University of Zurich[UZH]

Department of Geography

# Geographical analysis of canine cancer risk in Switzerland

## A spatial-temporal analysis of cancer risk of the canine population in Swiss municipalities

GEO 511 Master's Thesis

**Author**

Guidetti, Vanessa Sarah

11-925-690

**Supervised by**

Prof. Dr. Fabrikant, Sara Irina

**Co-Supervised by**

Msc. Boo, Gianluca

**Faculty representative**

Prof. Dr. Fabrikant, Sara Irina

Date of Submission: Friday, 21 April 2017

Department of Geography, University of Zurich

# Table of Contents

**List of Figures** iv

**List of Tables** vii

**Abbreviations** viii

**Acknowledgements** ix

**Abstract** x

**Part I - INTRODUCTION** 1

1 Introduction 1

2 Motivation and Context 1

3 Aims 3

4 State of Research and Research Gaps 4

4.1 Research Gaps 8

5 Research Questions 9

6 Structure of This Thesis 10

**Part II - THEORETICAL BACKGROUND** 11

7 Relevant Research Areas 11

7.1 Health Science and Public Health 11

7.1.1 Health and Health Science 11

7.1.2 Public Health and Veterinary Public Health 12

7.1.3 Epidemiology 13

7.1.4 Spatial Epidemiology 14

7.2 Geography, Medical Geography and the role of GIS 14

7.2.1 Geography 14

7.2.2 Medical Geography 14

7.2.3 Spatial Analysis in GIS 15

8 Cancer Disease 16

8.1 Definition of Cancer 16

8.2 Cancer Situation in Switzerland 18

8.3 Risk Factors of Cancer 19

9 Companion Animals in Epidemiology 19

9.1 Companion Animals As Sentinels 20

9.2 Canine Cancer Registries 22

9.3 Confounders in Canine Cancer Research 23

10 Spatial Epidemiology 24

10.1 Epidemiological Studies 24
10.2 Epidemiological Data 25
10.3 Methods of Spatial Cancer Epidemiology 26
10.3.1 Disease Mapping 27
10.3.1.1 Data standardisation 28
10.3.1.2 Disease Mapping Methods 30
10.3.2 Disease Clustering 32
10.3.2.1 Cluster definition 33
10.3.2.2 Cluster statistics 35
10.3.3 Ecological Analysis 37

**Part III - METHODOLOGY AND RESULTS 39**
11 Methodology 39
12 Data Preprocessing and Exploration 40
12.1 Data Sets 40
12.1.1 Canine Cancer Cases 40
12.1.2 Canine Population 41
12.1.3 Geodata of Municipalities 42
12.2 Data Preprocessing and Classification 42
12.2.1 Data Preprocessing 43
12.2.2 Data Classification 43
12.3 Data Exploration 45
13 Disease Mapping 47
13.1 SMR Calculation 48
13.2 Mapping Cancer Risk 49
13.3 Results 51
13.3.1 Disease Maps of the Specific Tumour Types 53
14 Cluster Analysis 58
14.1 Testing the Homogeneity of the Cancer Risks 58
14.1.1 Results 58
14.2 Scan Statistics 59
14.2.1 Scan Statistic for Continuous Data 60
14.2.2 Results 63
14.3 LISA 64
14.3.1 Results 67
14.4 Conclusion 75

**Part IV – DISCUSSION 76**
15 Discussion 76
15.1 Uncertainties in Methodology 76
15.1.1 Disease Mapping 77

15.1.2 Cluster Analysis 79
15.1.2.1 SaTScan™ 80
15.1.2.2 LISA 81
15.2 Results 82
15.2.1 Data and Interpretation 82
16 Answering the Research Questions 86
17 Conclusion and Outlook 88
17.1 Reflection on Achievements 88
17.2 Outlook for Future Work 89

**Bibliography 91**

**Appendix I**

**Personal Declaration XIII**

# List of Figures

Figure 7.1 Overview of the main research fields with relevant sub-fields. 11
Figure 10.1. Cartogram3. The cube of three dimensions that are important for accurate reading of cartograms: shape preservation, topology preservation and visual equalisation (Roth et al., 2010). 31
Figure 10.2. Two approaches to making value-by-alpha maps (Roth et al., 2010). 32
Figure 10.3. The three components of clustering in space-time: spatial, temporal and spatio-temporal (Lawson, 2006b). 34
Figure 10.4. The scale and zoning effects (Lloyd, 2010). 38
Figure 12.1. Dogs per age class. 45
Figure 12.2. Malignant canine tumour per age class. 45
Figure 12.3. Percentage of malignant tumour in population per age class. 45
Figure 12.4. Dogs per sex class. 46
Figure 12.5. Malignant canine tumour per sex class. 46
Figure 12.6. Percentage of malignant tumour in population per sex class. 46
Figure 12.7. Dogs per size class. 46
Figure 12.8. Malignant Canine tumour per size class. 46
Figure 12.9. Percentage of malignant tumour in population per size class. 46
Figure 12.10. Main malignant tumour morphology groups from 2008 to 2013. 47
Figure 12.11. Most common specific malignant tumour diagnosis codes according to their morphology from 2008 to 2013. 47
Figure 13.1. Colour scheme of the value-by-alpha-map. 51
Figure 13.2. Disease map of relative cancer risk for all cancer records in Swiss municipalities from 2008 to 2013. 52
Figure 13.3. Disease map of relative cancer risk of 'mastocytoma/ mast cell sarcoma' in Swiss municipalities from 2008 to 2013. 53
Figure 13.4. Disease map of relative cancer risk of 'adenoma/ adenocarcinoma' in Swiss municipalities from 2008 to 2013. 54
Figure 13.5. Disease map of relative cancer risk of 'lipoma/ liposarcoma' in Swiss municipalities from 2008 to 2013. 54
Figure 13.6. Disease map of relative cancer risk of mixed tumours in Swiss municipalities from 2008 to 2013. 55
Figure 13.7. Disease map of relative cancer risk of 'melanoma' in Swiss municipalities from 2008 to 2013. 56
Figure 13.8. Disease map of relative cancer risk of 'trichoepithelioma' in Swiss municipalities from 2008 to 2013. 57
Figure 13.9. Disease map of relative cancer risk of 'adenocarcinoma of anal glands' in Swiss municipalities from 2008 to 2013. 57
Figure 13.10. Disease map of relative cancer risk of 'dermatofibroma/ dermatofibrosarcoma, NOS' in Swiss municipalities from 2008 to 2013. 57

Figure 13.11. Disease map of relative cancer risk of non-Hodgkin in Swiss municipalities from 2008 to 2013. ..... 57
Figure 13.12. Disease map of relative cancer risk of 'hemangioma/ hemangiosarcoma' in Swiss municipalities from 2008 to 2013. ..... 57
Figure 14.1. Equation of the chi-square test (Bivand et al., 1993). ..... 58
Figure 14.2. Equation to calculate the likelihood under the null hypothesis (Kulldorff et al., 2009) ..... 61
Figure 14.3. Equation to calculate the log likelihood under the null hypothesis (Kulldorff et al., 2009). ..... 61
Figure 14.4. Equation to calculate the maximum likelihood estimate under the alternative hypothesis for the common variance (Kulldorff et al., 2009). ..... 62
Figure 14.5. Equation to calculate the log likelihood under the null hypothesis (Kulldorff et al., 2009). ..... 62
Figure 14.6. Equation to calculate the maximum LLR (Kulldorff et al., 2009). ..... 62
Figure 14.7. SaTScan™ cluster map of relative cancer risk for all cancer types in Swiss municipalities from 2008 to 2013. ..... 63
Figure 14.8. SaTScan™ cluster persistency map of relative cancer risk for all cancer types in Swiss municipalities from 2008 to 2013. ..... 63
Figure 14.9. Equation to calculate the local Moran's I (Levine, 2004). ..... 64
Figure 14.10 Approaches of contiguity: Rooks, Bishops and Queen's Case (Sawada, 2004). ..... 65
Figure 14.11. LISA cluster map of relative cancer risk for all cancer types in Swiss municipalities from 2008 to 2013. ..... 67
Figure 14.12. LISA cluster persistency map of relative cancer risk for all cancer types in Swiss municipalities from 2008 to 2013. ..... 67
Figure 14.13. LISA cluster map of relative cancer risk for 'mastocytoma/ mast cell sarcoma' sarcoma in Swiss municipalities from 2008 to 2013. ..... 68
Figure 14.14. LISA cluster persistency map of relative cancer risk for 'mastocytoma/ mast cell sarcoma' sarcoma in Swiss municipalities from 2008 to 2013. ..... 68
Figure 14.15. LISA cluster map of relative cancer risk for 'adenoma/ adenocarcinoma' in Swiss municipalities from 2008 to 2013. ..... 69
Figure 14.16. LISA cluster persistency map of relative cancer risk for 'adenoma/ adenocarcinoma' in Swiss municipalities from 2008 to 2013. ..... 69
Figure 14.17. LISA cluster map of relative cancer risk for 'trichoepithelioma' in Swiss municipalities from 2008 to 2013. ..... 69
Figure 14.18. LISA cluster persistency map of relative cancer risk for 'trichoepithelioma' in Swiss municipalities from 2008 to 2013. ..... 69
Figure 14.19. LISA cluster map of relative cancer risk for 'lipoma/ liposarcoma' in Swiss municipalities from 2008 to 2013. ..... 70
Figure 14.20. LISA cluster persistency map of relative cancer risk for 'lipoma/ liposarcoma' in Swiss municipalities from 2008 to 2013. ..... 70

Figure 14.21. LISA cluster map of relative cancer risk for mixed tumours in Swiss municipalities from 2008 to 2013. ..... 71
Figure 14.22. LISA cluster persistency map of relative cancer risk for mixed tumours in Swiss municipalities from 2008 to 2013. ..... 71
Figure 14.23. LISA cluster map of relative cancer risk for 'melanoma' in Swiss municipalities from 2008 to 2013. ..... 71
Figure 14.24 . LISA cluster persistency map of relative cancer risk for 'melanoma' in Swiss municipalities from 2008 to 2013. ..... 71
Figure 14.25. LISA cluster map of relative cancer risk for 'adenocarcinoma of anal glands' in Swiss municipalities from 2008 to 2013. ..... 72
Figure 14.26. LISA cluster persistency map of relative cancer risk for 'adenocarcinoma of anal glands' in Swiss municipalities from 2008 to 2013. ..... 72
Figure 14.27. LISA cluster map of relative cancer risk for 'dermatofibroma/ dermatofibrosarcoma, NOS' in Swiss municipalities from 2008 to 2013. ..... 73
Figure 14.28. LISA cluster persistency map of relative cancer risk for 'dermatofibroma/ dermatofibrosarcoma, NOS' in Swiss municipalities from 2008 to 2013. ..... 73
Figure 14.29. LISA cluster map of relative cancer risk for 'non-Hodgkin's lymphoma' in Swiss municipalities from 2008 to 2013. ..... 74
Figure 14.30. LISA cluster persistency map of relative cancer risk for 'non-Hodgkin's lymphoma' in Swiss municipalities from 2008 to 2013. ..... 74
Figure 14.31. LISA cluster map of relative cancer risk for 'hemangioma/ hemangiosarcoma' in Swiss municipalities from 2008 to 2013. ..... 74
Figure 14.32. LISA cluster persistency map of relative cancer risk for 'hemangioma/ hemangiosarcoma' in Swiss municipalities from 2008 to 2013. ..... 74

# List of Tables

Table 12.1. Example of the Canine Cancer Registry data structure with the most important variables. ..... 41
Table 12.2 Example of the reference population data structure with the most important variables. ..... 42
Table 12.3. Example of an incorrect entry in the Canine Cancer Registry. ..... 43
Table 12.4. Age classification. ..... 44
Table 12.5. Sex classification. ..... 44
Table 12.6. Indices of information source ..... 44
Table 12.7. Size classification. ..... 44
Table 12.8. Specific (ICD-O-3) tumour code and diagnosis of the most common cancer types. ..... 47
Table 13.1. Example of ratio calculation of cancer cases to population per strata according to confounding variables. ..... 48
Table 13.2. Example of strata population per Swiss municipality ..... 49
Table 13.3. Example of final SMR calculation per Swiss Municipality. ..... 49
Table 13.4. Classification of SMR values. ..... 50
Table 14.1. Results of homogeneity test (p-values) per cancer type and year. ..... 59

# Abbreviations

| | |
|---|---|
| CDC | *Centers for Disease Control and Prevention* |
| CSR | *Complete Spatial Randomness* |
| E | *Expected cancer cases* |
| FCI | *Fédération Cynologique Internationale* |
| FDHA | *Federal Department of Home Affairs* |
| FOPH | *Federal Office of Public Health* |
| FSO | *Federal Statistical Office of Switzerland* |
| GIS | *Geographic Information System* |
| GPS | *Global Positioning System* |
| IACR | *International Agency for Research on Cancer* |
| ICD-O | *International Classification of Diseases for Oncology* |
| IQR | *Interquartile range* |
| ITPA | *Institute of Animal Pathology* |
| IVPZ | *Institute of Veterinary Pathology Zurich* |
| LISA | *Local Indicators of Spatial Association* |
| LLR | *Log likelihood ratio* |
| lnL | *Log likelihood* |
| MAUP | *Modifiable areal unit problem* |
| NCI | *National Cancer Institute* |
| O | *Observed cancer cases* |
| OECD | *Organisation for Economic Co-operation and Development* |
| RQ | *Research question* |
| SCCR | *Swiss Canine Cancer Registry* |
| SMR | *Standardised mortality (or morbidity) ratio* |
| VPH | *Veterinary public health* |
| WCRF | *World Cancer Research Fund International* |
| WHO | *World Health Organisation* |

# Acknowledgements

This work only became what it is now thanks to the support of several people. I would especially like to thank my two supervisors Prof. Dr. Sara Irina Fabrikant and Gianluca Boo for their time, helpful inputs and encouragement throughout the whole research process of this master thesis. Without their support, this work would not have been possible. Special thanks go to Gianluca Boo for his support and patience with technical issues.

I would also like to take this opportunity to thank the team of the Collequium Helveticum for the provision of the cancer data used in this study, with a special mention for Ramona Graf, who helped with classification issues of canine cancer types and canine sizes. Last but not least, thanks go to Martina Henzi, Tanja Cervilla and Adrian Vetterli for proof-reading this master thesis.

# Abstract

Cancer is one of the leading causes of death in humans as well as companion animals. Since malignant tumours in companion animals also resemble those in humans, they can provide useful data in combating cancer. The overarching aim of this thesis is to contribute to a development in which the population can be healthier. Accordingly, the object is to localise spatio-temporal clusters of increased cancer risk that provide information for the prevention, treatment and combat of cancer in the Swiss Health System. For this purpose, the following research question is developed: "What is the spatial distribution of the occurrence of cancer from 2008 to 2013 in Switzerland?".

The research question is answered through two methods of spatial epidemiology: disease mapping and cluster analysis. Therefore, a total of 20,157 malignant tumours recorded in the Swiss Canine Cancer Registry are analysed for the period from 2008 to 2013. These diagnostic data are correlated with data on the Swiss canine population. The dogs are classified according to age, sex, size and place of residence and the standardised morbidity rate of cancer in dogs was investigated in a spatio-temporal context.

As a result, a set of spatio-temporal clusters is provided for all cancer records together. Considering the most common canine cancer types, spatio-temporal cluster centres can also be localised for the malignant tumour types 'melanoma' in the municipalities of Bern, Regensdorf, Bülach and Dübendorf and for 'lipoma/ liposarcoma' in the Zurich, Kloten und Maur. These areas may be of interest for further epidemiological studies, however, these findings on canine cancer risk should be carefully considered before being compared to humans. Further investigations are necessary, since the results of the analysis are highly influenced by the conducted methodology, the quality and aggregation level.

# Part I - INTRODUCTION

## 1 Introduction

Cancer is a disease that affects many people globally. Worldwide, almost one in six deaths is attributable to cancer and every year more than 14 million people are diagnosed with this disease (Ferlay et al., 2013). Cancer is also a main cause of death in Switzerland. In 2014 more than 16,000 humans died of cancer (FSO, 2016c). For the future, an increase in cancer incidence and cancer mortality is prognosticated (Ferlay et al., 2013). In order to counteract this development, an early prevention of cancer disease is important (FSO, 2016b). Since many known and unknown factors exist that affect the development of cancer, a better understanding of the spatial and temporal occurrence of cancer is crucial (Chammartin et al., 2016). Accordingly, the present study deals with this issue.

This chapter focuses on this thesis' motivation and context. Furthermore, it presents the state of research and gaps in the current research and highlights the aims of the thesis. The introduction concludes with the research questions and the structure of this work.

## 2 Motivation and Context

The main motivation of this work is to contribute to the development in the scientific field of public health. Winslow (1920) defines public health as a discipline aiming to prevent disease, prolong life and promote health through societal actions (Winslow, 1920). Put briefly, the aim of this science is to create conditions in which the population can be healthy (Schweikart & Kistemann, 2004). Since cancer is one of the leading causes for humans' morbidity and mortality worldwide (Ferlay et al., 2015), this master thesis focuses on cancer disease.

In 2012, approximately 14 million new cancer cases were diagnosed around the world and the number of new cancer cases is expected to rise. Experts forecast a rise by about 70 percent over the next two decades (Ferlay et al., 2015). In 2015, cancer was responsible for the death of 8.8 million people. Globally, it is the second most frequent cause of death. The main causes of cancer mortality are cancers of lung (1.69 million deaths), liver (788,000

deaths), colorectal (774,000 deaths), stomach (754,000 deaths) and breast (571,000 deaths) (WHO, 2017).

Cancer diseases are due to changes in the genome of the cells. In addition to genetic factors, risk factors are known that promote such changes and may increase a person's chances of cancer development (NIH, n.d.; WHO, 2017). These external factors may vary in geographic space. Therefore, the geographical location of cancer cases is an important information in cancer epidemiology (NIH, n.d.).

The number of people in our society and thus the incidence of chronic disease is growing. This development also applies to Switzerland. In addition to the high incidence and mortality rate, high treatment and care costs burden the health care system enormously (FDHA, 2014).

Existing structures of the Swiss healthcare are geared to acute care. Furthermore, the system is non-transparent and difficult to manage (FOPH, 2017). In 2011, the World Health Organisation (WHO) and the Organisation for Economic Co-operation and Development (OECD) criticised the organisation of the Swiss Health System in a report. The recommendations of the WHO and OECD include strengthening information systems, improving analytical capacities as well as increasing accountability and transparency (OECD & WHO, 2011).

In consideration of the recommendations of the WHO and OECD, priorities of Swiss health policy are compiled in the concept 'Health 2020'. It aims to improve all areas of the system and sustainably eliminate the weaknesses in the Swiss Health System. Referring to the Federal Department of Home Affairs (FDHA) and its Federal Office of Public Health (FOPH) of Switzerland, this is to ensure that people suffering from illnesses and the consequences of accidents continue to receive high-quality care (FDHA, 2014; FOPH, 2017). 'Health 2020' includes a nationwide cancer registry that also obligates the cantons to store spatial information of cancer occurrence. It aims to improve the prevention, early detection, care and treatment of cancer. The adopted draft law governs the collection, registration and transfer of the cancer data at the cantonal level. In addition, medical professionals and health care institutions have a duty to report cancer cases (FDHA, 2014). Such diagnostic records of cancer are an essential element of cancer epidemiology, prevention or control strategies for both humans and animals. The cancer registers store

information such as tumour type, malignancy and body location as well as spatial information about the occurrence (Grüntzig et al., 2015). Spatial epidemiology occupies with the investigation of health outcomes in geographic spaces; thus, spatial information is important for analyses of cancer registry data (Lawson, 2006b).

However, the data of the human cancer registry is currently not available for research purpose and has high restrictions due to data privacy (Federal Council of Switzerland, 2014).

Given that cancer is not only a main cause of mortality in man but also in dogs (Pinho et al., 2012), this paper focuses on the spatial occurrence of cancer cases in dogs. Studies indicate that many diseases can be found in humans as well as animals and that they could be linked to their shared environment (Reif, 2011; Van der Schalie et al., 1999). Furthermore, animals are sensitive indicators of environmental hazards and provide an early warning system for public health intervention. Accordingly, cancer data of companion animals can be used to investigate the cancer risk and environmental cancer risk factors (Reif, 2011).

# 3 Aims

The overarching aim of this master thesis is to connect the two research fields GIScience and spatial epidemiology to contribute to a development in which the population can be healthier. By connecting epidemiological methods with geographic information system (GIS) techniques, the goal is to better understand the disease occurrence in Switzerland. Three aims are associated with this goal. The first goal is to investigate the distribution of cancer diseases in Switzerland and provide a set of disease clusters that could also be valid for humans. In an exploratory manner, it will be studied whether patterns exist in the occurrence of canine cancer and how these patterns differ in space and time. Thus, the aim is to identify sites with very high cancer occurrence – so-called hot spots – over a longer period. These sites can provide information about possible external risk factors and may be fundamental in further epidemiological investigations. The second aim of this thesis is to promote the use and prove the effectiveness of epidemiological spatio-temporal analysis methods in studying companion animal disease risk. In this context, the present work aims to establish the use of companion animals as sentinels for diseases and further

endorse the values of their georeferenced disease registries in Switzerland. With reference to the wider context of this thesis, the third aim is to bring up the value of GIScience to product and exchange knowledge between disciplines such as health science and promote the use of geostatistical methods for a better understanding of the linkage between space, time and diseases. Using the example of this thesis, it is demonstrated that geostatistical methods can contribute to the development of the Swiss Health System.

# 4 State of Research and Research Gaps

The use of spatial information for a better understanding of disease is not new. Early on, spatial analyses of maps were used to associate diseases with spatial factors (Richardson et al., 2013). In the 18th century, Giovanni Maria Lancisi, a personal physician of the pope in Rome, led the decline of infectious diseases to improvements in hygiene and draining of swamps (Porta et al., 2014). However, the first recognised epidemiological study is the famous cholera map of London by Snow from the 19th century (Schweikart & Kistemann, 2004). In a map, Snow illustrated a connection of cholera cases and the drinking water supply. By spatially analysing the disease, Snow could detect spatial pattern of infections around the drinking water pumps (Snow, 1855). The studies of Snow leveraged the use of maps for spatial analysis in healthcare (Anthamatten & Hazen, 2011). In order to link health issues with possible external factors, knowledge about the spatio-temporal distribution of the health issues is essential (Jarup, 2004). Epidemiological studies on cancer risk factors increased over the years. However, many studies focused on influences of non-spatial factors like lifestyle variable such as diet and smoking (Elliott & Wartenberg, 2004). Geographic information systems were used late and have not yet been widely established in the health sector though the potential of GIS and the use of geostatistical methods in spatial epidemiology is high and can help improve health care (Anthamatten & Hazen, 2011). Schweikart and Kistemann (2004) argue that this is mainly due to the late linkage of the two disciplines geoscience and medicine and the lack of transfer knowhow. Further factors are the difficult handling of health data with respect to data protection and the scepticism about the availability and reliability of health-relevant data (Schweikart & Kistemann, 2004). Nonetheless, recent developments have stimulated dramatic growth in the application of spatial approaches to health issues and geostatistical methods in

epidemiological studies (Anthamatten & Hazen, 2011). As an example, Kinoshita, Wagatsuma and Okada (2007) examined the relation of the geographical distribution for malignant neoplasm of the pancreas to selected climatic factors in Japan. They conducted a multiple linear regression analysis and found correlations between the mortality rate of this cancer and low solar radiation as well as low temperature (Kinoshita, Wagatsuma & Okada, 2007).

Moreover, epidemiological studies based on georeferenced health data are carried out in Switzerland. A recent study investigated the main causes of human death including cancer in Switzerland from 2008 to 2012 by analysing gender and age-adjusted mortality data of the Swiss population. Among their most important findings regarding cancer death are the differences between the linguistic regions as well as urbanisation in Switzerland. For instance, the standardised mortality rate of prostate cancer was lower in the Italian-speaking part of Switzerland than in the German-speaking part. The standardised morbidity ratio (SMR) of people of 75 years and older, lower SMR could be identified in urban than in rural settings. They further showed that deaths due to cancers usually affected men more often than women and that external causes of death were responsible for six percent of the recorded deaths in the entire period. The authors argue that deepening the understanding of spatial variation of major causes of death such as cancer is crucial for targeting preventive measures, changing behaviours and a more cost-effective allocation of health resources (Chammartin et al., 2016).

However, according to Lam (2012), the interpretation of studies based on mortality data can be difficult since uncertainties exist whether the disease was actually the cause of death and cancer cases that were cured of the disease are excluded . Furthermore, mortality rates do not provide information about the time of the disease development and thus a link to possible environmental risk factors is difficult (Lam, 2012).

Human cancer registries are relatively novel and still under construction. For this reason, cancer data is (spatially) not very comprehensive. Since the access as well as the spatial resolution restricted due to data protection, human cancer data is currently not available for research purposes (Federal Council of Switzerland, 2014). Additionally, the latency time for cancer development in humans can be relatively long (Reif, 2011).

One solution of this problem is to analyse cancer data of animals. This concept that non-human organisms such as animals may serve as early warning systems for human health risk is not novel. One of the most widely-known applications of animals as sentinel for monitoring is the miner's canary. Because birds are more sensitive to the invisible lethal carbon monoxide than humans, miners used canaries as warn system of potentially concentrations of this gas in coal mines. The birds were taken into coal mines, whereby their death from exposure to it gave miners time to race to the surface (Burrell & Seibert, 1916; Schwabe, 1984).

