New York.

Prasannakumar, V., Vijith, H., Charutha, R., Geetha, N., 2011. Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment. Procedia - Soc. Behav. Sci., International Conference: Spatial Thinking and Geographic Information Sciences 2011 21, 317–325. https://doi.org/10.1016/j.sbspro.2011.07.020

Rahu, Kaja, Rahu, M., Zeeb, H., 2019. Sex disparities in premature adult mortality in Estonia 1995–2016: a national register-based study. BMJ Open 9. https://doi.org/10.1136/bmjopen-2018-026210

Rahu, K., Viiklepp, P., Villand, K., Pehme, L., Rahu, M., 2019. Respiratory tuberculosis incidence and mortality in Estonia: 30-year trends and sociodemographic determinants. Int. J. Tuberc. Lung Dis. 23, 112–118. https://doi.org/10.5588/ijtld.18.0388

Rajabi, M., Mansourian, A., Pilesjö, P., Åström, D.O., Cederin, K., Sundquist, K., 2018. Exploring spatial patterns of cardiovascular disease in Sweden between 2000 and 2010. Scand. J. Public Health 46, 647–658. https://doi.org/10.1177/1403494818780845

Rau, A., Munoz-Zanzi, C., Schotthoefer, A.M., Oliver, J.D., Berman, J.D., 2020. Spatio-Temporal Dynamics of Tick-Borne Diseases in North-Central Wisconsin from 2000–2016. Int. J. Environ. Res. Public. Health 17, 5105. https://doi.org/10.3390/ijerph17145105

Reile, R., Leinsalu, M., 2017. Ethnic variation in self-rated health–mortality association: Results from a 17-year follow-up study in Estonia. Medicina (Mex.) 53, 114–121. https://doi.org/10.1016/j.medici.2017.04.003

Rheenen, S. van, Watson, T.W.J., Alexander, S., Hill, M.D., 2015. An Analysis of Spatial Clustering of Stroke Types, In-hospital Mortality, and Reported Risk Factors in Alberta, Canada, Using Geographic Information Systems. Can. J. Neurol. Sci. 42, 299–309. https://doi.org/10.1017/cjn.2015.241

Santaularia, N.J., Larson, R., Uggen, C., 2021. Criminal punishment and violent injury in Minnesota. Inj. Epidemiol. 8, 11. https://doi.org/10.1186/s40621-021-00303-3

SaTScan, 2018.

Schneider, S., Vibo, R., Taba, N., Kõrv, J., 2020. Mortality in young adult patients with acute ischaemic stroke. Acta Neurol. Scand. 141, 242–249. https://doi.org/10.1111/ane.13217

Scott, L., Mobley, L.R., Il'yasova, D., 2017. Geospatial Analysis of Inflammatory Breast Cancer and Associated Community Characteristics in the United States. Int. J. Environ. Res. Public. Health 14, 404. https://doi.org/10.3390/ijerph14040404

Shah, S.A., Neoh, H.-M., Syed Abdul Rahim, S.S., Azhar, Z.I., Hassan, M.R., Safian, N., Jamal, R., 2014. Spatial Analysis of Colorectal Cancer Cases in Kuala Lumpur. Asian Pac. J. Cancer Prev. 15, 1149–1154. https://doi.org/10.7314/APJCP.2014.15.3.1149

Singh, G.K., Jemal, A., 2017. Socioeconomic and Racial/Ethnic Disparities in Cancer Mortality, Incidence, and Survival in the United States, 1950–2014: Over Six Decades of Changing Patterns and Widening Inequalities. J. Environ. Public Health 2017, e2819372. https://doi.org/10.1155/2017/2819372

Singh, H., Fortington, L.V., Thompson, H., Finch, C.F., 2016. An overview of geospatial methods used in unintentional injury epidemiology. Inj. Epidemiol. 3, 1–12. https://doi.org/10.1186/s40621-016-0097-0

Snow, J., Snow, J., Snow, J., 1936. Snow on cholera, being a reprint of two papers. The Commonwealth fund; London, Humphrey Milford, Oxford University Press, New York.

Song, J., Wen, R., Yan, W., 2018. Identification of Traffic Accident Clusters using Kulldorff's Space-Time Scan Statistics, in: 2018 IEEE International Conference on Big Data (Big Data). Presented at the 2018 IEEE International Conference on Big Data (Big Data), pp. 3162–3167. https://doi.org/10.1109/BigData.2018.8622226

Statistika andmebaas, 2020. Statistics Estonia 2020 [WWW Document]. URL http://andmebaas.stat.ee/Index.aspx?lang=en (accessed 10.30.20).

Syriopoulou, E., Rutherford, M.J., Lambert, P.C., 2021. Understanding disparities in cancer prognosis: An extension of mediation analysis to the relative survival framework. Biom. J. Biom. Z. 63, 341–353. https://doi.org/10.1002/bimj.201900355

Tiwari, N., Adhikari, C., Tewari, A., Kandpal, V., 2006. Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic. Int. J. Health Geogr. 5, 33. https://doi.org/10.1186/1476-072X-5-33

Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. Econ. Geogr. 46, 234–240. https://doi.org/10.2307/143141

Torres-Roman, J.S., Ruiz, E.F., Martinez-Herrera, J.F., Mendes Braga, S.F., Taxa, L., Saldaña-Gallo, J., Pow-Sang, M.R., Pow-Sang, J.M., La Vecchia, C., 2019. Prostate cancer mortality rates in Peru and its geographical regions. BJU Int. 123, 595–601. https://doi.org/10.1111/bju.14578

Tsai, P.-J., Lin, M.-L., Chu, C.-M., Perng, C.-H., 2009. Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. BMC Public Health 9. https://doi.org/10.1186/1471-2458-9-464

Uittenbogaard, A., Ceccato, V., 2012. Space-time Clusters of Crime in Stockholm, Sweden. Rev. Eur. Stud. 4, p148. https://doi.org/10.5539/res.v4n5p148

Vandenbroucke, J.P., Broadbent, A., Pearce, N., 2016. Causality and causal inference in epidemiology: the need for a pluralistic approach. Int. J. Epidemiol. 45, 1776–1786. https://doi.org/10.1093/ije/dyv341

Vaz, E., Shaker, R.R., Cusimano, M.D., 2020. A geographical exploration of environmental and land use characteristics of suicide in the greater Toronto area. Psychiatry Res. 287, 112790. https://doi.org/10.1016/j.psychres.2020.112790

Walter, S.D., 2001. Disease mapping: a historical perspective, Spatial Epidemiology. Oxford University Press.