The number of studies on companion animal cancer is constantly growing due to the increasing data availability (Bukowski & Wartenberg, 1997; Reif, 2011). For instance, recent investigations of Reif (2011) could link increased risk for malignant lymphoma and testicular and bladder cancer in dogs to exposures to pesticides. These canine cancer types are known models for 'non-Hodgkin's lymphoma' in humans (Reif, 2011). According to Bukowski and Wartenberg (1997) an important limitation in many veterinary epidemiological studies is the lack of population and disease data (Bukowski & Wartenberg, 1997). In Switzerland, however, cancer registries for both cats and dogs as well as the ANIS and AMICUS databases containing records of companion animals exist. Swiss investigations using companion animals as sentinels in cancer research are mainly carried out by the Collegium Helveticum. It comprises scientists from different disciplines and makes transdisciplinary research itself a research subject. The project "One Medicine – one Oncology" focuses on the incidence and geographical distribution of tumours in dogs and cats in Switzerland by evaluating data of the animal cancer registries (Collegium Helveticum, n.d.). To date, their research has focused on examining correlations between breed, sex and age for both cats and dogs based on cancer data from 1955 to 2008. Graf et al. (2015) conducted a retrospective study on tumours in felines in Switzerland based on cancer incidence data of the Swiss Feline Cancer Registry from 1965 to 2008 (Graf et al., 2015). They analysed the influence of sex, breed, neutering status, age and time on the development of the most common tumour types and locations by using a multiple logistic regression model. Among their most important findings are that differences in odds ratios for developing cancer exist between breeds. Further they proved that the odds of developing a tumour increases with age of a feline. Tumours were more frequent detected

for middle-aged and older than for young felines. Furthermore, they also showed that females have a higher risk of tumour development compared with males (Graf et al., 2015, 2016). Since the Canine Cancer Registry provides the most comprehensive and most durable animal cancer registry in Switzerland (Grüntzig et al., 2015), the present work is based on canine cancer.

Similar to the investigations on Swiss Feline Registry data, the study of Grüntzig et al. (2015) a retrospective study examined the occurrence of tumours in dogs in Switzerland from 1955 to 2008. In this study, the dogs were classified according to breed, age, sex, neuter status and place of residence and the incidence of cancer in dogs was investigated based on the Swiss general canine population. Considering all tumours, benign and malignant, the study showed that the annual incidence rates of tumours are relatively stable for the period from 1955 to 2008. They showed that most of the tumours were located in the skin, mammary gland and soft tissue and that a large variety of tumour types can be identified in one location. For instance, in the mouth and the pharynx seven different tumour types could be found. Almost half of all recorded tumours are malignant. Of these tumours, increased incidence could be identified for skeletal tumours, 'melanoma', 'gonadal germ cell tumours', 'epithelial tumours' and 'lymphoid tumours'. Furthermore, the age distribution of the cases was analysed. Most of the dogs diagnosed with tumours were between five and ten years of age. Another large group exists for dogs older than ten years of age. 15 percent of the dogs are aged between one and five years and only about seven percent were younger than one year old (Grüntzig et al., 2015). Currently, the researchers are investigating the new canine cancer data from 2008 to 2013. Boo (2016) conducted some first spatial analysis on the Swiss Canine Cancer data of 2008 to 2013 in his PhD project "Space-Time Analysis of Tumour Incidence: Companion Animals as Public Health Sentinels". This project includes the examination of the correlation between the 'melanoma' occurrence aggregated for the entire period and several explanatory variables. He showed that increasing distance to the veterinary results in less 'melanoma' incidences rate. Further he showed that increasing sun exposure, average income and average age can be associated with more 'melanoma' incidences rate (Boo, 2016). My work takes this spatial analysis further.

### 4.1 Research Gaps

In the new strategy 'Health 2020', spatial analyses of health data are given a subordinate role and the use of GIS in health care hasnot yet been established (FDHA, 2014; Schweikart & Kistemann, 2004). Even though epidemiological studies on environmental risk factors for both humans and animals have increased globally over the years, investigations concerning the influence of spatial information and the analysis of temporal variations remain rare. Studies on Swiss Canine Cancer Registry have rarely analysed the distribution in space and time as well as the relation to environmental risk factors. Furthermore, they have not yet distinguished the different cancer types of dogs (Collegium Helveticum, n.d.).

Many spatial clustering methods assess the spatial distribution of disease events within a fixed time period. This aggregation of data over time ignores the temporal variation within these time periods and is thus problematic, since some evidence of the variation in spatial variation will be lost (Lawson, 2006b; Marshall, 1991).

For these reasons, the present work focuses on the spatio-temporal analysis of the canine cancer distribution for cancer records as well as the most common cancer types in Switzerland from 2008 to 2013.

# 5 Research Questions

Having identified the research gaps, the research questions can be formulated. The central research questions (RQ) of this thesis address the spatial and temporal distribution of the relative cancer risk. According to this, the following questions are raised.

**Spatial distribution of canine cancer records**

RQ I *"What is the spatial distribution of the occurrence of cancer from 2008 to 2013 in Switzerland?"*

(a) *"Do significant **spatial** patterns exist in the geographical distribution of increased canine cancer risk or is the distribution random?"*

(b) *"How do these spatial patterns vary from 2008 to 2013, do significant **spatio-temporal** patterns exist in the geographical distribution of increased canine cancer risk?"*

Spatio-temporal patterns of the relative cancer risk are studied using analysis methods of spatial epidemiology and techniques of GIScience. RQ I (b) is based on RQ I (a). Consequently, answering the question RQ I (b) will highly depend on the results of the first question RQ I (a). RQ II addresses the distribution of the most common canine cancer types and is based on RQ I and focuses on the most common canine cancer types. Answering this second question will highly depend on the results of the first question.

**Spatio-temporal distribution of the most common canine cancer types**

RQ II *"How do these patterns of canine cancer risk differ between the most common cancer types?"*

# 6 Structure of This Thesis

This thesis is structured in four main parts. Subsequent to the introduction, the theoretical background is the focus of the second part of this thesis. In chapter 7, the work is embedded in research and the relevant fields of research are defined. Since this thesis focuses on cancer disease, chapter 8 addresses the definition of cancer, the current disease situation for the Swiss population and relevant risk factors. Subsequently, a chapter on the importance of companion animals in spatial epidemiological studies follows. The methods of spatial epidemiology that will be later used for the practical part of this thesis are introduced in chapter 10.

The third part of this thesis contains the analyses and the results. Chapter 11 deals with the data preprocessing and explores the data. In chapters 12, 13 and 14, the three methods are applied. At the end of each of these respective chapters, the results are presented and summarised. The final part concludes the thesis by discussing the methods and results in chapter 15 and answering the research questions in chapter 16. Chapter 17 comprises a reflection on the achievements and an outlook for future work.

# Part II - THEORETICAL BACKGROUND

This part of the master thesis presents the relevant scientific background of the related research fields. The first chapter embeds the work in research and defines the relevant fields of research "Health Science" and "Geography". Chapter 8 focuses on cancer disease. A definition of cancer, the current disease situation for the Swiss population and risk factors of cancer are presented. Subsequently, a chapter on the importance of companion animals in spatial epidemiological studies follows. In this chapter, the usage of companion animals as sentinels, canine cancer registries and confounding variables are introduced. Chapter 12 presents studies, data and the methods "disease mapping", "disease clustering" and "ecological analysis" of spatial epidemiology that will be later used for the practical part of this thesis.

## 7 Relevant Research Areas

This chapter provides an overview of the research areas which are relevant for this master thesis. As outlined in figure 7.1, this thesis touches the main fields of research "Health Science" and "Geography". It is an interdisciplinary work because it uses approaches and ways of thinking of the two disciplines. The research fields are explained below.

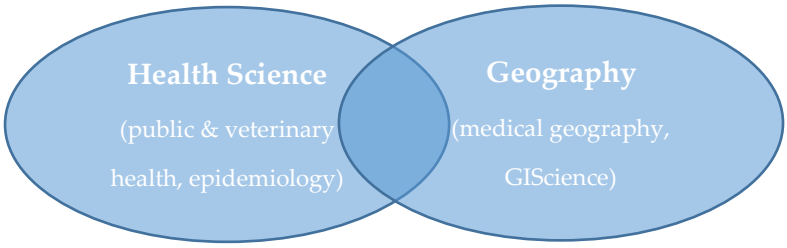



Figure 7.1 Overview of the main research fields with relevant sub-fields.

### 7.1 Health Science and Public Health

#### *7.1.1 Health and Health Science*

Health is largely culturally defined and can have many dimensions such as anatomical, mental and physiological (WHO, 2004). The WHO defines health as *"…the state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity."* (WHO, 2004, p. 28).

Reframing health in this way is not without problems since it is utopian and precludes any individual from actually achieving "health". Nonetheless, the definition of health encourages being open to alternative perspectives that consider health as holistic terms (Anthamatten & Hazen, 2011; Mayer & Meade, 1994).

Health science refers to the definition of health, as a generic term that includes many different disciplines such as basic science, medical geography, medicine and epidemiology (Last, 2007).

Health science research addresses complex themes such as the health status, health risks and healthcare as well as utilisation, costs and structures of the target populations. Thus, studies of health science also include elements such as climate, seasons, water quality and dietary habits of the population (Schweikart & Kistemann, 2004).

The research activities contribute to the professional expertise, knowledge and skills of members of the health professions (Last, 2007)

### *7.1.2 Public Health and Veterinary Public Health*

The discipline of public health is a main core of health science (Anthamatten & Hazen, 2011). In 1920, Charles-Edward Amory Winslow defined public health as follows:

> *"Public health is the science and the art of preventing disease, prolonging life and promoting physical health and efficiency through organized community efforts for the sanitation of the environment, the control of community infections, the education of the individual in principles of personal hygiene, the organization of medical and nursing service for the early diagnosis and preventive treatment of disease, and the development of the social machinery which will ensure to every individual in the community a standard of living adequate for the maintenance of health."* (Winslow, 1920, p. 30)

According to this definition, public health considers population-wide processes and emphasises health-promoting behaviours that generate benefit for a population group. Public health science often involves exploring the interaction of individuals with each other as well as with their environment (Anthamatten & Hazen, 2011). It includes all population-related measures (public or private) which contribute to disease prevention,

promotion of physical and psychological well-being and prolongation of life (Schweikart & Kistemann, 2004; Winslow, 1920). Put briefly, the overarching aim of public health is to create conditions in which the population can be healthy (Schweikart & Kistemann, 2004). Public health science is a collective term for all scientific activities (including epidemiology) that form the scientific base for public health practice, services and systems (Porta et al., 2014). Veterinary public health (VPH) is part of public health and can be defined as "*the sum of all contributions to the physical, mental and social well-being of humans through an understanding and application of veterinary science*" (WHO, 1999).

### *7.1.3 Epidemiology*

Epidemiology is a core area of public health and an interdisciplinary field combining methodological components of different disciplines. The roots of the word “Epidemiology” come from the Greek. It is a combination of the three word components "epo" (= over), "demos" (= the people) and "logos" (= the word, transcending the doctrine) (Schweikart & Kistemann, 2004).

According to Porta et al. (2014) , the term can be defined as the study of the occurrence and distribution and the determinants of health-related events, states, and processes in specified population groups and the control of health problems by applying the acquired knowledge. Epidemiology may also include the study of disease in populations of animals and plants. The analysis by time, place (or space), and population is essential to study the occurrence and distribution. Determinants influencing health can be of geophysical, biological, behavioural, social, cultural, economic or political nature. Studies of epidemiology can include different approaches such as surveillance, observation, hypothesis testing, analytic research or prediction. The main objective of epidemiology is to promote, protect and restore health and advance scientific knowledge (Porta et al., 2014).

In accordance with the above definition, cancer epidemiology is a sub-specialty of epidemiology, analysing the distribution of cancer occurrence and cancer-specific deaths taking into account social, economic, ecological and demographic conditions in space and time (Schweikart & Kistemann, 2004).

### *7.1.4 Spatial Epidemiology*

Spatial epidemiology is part of descriptive epidemiological studies. It is also termed as geographical epidemiology, which underlines the connection to (medical) geography. It can be defined as "…*the description of spatial patterns of disease morbidity and mortality…with the aim of formulating hypotheses about the aetiology*[1] *of diseases*" (Bailey, Fotheringham & Rogerson, 1994, p. 15). Although epidemiological investigations with spatial aspects are among the earliest epidemiological studies, spatial epidemiology, in particular veterinary spatial epidemiology, is a very young discipline within epidemiology. The methodical development of spatial epidemiology began only about 50 years ago (Berke, 2003).

## 7.2 Geography, Medical Geography and the role of GIS

### *7.2.1 Geography*

Geography is a broad discipline and many definitions exist. Geography uses a spatial, or geospatial perspective to study what kind of phenomena where and why occur (CDC, 2016). Moseley et al. (2007) define geography as a discipline that "…*seeks to understand and study the spatial organization of human activity and of people's relationships with their environment*" (Moseley, Lanegran & Pandit, 2007, p. 2). Geographical studies also investigate the interdependence among places and regions while respecting the individuality and uniqueness of specific places (Moseley, Lanegran & Pandit, 2007). Different sub-disciplines of geography exists. This work takes place in GIScience, briefly defined as a research area that uses technologies such as GIS to understand the world (Clarke, 2003). The focus is on Geographical Information, Visualisation and Analysis (GIVA).

### *7.2.2 Medical Geography*

Scientists try to answer health questions through instruments and approaches of geography, focusing on differences in space and using maps and geostatistical methods

[1] Aetiology defines the study of causes, causation or causality, usually applying to disease (WHO, 2004).

to identify spatial patterns and influences (Anthamatten & Hazen, 2011; Meade, 2014). This sub-discipline of geography is termed as medical geography, or health geography. The discipline studies the spatial aspects of health and illness. Many subjects of medical geography studies overlap with public health research, but medical geographers focus on place to move beyond traditional questions of public health (Anthamatten & Hazen, 2011). The geographic aspects of health care are mainly concerned with the question "where?". Medical geography looks at questions such as: "Where do diseases affect people, animals or plants?", "Where are the agents located that cause the disease?", "Where does something occur and why is it significant?" or " Where can we intervene to eliminate risks or to improve health services delivery?" (Cromley & McLafferty, 2011).

One objective of medical geographers is to teach us about health. Thus, they focus on the development of methods for a clear and useful communication of health data. Anthamatten & Hazen (2011) recommend that geographical studies on health topics focusing on how and why things vary across space, should consider both, spatial patterns of disease distribution as well as how people organise themselves across space in relationship with their environments. The use of computerised mapping technologies such as GIS have made it easier answering spatial questions in health sciences (Anthamatten & Hazen, 2011).

### *7.2.3 Spatial Analysis in GIS*

Most of the objects in public health are related to space (Anthamatten & Hazen, 2011; Meade, 2014). Consequently, information such as new disease cases, deaths caused by disease or also risk factors can be geocoded and mapped. The potential of these geodata can only be exploit by proceeding data in GIS (Hirschfield et al., 1995; Twigg, 1990). A variety of definitions of different approaches of GIS exist. In summary, GIS is a computer system comprising hardware and software with the possibility to collect, manage, analyse and visualise large amounts of data for several users (Maguire, 1991; Schweikart & Kistemann, 2004). Further advantages of GISs include the ability to operate repetitive tasks and quickly compare spatial data (Rezaeian et al., 2007).

Due to recent developments in geospatial technologies, the application of spatial approaches to health issues is significantly grown. The ability of epidemiologists and public health specialists to work with spatial data has changed dramatically by GIS. Health researcher, administrators and policy-makers broadly use GIS to store, analyse and communicate spatial health data (Anthamatten & Hazen, 2011).

It is difficult to specifically define spatial analysis. In accordance with Bailey et al. (1994), spatial analysis is the quantitative study of phenomena in space with the ability of spatial data manipulation into different forms and of extracting additional meanings as a result (Bailey et al., 1994).

The spatial analysis in GIS can be sub-divided into geometric and geostatistical spatial analysis. Geometric spatial analysis focused, as its title suggests, on the geometry of geographic features (such as patterns and shapes). Geostatistical spatial analysis methods deal with spatially distributed variables to which spatial coordinates are assigned to each measurement or observation value. These variables can be described and examined by statistical tests and methods in GIS. The aim of the analysis is to describe spatial trends and construct forecast values for locations without measures (Schweikart & Kistemann, 2004). These spatial analysing methods can be used to investigate cancer disease.

# 8 Cancer Disease

Since the present study aims to contribute to the development in public health by analysing spatial data of the Swiss Canine Cancer Registry, this chapter introduces the disease. An understanding of the disease is important for correct data processing and interpretation of the results. First, a definition of cancer is elaborated, before an overview of the recent cancer situation in Swiss human population is provided and the role of risk factors presented.

## 8.1 Definition of Cancer

Cancer cannot be defined as one disease, it is a general term given to a collection of different related diseases that can affect any part of the body. Cancer can vary greatly in

terms of the pathogenesis[2], course of the disease and its treatment. Often, other terms such as malignant tumours and neoplasms are used (WHO, 2017).

The National Cancer Institute of USA (NCI) defines cancer as a non-infectious disease of the genes in the cells of our body (NIH, n.d.). In all types of cancer, cancer arises from the transformation of normal cells into tumour cells. Some body cells begin to divide without stopping and create abnormal cells. This abnormal growth is termed neoplasms. These abnormal cells can become cancer (malignant tumour cells), when they grow beyond their usual boundaries and spread into surrounding tissues. Otherwise they are called benign tumours. When cells spread to other organs, it is referred to as metastasising that is a major cause of death from cancer (IACR, 2017; NIH, n.d.; WHO, 2017).

Put briefly, it is a cell growth that is out of control and spreads tumour cells. Talking about cancer risk describes the chance that a person or animal will develop cancer (American Society of Clinical Oncology, 2015). Cancer diseases have an astonishing geographic variation in their appearance. As an example, spatial observations were made of different incidence rates between sex classes, age classes and countries. These patterns of cancer incidence can (and have) changed with time (Meade, 2014). Worldwide, more than 100 types of cancer are known. The formation of classes of cancer cases is a problem for the analysis in GIS due to this variety of cancer cases (Schweikart & Kistemann, 2004). The WHO and the International Agency for Research on Cancer (IACR) elaborated the International Classification of Diseases for Oncology (ICD-O) to classify neoplasms. Worldwide, cancer registries use this classification guideline to record incidence of malignancy and survival rates for cancer control. Two aspects are used to code and describe the tumour in the ICD-O: the morphological code that describes the cell type and the behaviour (malignant or benign) of the tumour and the topographical code that describes the anatomical site of origin of the tumour (IACR, 2017).

Since cancer diseases are non-infectious diseases, the number of present cases has no direct impact on the number of cases in the future. The pathogenesis of cancer can have various causes and genetic factors of a person play an important role. In addition to a person's

[2] Pathogenesis defines the mechanisms (the production and development) that causes a disease (Porta et al., 2014).

genetic factors, external factors are known that promote changes in cell growth. It has been shown that interaction between the genetic factors and a number of external factors promote the development of cancer (NIH, n.d.). These factors are known as risk factors and are discussed in a following section 8.3.

### 8.2 Cancer Situation in Switzerland

The cancer situation of the Swiss population is presented in the cancer report "Krebs in der Schweiz 2015" by the FSO (FSO, 2016b).

From 2008 to 2012 approximately 20,800 men and 17,650 women of all age categories were diagnosed with cancer. The most common cancer cases are prostate, breast, colorectal and lung cancer. These cancers account for slightly more than half of all new cancer cases within this period. Although cancer can occur at any age, the risk increases with age. According to the FSO, the risk of developing cancer before the age of 70 is about 25 percent for men and 21 percent for women. Almost every second man and at least one in three women have to expect to be diagnosed with cancer in the course of their lives. In addition to the high incidence rate of cancer, the disease is one of the most frequent causes of human death in Switzerland. The death of every fourth man and fifth woman is caused by cancer. Differences in the probability of developing cancer do not exist only between men and women and different age classes. In Switzerland, the probability of developing cancer is unequally distributed between social groups as well as cantons and regions. As an example, in Western Switzerland and Tessin, men and women are significantly more likely to develop cancer than in the German-speaking part of Switzerland. The interpretation of these social and spatial differences is difficult. They may be associated with risk factors and exposures but could also be explained by factors of the health system such as the utilisation of medical services. A positive development of the cancer situation in Switzerland is that in the past 30 years, the age standardised mortality rates has declined by 27 percent among women and 36 percent among men. Nonetheless, incidence rates are still slighting increasing. As a result of cancer, one out of thirteen persons is hospitalised. In an international comparison, Switzerland is one of the countries with the highest survival rate across all types of cancer (FSO, 2016b). However, the treatment cost of cancer disease are high and have a significant impact on government spending on

health care (OECD & WHO, 2011). To fight against cancer, prevention is an important strategy (FSO, 2016b). For this purpose, a better understanding of cancer diseases, their spatial distribution and their influence factors is essential (Schweikart & Kistemann, 2004).

### 8.3 Risk Factors of Cancer

The changes in cell growth are usually the result of an interaction between genetic factors and external carcinogenic factors, also termed as cancer risk factors. In individual cases, the cause of a cancer is usually not known. Cancer risk factors include exposure to physical carcinogens such as ultraviolet and ionising radiation or chemical carcinogens such as arsenic. Cancer risk factors also include biological carcinogens such as infections of certain viruses, bacteria or parasites (FSO, 2016b; NIH, n.d.).

Studies have also shown that certain behaviours such as tobacco consumption, alcohol consumption, obesity, unhealthy diet and lack of activity increase the risk of developing cancer (NIH, n.d.). Cancer risk factors also include things people cannot control. As previously indicated in the cancer situation of Switzerland, ageing is another fundamental factor for the development of cancer. This can be explained by the fact that the exposures against carcinogens accumulate and the tendency for cellular repair mechanisms is less effective as a person grows older. Genetic predispositions in hereditary material can also promote the pathogenesis of cancer (FSO, 2016b; NIH, n.d.).

In about 20 percent of cancer cases in Switzerland, genetically determined metabolic variants were responsible for cancer (FSO, 2016b). The WHO argues that modifying or avoiding of key risk factors can significantly reduce the burden of cancer (WHO, 2017).

## 9 Companion Animals in Epidemiology

The analysis of cancer occurrence and cancer risk of companion animals such as dogs is interesting for human cancer epidemiology (Reif, 2011). In this part of the thesis, the role of companion animals in cancer epidemiological studies is examined and canine cancer registers introduced.

### 9.1 Companion Animals As Sentinels

Several studies and examples have shown the usefulness of animals to predict human illness (Reif, 2011). The idea of using animal sentinels[3] as models for epidemiologic studies of human diseases and environmental exposures is not novel. In 1991, the National Research Council of America argued the following:

> *"Like humans, domestic animals and fish and other wildlife are exposed to contaminants in air, soil, water, and food, and they can suffer acute and chronic health effects from such exposures. Animal sentinel systems—systems in which data on animals exposed to contaminants in the environment are regularly and systematically collected and analysed—can be used to identify potential health hazards to other animals or humans."*
>
> (National Research Council, 1991, p. 1)

Many veterinary epidemiological studies focus on cancers in companion animals (Reif, 2011). In 'Saunders comprehensive veterinary dictionary', companion animals – also termed as domestic animals or pets – are defined as all animals such as dogs, cats, mice or birds that are kept by humans for company, amusement, psychological support, extrovert display and all of the other functions that humans need to share with animals of other species (Studdert, Gay & Blood, 2011). Companion animals with a high level of veterinary health care and subsequent data available, especially dogs and cats, are considered the most valuable sentinels species for several reasons (Brønden, Flagstad & Kristensen, 2007). Companion animals share the environment with their human companions. Consequently, they are exposed to many of the same agents as their owners (Reif, 2011). Since many animals suffer a similar spectrum of disease as humans and have similar pathologies, they can be sensitive indicators of environmental hazards. Hence, companion animals are used in the detection of hazards (e.g. pollution or infectious diseases) (Pinho et al., 2012; Van der Schalie et al., 1999).

[3] Stahl Jr (1997) defines animal sentinels as animals that can react to environmental contaminanta before humans are affected (Stahl Jr, 1997).

Cancer and other diseases have relatively short latent periods in animals compared with those for humans . This allows the investigation of spontaneous diseases in companion animal models (Kelsey, Moore & Glickman, 1998; Reif, 2011). Furthermore, animals ageing process is much faster compared to that of humans. Thus, they can provide an early warning system for public health intervention and provide information about potential risks for humans at an early stage (Reif, 2011). Other advantages are the restricted daily mobility, fewer changes in residence over an animal's shorter lifespan and the relative freedom from concurrent exposures. These aspects contribute to the likelihood that exposure assessment can be carried out more accurately in studies of animal diseases (Kelsey, Moore & Glickman, 1998).

Companion animals can be used in well-designed epidemiological studies as an independent approach to cancer research. Bukowski and Wartenberg (1997) argue that these studies can provide additional insights that are not available from laboratory-based studies of experimental animals and they can reduce the number of laboratory animal studies needed to build a body of evidence of cancer carcinogenicity (Bukowski & Wartenberg, 1997). Furthermore, studies have also shown that bias due to confounding by certain variables and due to exposure misclassification are minimal in epidemiologic studies in dogs compared with epidemiologic studies in humans (Kelsey et al., 1998; Reif, 2011).

The advantages of using cancer of companion animals as spontaneous animal models and sentinels are underpinned by many studies based on data of cats and dogs (Van der Schalie et al., 1999). This work focuses on the investigation of canine cancer. Studies shown that cancer is the most common fatal disease and the cause of approximately 15-30 percent of all deaths of dogs (Brønden et al., 2007). Dogs share their environment intimately with humans and many canine cancers resemble those in humans in biological behaviour, pathologic features, proportional morbidity, and recognised risk factors (Kelsey et al., 1998; Reif, 2011). Since dogs have a much faster ageing process than humans and a life span that allows equivalent human cancers in their morphology and biological behaviour, the development of cancer in dogs can be used as a model for comparison with human patients (Boo, 2014; Van der Schalie et al., 1999).

Brønden et al. (2007) argue that variations in cancer distribution between different geographic areas and correlations between neoplasms in companion animals and humans can support and develop cancer research (Brønden et al., 2007).

### 9.2 Canine Cancer Registries

Cancer data of companion animals is provided by companion animal cancer registries. Dos Santos (1999) defines cancer registries as organisations for the systematic collection, storage, analysis, interpretation and reporting of data on subjects with cancer (dos Santos Silva, 1999). Two main types of cancer registries exist: hospital-based or population-based cancer registries. In hospital-based cancer registries all cases in a given hospital or diagnostic laboratory are recorded. These records do not contain quantified knowledge of the catchment area or the population at risk from which the cases arise. In population-based cancer registries, all new cases in a well-defined and enumerated population, generally in a specific geographical area, are recorded. Generally, the population is based on in a specific geographical area. These registries enable the estimation of tumour incidence rates (dos Santos Silva, 1999; Nødtvedt et al., 2012).

The analysis of these data of cancer registries can provide information for the evaluation of incidence and relative risk estimates for epidemiological studies of spontaneous companion animal cancers. Examples are information about risk factor identification, treatment evaluation or patients location for clinical trials as well as cases in case-control studies (Brønden et al., 2007).

The use of these canine cancer data has several advantages over the use of human cancer data. In Switzerland, human cancer registries have high access restrictions for research due to strict privacy regulations (Federal Assembly of the Swiss Confederation, n.d.) and a low spatial resolution (cantonal or regional level) (NKP, 2014; Schweikart & Kistemann, 2004). The Swiss Cancer Registries for companion animals have lower access restrictions and a finer spatial resolution (municipal level) (Collegium Helveticum, n.d.).

However, the usefulness of cancer registries strongly depends on the quality of data of the population at risk. Information on characteristics such as age and sex of canine cancer data is of little use if the information is not available for the population (Brønden et al., 2007).

In Switzerland, research on Canine Cancer Registry data is mainly conducted by the Collegium Helveticum, which was founded in 1997 and comprises scientists from different disciplines and makes transdisciplinary research itself a research subject. The project "One Medicine – one Oncology" focused on the incidence and geographical distribution of tumours in dogs and cats in Switzerland from 1953–2012 by evaluating data of the animal cancer registries (Collegium Helveticum, n.d.).

### 9.3 Confounders in Canine Cancer Research

In the previous chapter 8.3, various risk factors were presented. Several risk factors such as age must be considered in the analysis and interpretation of results to minimise the generation of spurious correlations. Many variables may result in bias in epidemiological studies due to confounding. These variables are termed confounders and are presented in this part of the work (Bukowski & Wartenberg, 1997).

Bukowski and Wartenberg (1997) argue that potential confounders were limited to sex, age and breed (Bukowski & Wartenberg, 1997). Research of the Collegium Helveticum focused on examining correlations between breed, sex and age for both cats and dogs on the data of 1955 to 2008. These studies underpin the cofounders. It has to be mentioned that these research studies did not distinguish canine cancer types and the geographical location (Pospischil et al., 2015). Obviously, as an example, given that only female dogs can develop cancer of female reproductive organs, cancer data should be adjusted for sex. Similar to humans, the age of the dogs has a distinct influence on the incidence of specific tumours (Pospischil et al., 2015). Another related issue to the adjustment of age is that the lifespan of large-size canine breeds is generally much shorter than that of small dogs. For this reason, data may also be adjusted for body size of dogs (Kelsey et al., 1998).
Several studies further show variations between countries and significant differences between different breeds (Brønden et al., 2007). For instance, investigations on the Swiss Canine Cancer Registry have shown that canine breeds like Boxer, Poodle or Retriever were at higher risk and canine breeds such as Great Dane, Bulldog or Yorkshire Terrier were at lower risk of developing a tumour compared with crossbreeds. These investigations also examined a higher risk of developing tumours outside of the genital

organs for castrated dogs (Pospischil et al., 2015). The presented variables age, sex and size are involved later in conducting the analysis.

# 10 Spatial Epidemiology

As shown in the previous chapter 7, this work can be embedded in both GIScience and (spatial) epidemiology. Therefore, the main features of epidemiological studies, the relevant data as well as the methods of the spatial epidemiology are presented, which are the basis of the spatial analysis with GIS.

## 10.1 Epidemiological Studies

Cancer epidemiology provides the methods to describe cancer events in a population group. Their primary objective is to determine the causes of cancer. It also allows the identification and quantification of possible risk factors and options of intervention. The aim of research is to apply the findings directly to the control and prevention of cancer (dos Santos Silva, 1999). For this purpose, information that are not directly related to cancer disease must also be considered in research. Such information could include aspects that are related to the status of health or to life quality. The temporal classification of epidemiological questions is an important part of the spatial information and of great importance in epidemiology. It allows the creation of predictions of disease trends or the evaluation of intervention programs and helps to identify potential risk factors. However, the dimension time makes the application of GIS also particularly difficult (Schweikart & Kistemann, 2004).

In epidemiological studies the relationships between the health status of individuals and the risk exposure are described and analysed (National Research Council, 2012; M Porta et al., 2014). The aim is to gain knowledge of the disease distribution, the causes of disease and the conditions that lead to the disease (Schweikart & Kistemann, 2004). This work can be termed as a descriptive study, since the collected data are used to investigate the distribution of the disease. The results can be used in further analytical studies to investigate the causal relationship between cancer and factors (WHO, 2015).

More specifically, it can be seen as part of a case-control study. In case-control studies, persons (or in this case animals) with the diagnosis of a disease are identified as a case.

The study investigates whether cases differ from control subjects in terms of risk factors. Control subjects are individuals without the diagnosis of the disease, but are similar in other factors. Using statistical analysis methods, the significant relationships are examined. All exposure to relevant biological, chemical or physical factors, as well as social, economic and demographic conditions in space and time can be identified as risk factors. Epidemiological data, especially data from cancer registries, play an important role in epidemiological studies. Reliable population-related cancer information is required to better understand the spatial distribution of cancer (Schweikart & Kistemann, 2004).

### 10.2 Epidemiological Data

Epidemiological studies investigating human or canine cancer use data from various sources (often of different scale). These data are mainly the three types of data:

- population statistics;
- data on influencing factors; and
- epidemiological health and disease indicators.

Population statistics are mostly provided by public authorities. Data on influencing factors can be of different origin. Data on epidemiological health and disease indicators are recoded in cancer registries. They include prevalence, incidence, mortality and survival data as well as data on health-relevant factors (Schweikart & Kistemann, 2004).

Disease prevalence is a measure of disease occurrence of existing disease cases. It can also be a measure of the occurrence of any type of health condition, exposure or factor related to health (e.g. prevalence of smoking). It is determined by the division of the total number of individuals who have the condition at a particular time (during a particular time period) by the population at risk of having the condition at that time (or midway through the period) (Porta et al., 2014). In accordance with this definition, cancer prevalence is determined from the total number of persons (or animals) with the diagnosis cancer at a certain time relative to the population at risk. The frequency of diseases or symptom-complexes in population groups can be established by representative population sampling (Schweikart & Kistemann, 2004).

The incidence of disease is the number of instances of new health-related events in a defined population (or area) within a specified period and can be measured as a frequency

count, rate or proportion (Porta et al., 2014). Thus, cancer incidence is the number of new appearances of cancer cases in a specified population during a given period. As a rule, incidence rates express the annual number of cancer cases per 100,000 people at risk (Schweikart & Kistemann, 2004). According to World Cancer Research Fund International (WCRF), cancer density is very high for administrative units with a cancer incidence rate of more than 300 cancer cases to 100,000 inhabitants (WCRF, n.d.).

The mortality of disease is determined from the total number of deaths caused by a disease in a population. Data on cancer mortality informs about the overall mortality of cancer of a population. Data on health-relevant factors includes various indicators of the health system and further indicators (such as age) that are relevant for cancer epidemiology (Schweikart & Kistemann, 2004). Given that the present study addresses the occurrence of cancer disease, only prevalence and incidence data are relevant. Geocoding of the data is a prerequisite for the application of the methods of spatial epidemiology and spatial analysis in GIS (Beale et al., 2008).

### 10.3 Methods of Spatial Cancer Epidemiology

In any spatial epidemiological analysis, the study focus specifies the nature and style of the spatial epidemiological methods. Generally, the study focus comprises at least one hypothesis about the nature of the spatial distribution of disease that is to be investigated. The distribution of cases of disease are usually thought to follow an underlying model. The observed disease data may also contain extra noise, which can be in the form of random variation around the model of interest. The model often includes aspects of the null hypothesis of the spatial distribution of the cases, capturing the "normal" variation that is expected of the distribution of the cases in space. Since the disease cases are thought to arise in relation to the local variation in the underlying population distribution of the study region, the null hypothesis is often defined by this distribution. However, the model can also include aspects of an alternative distribution (Lawson, 2006b).

According to Lawson (2006b), in many spatial epidemiological studies the focus of interest involves identifying features of the spatial distribution that are captured by an alternative distribution. Lawson argues that this is mainly related to excess the spatial aggregation of disease cases in areas of the map. Mostly, there is need to examine areas of higher

aggregation that would be normally expected. Consequently, once the normal variation is allowed for, the focus lies on the residual spatial incidence above the normal incidence. The underlying hypotheses that are relevant in spatial cancer epidemiology can be categorised according to the three predominant methods of spatial epidemiology: disease mapping, disease clustering and ecological analysis (Lawson, 2006b).

Subsequently, the three categories are introduced focusing on disease mapping and disease clustering, since these two methods are necessary to answer the research questions. The ecological analysis will be briefly presented.

### *10.3.1 Disease Mapping*

Disease mapping uses models to describe the overall disease distribution on the map (Lawson, 2006b). They provide a rapid visual summary of complex geographic information and can uncover the underlying structure in the data that are missed in tabular presentations (Lawson, 2006b; Wilkinson et al., 1997).

Representing and analysing maps of disease incidence of mortality data is a basic tool in the analysis of regional (public) health (Rezaeian et al., 2007). In the recent years, methods for disease mapping has constantly progressed. Many and various intended uses of disease maps exist (Lawson, 2006b; Rezaeian et al., 2007; Wilkinson et al., 1997). According to Wilkinson et al. disease maps can be used to generate hypotheses as to aetiology, for surveillance to highlight areas at apparently high risk, as well as aiding policy formation and resource allocation. They are also useful to put specific disease clusters and results of studies on point data in the right context (Wilkinson et al., 1997).

Lawson (2006) categorised the uses of disease maps as follows. First, they can be used to assess the need for geographical variation in the allocation of health resources. In this case, the object of disease mapping is to "clean" the disease map of any random noise and any artefacts of population variation. Second, disease mapping can be useful in research studies and contribute to a better understanding of the relation of incidence to explanatory variables. This case assesses specific hypotheses about the disease incidence and includes additional information in the analysis, such as covariates (Lawson, 2006b).

#### 10.3.1.1 Data standardisation

Maps of raw incidence counts inaccurately reflect the spatial pattern of the mapped phenomenon (Roth, Woodruff & Johnson, 2010). For instance, Swiss municipality units containing large populations are capable of having higher incidence counts than those containing small populations. Thus, spatial disease analysis should focus on determining the risk of developing cancer for each unit. In order to determine the risk status of an area regarding the disease incidence a convenient approach is to first assess what disease incidence should be locally 'expected' in the area, before subsequently comparing the observed incidence to the 'expected' incidence (Lawson, 2006b). In this process, it must be considered that potential cancer risk factors such as age, sex or other variables can vary from one place to another. Given that the observation of differences in risk of cancer is likely to be confounded by these variables, the comparisons of risk must take these variables into account (Rezaeian et al., 2007; Selvin, 1996). The process of adjustment for potential confounding variables is essential for evaluating spatial variations in epidemiological rates. Thus, it is important to standardise the data according to relevant social risk factors before mapping and analysing the disease (Selvin, 1996)(Selvin, 1996).
A standardisation is a set of techniques which is used in comparing two or more populations to adjust rates and any other measures of occurrence. The techniques are based on weighted averaging and have the object to remove as much as possible effects of differences in confounding variables such as age, ethnicity, area of residence, sex, etc. Common methods use weighted averaging of rates specific for age, sex or some other potential confounding variables according to some specified distribution of these variables (Porta et al., 2014). The two most widely used standardising methods are the direct and indirect standardisation. The choice of the appropriate method strongly depends on the study data (Gail & Benichou, 2000; Rezaeian et al., 2007). The direct method averages the specific rates from a study population by using the distribution of the specified reference population as weights. The result is an estimate of the expected number of deaths or new disease cases in the underlying population. As the direct method represents what a crude rate would have been in the study population if that population had the same distribution as the previously specified standard population, it can be useful in situations in which disease rates in the standard population are not available. This is

always undertaken with respect to the variable(s) for which the standardisation was carried out (Gail & Benichou, 2000). A ratio that is based on this direct standardisation and can be easily interpreted is termed the comparative mortality figure or standardised incidence rate ratio. It is calculated by dividing the expected number (E) by the observed number (O) of disease cases or deaths in the reference population within the same period (Gail & Benichou, 2000; Rezaeian et al., 2007).

Once the specific rates of the study population are either statistically unstable or unknown, the indirect method is used. For example, when only the age-specific rates are available but not the age-specific incidence counts for a reference population. In this case, the specific rates in the standard population are averages by using the distribution of the study population as weights. The ratio of the crude rate for the study population to the weighted average so obtained is the standardised mortality or standardised morbidity ratio (SMR). The SMR is commonly used in etiological studies. The calculation of the SMR assumes that the specific rates in the reference population also applies to the population under study (Rezaeian et al., 2007).

Consequently, the standardised morbidity ratio can be defined as "…*the ratio of the incident number of cases of a specified condition in the study population to the incident number that would be expected if the study population had the same incidence rate as a standard or other population for which the incidence rate is known…*" and the standardised mortality ratio as "…*the ratio of the number of deaths observed in the study group or population to the number that would be expected if the study population had the same specific rates as the standard population*."(Porta et al., 2014, p. 269).

The indirectly standardised rate itself is calculated by multiplying the SMR and the crude rate for the standard population. However, it is rarely used in etiological studies (Rezaeian et al., 2007).

When using one of the two methods, on should take into account that both methods have problems. In 1987, a study on statistical methods in cancer research of Heseltine, Day and Breslow has shown that the direct method may provide less stable estimates since the standard error of the rates depends on variations in the variable specific number of cases rather than the total number of cases. They also showed that the indirect method of

standardisation yields different rate ratios for cohorts with a different demographic structure even though the incidence rates within the demographic strata are identical. Furthermore, they showed that the comparison of indirectly adjusted ratios may have a potential bias comparable with statistical confounding (Heseltine, Day & Breslow, 1987; Rezaeian et al., 2007).

An example regarding the age distribution is given by the Rezaeian et al (2007). They argue that, when compared with an external reference population, the indirect standardisation method results in different rate ratios for cohorts with a different demographic structure even though the incidence rates within the demographic strata are identical. Despite these raising cautions in the literature, well-known statisticians recommend the SMRs for disease mapping (Rezaeian et al., 2007).

#### 10.3.1.2 Disease Mapping Methods

Traditionally, choropleth maps are used to visualise disease. It is a popular cartographic technique to visualise social data in a map. Commonly, for choropleth disease maps individual incidences are aggregated according to a set of data collection units. The units are subsequently coloured by its aggregated value (Slocum et al., 2005).

However, this technique is limited in displaying additional variables in the map such as time. A common solution to the problem above are cartograms – also termed as value-by-area maps – which are often used to display disease in a map (Roth et al., 2010). In these value-by-area maps, subdivisions of regions or countries are represented by rectangles proportionated in area to the value that it represents in certain statistical distributions (Raisz, 1938). This means that "*the maps visually equalize a basemap before mapping a variable by adjusting the size of each enumeration unit by a second related variable*" (Roth et al., 2010, p. 130).

In 1998, Keim and Herrmann argued that the ideal cartogram is one that visually equalises the basemap while completely preserving shape and topology (Keim & Herrmann, 1998). According to this, Roth et al. (2010) presented the cube of Cartogramm$^3$ (shown in figure 12.1) with its three dimensions that are important for accurate reading of cartograms. These dimensions are shape preservation, topology preservation and visual equalisation (Roth et al., 2010).

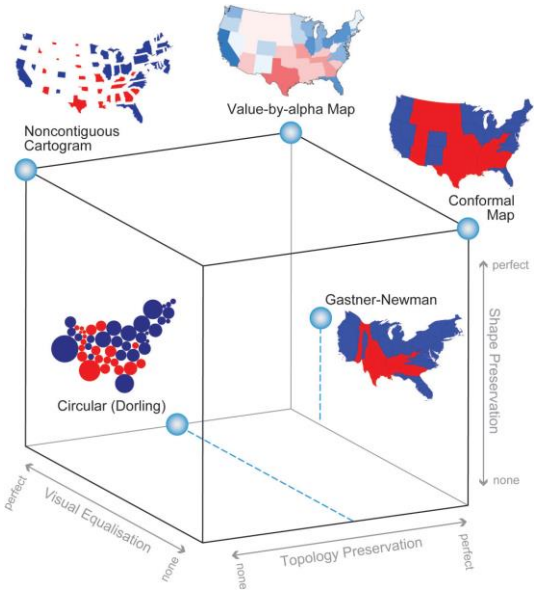



Figure 10.1. Cartogram3. The cube of three dimensions that are important for accurate reading of cartograms: shape preservation, topology preservation and visual equalisation (Roth et al., 2010).

The figure illustrates that a value-by-area map can be placed somewhere within the bounds of this cube. Since scaling the base map units according to an equalising variable always implies a distortion of either shape or topology of the original geography, any value-by-area cartogram, whether a non-contiguous cartogram[4] or a conformal map[5], is a compromise across these three dimensions. For general map-reading tasks these compromises reduce the effectiveness of the visualisation. Accordingly, they presented a new mapping technique termed as a value-by-alpha map. These maps are able to perfectly achieve each of the three dimensions without compromise, equalising the base map while preserving shape and topology. The difference to a value-by-area map is that the value-by-alpha map adjusts the alpha channel when visually equalise the base map, rather than the size of each enumeration unit (Roth et al., 2010). This allow the visualisation of the temporal variation of disease in the alpha channel, which is the reason why this method is used in the present thesis (chapter 13.2).

Two different approaches exist for generating value-by-alpha maps, as illustrated in figure 12.2. Compared to existing methods for generating cartograms, both approaches are

[4] In non-contiguous maps individual units are arranged in an approximately geographically correct way but re-sized and spaced apart to depict a quantifiable difference between values of data in the unit areas (Bortins & Demers, 2002).

[5] In conformal maps, the size of local angles is preserved In the representation (Furuti, 2016).

relatively straightforward and can be applied in this study. For both methods, the modifying colour is found in the layer without the original colour assignment (Roth et al., 2010).

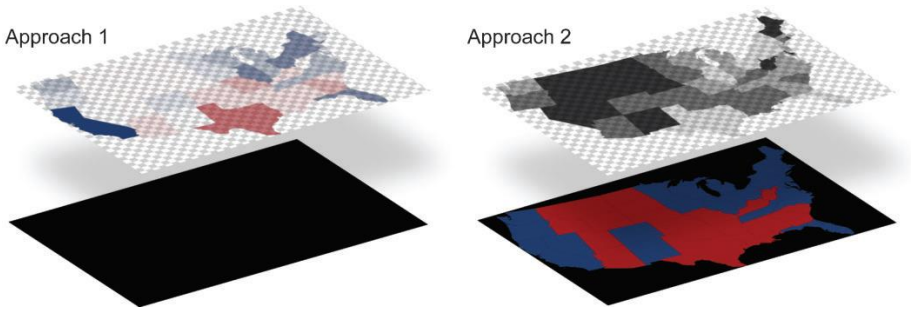



Figure 10.2. Two approaches to making value-by-alpha maps (Roth et al., 2010).

Fot the first approach, the original colour is assigned to the data layer on top. This method assigns the alpha channel of a single data layer according to the equalising variable. The alpha value of the enumeration unit increases parallelly to the increase of the value of the equalising variable. Differently to the first approach, in the second approach the original colour is assigned to the bottom of the two layers. Thus, the alpha channel of a top data layer is assigned according to the equalising variable. The alpha value of the enumeration unit in the top layer decreases, as the value of equalising variable increases. However, when interpreting the maps, it must be taken into account that this form of a map is not free from limitations (Roth et al., 2010).

### *10.3.2 Disease Clustering*

Spatial cluster analysis is an important method not only in spatial epidemiology. It plays an essential role in quantifying geographic variation patterns and is thus used in many other fields such as disease surveillance, landscape ecology, population genetics or crime analysis (Jacquez, 2008). The overarching aim of disease clustering is to detect unusual aggregations of the disease (Openshaw, Charlton & Craft, 1988).

Due to growing concerns of the population about environmental effects on the health status during 1980s, the interest to cluster analysis in public health has grown . As a consequence, the development of methods for the evaluation of diseases clusters has been driven (Lawson, 2006b; Wartenberg, 2001). Lawson (2006) argues that for many diseases the spatial incidence will naturally display clustering at some spatial scale, even after the

'at-risk' population effects are considered. The reasons for such clustering of disease are various. Examples are viral pathogens for non-infectious diseases or other common but unobserved factors which could lead to observed clustering in maps (Lawson, 2006b).
In cancer research, spatial patterns contain the geographic trace of processes, covariates, and factors. These spatial patterns determine how cancer risk varies across and is expressed within populations (Jacquez, 2008).
This sub-chapter provides the underlying principles of spatial cluster analysis. First, it presents the definition of a cluster regarding the type of data, before the clustering typology of global, local and focused methods are detailed.

#### 10.3.2.1 Cluster definition

A wide range of clustering definition exist and are discussed in the following part. A very basic definition was given by Knox (1989), defining a cluster as a geographical bounded group of occurrences that is of sufficient size and concentration and unlikely to have occurred by chance (Knox, 1989). Lawson distinguishes two basic types of clusters, namely hot-spot clusters and parametric clusters. Hot-spot clustering can be defined as any area within the study region of significant elevated risk, regardless of shape or extent of the cluster. A prerequisite for a cluster is that the area meets some statistical criteria. In disease clustering, areas of significantly high risks usually are of interest. Parametric clusters are clusters for which the study region has a pre-specified cluster structure. This usually implies region-wide parameters that control the cluster form as well as some stronger restrictions on the cluster form (Lawson, 2006b).
A more specific definition was developed by Jacquez (2008). He argues that the definition of cluster must also consider the type of data that are being studied. Thus, clusters may be event-based, population-based, field-based, or feature- based (Jacquez, 2008). Event-based data always include point locations (such as the places of residence and time of cancer diagnosis) and counts (such as cancer cases at the location). Population-based data comprise information on the reference population on which the events arose and further include disease rates such as morbidity ratio or mortality ratio. Disease rates include case counts in the numerator and size of the at-risk population in the denominator. Field-based data comprise continuously distributed observations (such as concentrations and

temperatures) over space. Finally, feature-based data include boundaries and polygons that may be derived from field-based data, such as zones of rapid change in an attribute's value (Jacquez, 2008). Based on these data types, Jacquez defines a spatial cluster as "…*an excess of events (for event- and population- based data, such as a cancer cluster) or of values (for field-based data, such as a grouping of excessively high concentrations of cadmium in soils) in geographic space.*" (Jacquez, 2008, p. 396).

The Centers for Disease Control and Prevention (CDC) of the Unites States (1990) define disease clusters as an unusual aggregation of disease cases and high incidence rates that are grouped together in space or time or both and are beyond what would normally be expected (CDC, 1990). The term "cluster" in spatial disease analysis is generally used for relatively uncommon events and non-infectious disease such as cancer (Elliott et al., 2000; Porta et al., 2014).

In accordance with this definition of disease clusters, temporal and spatial information must be considered in disease clustering. Lawson (2006b) argues that spatial clustering is one component of the dynamic behaviour of disease within a framework of spatial and temporal variations (Lawson, 2006b). Figure 12.3 illustrates the space-time cube with its three cluster types: spatial cluster, temporal cluster and spatio-temporal cluster.

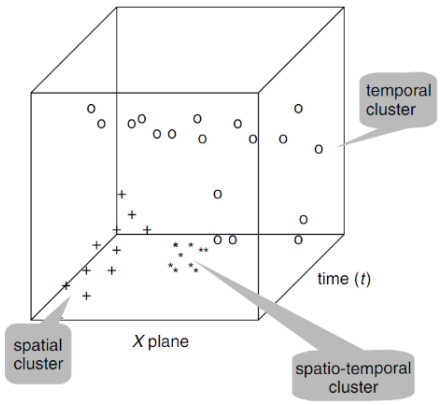



Figure 10.3. The three components of clustering in space-time: spatial, temporal and spatio-temporal (Lawson, 2006b).

It is possible to extend parametric models defined for spatio-temporal clustering in disease; for example, by including temporal and spatio-temporal cluster terms. One way to achieve a suitable cluster definition in the three different domains is through the notion of the cluster's persistence, whereby three cluster types can be distinguished as follows: a purely spatial cluster must occur throughout the time domain (i.e. persist through time); a purely

temporal cluster must persist through space; and a spatio-temporal cluster must not persist in either domain (Lawson, 2006b).

#### 10.3.2.2 Cluster statistics

Several approaches exist for spatial pattern recognition. Most commonly used methods are founded on statistics and are summarised under the term statistical pattern recognition. In these methods, a statistic is calculated that quantifies a relevant aspect of spatial pattern in event-based, population-based, field-based or feature-based data (Jacquez, 2008).