Wang, Y., Du, Q., Ren, F., Liang, S., Lin, D.-N., Tian, Q., Chen, Y., Li, J.-J., 2014. Spatio-temporal variation and prediction of ischemic heart disease hospitalizations in Shenzhen, China. Int. J. Environ. Res. Public. Health 11, 4799–4824. https://doi.org/10.3390/ijerph110504799

World Health Organization, 2020. The top 10 causes of death [WWW Document]. URL https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (accessed 1.10.21).

World Health Organization, 2014. Injuries and violence: the facts 2014. World Health Organization.

Annexes

Annex1: Age distribution of European Standard Population, 2013

Total	100 000
0	1000
1–4	4000
5–9	5500
10–14	5500
15–19	5500
20–24	6000
25–29	6000
30–34	6500
35–39	7000
40–44	7000
45–49	7000
50–54	7000
55–59	6500
60–64	6000
65–69	5500
70–74	5000
75–79	4000
80–84	2500
85–89	1500
90–94	800
95–99	175
100+	25

Annex 2 Global Moran's I

Table A. Spatial autocorrelation patterns of deaths caused by diseases of the circulatory system in Estonia during 2005-2019

Year	Moran Index	Expected Index	Variance	z-score	p-value
2007	-0.19	-0.07	0.02	-0.88	0.378678
2010	-0.23	-0.07	0.02	-1.20	0.229428
2013	-0.26	-0.07	0.02	-1.42	0.154411
2016	-0.26	-0.07	0.02	-1.43	0.152501
2019	-0.24	-0.07	0.02	-1.08	0.277195

Table B. Spatial autocorrelation patterns of deaths caused by injury and poisoning in Estonia during 2005-2019

Year	Moran Index	Expected Index	Variance	z-score	p-value
2007	0.12	-0.07	0.019	1.36	0.173784
2010	0.33	-0.07	0.019	2.94	0.003281*
2013	-0.07	-0.07	0.019	0.01	0.989370
2016	-0.05	-0.07	0.018	0.17	0.863322
2019	0.07	-0.07	0.019	1.06	0.289581

*Statistically significant spatial autocorrelation

Table C. Spatial autocorrelation patterns of deaths caused by Malignant neoplasm in Estonia during 2005-2019

Year	Moran Index	Expected Index	Variance	z-score	p-value
2007	-0.19	-0.07	0.019	-0.88	0.378678
2010	-0.23	-0.07	0.019	-1.20	0.229428
2013	-0.26	-0.07	0.018	-1.42	0.154411
2016	-0.26	-0.07	0.017	-1.43	0.152501
2019	-0.22	-0.07	0.018	-1.09	0.277195

Annex 3

Table A. Spatial variation in temporal trends of deaths caused by diseases of the circulatory system in Estonia during 2005-2019

Cluster Type	County	Observed	Expected	R.R.	Inside time trend	Outside time trend	LLR	p-value
Lower*	Ida-Viru	4506	6495.65	0.67	10.72	3.2	155.52	0.001
Higher	Hiiu, Saare	8225	1945.06	4.75	1.08	4.05	57.58	0.001
Lower	Tartu, Põlva	7858	8269.74	0.94	1.9	4.14	28.88	0.001
Lower	Harju, Rapla, Järva	11180	28057.95	0.26	3.25	4.2	8.73	0.002

Note: Most likely clusters are noted with *

Table B. Spatial variation in temporal trends of deaths caused by injury and poisoning in Estonia during 2005-2019

Cluster Type	County	Observed	Expected	R.R.	Inside time trend	Outside time trend	LLR	p-value
Higher*	Ida-Viru, Lääne-viru	1580	1253.3	1.34	5.43	1.35	16.10	0.001
Lower	Harju, Lääne, Rapla, Järva, Jõgeva	1731	3374.39	0.35	0.29	2.89	7.61	0.006

Note: Most likely clusters are noted with *

Table C. Spatial variation in temporal trends of deaths caused by Malignant neoplasm in Estonia during 2005-2019

Cluster Type	County	Observed	Expected	R.R.	Inside time trend	Outside time trend	LLR	p-value
Lower*	Ida-Viru	1423	2458.96	0.55	15	7.68	42.02	0.001
Higher	Valga, Järva, Pärnu, Jõgeva, Tartu, Viljandi, Põlva, Võru	12010	7783.48	2.16	7.12	9.15	18.09	0.001

Note: Most likely clusters are noted with *

Non-exclusive licence

Non-exclusive licence to reproduce thesis and make thesis public

I, Mika Kuno,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, “Cause specific death rates and their geographic aspect in Estonia” supervised by Evelyn Uuema, Ph.D., Katrin Lang, Ph.D., and Hans Orru, Ph.D.
2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.
4. I certify that granting the non-exclusive licence does not infringe other persons’ intellectual property rights or rights arising from the personal data protection legislation.

Mika Kuno

24/05/2021