Waller and Jacquez (1995) introduced the followed five-component mechanism that underpins most tests used in spatial cluster analysis (Jacquez, 2008; Waller & Jacquez, 1995):

- Test statistic
- Alternative hypothesis
- Null hypothesis
- Null spatial model
- Reference distribution

A relevant aspect of spatial pattern is quantified in the test statistic. Examples of methods are Moran's I, Geary's c, LISA, etc. The alternative hypothesis describes the spatial pattern (for whose detection the test was designed). The null hypothesis describes the spatial pattern that is expected when the alternative hypothesis is false. This spatial pattern could – for instance – be a uniform cancer risk. If the null hypothesis is true, the distribution of the test statistic is termed the reference distribution. The null spatial model is a mechanism used to generate this reference distribution. Many null spatial models use Monte Carlo as a randomisation technique (Waller & Jacquez, 1995). To provide a probabilistic assessment of how unlikely an observed spatial pattern is under the null hypothesis, the numerical value of the test statistic is compared to the reference distribution (Jacquez, 2008; Waller & Jacquez, 1995).

Debates exist how the null hypothesis and null spatial model should be specified and many implementations of spatial cluster tests use the Complete Spatial Randomness (CSR) to employ the null hypothesis. However, most geographic systems are highly complex

and spatial randomness rarely – if ever – occurs (since some spatial pattern is almost always present) (Jacquez, 2008). Thus, CSR is not a relevant null hypothesis for highly complex and organised systems such as those encountered in geography, spatial epidemiology, and exposure assessment (Liebisch, Goovaerts & Kaufmann, 2002). Alternatively, neutral models, which are more plausible than CSR, can be used as mechanisms for constructing null hypotheses. These models account for differences in underlying population sizes and for regional and local variation in mean values (Goovaerts & Jacquez, 2004; Jacquez, 2008).

In order to identify the spatial patterns above and beyond the incorporated neutral model, a cluster analysis can be conducted (Jacquez, 2008).

According to Porta et al. (2014), cluster analysis can be defined as a set of statistical methods which is used to group variables or observations into interrelated subgroups (Porta et al., 2014).

Three types of cluster statistics can be distinguished (Jacquez, 2008):

- Global statistics
- Local statistics
- Focused statistics

**Global cluster statistics** such as Moran's I are sensitive to the occurrence of departures from the null hypothesis or spatial clustering anywhere in the study area. Thus, global statistics can identify whether spatial structure such as clustering, autocorrelation or uniformity exists. However, they do not identify where the clusters are and they do also not quantify how spatial dependency varies from one place to another. **Local cluster statistics** such as Local Indicators of Spatial Association (LISA) quantify spatial autocorrelation and clustering within small areas. These small areas together comprise the study area. Thus, local statistics can both offer information about the nature of spatial dependency in a given locality and provide a global test. The third type of cluster statistics are the **focused cluster statistics**. These cluster statistics quantify clustering around a focus (a specific location). Focused cluster statistics are mostly used to quantify disease clusters around potential sources of environmental pollutants. Looking for the "one" suitable cluster test is only appropriate with prior knowledge of the cluster shape (Jacquez, 2008).

### *10.3.3 Ecological Analysis*

The third essential method category in spatial epidemiology is ecological analysis, also termed geographic correlation studies. This method is not used to answer the research questions, although it is an essential method for spatial epidemiology and might be interesting in studies that further develop this work. Thus, it is briefly presented in this part.

The aim of ecological analysis – first expressed by Hippocrates 400b.C. – is to assess the relationship between the geographic variations of disease frequency across population groups in exposure to explanatory factors (e.g. particular agent) (dos Santos Silva, 1999; Jones, 1923). These agents could be environmental variables (e.g. measured in air, water, or soil), socioeconomic and demographic measures (e.g. income), or lifestyle factors (e.g. diet) which are measured on a geographic (ecologic) scale (Elliott et al., 1992; Rezaeian et al., 2007).

The analysis is based on aggregated or grouped data and crucially depends on scale (Lawson, 2006b). The process of data aggregation is difficult and essentially impact on the inference of ecological analysis results (Elliott et al., 1998; Porta et al., 2014).

The groups (regions) of the population must be sufficiently large to have stable rate estimates but also sufficiently small to make them homogeneous regarding their (socioeconomic) characteristics. For to large regions a greater possibility exists that associations are artifactually created or masked, which can result in errors in inference, a problem known as ecological fallacy. For to small regions, random variation in small numbers of events can lead to spurious spatial patterns (Morgenstern, 1982; Richardson & Monfort, 2000). A further problem arises due to the scale dependency of data, called the modifiable areal unit problem (MAUP), introduced by Stan Openshaw. The MAUP arises from the uncertainty induced by the aggregation procedure (Openshaw, 1984). Figure 10.4 illustrates the two distinct types scale and zone of MAUP. The scale effect is based on the fact that the results of statistical analyses on aggregated data over areas of different sizes may vary. The zoning effect of MAUP is based on the idea that different space settings or different sets of zones with the same or similar areas but different forms may also provide different results in the analysis (Lloyd, 2010).

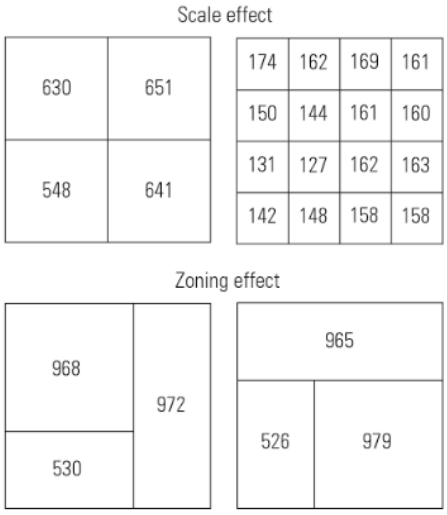



Figure 10.4. The scale and zoning effects (Lloyd, 2010).

To interpret results of ecological (and spatial) analysis, it is necessary to take the scale of the analyses into account. An approach to deal with the MAUP is to repeat the analysis with several levels of aggregation. To deal with the problem of aggregation, observations will usually need validation and replication at the individual level. Nonetheless, ecologic studies are important and can (as shown in the past) contribute to the development and exploration of major hypotheses of public health importance (Elliott & Wartenberg, 2004).

# Part III - METHODOLOGY AND RESULTS

The third part of this thesis encompasses the spatial analyses and the results. Chapter 11 describes the methodology that is applied. Chapter 12 presents and explores the data basis. The analytical steps are performed in chapter 13 and chapter 14. In chapter 13 assesses the disease mapping of cancer risk. In chapter 14 homogeneity tests and two approaches of cluster analyses are conducted. At the end of this part, a conclusion is presented.

## 11 Methodology

The procedure of the methodology is structured as follows:

1 Step: Data preprocessing and exploration

2 Step: Disease Mapping

3 Step: Spatio-Temporal Cluster Analysis

The data are aggregated for each year to analyse the changes over the period. In a first step, the data are processed and explored. This is an important step for the further methodology and the subsequent interpretation of the data. In order to answer the research question, the data are standardised and the cancer risk is determined. This is conducted for all cancer records together as well as for the most common malignant tumour types. Subsequently, the results are presented in a disease map and interpreted. In order to determine spatial and temporal patterns in the cancer risks, cluster analysis is carried out for all cancer types as well as for the most common cancer types. Before doing this, the data is tested for heterogeneity. This global test shows whether the risks actually differ or if they are homogeneous. The results of the cluster analysis are finally visualised and analysed. The choice of methodology is explained in the respective parts.

The processing of the data is mainly performed in the software RStudio with the programming language R. The structure of the R code is based on the methodology structure of this work and the different parts are commented directly in the script. For the cluster analysis, additional software is required which is presented at the respective step. All the maps and the R script can be found on the supplied CD-ROM.

# 12 Data Preprocessing and Exploration

Before choosing the analysis method, in this chapter the used data is presented, preprocessed and explored.

## 12.1 Data Sets

Two kinds of data are used for the analysis. The Canine Cancer Registry contains information on cancer cases. In the analysis, this data is compared with data on the canine reference population provided by the ANIS database (since 2016 named AMICUS). Subsequently, the two data sources as well as the related geodata are presented.

### *12.1.1 Canine Cancer Cases*

For the spatial analysis, canine cancer data provided by the Swiss Canine Cancer Registry (SCCR) was used for the time period from 2008 to 2013. The canine tumour registry encompasses malignant and benign tumour diagnostic records since 1955. These records are provided by three veterinary diagnostic laboratories in Switzerland. The Vetsuisse Faculty at the Institute of Veterinary Pathology in Zurich (IVPZ) provided three sets of diagnostic records from canine post-mortem, biopsy and cytology samples of different origin. The IVPZ provided three sets of diagnostic records from canine post-mortem, biopsy and cytology samples of different origin. The first data set provides digitalised data from originally handwritten documents of canine post-mortem samples from 1955 to 1964. The second data set provides digitalised data of originally transcribed diagnostic key words onto punch cards. These records comprise both, canine post-mortem and biopsy samples from 1964 to 1988. The third data set stored the electronic patient record from canine post-mortem, biopsy and cytology samples in the system of the IVPZ. All the datasets were recorded in only one system to reduce overlapping. The second veterinary diagnostic laboratory is the Institute of Animal Pathology at the University of Bern (ITPA) that issued a set of diagnostic records from canine post-mortem and biopsy samples since 1983. The third veterinary diagnostic laboratory is the private veterinary diagnostic laboratory "*Zyto-Histo Diagnostics*" that is based in Rorbas, Freienstein. This laboratory provided diagnostic records from canine biopsy samples since 2007 (Grüntzig et al., 2015).

Table 12.1. presents the canine cancer data structure with the most important variables. The tumour diagnoses were already coded according to the tumour topographical and morphological keys of the ICD-O-3 (Grüntzig et al., 2015). Additionally, a list containing the tumour topographies and morphologies is provided in a separate file for each of them. The canine cancer data used for the present analysis provide further information such as age, sex and breed group.

Table 12.1. Example of the Canine Cancer Registry data structure with the most important variables.

| FSO-No. | Postcode | Locality | Year | Age | Sex | Breed Group | Tumour Location | Tumour Code | Tumour Group |
|---|---|---|---|---|---|---|---|---|---|
| **5586** | 1000 | Lausanne | 2008 | 11 | Female | Crossbreed | 44.0 | 8800 | Skin |
| **5586** | 1003 | Lausanne | 2010 | 8 | Male | Boxer | 44.0 | 9740 | Skin |

The Swiss Cancer Registry summarises tumour prevalence per year for Swiss municipalities (including enclaves in Switzerland). This means that a dog diagnosed with a tumour is recorded every year (from the year of diagnosis) until he is cured or he died. The Canine Cancer Registry for the study period from 2008 to 2013 comprises 39,982 records (IVPZ n=15,197, ITPA n=8,334, "*Zyto-Histo Diagnostics*" n=16,451). Since the study focuses on cancer risk, only malignant tumours (20,157 records) are used for further analyses.

### *12.1.2 Canine Population*

Data on the canine reference population was provided by the ANIS database. For the time period of interest, the registration of dogs was no obligation. However, the data set comprises 3,189,207 records providing many information about the canine population in Switzerland. Table 12.2. presents the data structure of reference population data set with the most important variables.

Table 12.2 Example of the reference population data structure with the most important variables.

| FSO-No. | Postcode | Locality | Year | Age | Birth | Death | Sex | Breed | Breed Group |
|---|---|---|---|---|---|---|---|---|---|
| **437** | 2723 | Mont-Tramelan | 2012 | 2 | 2010 | 0 | female | Affenpinscher | Pinscher |
| **4761** | 8370 | Sirnach | 2010 | 1 | 2009 | 0 | female | Affenpinscher | Pinscher |
| **118** | 8630 | Rüti ZH | 2010 | 1 | 2009 | 0 | female | Affenpinscher | Pinscher |

Since post codes are not unique, the spatial information of both data sets is provided by the municipality number (FSO-No. or also termed GEOSTAT-No.). The Federal Statistical Office (FSO) assigns a number (a key) to each municipality (including enclaves in Switzerland). These entries can change; for instance, due to mutations of municipalities, changes to municipal boundaries or changes in the districts or comparable administrative entities of the canton. For this reason, one of the responsibilities of the FSO is the creation, management and publication of the official commune register in Switzerland. The municipality names and numbers of this official commune register must be used by all public authorities (FSO, 2017). This allows the spatial comparison of the data sets.

### *12.1.3 Geodata of Municipalities*

The spatial information of the municipalities is provided by the official geodata of FSO is freely available on their webpage:
https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/geostat/geodaten-bundesstatistik/administrative-grenzen/generalisierte-gemeindegrenzen.assetdetail.453578.html
The data includes all generalised boundaries for Swiss municipalities as well as enclaves in 2016 and is based on the Swiss coordinate system CH1903 / LV03 (FSO, 2016a).

## 12.2 Data Preprocessing and Classification

Before the analysis, the data is preprocessed and categorised for standardisation purpose. This procedure is important to reduce uncertainties and errors in the analysis. The steps of data preprocessing and data classification are explained below.

### *12.2.1 Data Preprocessing*

Both records contain names and labels with umlauts and special characters, typing errors and several formats. In order to correctly display the entries of both datasets in visualisations, the records were revised by selecting the encoding type, eliminating errors and adjusting formats. In reviewing the records, entries were found where several single records where aggregated in one entry. Such mistakes arise due to errors in the files. The records were detached from each other and listed in single entries. Furthermore, the breed groups of both data sets were compared and differences in spelling removed. Since the entries of the official commune register are not static, the FSO numbers and names of the municipalities have to be updated. For this purpose, the FSO numbers and names must be compared with the records of the official commune register. Since data from 1955 to 2008 were already used by the Collegium Helveticum in spatial analyses, all mutation reports of the FSO since 2008 have been collected in an excel file. This allowed an automatic comparison of the data with the actual FSO numbers and names. Additionally, FSO numbers which have not been modified since 2008, have been reviewed and corrected manually. In addition to this update of spatial information, incomplete entries have been completed and errors corrected. An example of an error is illustrated in table 12.3 where FSO-No. and postcode are correct. The incorrect old municipality name was adjusted with the correct new name. Incorrect entries that could not be corrected were excluded from the analysis.

Table 12.3. Example of an incorrect entry in the Canine Cancer Registry.

| FSO-No. | Postcode | Old Locality | New Locality |
|---|---|---|---|
| **5747** | 1353 | Borovnica, Slowenia | Bofflens |

### *12.2.2 Data Classification*

In accordance with the literature (chapter 9.3), the present study classifies dogs according to their age, sex and size. For comparison purposes with earlier studies, the age information of the records is classified according to the specification in table 12.4. The Canine Cancer Registry also comprises records with missing age information or unrealistic values (e.g. 100,000,000). These records derive the value ‘unknown’ and are

indicated with an age index of zero. Since both data sets provide information about the sex of dogs, this information is classified and indicated according to table 12.5.

Table 12.4. Age classification.

| Age Index | Age class |
|---|---|
| **0** | Unknown |
| **1** | 0-5 |
| **2** | 6-10 |
| **3** | 11-15 |
| **4** | 16-20 |
| **5** | >20 |

Table 12.5. Sex classification.

| Sex Index | Sex |
|---|---|
| **0** | Unknown |
| **1** | Female |
| **2** | Male |

Since information about the size of the dogs is not recorded in the registries, they have to be derived from literature. The standards of the World Canine Organisation "*Fédération Cynologique Internationale*" (FCI) were used as the main source. However, not all breed groups are officially recognised by FCI or the descriptions do not contain size information. Thus, further literature was used to complete the information. The source of the size information is indicated in the data with an index. A list of the used literature and its indices is presented in table 12.6.

Depending on the breed group and the sex, the size can vary greatly. For this reason, the determined size information is not a single value, but rather value ranges. For the classification, the mean value was calculated and classified according to the definition of size classes of Alderton, Morgan and Schmitz et al. (2001), illustrated in table 12.7 (Alderton, Morgan & Schmitz, 2001).

Table 12.6. Indices of information source.

| Source Index | Source |
|---|---|
| **1** | (FCI, n.d.) |
| **2** | (Alderton et al., 2001) |
| **3** | (Lehari, 2004) |
| **4** | (KUSA, 2016) |
| **5** | (AKC, 2017) |
| **6** | Wikipedia |

Table 12.7. Size classification.

| Size Index | withers height [cm] | Size category |
|---|---|---|
| **1** | < 46 | Small |
| **2** | 46 - 61 | Medium |
| **3** | > 61 | Large |

After completion of the data preprocessing and classification, the data exploration can start.

### 12.3 Data Exploration

A first exploration of the data shows interesting observations. In a first step, the age structure of the data is explored. Figures 12.1-12.3 illustrate the main information of the age structure. Looking at the distribution of age in the population, the age structure shows that in Switzerland many very young dogs are registered and that the number of registered dogs decreases with increasing age. However, the age structure in the records of the cancer registry shows the opposite. Malignant tumours were diagnosed in comparatively few very young (0-5) and old dogs (> 15). Most malignant tumours were recorded for dogs between five and ten years and ten and 15 years, respectively. It should be noted here that the Canine Cancer Registry contains about three percent of records without age information. When focusing on the percentage of the tumour diagnoses per age class, this shows that dogs between five and ten years and dogs between ten and 15 years have with 0.8 and 1.2, respectively, the highest malignant tumour diagnosis rate.

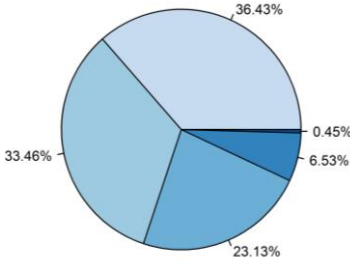



Figure 12.1. Dogs per age class.

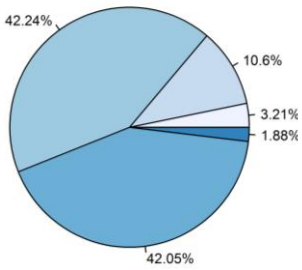



Figure 12.2. Malignant canine tumour per age class.

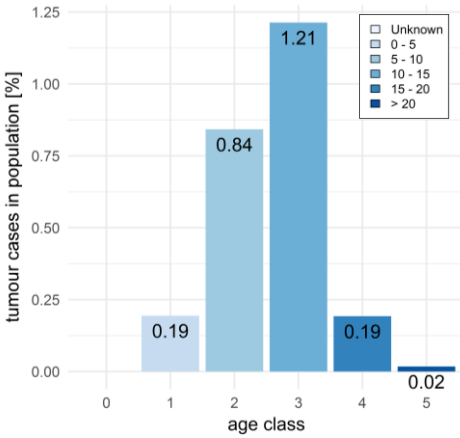



Figure 12.3. Percentage of malignant tumour in population per age class.

Figures 12.4.-12.6 illustrate the main information of the sex structure. The sex structure in both data sets show similar observations, female and male dogs have about the same percentage in the registries. Focusing on the percentage of the malignant tumour diagnoses per sex class shows that the cancer diagnosis rate is slightly higher for female compared with male dogs. One important point presented in the figures is that many registered malignant tumours have a lack of sex information. The Canine Cancer Registry contains about three percent of records without sex information, compared with the

population registry at only approximately 0.5 percent. Consequently, they cannot be linked directly to the population and the cancer diagnosis rate for this class is very high.

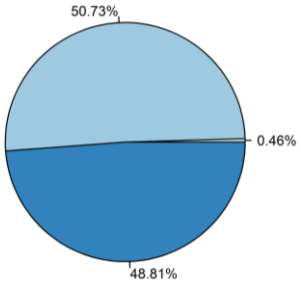



Figure 12.4. Dogs per sex class.

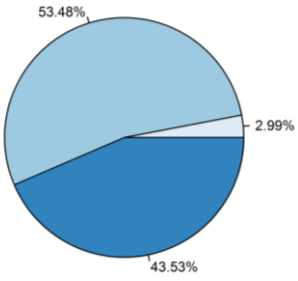



Figure 12.5. Malignant canine tumour per sex class.

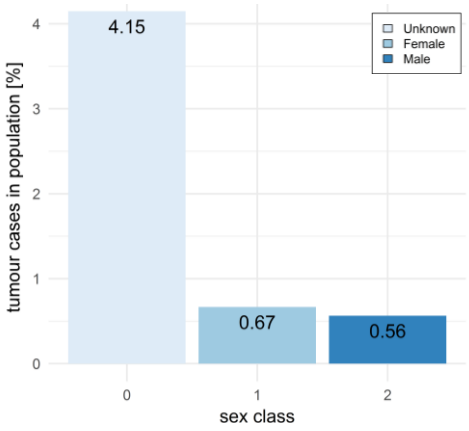



Figure 12.6. Percentage of malignant tumour in population per sex class.

Figures 12.7.-12.9 illustrate the main information of the size structure. Comparing the size structure of the data sets is difficult since the percentage of records without size information is high. This uncertainty must be considered when interpreting the results of the analysis. One interesting observation for records with size information is that the taller the dogs, the higher the cancer diagnosis rate.

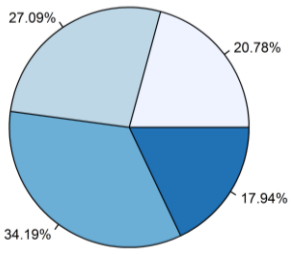



Figure 12.7. Dogs per size class.

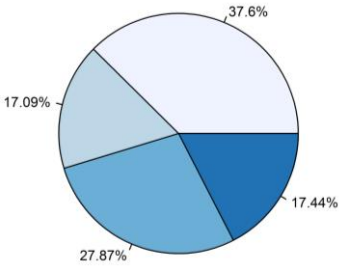



Figure 12.8. Malignant Canine tumour per size class.

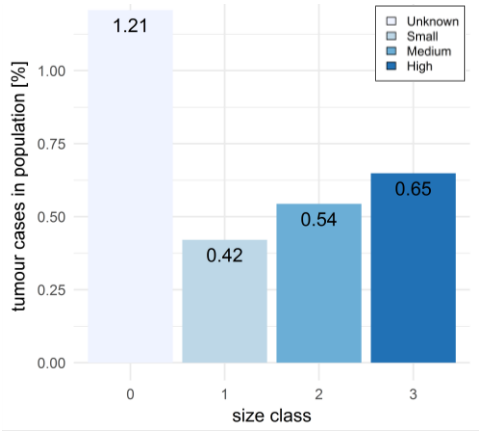



Figure 12.9. Percentage of malignant tumour in population per size class.

The tumour morphology of the diagnoses was used to determine the most common cancer types. This provides information about the shape and structure of the malignant tumours (IACR, 2017).

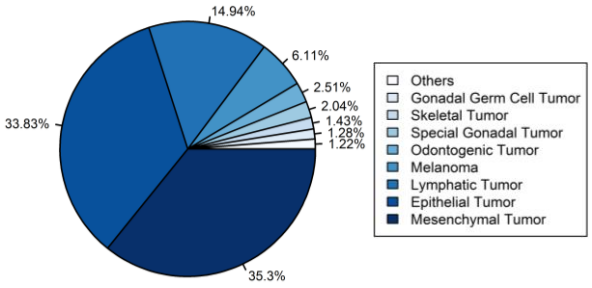



Figure 12.10. Main malignant tumour morphology groups from 2008 to 2013.

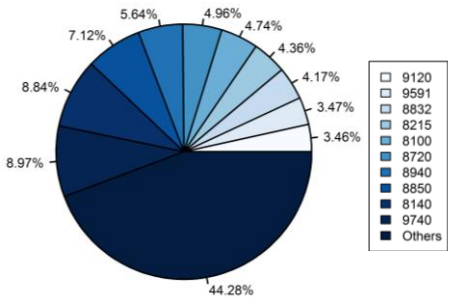



Figure 12.11. Most common specific malignant tumour diagnosis codes according to their morphology from 2008 to 2013.

Figure 12.10 presents the main morphology groups for malignant tumours from 2008 to 2013. The groups summarise the specific morphologies of the tumours. 'Mesenchymal tumour', 'epithelial tumour', 'lymphatic tumour' and 'melanoma' are among the most common morphology groups. A more specific diagnosis is illustrated in figure 12.11 which presents the most important specific tumour diagnosis codes according to their specific morphology. Since most of the specific ICD-O-3 tumour codes account for less than three percent of all diagnosis, this value is taken as the limit for the determination of the most common cancer types. Table 12.8 presents these ten specific tumour codes with the diagnosis name.

Table 12.8. Specific (ICD-O-3) tumour code and diagnosis of the most common cancer types.

| Specific tumour code | Specific diagnosis |
|---|---|
| **9120** | Hemangioma/ Hemangiosarcoma |
| **9591** | Non-Hodgkin Lymphoma, NOS |
| **8832** | Dermatofibroma / Dermatofibrosarcoma, NOS |
| **8215** | Adenocarcinoma of Anal Glands |
| **8100** | Trichoepithelioma |
| **8720** | Melanoma |
| **8940** | Mixed Tumour |
| **8850** | Lipoma /Liposarcoma |
| **8140** | Adenoma / Adenocarcinoma |
| **9740** | Mastocytoma / Mast Cell Sarcoma |

# 13 Disease Mapping

In order to describe the overall disease distribution, the prepared and cleaned data is used to examine and map cancer risk. The procedure is performed per year for all cancer types together as well as for each of the ten selected cancer types. The results of the SMR calculations are visualised in alpha-by-value maps and summarised at the end of this part.

## 13.1 SMR Calculation

The literature review has indicated that age, sex and size are known confounding variables of canine cancer. In the previous chapter, the variables have already been categorised in both data sets. Before the SMR can be calculated, the strata are determined. The strata are determined by combining the categories of the three variables. Each possible combination represents one strata. Thus, each dog can be uniquely assigned to strata since the variables also contain a category for unknown values. In a next step, the canine population as well as the registered cancer cases per strata (age, size, sex) is determined and the strata ratio, the ratio of the strata cancer cases divided by the strata population, is calculated (as illustrated in table 13.1.).

Table 13.1. Example of ratio calculation of cancer cases to population per strata according to confounding variables.

| **Strata** | **Strata specification** | | | **Strata cancer cases** | **Strata population** | **Strata ratio** |
|---|---|---|---|---|---|---|
| | Age | Sex | Size | | | |
| **Strata 1** | 3 | 1 | 1 | 177 | 19821 | 0.0089 |
| **Strata 2** | 3 | 1 | 2 | 285 | 34620 | 0.0082 |

Since the spatial information is available at municipal level, the population of strata is determined per municipality. Now the percentage of the Swiss canine population for the strata and the canine population of the strata per municipality are available. With these standardised data, the actual SMR calculation can start. According to Rezaeian et al. (2007), the SMR is calculated by dividing the observed number (O) by the expected number (E) of cancer diagnoses within the study population (Rezaeian et al., 2007). The information of the observed cancers per municipality is already available in the Canine Cancer Registry. Thus, the records per municipality only have to be summed up. At this point it has to be mentioned that for the SMR calculation of the most common cancer types, only the records with the specific tumour codes were considered (see table 12.8). The expected cancer cases per municipality can be determine as follows. By multiplying the strata population of the municipalities with the previously determined strata ratio, the expected cancer cases per municipality and strata can be calculated. This step is illustrated in table 13.2.

Table 13.2. Example of strata population per Swiss municipality.

| FSO-No. | Strata | Strata Population in municipalities | Strata ratio | Expected |
|---|---|---|---|---|
| **1** | Strata 1 | 304 | 0.0089 | 2.715 |
| **1** | Strata 2 | 379 | 0.0082 | 3.120 |

Subsequently, by summing up all the expected cancer cases of strata per municipality, the expected cancer cases per FSO-No. are determined. Table 13.3 shows an example of the final SMR calculation. The SMR presents the ratio of observed cancer cases to the expected cancer cases. The subsequent part visualises the results of cancer risk in the Swiss municipalities.

Table 13.3. Example of final SMR calculation per Swiss Municipality.

| FSO-No. | Population | Observed | Expected | SMR |
|---|---|---|---|---|
| **1** | 9395 | 151 | 63.63 | 2.37 |
| **2** | 854 | 5 | 5.90 | 0.85 |
| **3** | 512 | 6 | 3.52 | 1.71 |
| **4** | 422 | 2 | 2.77 | 0.72 |

## 13.2 Mapping Cancer Risk

In this part of the thesis, the cancer risks are illustrated graphically. In a first step the spatial distribution of the SMR are visualised in a map. Overall, disease maps of all cancer types together as well as for the ten selected cancer types were created per year from 2008 to 2013. The maps are in the appendix as well as on the supplied CD-ROM and are used as a help of the result interpretation. In order to compare the municipalities, the SMR values were categorised in four classes, according to the recommendations of Roth et al. (2010). The categories are presented in table 13.4. A value of one means that just as many tumours were recorded, as expected. Values higher than one mean that more tumours were diagnosed than expected and thus the cancer risk is increased. Values lower than one mean that fewer cancer cases were recorded than expected. Therefore, the cancer risk is low. A municipality with a SMR of 0.5 means that 50 percent fewer cancer cases were diagnosed than expected (1.5: 50 percent more than expected).

Table 13.4. Classification of SMR values.

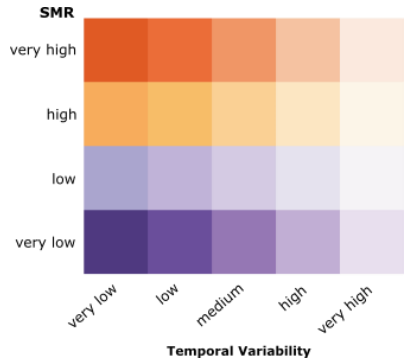
| SMR class | SMR category |
|---|---|
| **0 – 0.5** | Very low risk |
| **0.5 – 1** | Low risk |
| **1 – 1.5** | High risk |
| **> 1.5** | Very high risk |

The aim is, to visualise the spatial distribution of cancer risk together with the temporal variation from 2008 to 2013. For this purpose, the method alpha-by-value mapping was used. More precisely, the first approach of alpha-by-value mapping was utilised. For this approach, the alpha value of the enumeration unit increases parallelly to the increase of the value of the equalising variable. In this case, the SMR is the variable of interest that is being mapped with colour. It is also the variable that is weighted by the equalising variable. To identify the SMR value representing the time period from 2008 to 2013, the median value was calculated and classified according to the four categories in table 13.4. The median is the 50$^{th}$ percentile[6] and is used to determine the central value of a set of numbers, in this case the central value of the SMR (McPherson, 2013). The equalising variable is the variability of the SMR values over the time period. This variable is symbolised by alpha and thus visually equalises the map. Since the SMR value for 2008 to 2013 is represented by the median, the interquartile range (IQR) was determined to measure these SMR variation over time. The IQR is a measure of variability and represents the difference between the 75$^{th}$ and the 25$^{th}$ percentile. The median and the IQR provide information on the location and spread of values. Alternatively, the mean and standard deviation could have also been used. The problem with mean and standard deviation is that they are not robust against outliers and are significantly affected by errors (Upton & Cook, 1996). For this reason, the median and IQR were used. According to the recommendations of Roth et al. (2010), the IQR values are classified into five classes (very low, low, medium, high and very high temporal variability). The upper limit of the classification is defined by the maximum of IQR. For the comparison of the maps, the other class breaks are identical for all maps. These breaks are specified based on the SMR of all

[6] A percentile is a measure indicating the number where a certain percentage of values fall below (McPherson, 2013).

cancer types together by determining the statistical values minimum, 25th percentile, median, mean and 75th percentile of the IQR. The lower limit is defined by both, minimum IQR and 25th percentile of the IQR, since both have a value of zero. White was chosen as the colour that modifies the original unit colour in accordance with the alpha value changes. The combination of the categories of median and IQR determines the colour which represents the SMR changes in time in the map (the colour scheme is shown in figure 13.1).

Figure 13.1. Colour scheme of the value-by-alpha-map.




### 13.3 Results

The results of disease mapping are described in this part. The description refers mainly to the alpha-by-value maps. For a better understanding of individual results the choropleth maps were consulted. These are presented in the appendix of this work. As illustrated in the map in figure 13.2, a majority of Swiss municipalities exhibit a low to very low cancer risk. High risks of cancer can particularly be observed in the midland of Switzerland, mainly for municipalities in the region around Zurich up to Aargau and Schaffhausen. In these region, some municipalities can be identified with a very cancer risk (up to seven times higher than normally expected). Increased cancer risk is also observed in municipalities located in the border region of the cantons of St. Gallen, Glarus and Graubünden (from Valsot to Davos). The canton of Bern (esp. the municipalities Bern and Thun and its surrounding municipalities) is a further region where municipalities with increased cancer risk are located. What is striking about the cancer risk distribution is that many municipalities with a high risk of cancer are located

in the area of Zurich and Bern, where the three veterinary diagnostic laboratories are also localised.

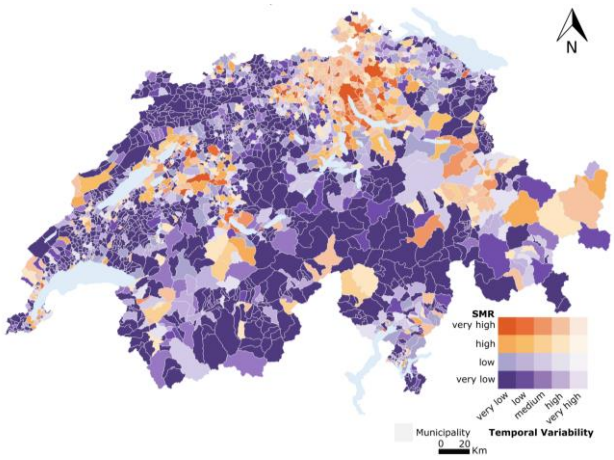



Figure 13.2. Disease map of relative cancer risk for all cancer records in Swiss municipalities from 2008 to 2013.

The other regions around Basel, Jura, Solothurn down to Ticino or from the Valais to Vaud mainly show low to very low cancer risk. In these regions, few municipalities are localised with increased cancer risk. These municipalities can mainly be observed south and south-west of Lake Neuchâtel and Lake Geneva, in the Valais, in Vaud (around the municipality of Bex) and close to the national border to Italy from Reckingen-Gluringen to Cevio.

Consulting the choropleth maps of each year (appendix A - K) show that the cancer risk of municipalities can change significantly between the years. This can be illustrated by the example 'Wohlen bei Bern' (FSO-No. 360). For this municipality, a very high cancer risk is determined in 2008, 2010 and 2011. However, the cancer risk is low in 2009 and 2012 and very low in 2013.

The variability of the SMR values is shown in the map (figure 13.2). The map illustrates that most municipalities with a very low and low risk of cancer also have a low SMR variability. This means that the low cancer risk has not changed significantly for these municipality from 2008 to 2013. Spatio-temporal clusters are expected in regions, where the ratio between expected and observed cancer cases is high and the temporal variability of this increased cancer risk low.

### *13.3.1 Disease Maps of the Specific Tumour Types*

The disease maps of the cancer risk for the specific cancer types present a predominantly low cancer risk in Switzerland. This is unsurprising since the number of cancer records for the specific malignant tumour types is small and thus for many municipalities no cancer cases are recorded (this issue is further discussed in the final part of the thesis). However, the spatial distribution is similar to that of all cancer types together. Increased cancer risk for the specific tumour types is observed in many regions as for all cancer records. For each of the ten cancer types, high cancer risks can be found especially in urban areas (Zurich, Winterthur, Bern etc.). Subsequently, the spatial distribution of the risks of the individual cancers is described below.

**Mastocytoma/ Mast Cell Sarcoma** (~ 9% of all cancer records)

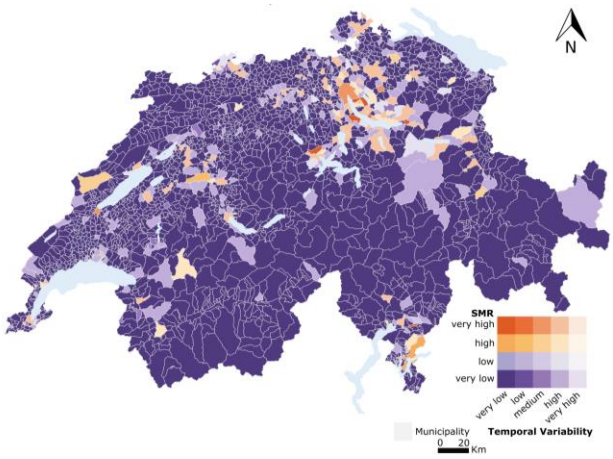



Figure 13.3. Disease map of relative cancer risk of 'mastocytoma/ mast cell sarcoma' in Swiss municipalities from 2008 to 2013.

For the malignant tumour type 'mastocytoma/ mast cell sarcoma', very high values up to 11.4-fold increased cancer risk exist (see figure 13.3). Most municipalities with an increased risk of cancer can also be observed in the region of Lake Zurich. Some municipalities with high cancer risk can be found especially in urban areas of Bern, Thun, Geneve and Lugano etc. A further region with high cancer risk is the north-western lakeside of Lake Lucerne (esp. the municipality Emmen). The cancer risks of most of these municipalities show a high temporal variability. However, when looking at the cancer risk of the entire period, few municipalities such as Zurich, Horgen, Zollikon, Maur and Schmerikon have a very high cancer risk and low temporal variability. For these areas, spatio-temporal clusters are expected.

**Adenoma/ Adenocarcinoma** (~ 8.8% of all cancer records)

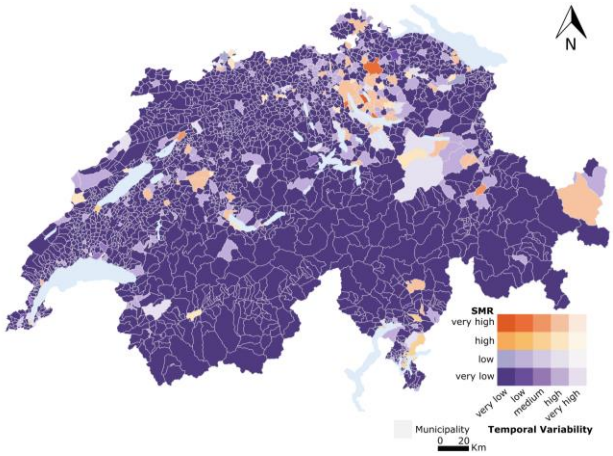



Figure 13.4. Disease map of relative cancer risk of 'adenoma/ adenocarcinoma' in Swiss municipalities from 2008 to 2013.

The disease map of 'adenoma/ adenocarcinoma' shows a similar spatial distribution as the one of 'mastocytoma/ mast cell sarcoma'. However, almost all municipalities with increased cancer risk also have a high temporal variability. Very high temporal variability is also observed in municipalities with low cancer risk such as Glarus (South and North). For this type of cancer large changes in cancer risks can be observed over the years. Therefore, purely spatial patterns and fewer temporal patterns are to be expected. Exceptions are the municipalities of Winterthur, Maur, Adliswil and Stäfa which show a very high cancer risk and low temporal variability.

**Lipoma/ Liposarcoma** (~ 7% of all cancer records)

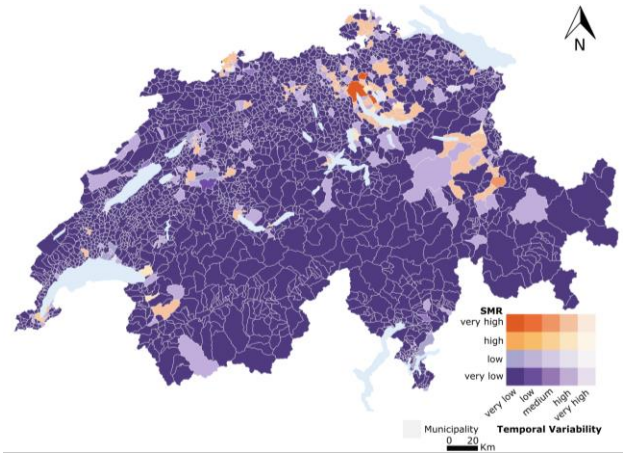



Figure 13.5. Disease map of relative cancer risk of 'lipoma/ liposarcoma' in Swiss municipalities from 2008 to 2013.

An increased cancer risk for the malignant tumour type 'lipoma/ liposarcoma' can mainly be localised in municipalities of the region of Zurich and Lake Zurich. In addition, small regions with increased cancer risk exist around Basel, Schaffhausen and Geneva as well as

around Glarus (South and North) and from Trimmis to Domat/Ems. All municipalities with increased cancer risk exhibit a very high temporal variability. Exceptions are Zurich, Maur and Kloten, which show a very high risk of cancer (up to 26-fold increased) and a low temporal variability.

**Mixed Tumour** (~ 5.5% of all cancer records)

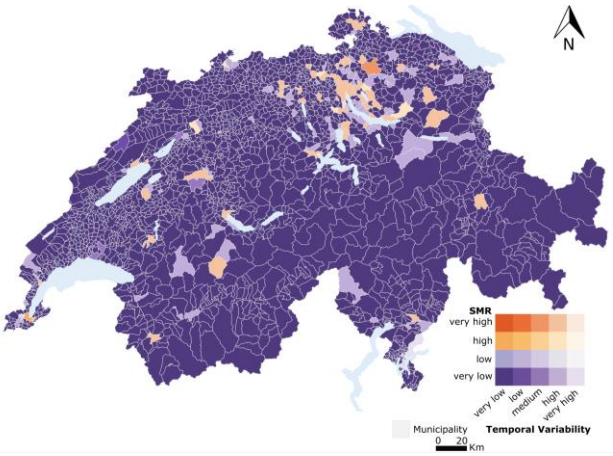



Figure 13.6. Disease map of relative cancer risk of mixed tumours in Swiss municipalities from 2008 to 2013.

For the group of mixed tumours, no municipalities with very high cancer risks and only a few with high cancer risk exist. Similar as for the previous cancer types, municipalities with high cancer risk are mainly located in the region of Zurich. Of these, the municipality of Winterthur is the only municipality that has a high cancer risk with a medium temporal variability.

**Melanoma** (~ 5% of all cancer records)

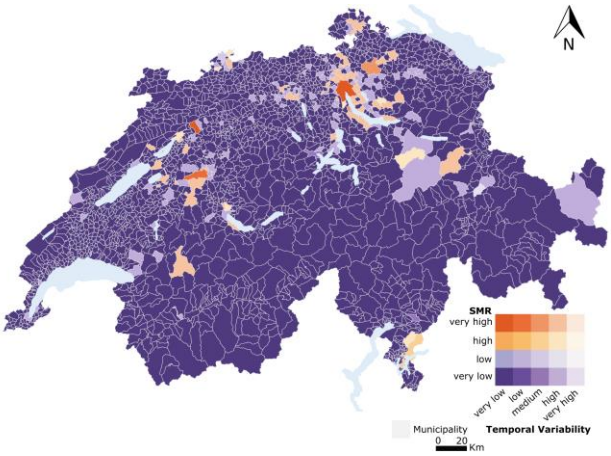



Figure 13.7. Disease map of relative cancer risk of 'melanoma' in Swiss municipalities from 2008 to 2013.

For the cancer type 'melanoma', municipalities with increased cancer risk are also predominantly located in the area of Zurich. Again, the cancer risks between the years strongly differ. Only the municipalities of Zurich, Bern and Grenchen provide a very high cancer risk with low temporal variability.

The five tumour types 'trichoepithelioma' (~4.7% of all cancer records), 'adenocarcinoma of anal glands' (~4.3% of all cancer records), 'dermatofibroma/ dermatofibrosarcoma, NOS' (~4% of all cancer records), non-Hodgkin (~3.5% of all cancer records) and 'hemangioma/ hemangiosarcoma' (~3.5% of all cancer records) a very similar map for the spatial cancer risk distribution in Switzerland (figures 13.8-13.12). All of them have an increased risk with high variability, mainly in urban municipalities. Some exceptions can be observed for municipalities with high cancer risk and low temporal variability; for instance, for the tumour type 'trichoepithelioma' in the municipality of Zurich. For the non-Hodgkin species, the municipalities of Bern and Lugano show a rather low temporal variability. Lugano even has a very high cancer risk. For the 'hemangioma/ hemangiosarcoma' type of cancer, very few municipalities exist with an increased cancer risk. This is unsurprising, since it has the lowest number of cancer registrations compared with the other nine cancer types.

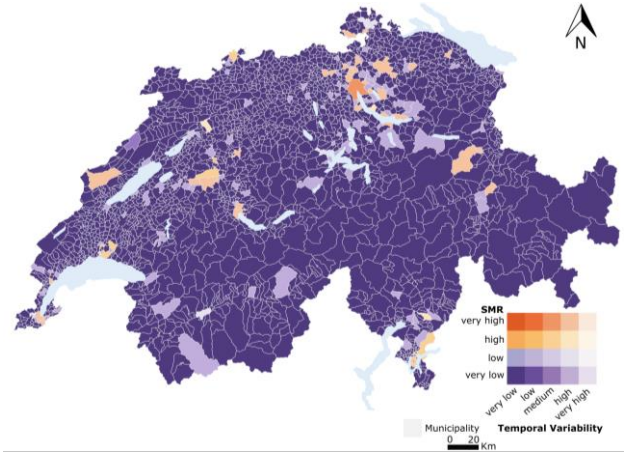



Figure 13.8. Disease map of relative cancer risk of 'trichoepithelioma' in Swiss municipalities from 2008 to 2013.

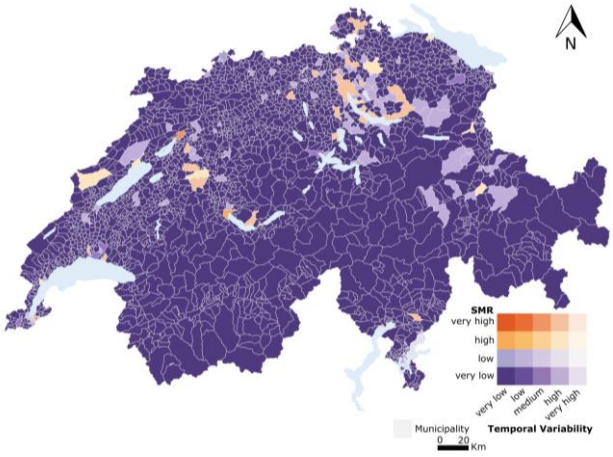



Figure 13.9. Disease map of relative cancer risk of 'adenocarcinoma of anal glands' in Swiss municipalities from 2008 to 2013.

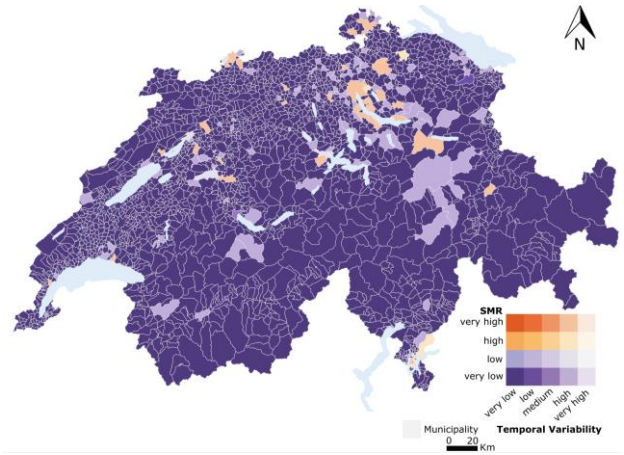



Figure 13.10. Disease map of relative cancer risk of 'dermatofibroma/ dermatofibrosarcoma, NOS' in Swiss municipalities from 2008 to 2013.

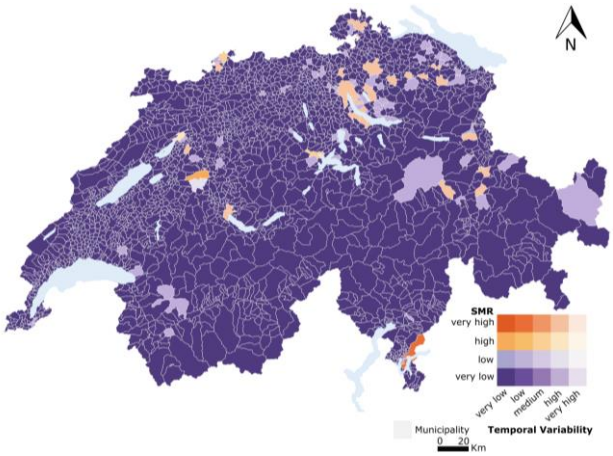



Figure 13.11. Disease map of relative cancer risk of non-Hodgkin in Swiss municipalities from 2008 to 2013.

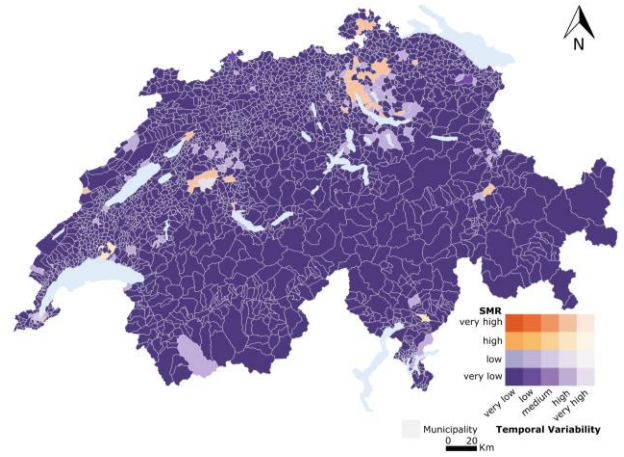



Figure 13.12. Disease map of relative cancer risk of 'hemangioma/ hemangiosarcoma' in Swiss municipalities from 2008 to 2013.

# 14 Cluster Analysis

In order to detect spatial as well as temporal disease patterns, a cluster analysis was conducted. Prior to the analysis, the heterogeneity of the SMR calculations is assessed. The cluster analysis itself is performed by using the software SaTScan™ of Kulldorff et al. (2015), which provides methods for spatio-temporal analysis of disease events (Kulldorff, 2015). Since the method did not provide results for the specific cancer types, the cluster analysis was further performed by using the local indicator of spatial association (LISA). The cluster results are visualised in maps according to chapter 13.2 and presented at the end of each analysis.

## 14.1 Testing the Homogeneity of the Cancer Risks

Conducting analysis of the presence of clusters is only useful if differences among the cancer risks exist. Since the expected and observed number of cancer cases are known, a chi-square test can be carried out, which tests for significant differences between the two values (Bivand, Pebesma & Gomez-Rubio, 1993). The test was performed for all SMR calculations (for all tumour types together and for the ten selected tumour codes per year). The chi-square is computed by the equation in figure 14.1 (Bivand, Pebesma & Gomez-Rubio, 1993).

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - \theta E_i)^2}{\theta E_i},$$

Figure 14.1. Equation of the chi-square test (Bivand et al., 1993).

The null hypothesis of this test asserts that the populations are homogeneous with respect to the observed and the expected cancer cases. The test performs 999 permutations. If the p-value of the test is less than the significance level of 0.05, the null hypothesis has to be rejected. In this case, heterogeneity is proved (Backhaus et al., 2011; Bivand et al., 1993).

### *14.1.1 Results*

The results of the homogeneity test in table 14.1 show that spatial clusters can be expected for all cancer registrations together as well as for the specific cancer types, since differences among the cancer risks exist ($p < 0.05$). However, for the specific cancer types – with the

exception of the malignant tumours 'mastocytoma/ mast cell sarcoma' and 'adenoma/ adenocarcinoma' – clusters cannot be expected for all years. This is because the p-values of the individual years are above the significance level of 0.05 (highlighted in red). This means that the cancer risks do not differ significantly for these years. Whether spatio-temporal clusters can actually be localised will be examined in the cluster analysis. However, the test already provides information about the expectance of spatio-temporal clusters. For instance, no space-time clusters are expected for the cancer types of 'non-Hodgkin lymphoma, NOS' and 'hemangioma/ hemangiosarcoma' since the cancer risks are not heterogenous for two consecutive years. Nonetheless, the cluster analysis is carried out for all tumour types and years and the results are compared later with the cluster analysis.

Table 14.1. Results of homogeneity test (p-values) per cancer type and year.

| Cancer types | Entire period | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|---|---|
| **All** | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| **9740** | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.013 | 0.012 |
| **8140** | 0.001 | 0.003 | 0.002 | 0.012 | 0.008 | 0.001 | 0.013 |
| **8850** | 0.001 | **0.097** | 0.002 | 0.002 | **0.109** | 0.043 | **0.069** |
| **8940** | 0.002 | 0.007 | 0.008 | 0.007 | 0.004 | 0.002 | 0.002 |
| **8720** | 0.001 | **0.053** | 0.003 | 0.002 | 0.009 | 0.001 | **0.134** |
| **8100** | 0.001 | **0.285** | 0.001 | 0.006 | 0.016 | 0.004 | **0.101** |
| **8215** | 0.001 | 0.017 | 0.001 | 0.018 | **0.09** | **0.683** | 0.021 |
| **8832** | 0.001 | 0.029 | **0.285** | 0.021 | 0.009 | **0.086** | 0.005 |
| **9591** | 0.004 | **0.186** | 0.004 | **0.753** | **0.91** | **0.134** | 0.033 |
| **9120** | 0.003 | **0.056** | 0.012 | **0.087** | **0.174** | 0.03 | **0.33** |

## 14.2 Scan Statistics

To date, previous analyses have assessed the presence of heterogeneity of cancer risk in Switzerland and offered a general assessment of the presence of clusters. However, given that the actual locations of these clusters remain unknown, a different approach must be followed. The definition of clusters and the general underlying statistics were presented in the theoretical background of this work. Many different methods and analysing machines of detecting cluster locations exists. According to Hjalmars et al. (1996) very

useful methods to detect spatio-temporal disease clusters are scan statistics (Hjalmars et al., 1996). Scan statistic methods are based on moving windows covering only a few areas each time and for which a local test of clustering is carried out. By moving the windows and repeating the procedure the whole study area can be covered and the detection of the locations of clusters of disease is possible (Bivand et al., 1993).

For this master thesis, the cluster analysis is conducted with a scan statistic method provided by Kulldorff's statistic software SaTScan™. This free software is designed for spatial, temporal and space-time disease data analysis. Kulldorff's statistic distinguishes six different models (Kulldorff, 2015).

### *14.2.1 Scan Statistic for Continuous Data*

A scan statistic for continuous data was developed in 2009 by Kulldorff, Huang and Konty. This method is based on the normal probability model (Kulldorff, Huang & Konty, 2009). Since the SMR data from Switzerland have continuous values, this model was chosen for cluster location detection. The test is performed based on the following null and alternative hypotheses:

***Null hypothesis***: "*All observations come from the same distribution*"
(Kulldorff et al., 2009, p. 8).

***Alternative hypothesis***: "*There is one cluster location where the observations have either a larger or smaller mean than outside that cluster*"(Kulldorff et al., 2009, p. 8).

Since the research questions of this thesis also address temporal patterns, the test scans for spatio-temporal cluster locations. For this purpose, the test compares the number of observed cases in a cluster to what would have been expected if the spatial and temporal locations of all cases were independent of each other. In accordance with this, the test detects a cluster in a geographical area if that area has a higher proportion of its cases during a specific time period compared to the remaining geographical areas. This means that if one geographical area has more observed cases compared to normal while other areas have a normal number during a one-year period, there will be a cluster in that first area (Kulldorff, 2015; Kulldorff et al., 2009). Since it makes more sense to only search for

clusters with exceptionally high values, the default parameter setting was adjusted. The space-time permutation model automatically adjusts for both purely spatial and purely temporal clusters. The moving cylindrical window of this test has a circular geographic base with a height that corresponds to the time period of potential clusters. This cylindrical window is moved in space and time, whereby it visits the entire study area at each possible time period. Each of these cylindesrs reflect a possible cluste*r* (Kulldorff, 2015).

The data that is analysed comprises a number of continuous SMR observations with values $x_i$ (i = 1,…,N). Each observation is at a spatial location s (s = 1,...,S), with spatial coordinates (latitude lat(s) and longitude long(s)) (Kulldorff et al., 2009). For this purpose, the centroids of the municipalities were determined and its coordinates transformed from CH1903 / LV03 to WGS84. Thus, the locations of the study area are the centroid points of Swiss municipalities. These locations can have one or more observations (maximum one observation per year) so that S ≤ N. In a next step, the sum of the observed SMR values $x_s$ are defined for each location s as $x_s = \sum_{i \in s} x_i$. Further the number of observations in the location is defined as ns. X is the sum of all the observed SMR values $X = \sum_i x_i$. These observations and locations are used for the likelihood calculations of the cycles (Kulldorff et al., 2009).

Now, the log likelihood ratio LLR(z) is calculated for each circle z. For this purpose, the maximum likelihood under the null hypothesis as well as under the alternative hypothesis was calculated according to the equations in figures 14.2 and 14.3 (null hypothesis) and figures 14.4 and 14.5 (alternative hypothesis). Note that the calculation under the alternative hypothesis compares the mean inside the circle with the mean outside the circle (Kulldorff et al., 2009).

$$L_0 = \prod_i \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Figure 14.2. Equation to calculate the likelihood under the null hypothesis (Kulldorff et al., 2009).

$$lnL_0 = -Nln(\sqrt{2\pi}) - Nln(\sigma) - \sum_i \frac{(x_i-\mu)^2}{2\sigma^2}$$

Figure 14.3. Equation to calculate the log likelihood under the null hypothesis (Kulldorff et al., 2009).

$$\sigma_z^2 = \frac{1}{N}\left( \sum_{i \in z} x_i^2 - 2x_z\mu_z + n_z\mu_z^2 + \sum_{i \notin z} x_i^2 - 2(X - x_z)\lambda_z + (N - n_z)\lambda_z^2 \right)$$

Figure 14.4. Equation to calculate the maximum likelihood estimate under the alternative hypothesis for the common variance (Kulldorff et al., 2009).

$$lnL(z) = -Nln(\sqrt{2\pi}) - Nln(\sqrt{\sigma_z^2}) - N/2$$

Figure 14.5. Equation to calculate the log likelihood under the null hypothesis (Kulldorff et al., 2009).

The test statistic is defined as the maximum likelihood ratio (max LLR) over all circles that is later used to identify the most likely clusters (see equation in figure 14.6).

$$\max_z(lnL_z / lnL_0)$$

Figure 14.6. Equation to calculate the maximum LLR (Kulldorff et al., 2009).

The radius of the moving circle varies continuously in size from zero to the upper limit. The usage of upper limits of clusters ensures that both small and large clusters can be found (Kulldorff et al., 2009). The tests were carried out for various upper limit settings with respect to the recommendations of the authors. However, the results did not differ significantly. For this reason, the upper limit is defined by circles containing at most 50 percent of all observations.

In order to evaluate the statistical significance of the most likely cluster, SaTScan™ uses the Monte Carlo hypothesis testing. For this purpose, a large set of random data sets are created by random permutation of the observed values and their corresponding locations. Subsequently, for each random data set the most likely cluster is found by calculating the log likelihood lnL(z) for each circle (Kulldorff et al., 2009). For this test, 999 permutations are chosen. Thus, the most likely cluster of the real data set is statistically significant (p= 0.05), if the log likelihood ratio from the real data set is among the five percent highest of all the data set. The permutation based Monte Carlo hypothesis testing procedure makes possible that the statistical inference remains valid even if the true distribution is not normal (Kulldorff et al., 2009).

Finally, the results of the cluster analysis are presented in two maps. The alpha-by-value map represents the median and IQR values of the SMR calculations for all municipalities

located in the clusters. The temporal persistency map represents the duration for which the detected cluster is valid.

### *14.2.2 Results*

The cluster analysis of all cancers shows, as expected in chapter 13.3, a large cluster of increased cancer risk in the Zurich region. This spatio-temporal cluster occurs for three years. In addition, five small spatio-temporal clusters are localised for one to two years in the region south of Lake Neuchâtel and close to the national border to Italy. Considering the entire time period, the map shows that the cancer risk of almost all communities in the cluster is increased and partly also very high. For the regions of Bern and the border region of the cantons St.Gallen and Graubünden no clusters could be detected.

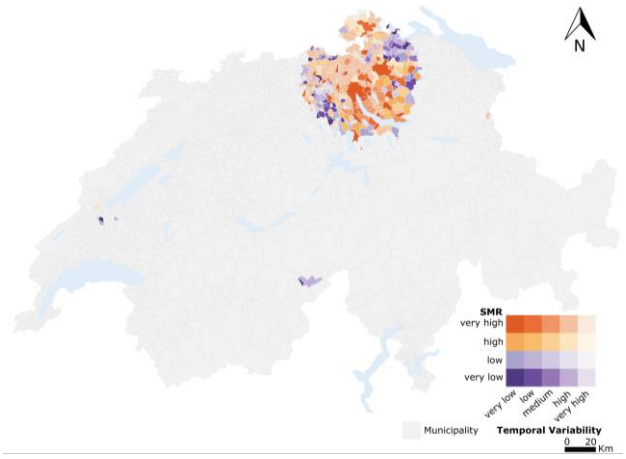



Figure 14.7. SaTScan™ cluster map of relative cancer risk for all cancer types in Swiss municipalities from 2008 to 2013.

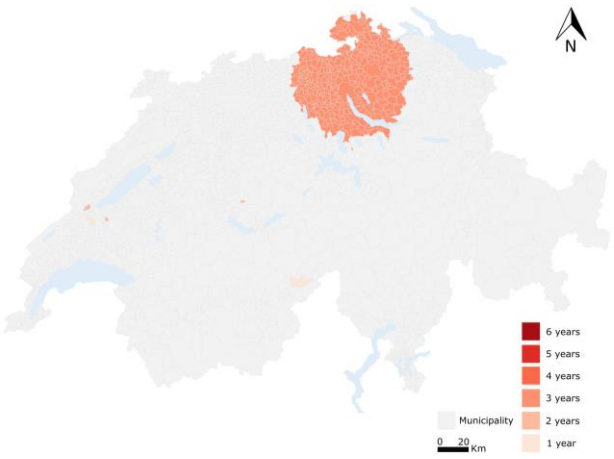



Figure 14.8. SaTScan™ cluster persistency map of relative cancer risk for all cancer types in Swiss municipalities from 2008 to 2013.

As illustrated in the disease maps in chapter 13.3, some regions exist where clusters for the specific cancer types would have been expected. However, the analysis of the specific cancer types does not yield any realistic spatio-temporal clusters. SaTScan™ only identifies single municipalities (usually only one) as a cluster that occur for a maximum of two years. When looking at the choropleth maps for the SMR from 2008 to 2013, it is shown that these are the municipality with the maximum SMR values (up to 87-fold cancer cases than expected). For this reason, the results are not further described and the few maps with detected clusters are on the supplied CD-ROM.

The unrealistic results could mainly be explained by the underlying methodology of SaTScan™. As already mentioned in the results of the disease maps, since only a few cancer cases are recorded for the specific cancer types, many municipalities have an SMR value of zero. Since the method compares the max LLR inside and outside the circles and determines the p-value based on a random data set of 999 permutations, the method has its difficulties with determining significant clusters for the specific cancer types. This is further discussed in the final part of this thesis.

### 14.3 LISA

Based on the results of the analysis with SaTScan™, another cluster analysis method is used. As the further cluster analysis, the LISA of Anselin (1995) is conducted because it is based on local statistics and spatial autocorrelation. Considering the underlying equation of LISA (figure 14.9), this method can better deal with the present SMR data, especially data of the specified cancer types. The analysis is conducted with the free software programm for spatial analysis GeoDa™ by using the method 'univariate Local Moran's I', which is based on the local Moran statistic of Anselin (1995) (Anselin, 1995; Anselin, Syabri & Kho, 2006). However, a temporal cluster analysis with LISA is not directly possible.

Similar to the Gi and G statistics of Getis and Ord (1992), LISA may indicate local spatial clusters of non-stationary (or hot spots) (Anselin, 1995; Getis & Ord, 1992). Furthermore, LISA can be used to identify outliers by assessing the influence of single locations on the magnitude of the global statistics. While global Moran's I quantifies the spatial autocorrelation for the whole study area, the LISA measure the degree of spatial autocorrelation at each specific location by using local Moran's I. The sum of LISAs for all the observations together is proportional to the global Moran's I. LISA indicates the extent of significant spatial clustering around an observation location, which have a similar value as the observation (Anselin, 1995). Local Moran's I index can be calculated according to the equation in figure 14.7.

$$I_i = \frac{z_i - \bar{z}}{\sigma^2} \sum_{j=1, j \neq i}^{n} \left[ W_{ij}(z_j - \bar{z}) \right],$$

Figure 14.9. Equation to calculate the local Moran's I (Levine, 2004).

In the present work, $z_i$ is the SMR value at the municipality location i. $\bar{z}$ is the mean and $\sigma^2$ the variance of all SMR samples. $z_j$ is the SMR value at other locations. $W_{ij}$ can be defined as the inverse of the distance since it is a distance weighting between the SMR value at the municipality location i ($z_i$) the values at other locations j ($z_j$) (Fu et al., 2014; Levine, 2004). In order to calculate LISA, the neighbourhood for each SMR observation ($z_i$) must be defined and can be formalised by means of a spatial weight matrix (or contiguity matrix) $W_{ij}$. The spatial weights matrices provide information about the neighbourhoods of a location (Anselin, 1995; Fu et al., 2014). Neighbourhood can be defined in various ways, whereby the most common definitions are based on distances (such as distance band or k-nearest neighbours) or contiguity of boundaries (Sawada, 2004).

Since the point data of cancer records is only available at the aggregated municipality level, methods that are based on distance measures are less useful. For this reason, an approach of contiguity was chosen to define the spatial weight matrices.

Contiguity has a rather broad definition depending on the research question, although most analyses in spatial autocorrelation define contiguity as the neighbourhood relations . These most common forms of contiguity used in spatial autocorrelation are rooks case, bishops case or queens case (Sawada, 2004). The three approaches are illustrated in figure 14.9. Rooks case contiguity considers only the locations of the four neighbourhood cells and bishops case the cell locations in the diagonals of the relation. Queen's case is the only case that considers all the eight cells of the neighbourhood (Sawada, 2004).

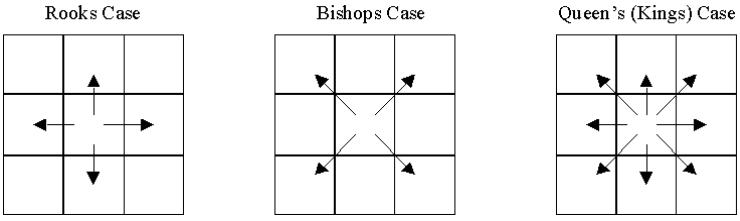



Figure 14.10 Approaches of contiguity: Rooks, Bishops and Queen's Case (Sawada, 2004).

To define the spatial weights matrix, the first-order (direct neighbourhood) contiguity of queen's case was used. Cells that are not directly related receive a weight of zero and related cells a weight of one. All column with non-zero values indicate the relevant neighbours of the SMR observation (Sawada, 2004).

Similar to SaTScan™, GeoDa™ defines the significance of local Moran's I through a permutation test. 999 permutations are used to determine the likelihood of observing the actual determined local Moran's I value under conditions of spatial randomness. Thus, each SMR observation is assigned a vector of randomly generated numbers. This vector is used to randomly relocate each SMR observation in space. For each permutation, similar to SaTScan™, the statistic is computed each time with a different set of random numbers to generate a random reference distribution of the local Moran's I (Anselin et al., 2006).
The significance level of 0.05 was chosen. Consequently, the detected spatial cluster centres are statistically significant, if the likelihood is among the five percent.

The results of the univariate local Moran's I test can be interpreted as follows. On the one hand, positive local Moran's I values imply that the target value is similar to its neighbourhood. This method allows the identification of location of high-high (high values in a high value neighbourhood) spatial cluster centres as well as low-low (low values in a low value neighbourhood) spatial cluster centres. On the other hand, negative local Moran's I values imply potential spatial outliers that are obviously different from the values of its neighbourhood. Spatial outliers can have a high value in a low value neighbourhood (high–low outliers) or a low value in a high value neighbourhood (low–high outliers) (Fu et al., 2014).
The cluster maps of all years and SMR calculations can be found on the supplied CD-ROM. In the present work, only high-high clusters are of interest, since they identify areas with a health problem.

Even if the LISA method does not identify space-time clusters, the variability in time can be assessed by comparing two maps. The first map assesses the changes of the cluster location detection during the time period by counting the years in which the cluster occurs at a location. Since the study period includes six years (from 2008 to 2013), a location could have a value of maximum six, showing that the cluster significantly exists over the entire period of six years. A value of zero identifying locations for which no signification clusters were detected. The second map assesses the variability of the SMR values within the clusters by alpha-by-value mapping according to chapter 13.2.

### *14.3.1 Results*

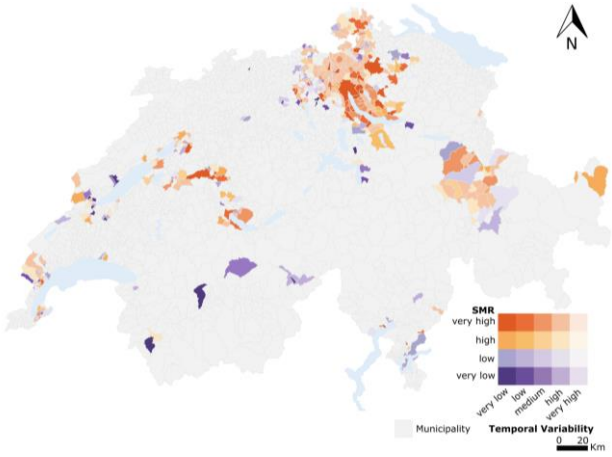



Figure 14.11. LISA cluster map of relative cancer risk for all cancer types in Swiss municipalities from 2008 to 2013.

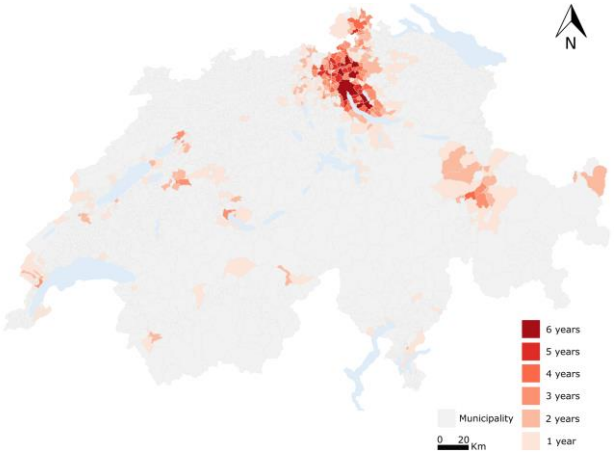



Figure 14.12. LISA cluster persistency map of relative cancer risk for all cancer types in Swiss municipalities from 2008 to 2013.

As previously with SaTScan™, cluster centres were detected in the region of Zurich (figures 14.11 and 14.2). These are partly detected (unlike SaTScan™) for the entire period from 2008 to 2013. More spatial clusters (that were not detected with SaTScan™), are mainly located in the region around Glarus (north and south), Bern and Biel as well as around the Lake Thun and west of the Lake Geneva. Many of the municipalities located in these regions only appear as a cluster centre for one year, although a few appear up to four years and have a high cancer risk with medium temporal variability for the entire period. These regions are interesting for further epidemiological studies.

The analysis with LISA also results in cluster for the specific cancer types. Subsequently, these results are presented per malignant tumour type.

**Mastocytoma/ Mast cell sarcoma** (~9% of all cancer records)

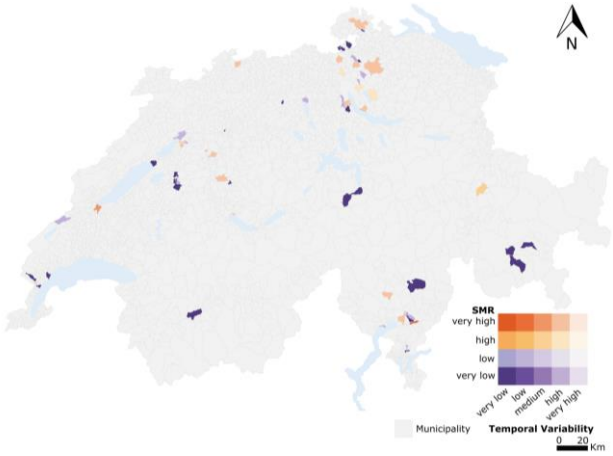



Figure 14.13. LISA cluster map of relative cancer risk for 'mastocytoma/ mast cell sarcoma' sarcoma in Swiss municipalities from 2008 to 2013.

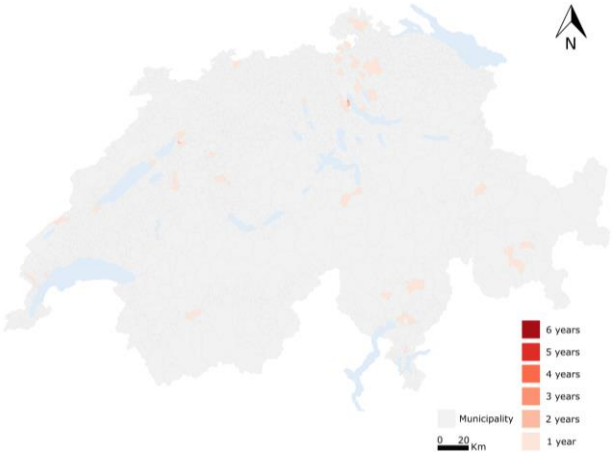



Figure 14.14. LISA cluster persistency map of relative cancer risk for 'mastocytoma/ mast cell sarcoma' sarcoma in Swiss municipalities from 2008 to 2013.

The malignant tumour type 'mastocytoma/ mast cell sarcoma' has with nine percent the largest share of cancer diagnoses in dogs in Switzerland. Considering the spatial distribution, few cluster centres are detected in the cantons of Zurich, Ticino, Bern and Graubünden, as well as in Geneva, Valais, and Schaffhausen. Although the homogeneity test has proved heterogeneity for the cancer risk of the single years, the patterns are mainly of spatial nature. Municipalities with increased cancer risks, the cluster centres only occur in one year. For this reason, this tumour type is rather uninteresting for further analyses since they only occur in one year.

**Adenoma/ Adenocarcinoma** (~8.8%) and **Trichoepithelioma** (~4.7% of all cancer records)

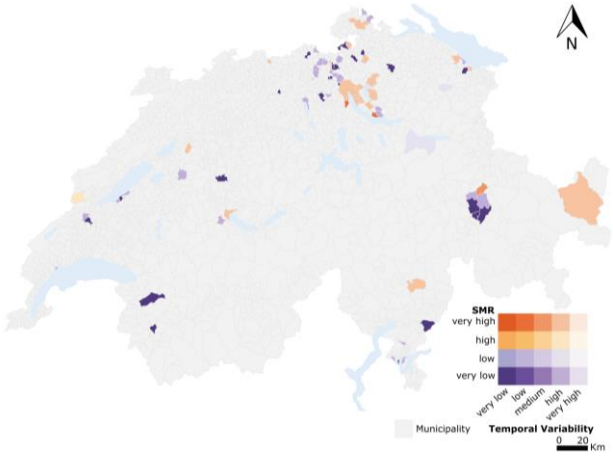



Figure 14.15. LISA cluster map of relative cancer risk for 'adenoma/ adenocarcinoma' in Swiss municipalities from 2008 to 2013.

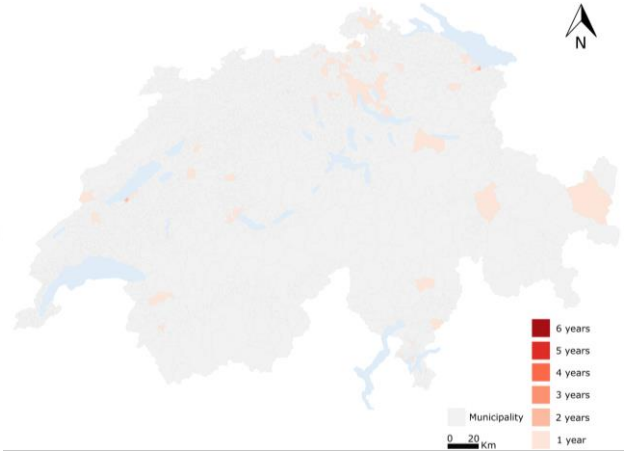



Figure 14.16. LISA cluster persistency map of relative cancer risk for 'adenoma/ adenocarcinoma' in Swiss municipalities from 2008 to 2013.

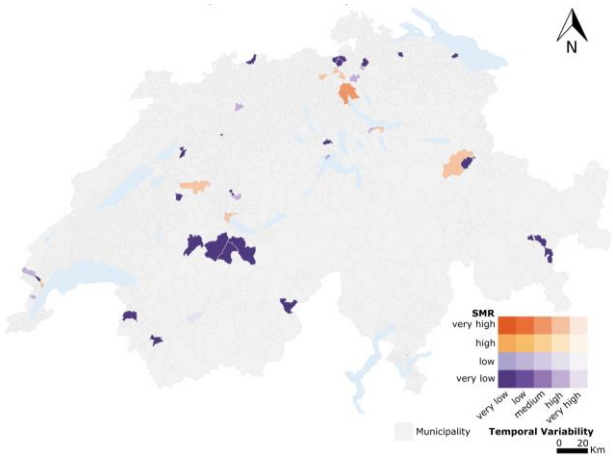



Figure 14.17. LISA cluster map of relative cancer risk for 'trichoepithelioma' in Swiss municipalities from 2008 to 2013.

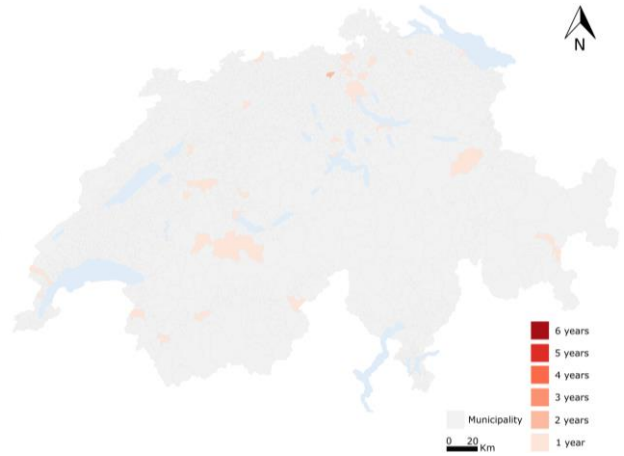



Figure 14.18. LISA cluster persistency map of relative cancer risk for 'trichoepithelioma' in Swiss municipalities from 2008 to 2013.

Both tumour types, 'adenoma/ adenocarcinoma' and 'trichoepithelioma', show a similar pattern and are relatively uninteresting for further analyses. Spatial patterns can be observed in the area of Zurich up to Schaffhausen and in the canton of Graubünden. However, these clusters area (similar to 'mastocytoma/ mast cell sarcoma') mainly spatial only available for one year (maximum two years).

**Lipoma/ Liposarcoma** (~7% of all cancer records)

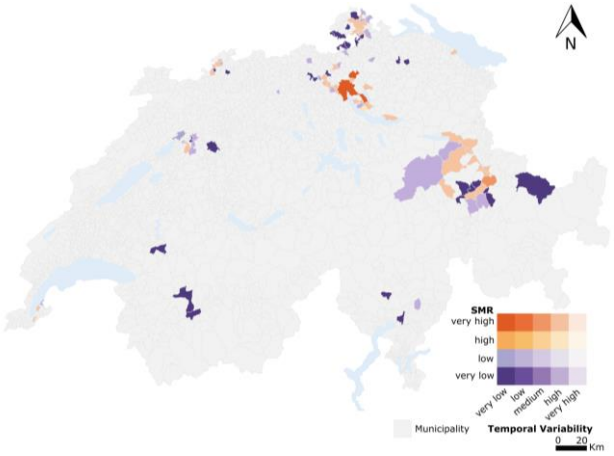



Figure 14.19. LISA cluster map of relative cancer risk for 'lipoma/ liposarcoma' in Swiss municipalities from 2008 to 2013.

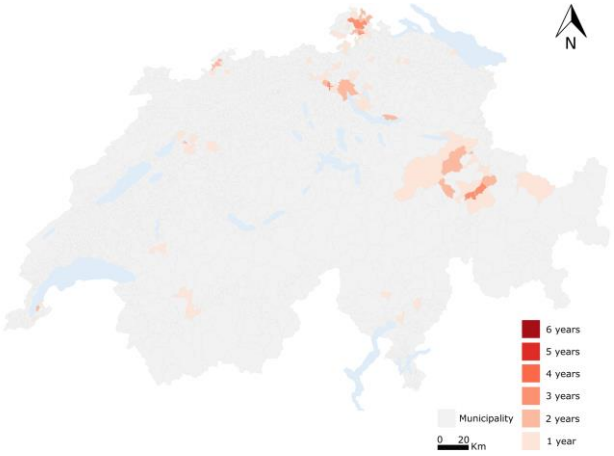



Figure 14.20. LISA cluster persistency map of relative cancer risk for 'lipoma/ liposarcoma' in Swiss municipalities from 2008 to 2013.

The cancer type of 'lipoma/ liposarcoma' is interesting for further analysis since some cluster centres occur over several years and, when considering the entire period, show an increased cancer risk. These cluster centres are mainly located around the municipalities of Zurich and Schaffhausen as well as in the border region of Glarus to Graubünden. Cluster centre that consists over three years are the municipalities Domas/Ems and Chur. Considering the entire period, the alpha-by-value map illustrates that some of the cluster centres in the identified regions even show a very low cancer risk. However, the municipalities Zurich, Kloten und Maur provide a very high cancer risk of with low temporal variability. Thus, they can be identified as spatio-temporal clusters.

**Mixed Tumour** (~5.5% of all cancer records)

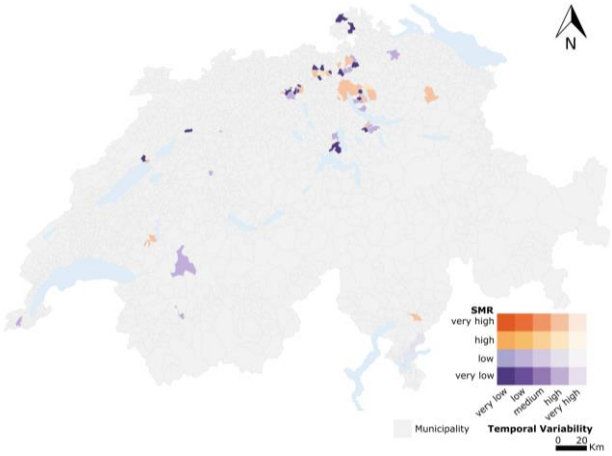



Figure 14.21. LISA cluster map of relative cancer risk for mixed tumours in Swiss municipalities from 2008 to 2013.

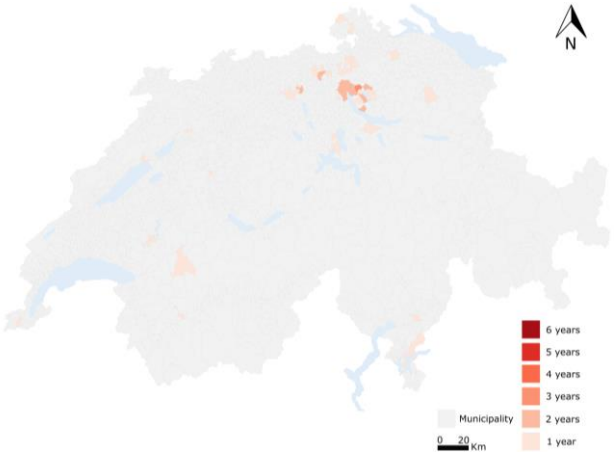



Figure 14.22. LISA cluster persistency map of relative cancer risk for mixed tumours in Swiss municipalities from 2008 to 2013.

Cluster centres of mixed tumours can be localised mainly in the northern Switzerland from Lake Lucerne to Schaffhausen. Most of these centres only exist for one year, although some can also be observed for up to three years (for example, Zurich, Maur, Dübendorf). Considering the SMR for the entire period, these municipalities provide a high cancer risk with medium temporal variability. However, for further analysis, this group of tumours is not very interesting, because it comprises mixed tumour types.

**Melanoma** (~5% of all cancer records)

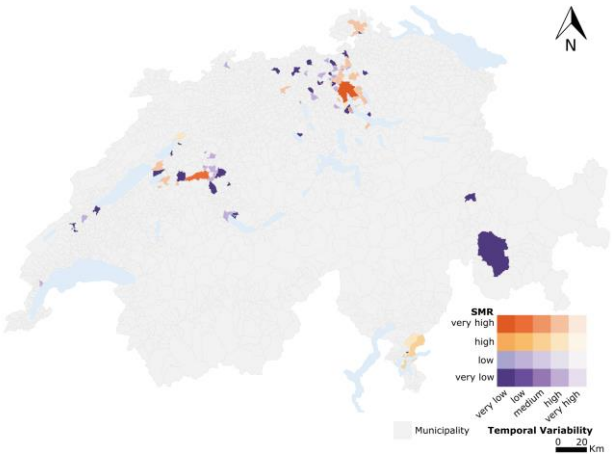



Figure 14.23. LISA cluster map of relative cancer risk for 'melanoma' in Swiss municipalities from 2008 to 2013.

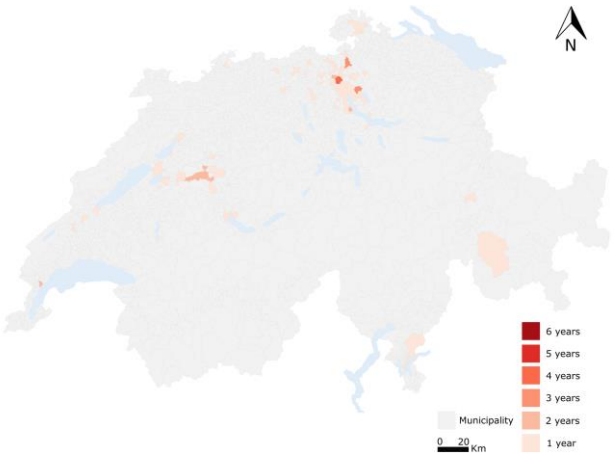



Figure 14.24 . LISA cluster persistency map of relative cancer risk for 'melanoma' in Swiss municipalities from 2008 to 2013.

For the cancer type of 'melanoma' tumours, some cluster centres can be identified in the regions around Bern and Zurich as well as south of Lake Neuchâtel. Single cluster centres

are also located in the cantons of Graubünden and Ticino. Most of the cluster centres only occur for one year. Significant cluster centres for three to four years can be observed in the municipalities of Bern, Regensdorf, Bülach and Dübendorf. These cluster centres provide a high cancer risk with low temporal variability. Thus, they are interesting for further epidemiological studies.

**Adenocarcinoma of Anal Glands** (~4.3% of all cancer records)

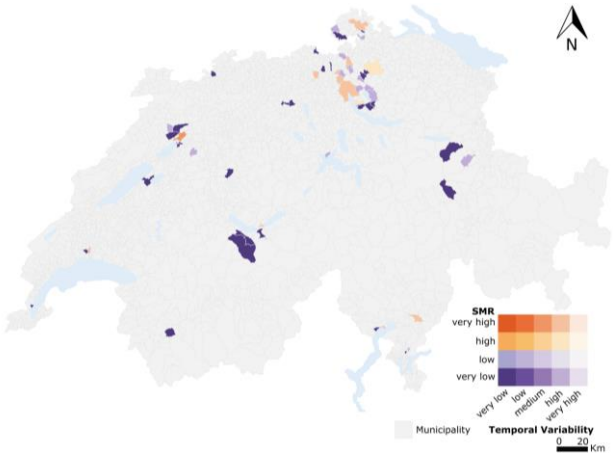



Figure 14.25. LISA cluster map of relative cancer risk for 'adenocarcinoma of anal glands' in Swiss municipalities from 2008 to 2013.

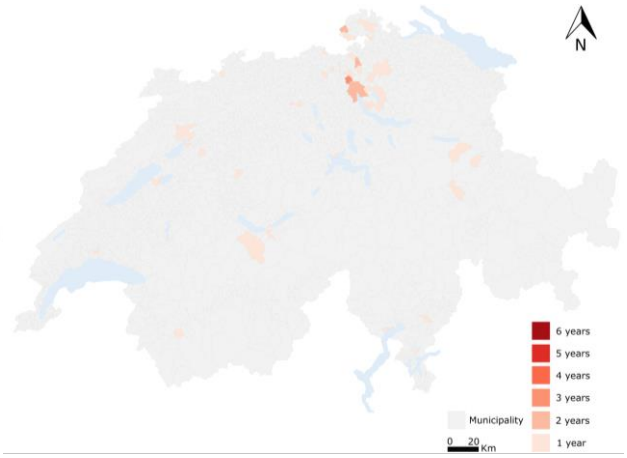



Figure 14.26. LISA cluster persistency map of relative cancer risk for 'adenocarcinoma of anal glands' in Swiss municipalities from 2008 to 2013.

For the cancer type of 'adenocarcinoma of anal glands' tumours, single cluster centres are distributed throughout Switzerland. Similar to the cancer type of 'melanoma', the cluster centres in the municipalities of Zurich, Regensdorf, Bülach and Hallau can be identified as spatio-temporal cluster centres. They are significant for two to three years and show an increased cancer risk over the entire period. However, they show a rather high temporal variability, which means that the SMR value from 2008 to 2013 can strongly change.

**Dermatofibroma / Dermatofibrosarcroma NOS** (~4% of all cancer records)

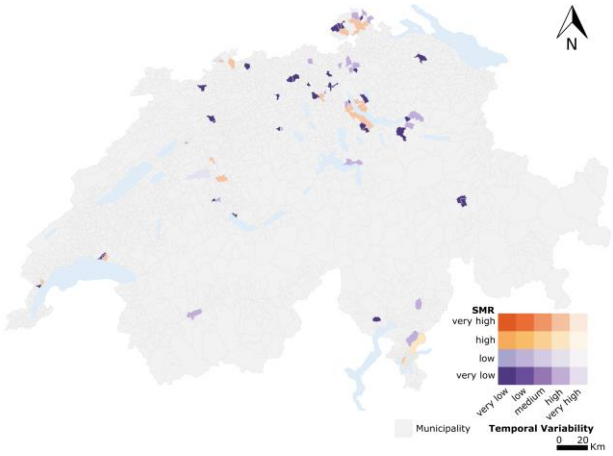



Figure 14.27. LISA cluster map of relative cancer risk for 'dermatofibroma/ dermatofibrosarcoma, NOS' in Swiss municipalities from 2008 to 2013.

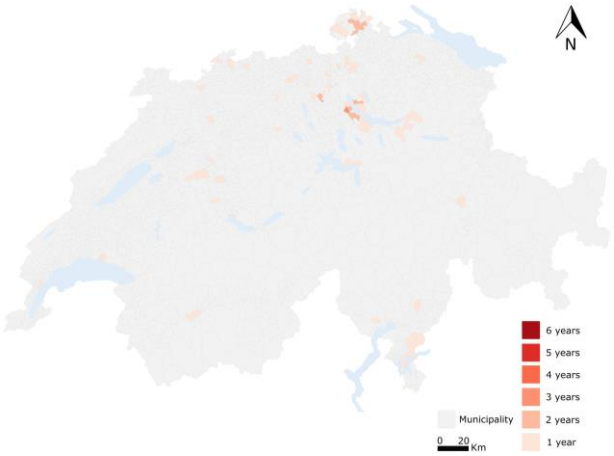



Figure 14.28. LISA cluster persistency map of relative cancer risk for 'dermatofibroma/ dermatofibrosarcoma, NOS' in Swiss municipalities from 2008 to 2013.

The results of the cluster analysis for the cancer group is not very interesting. Spatial clusters can be observed mainly in the northern midland. Single clusters are also located in the canton of Ticino and on Lake Geneva. However, most cluster centres significantly occur for one year. Only the municipalities of Schaffhausen and Neuhausen on the Rhinefall as well as Horgen, Langnau am Albis and Bremgarten (AG) are localised cluster centres that occur over several years (up to three years) and, considering the entire period, also show an increased cancer risk. However, these municipalities also provide an increased temporal variability of the SMR values.

**Non-Hodgkin's lymphoma** (~3.5% of all cancer records)

For the cancer type 'non-Hodgkin's lymphoma', similar to the cancer type of 'lipoma/ liposarcoma', cluster centres are located mainly in the border area of the cantons of Graubünden to St. Gallen and Glarus as well as around the northern basin of Lake Zurich. However, most of these cluster centres are not interesting as they only significantly occur for one year and / or do not show any increased cancer risk over the entire period. Only the municipalities of Schaffhausen and Eglisau as well as Meilen, Zollikon, are cluster centre for two years and, considering the entire period, show an increased cancer risk. However, these municipalities also provide an increased temporal variability of SMR values, which coincides with the findings of the homogeneity test.

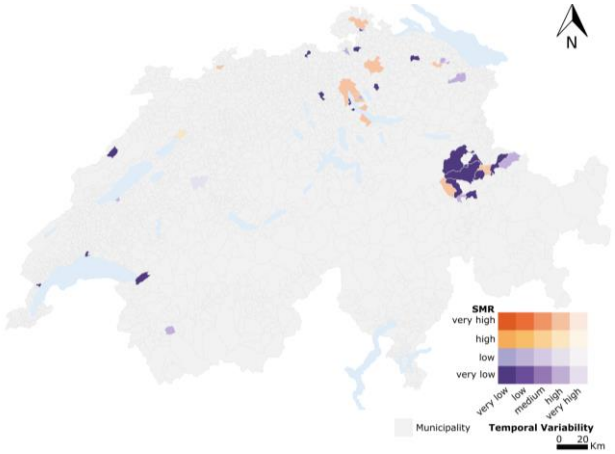



Figure 14.29. LISA cluster map of relative cancer risk for 'non-Hodgkin's lymphoma' in Swiss municipalities from 2008 to 2013.

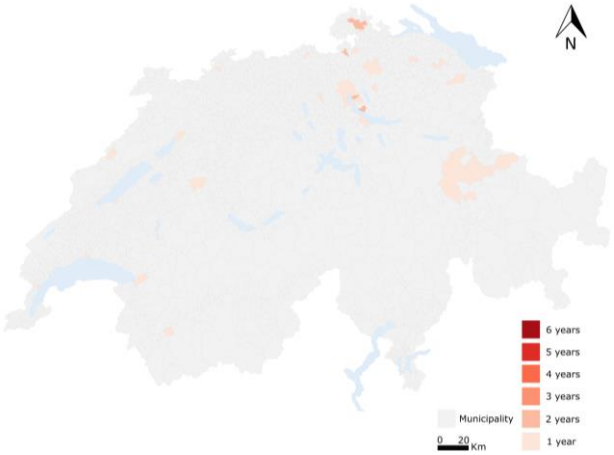



Figure 14.30. LISA cluster persistency map of relative cancer risk for 'non-Hodgkin's lymphoma' in Swiss municipalities from 2008 to 2013.

**Hemangioma/ Hemangiosarcoma** (~3.5% of all cancer records)

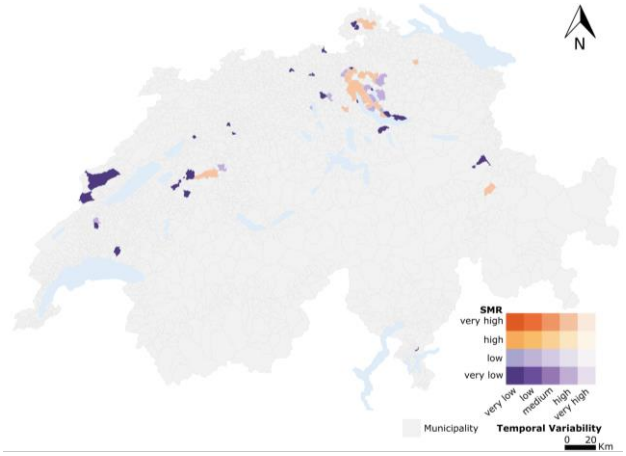



Figure 14.31. LISA cluster map of relative cancer risk for 'hemangioma/ hemangiosarcoma' in Swiss municipalities from 2008 to 2013.

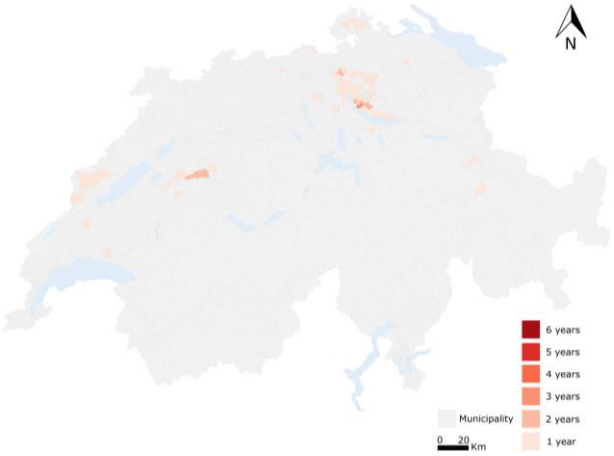



Figure 14.32. LISA cluster persistency map of relative cancer risk for 'hemangioma/ hemangiosarcoma' in Swiss municipalities from 2008 to 2013.

For the malignant tumour types 'hemangioma/ hemangiosarcoma', cluster centres exist around Zurich and along the eastern lakeside of Lake Zurich. In addition, cluster centres are located near Schaffhausen, west of Lake Neuchâtel and near Bern. Considering the whole period, the SMR values strongly change over time. This was already expected due to the results of the homogeneity test. Only the cluster centres of the municipalities of Niederhasli, Bern and the municipalities of Erlenbach (ZH), Herrliberg, Küsnacht (ZH), Egg exist for several years and show a high cancer risk over the entire period.

## 14.4 Conclusion

Considering all cancer records, the analysis shows that a large spatio-temporal cluster can be localised in the area of Zurich. However, it is very unlikely that risk factors for the different cancer types occur exclusively in these regions. This is further discussed in the final part of the thesis. In the analysis of the specific types of cancer, the different strengths and weaknesses of the methods SaTScan™ and LISA are present. The methodology SaTScan™ cannot identify real space-time clusters for the specific tumour types. However, several regions could be identified by means of the LISA method and may hold interest for further epidemiological studies. Spatial-temporal clusters could be localised for the malignant tumour types 'melanoma' and 'lipoma/ liposarcoma'. The causes of these spatio-temporal clusters can be of differing nature and must be verified with further studies.

# Part IV – DISCUSSION

In this part of the thesis, the methodology and the results are discussed. First, the strengths and limitations of the methods are highlighted, before the results are interpreted with regard to the quality of data and the research questions of the present thesis are answered. Finally, a conclusion is given by reflecting on the achievements of this thesis and by providing an outlook for future works.

## 15 Discussion

The results of this study should be considered with caution, as GIS and spatial analysis methods can play a major role in shaping an answer to spatial epidemiological questions. In this regard, studies that work with different data and methods are reported as prone to errors and uncertainties, which can rarely be prevented. However, these uncertainties should be discussed in order to provide information of spatial epidemiological study results. Therefore, interpretations und assumptions based on the results of the present study are accompanied by considerations on the degree of uncertainty (Lam, 2012). Unfortunately, uncertainty is difficult to quantify. According to Lam (2012), this issue can be of different origin and occur at various steps of the analysis (Lam, 2012).
With this in mind, the employed methodology and the results, including the original data source, are subsequently discussed.

### 15.1 Uncertainties in Methodology

Given that the use of different methods or the same methods with different parameter settings can lead to varying results, the applied geospatial methods to analyse the cancer data can add uncertainty to the final results (Cressie, 1993). In addition, spatial models might not be applicable everywhere since topologies vary across the world (Lam, 2012). Subsequently, the weakness and sources of uncertainty for the applied methods is discussed.

### *15.1.1 Disease Mapping*

For all cancer types, the same standardisation procedure was applied. However, this strategy can lead to uncertainties and loss of precision of the SMR since the three variables age, sex and size are not necessarily confounding variables for all cancer types. For instance, sex is not expected to be a confounding factor for melanoma (Higginson, Muir & Munoz, 1992), and in this case the weighting scheme could be simplified.
In addition, according to Porta et al. (2014), the interpretation of SMR values is problematic because such indirect standardisation is based on different weighting schemes across each study population – i.e., municipal unit. For this reason, comparing SMRs across different municipal units could result in uncertain outcomes as confounded by the weighting schemes relative to each population (Australian Institute of Health and Welfare, 2011; Porta et al., 2014). Even for the same population and the same municipal unit, the age, sex and size distribution may change over time. In addition, the use of the indirect standardisation method for time series analysis is problematic because of possible temporal inconsistencies of cancer rates (Australian Institute of Health and Welfare, 2011).

In spite of this issue, the indirect method is still reported as an appropriate method to assess cancer risk (Rezaeian et al., 2007). This because the indirect standardisation allows a comparison of observed and expected age-sex-size-specific cancer events, whereby the direct standardisation may provide less stable estimates because the standard error of the rates depends on the variations in the age-sex-size-specific number of cases rather than the total number of cases (Australian Institute of Health and Welfare, 2011; Rezaeian et al., 2007).

Furthermore, calculating the SMR for small-areal units with population of 5,000 or less (such as most Swiss municipalities) is reported to be problematic, as it is expected to introduce extreme fluctuations in the SMR that may be explained by sample variability (Elliott, Martuzzi & Shaddick, 1995). For this reason, small-scale geographic studies are often characterised by uncertainty resulting from methodology that can lead to the misinterpretation of results. An example is that small and sparsely populated areas are often overestimated in statistical test procedures (Elliott & Wartenberg, 2004). Further,

although cancer cases are standardised, sparsely populated areas with few (or zero) cases can generate extreme SMR values. This because the variance of the SMR is inversely related to the expected cancer cases and thus small populations have large variability in the estimated rates. Given that these sparsely populated municipalities often have a larger area than densely populated municipalities, they tend to dominate the map visually despite producing the most uncertain risk estimates (Elliott et al., 1995; Elliott & Wartenberg, 2004). This involves that artificially inflated SMR will be found where de facto no precise risk can be estimated. On the other hand, areas with significantly higher risk might not be visually detected.

For mapping SMR, value-by-alpha maps have essentially the same purpose of a classic choropleth map, that is visualising variations of the measures over the study area, but with the possibility of embedding additional information through changes of opacity (Roth et al., 2010). This has the advantage over other bivariate mapping methods, such as cartograms, of comparing of holding shape, size and arrangement of units constant across maps is easy. For this reason, value-by-alpha maps are an appropriate technique for understanding complex geographic patterns. However, this technique involves several limitations and sources of uncertainty. Firstly, value-by-alpha maps (as well as choropleth maps) must classify the equalising variable before mapping (Roth et al., 2010).

Therefore, the results can be different depending on the number of classes and classification method (Lawson & Williams, 2001). As an example, the present study uses four classes for the SMR, defined according to chapter 13.2. Using four classes, defined according to another classification method, would result in a different understanding of the cancer risk distribution (Lawson & Williams, 2001).
Secondly, in value-by-alpha maps municipalities are assigned to a classes of cancer risk (e.g. high cancer risk with low temporal variability), however an exact SMR value cannot directly be determined from the map. This complicates the identification of temporal trends within the study period; for instance, if SMRs are very high for a period of three years and low to very low for the remaining three years.

Thirdly, a further limitation that is specific of alpha-by-value maps is readability. According to Roth et al. (2010), alpha-by-value maps are dominated by enumeration units with large values in the equalizing variable since they are emphasized indirectly and becoming more noticeable due to value changes to surrounding units. As a result, small but thematically important enumeration units may still be difficult to read. Finally, the readability of alpha-by-value maps is also limited by the graphical resolution of the support that may dramatically influence the perception of colours (Roth et al., 2010).

### *15.1.2 Cluster Analysis*

As shown through the analytical effort, both cluster methods have their strengths and weaknesses. The two methods have in common that the selected number of permutations in the cluster statistic can result in major uncertainties. Although permutations tests reduce the effect of biased data and allow for the localisation of significant clusters, these tests can involve uncertainties. Since the statistic is computed each time with a different set of random reference distribution, the clusters may differ for each permutation and are thus not exactly replicable (Bivand et al., 1993). This is also the case when conducting a homogeneity test.

In addition, the use of a confidence level of 95 percent is also problematic since all correlations of SMR values below this level are completely excluded (Baker, 2016; Lam, 2012). As indicated in chapter 10.3.3, the results of a cluster analysis strongly depend on the aggregation level. In the cluster analyses performed in this study, cancer data are aggregated at the municipal level. Analyses with higher spatial resolution e.g., address-point data for the individual subjects, could lead to different results (Lawson, 2006b; Lloyd, 2010). In addition, the conducted cluster analyses only consider data in a two-dimensional space (Anselin et al., 2006; Kulldorff, 2015). This could result in further sources of uncertainties, especially in mountainous regions, which are characterised by great differences in altitude between municipalities.

#### 15.1.2.1 SaTScan™

As previously mentioned, in small and sparsely populated areas, SMR values are generally overestimated and spatial accumulations of high SMR values occur rather randomly. Cluster analysis methods help to minimise these uncertainties. However, not all methods deal equally well with this problem. This is also a weakness of SaTScan™ (Kulldorff, 2015; Lawson & Williams, 2001). In our study, for the specific cancer types, the SMR values present few very high values. Most values are low or even zero.
Since the data set generated through random permutations is created from the existing SMR values, these results also consist predominantly of low values. As shown in the methodology SaTScan™ (chapter 14.2.1), the calculation of the LLR requires the mean value. Consequently, the LLR of a circle is strongly influenced by outliers (Kulldorff, 2015). For this reason, SaTScan™ detects statistically increased rates only for municipalities with extreme SMR values where actually no increased risks exist. Areas with significantly increased risk could, therefore, not be detected.

Because the SaTScan™ clustering method requires point geometry as spatial input data, the centroid point of the polygons had to be determined beforehand. However, the point location depends strongly on the applied method and the size and shape of the municipalities. This may influence the cluster analysis (Lawson, 2006a).
Usually, the using SaTScan™ for disease data enumerated within spatial units is problematic, since this clustering method considers point data as single cases (Kulldorff, 2015). However, the analysed cancer data comprises several cancer cases that are aggregated at the municipal level. This is particularly challenging for large and sparsely populated municipalities, where large distances may occur between the individual cases.

A great strength of SaTScan™ is the possibility to detect spatio-temporal clusters. However, spatio-temporal clusters can only be detected if the significantly increased cancer risks have no temporal interruption (Kulldorff, 2015; Kulldorff et al., 2005). As an example, if the cancer risk is significantly increased from 2008 to 2011 and in 2013, a spatio-temporal cluster is only detected from 2008 to 2011. This because the cancer risk in 2012 is not significantly increased. However, the low cancer risk in 2012 could result from poor

data quality. Furthermore, visualizing the resulting clusters may be difficult. By combining the alpha-by-value-mapping technique for mapping the different clusters with a cluster persistency map, the temporal persistence of a cluster centre across the period of study can be shown. Still, the exact period for which a cluster is observed cannot be determined from the maps.

#### 15.1.2.2 LISA

When using local statistic LISA to determine spatial clusters, the results are often reported to be affected by the definition of weight function, data standardisation and classification and extreme values (Zhang et al., 2008). However, as the results of this thesis show, the local Moran's I index is less affected by extreme values of small and sparsely populated municipalities in combination with a permutation test compared to scan statistics of SaTScan™. Due to this differences in the underlying methodology, conducting a cluster analysis with LISA also results in cluster centres in Swiss municipalities for the most common cancer types.

By allowing ad-hoc definitions of the neighbourhood structure, the aggregation of data at municipal level is less problematic. With this regard, various spatial weight matrices can be designed based on different methods and neighbourhood definitions. The problem with using distance measures to assess neighbourhood is obviously that the data is already aggregated at the municipal level and only areal data is available. For this reason, in the present study, measures of contiguity across municipal borders provide a better method to design spatial weight matrices. Depending on the definition of contiguity (rocks, bishops or queen's contiguity), the results of the local statistic may vary (Sawada, 2004). In addition, defining the contiguity order also affects the results of the analysis. Since the municipalities of Switzerland are of very different in size, shape and spatial arrangement, the choice of contiguity order is rather complex. As an example, when using a first order contiguity level, the municipal units of Zurich and Kloten are not considered as neighbouring, even though they are geographically very close. For such small-area units, the use of the second order contiguity could be seen as appropriate. However, for larger units this would result in neighbourhood relations for geographical distant municipalities, for instance Chur and Scuol.

One weakness of LISA is, that the method does not allow for spatio-temporal cluster statistics, by complicating the temporal analysis of cancer risk. However, a workaround consisted in using value-by-alpha maps and persistency maps to present temporal cluster persistency across the period of study. Therefore, the resulting spatio-temporal clusters strongly dependent on the alpha-by-value mapping technique and the used definition of spatio-temporal clusters. Further, as mentioned, the resulting maps only allow insight into the overall temporal duration and not into the specific period. Consequently, our maps cannot provide detailed information on spatio-temporal trends, since cluster centres might exist with temporal interruptions.

Obviously, the generated results can only be as good and reliable as the underlying data. Thus, the spatial model and its analytical methods are confined by the quality of data. Subsequently, the results and the quality of the data are discussed.

## 15.2 Results

The analysis of all cancer records showed that significant spatio-temporal clusters can be detected in the region of Zurich and partly in the region of Bern. As mentioned, it is very unlikely that risk factors for the different cancer types occur exclusively in these regions. In addition to the uncertainty of the methodology, the data quality may also affect the analysis. Common sources of uncertainty in the analysis of disease data are data availability, confidentiality, changes in disease definition, and the time and spatial scales used for reporting (Pickle et al., 2006). These issues are discussed in the context of the present study.

### *15.2.1 Data and Interpretation*

According to Shi (2009) the main uncertainties in disease data consist in the address location and attribute uncertainty, sampling variability, and further spatial data quality issues (Shi, 2009). In addition, specific problems that are commonly associated with health data are underreporting, coding errors, diagnostic errors and the relevance for research (Twigg, 1990).

In Switzerland, canine cancer registration is not mandatory. Moreover, the registration of the canine population is prescribed by law only since 2016 (Swiss Confederation, 2015). Further, the data recorded in cancer registries is not originally meant for spatial analysis purpose (Jerrett et al., 2003) and its access is restricted because of confidentiality policies (Lawson, 2006b). Consequently, the exact place of residence of dogs diagnosed with cancer is not available, and thus the spatial scale is limited.

In addition to the data access, the frequency of visiting to the veterinary may also influence the probability of cancer detection. Studies have shown that a relationship exists between the frequency of visiting to the veterinary and the age of a dog (Verma, 2015). This could be an explanation why the age distribution of the canine cancer records deviate from existing findings reporting higher prevalence of cancer in older dogs.

Uncertainties in the canine cancer data also arise from incomplete and inconsistent management of the records. For example, in many records information about sex, age or breed of dogs is missing and had to be classified as unknown. Unrealistic values in the data sets also had to be classified as unknown, for instance, where the age would be reported as 100,000,000. Furthermore, some records contained spatial information that would not be relevant for our study, for instance about dogs that are based abroad but checked for cancers in Switzerland. In addition, data on the size of dogs was not reported in the database and had to be determined based on the ancillary information retrieved from the literature. Since not all entries contain explicit breed names and/or no consistent specification on the size was found, many entries had to be classified as unknown (see chapter 12.3). Despite data pre-processing, not all uncertainties could be eliminated from the data.

In addition data accuracy could be liked to different measuring and recording techniques (Lam, 2012). In the diagnosis of tumours, this issue could be linked to the fact that data is based on different diagnostic techniques carried out by different veterinaries. Further, the common cancer types might be diagnosed at locations where canine tumours are easier to detect (Graf et al., 2016; Grüntzig et al., 2015). Although the classification of malignant

tumour is based on the ICD-O-3 standards, problems also arise in the classification of the cancer diagnoses: first, different cancer types that are caused by the same agents are classified and analysed separately; second, a problem in classification exists if a dog or human is diagnosed with several diseases or cancer types at the same time; and third, criteria for cancer diagnoses may also have changed over time (Grüntzig et al., 2015; Meade, 2014).

Cancers can be considered as chronical diseases because they develop over a long period of time. Cancers' aetiologies and risk factors are not always fully known and understood and they can have impact over different time scales. However, cancer data such as cancer prevalence, incidence or mortality only provide information of a specific moment in time, normally the year of diagnosis. The results of cluster analyses are therefore difficult to interpret, since linking cancer risk and potential risk factors highly depends on the accuracy of exposure assessment and the latency time from the initial exposure. For this reason, changes in the population as well as exposure to risk factors over time due to migratory patterns complicate the analysis, especially for diseases with longer latency time (Jarup, 2004). In the present case, cancer diagnoses are only recorded for dogs living in Switzerland at the time of the diagnosis, which means also for dogs that have been previously living abroad.

Several uncertainties exist that address the quality and scale of georeferenced data (Elliott & Wartenberg, 2004). The cancer and demographic dataset employed in this study are only available in a spatially aggregated form, at the municipal level. As mentioned, sparsely populated municipalities can artificially result in high SMR values. Further, these units are generally also more sensitive to statistical fluctuations (Elliott et al., 1995; Elliott & Wartenberg, 2004). For this reason, areas and study periods in geospatial analysis should be sufficiently large to achieve stable rate estimates. Considering larger areas or long periods in geospatial analysis, however, leads to the problem that geographical variations or temporal trends cannot be detected (Elliott & Wartenberg, 2004). In this regard, Monte Carlo simulation can help to deal with scale effects through sensitivity analyses (Schneider et al., 1993).

Further, the question arises whether partitioning space artificial administrative units, such as municipal units, is appropriate for reporting the population and disease. The spatial structure has an impact on the detection of cancer clusters, e.g. by obscuring existing spatial patterns. Consequently, the MAUP and scale effects are not only problems of ecological analysis but rather of all spatial analysis including disease mapping and cluster analysis (Jarup, 2004; Lam, 2012). Furthermore, administrative boundaries can change with time which poses additional uncertainties for the study (Lam, 2012). These changes must be considered in disease and demographic data registries. However, it is unclear whether the actual FSO-No. were used when recording cancer cases and population and whether the registries are kept up to date. Furthermore, there is some uncertainty surrounding the recording practice of the spatial information.

For previous reasons, although earlier studies have shown that dogs can be used as sentinels in comparative cancer research, the present findings on canine cancer risk should be carefully considered before being compared to humans.

# 16 Answering the Research Questions

In this part, the research questions are answered. First, research question I is answered regarding the spatial distribution of the occurrence of all cancer records from 2008 until 2013 in Switzerland.

RQ I (a) *"Do significant **spatial** patterns exist in the geographical distribution of increased canine cancer risk or is the distribution random?"*

The test of homogeneity examined the presence of heterogeneity and proved that spatial clusters exist in the occurrence of cancer record with a level of significance of five percent. Both cluster detection approaches confirmed this existence of clusters in cancer risk for the cancer records in the Canine Cancer Registry. The cluster analysis resulted in a large cluster of increased cancer risk in area of Zurich to Aargau and Schaffhausen for the period of interest. However, the results of the two methodological approaches cannot be treated as equal due to the underlying methodology. With the scan method, spatial clusters in the region south of Lake Neuchâtel and close to the national border to Italy were detected. Further spatial cluster centres were detected wit LISA mainly around Glarus North and South, Bern, Biel, around the Lake Thun and some more areas. Thus, the distribution of cancer risk is not random.

RQ I (b) *"How do these spatial patterns vary from 2008 to 2013, do significant **spatio-temporal** patterns exist in the geographical distribution of increased canine cancer risk?*

Temporal patterns were assessed by comparing the temporal variability of the SMR for the entire period with the results of the two clustering approaches. Both approaches showed that temporal variation exist in the detection of significant clusters and many cluster centres appear for only one year. The large cluster in the area of Zurich appear three up to six years depending on the underlying method. However, many of the

municipalities located in the clusters provide a relatively high temporal variability of the SMR. Few cluster centres could be detected that appear up to four years and have a high cancer risk with medium temporal variability for the entire time period. With this knowledge about the general distribution of cancer risk in Switzerland, the spatio-temporal distribution of the most common canine cancer types can be assessed by answering RQ II:

RQ II *"How do these patterns of canine cancer risk differ between the most common cancer types?"*

Considering the specific cancer types, the test of homogeneity proved the presence of heterogeneity for all the ten cancer types but the cancer risks are not heterogenous in all years. The test already provided information that no spatio-temporal clusters could be expected for the malignant tumour types 'non-Hodgkin lymphoma' and 'hemangioma/ hemangiosarcoma'. Spatial clusters could be identified for all specific cancer types. The analysis has shown that even though the identified municipalities differ from each other, many cluster centres are located in similar regions. However, the detected clusters mostly occur for only one year or – considering the entire period – provide a low cancer risk and/or high temporal variability. However, spatial-temporal cluster centres could be localised for the malignant tumour types 'melanoma' in the municipalities of Bern, Regensdorf, Bülach and Dübendorf and for 'lipoma/ liposarcoma' in the Zurich, Kloten und Maur. These areas may be of interest for further epidemiological studies.

# 17 Conclusion and Outlook

This chapter reflects on the achievements of this study and offers an outlook for future work.

## 17.1 Reflection on Achievements

This master thesis has analysed the spatial distribution of cancer from 2008 to 2013 in Swiss municipalities and thus contributed to a better understanding of the cancer occurrence. Despite the existence of many uncertainties regarding the methodology and results, important findings can be drawn from the spatio-temporal analyses. Thus, the analysis provides a set of spatio-temporal clusters for the two cancer types 'melanoma' and 'lipoma/ liposarcoma' which can be used in further studies on human cancer research and ecological studies.

By using Canine Cancer Registry data, the work helps to establish the use of companion animal in epidemiological studies as well as to show the importance of disease registries. The application of spatio-temporal GIScience approaches identified strengths and weaknesses of the underlying geostatistical methods with respect to the topology of Switzerland. This can contribute to the development of new analysing methods and approaches and can help in future studies in the field of GIVA. Furthermore, connecting the methods of the two research fields spatial epidemiology and GIScience, the analysis could identify regions with increased cancer risk, where cancer prevention and treatment strategies in Switzerland should focus on. This can be seen as an example how the collaboration between different research areas can advocate new health care strategies and improve the Swiss Health System. As a result, the connection of methods from spatial epidemiology with approaches and tools of GIScience can contribute to the developments in public health. This demonstrates the potential of GIScience to product and exchange knowledge between other disciplines than geography. Consequently, all aims of this master thesis were achieved.

## 17.2 Outlook for Future Work

This master study has identified spatio-temporal cluster in Swiss municipalities for the two cancer types 'melanoma' and 'lipoma/ liposarcoma'. However, these results are affected by several uncertainties. In order to validate the findings, the distribution of cancer in Switzerland should be addressed in future studies with multi-level analysis (Lam, 2012; Richardson et al., 2013). For this purpose, spatial data mining should be further promoted (Richardson et al., 2013); for example, through the obligation to register cancer diagnoses. A first step in this direction has already been taken with the new law (in force since January 2016) that obligates the registration of dogs in the national canine database (Swiss Confederation, 2015).

In future studies, the results could further be examined in cohort studies, ecological analysis and agent-based modelling. For the detected clusters of 'melanoma' such ecological analysis could address the exposure to solar radiation. For 'lipoma/ liposarcoma', suggestions exist that the tumours are not significantly affected by general environmental factors (Higginson, Muir & Munoz, 1992). This could be investigated – for example – by agent-based modelling.

Due to the topology of Switzerland and the differences in altitude, a development of three-dimensional spatio-temporal cluster analysis should be strived. Future studies on human cancer should examine these findings and prove the use of companion animals as sentinels.

Additionally, institutional and educational models should be improved in health research and practice for future spatial analysis on health issues. With an improved density, accuracy, and specificity of geospatial health data, spatial and spatio-temporal analysis of complex spatial health processes would be facilitated and also allow analysis at the level of the individual (Richardson et al., 2013).

Thanks to new technologies, increasingly detailed health-related GIS data can be generated; for instance, interactive real-time Global Positioning System (GPS) and GIS functionalities are increasingly embedded in cell phones and other mobile devices. This allows growing participatory geodata generation by citizens and will enable illustrating life paths in the future, consequently providing a better assessment of the ecological or social risk factors (Richardson et al., 2013; Zenk et al., 2011).

The growing amount and increasing quality of health-related data will allow further spatial and temporal analyses of health issues and related risk factors. To facilitate the generation, management and use of the growing geospatial disease data, underlying technical and institutional challenges such as scientific access, international standards, common terminology, interoperability and data confidentiality must be globally fostered. Since the distribution of diseases do not stop at national borders, international cooperation on health projects using GIS is also of interest for Switzerland (Richardson et al., 2013).

# Bibliography

AKC. (2017). American Kennel Club. Retrieved March 27, 2017, from http://akc.org/

Alderton, D., Morgan, T., & Schmitz, S. (2001). Hunde-Rassen: der kompetente Führer mit über 1000 Farbfotos und Bestimmungsübersicht. Blv.

American Society of Clinical Oncology. (2015). Understanding Cancer Risk. Retrieved April 1, 2017, from http://www.cancer.net/navigating-cancer-care/prevention-and-healthy-living/understanding-cancer-risk

Anselin, L. (1995). Local indicators of spatial association — LISA. Geographical Analysis, 27(2), 93–115. Retrieved from www.drs.wisc.edu/people/faculty/curtis/documents/RS977/Anselin1995.pdf%5Cn www.spatialanalysisonline.com/output/html/LocalindicatorsofspatialassociationLISA.html

Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis, 38(1), 5–22.

Anthamatten, P., & Hazen, H. (2011). An introduction to the geography of health. (H. Hazen, Ed.). London: London : Routledge.

Australian Institute of Health and Welfare. (2011). Principles on the Use of Direct Age-standardisation in Administrative Data Collections: For Measuring the Gap Between Indigenous and Non-Indigenous Australians. Australian Institute of Health and Welfare.

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2011). Backhaus-Erichson-Plinke-Weiber. Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. Berlin. Springer.

Bailey, T. C., Fotheringham, S., & Rogerson, P. (1994). A review of statistical spatial analysis in geographical information systems. Spatial Analysis and GIS, 13–44.

Baker, M. (2016). Statisticians issue warning on P values. Statement Aims to Halt Missteps in the Quest for Certainty, Nature, 531, 151.

Beale, L., Abellan, J. J., Hodgson, S., & Jarup, L. (2008). Methodologic Issues and Approaches to Spatial Epidemiology. Environmental Health Perspectives, 116(8), 1105–1110.

Berke, O. (2003). Statistische Methoden der räumlichen Epidemiologie.

Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V. (1993). Applied Spatial Data Analysis with R. Canadian Journal of Psychiatry (Vol. 38).

Boo, G. (2014). Companion Animals as Public Health Sentinels. Space-Time Analysis of Tumor Incidence in Switzerland. In PhD Concept Presentation.

Boo, G. (2016). Spatial Distribution of Canine Cancer in Switzerland. A case study of skin tumors for the period 2008–2013.

Bortins, I., & Demers, S. (2002). Cartogram Types. Retrieved April 1, 2017, from http://www.ncgia.ucsb.edu/projects/Cartogram_Central/types.html

Brønden, L. B., Flagstad, A., & Kristensen, A. T. (2007). Veterinary cancer registries in companion animal cancer: a review. Veterinary and Comparative Oncology, 5(3), 133–144.

Bukowski, J. A., & Wartenberg, D. (1997). An alternative approach for investigating the carcinogenicity of indoor air pollution: Pets as sentinels of environmental cancer risk. Environmental Health Perspectives, 105(12), 1312–1319.

Burrell, G. A., & Seibert, F. M. (1916). Gases found in coal mines (Vol. 14). USGPO.

CDC. (1990). Guidelines for Investigating Clusters of Health Events. Retrieved March 9, 2017, from https://www.cdc.gov/mmwr/preview/mmwrhtml/00001797.htm

CDC. (2016). GIS and Public Health at CDC. Retrieved March 23, 2017, from https://www.cdc.gov/gis/

Chammartin, F., Probst-Hensch, N., Utzinger, J., & Vounatsou, P. (2016). Mortality atlas of the main causes of death in Switzerland, 2008-2012. Swiss Medical Weekly, 146(February), 13.

Clarke, K. C. (2003). Getting Started with Geographic Information Systems (4th ed.). Pearson Education.

Collegium Helveticum. (n.d.). Collegium Helveticum. Retrieved February 27, 2017, from http://www.collegium.ethz.ch/de/home/

Cressie, N. (1993). Statistics for Spatial Data (2 edition). Hoboken, NJ: Wiley-Interscience.

Cromley, E. K., & McLafferty, S. L. (2011). GIS and public health. Guilford Press.

dos Santos Silva, I. (1999). Cancer Epidemiology: Principles and Methods. International Agency for Research on Cancer. Lyon.

Elliott, P., Best, N., Armitage, P., & Colton, T. (1998). Geographical patterns of disease. International Encyclopaedia of Biostatistics. Chichester: Wiley, 1(998), 1.

Elliott, P., Cuzick, J., English, D., & Stern, R. (1992). Geographical and Environmental Epidemiology: Methods for Small Area Studies. Oxford: Oxford University Press.

Elliott, P., Martuzzi, M., & Shaddick, G. (1995). Spatial statistical methods in environmental epidemiology: a critique. Statistical Methods in Medical Research, 4(2), 137–159.

Elliott, P., Wakefield, J. C., Best, N. G., & Briggs, D. J. (2000). Spatial epidemiology: methods and applications. Oxford University Press.

Elliott, P., & Wartenberg, D. (2004). Spatial epidemiology: Current approaches and future challenges. Environmental Health Perspectives (Vol. 112).

FCI. (n.d.). Federation Cynologique Internationale for Dogs worldwide - Breeds. Retrieved March 27, 2017, from http://www.fci.be/en/Nomenclature/

FDHA. (2014). Bessere Daten helfen Krebserkrankungen besser zu verstehen. Medienmitteilung Bundesrat, 2.

Federal Assembly of the Swiss Confederation. (n.d.). Bundesgesetz über die Registrierung von Krebserkrankungen. Retrieved from https://www.admin.ch/ch/d/gg/pc/documents/2151/121106_KRG_Entwurf_de.pdf

Federal Council of Switzerland. (2014). Botschaft zum Krebsregistrierungsgesetz.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., … Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International Journal of Cancer, 136(5), E359-86.

Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., … Bray, F. (2013). GLOBOCAN 2012 v1. 0, Cancer Incidence and Mortality Worldwide 2013. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase.

FOPH. (2017). Health2020. Retrieved February 15, 2017, from https://www.bag.admin.ch/bag/en/home/themen/strategien-politik/gesundheit-2020.html

FSO. (2016a). Generalisierte Gemeindegrenzen: Geodaten. Retrieved October 10, 2016, from https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/geostat/geodaten-bundesstatistik/administrative-grenzen/generalisierte-gemeindegrenzen.assetdetail.453578.html

FSO. (2016b). Krebs in der Schweiz 2015. Statistik Der Schweiz, (Gesundheit), 1177–1500. Retrieved from https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/gesundheitszustand/krankheiten/krebs.html

FSO. (2016c). Sterblichkeit, Todesursachen - Daten, Indikatoren.

FSO. (2017). Amtliches Gemeindeverzeichnis der Schweiz. Retrieved March 27, 2017, from https://www.bfs.admin.ch/bfs/de/home/grundlagen/agvch.html#mutationsmeldungen__content_bfs_de_home_grundlagen_agvch_jcr_content_par_tabs

Fu, W. J., Jiang, P. K., Zhou, G. M., & Zhao, K. L. (2014). Using Moran's I and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China. Biogeosciences (Vol. 11). Copernicus GmbH.

Furuti, C. (2016). Map Projections: Conformal Projections. Retrieved April 1, 2017, from http://www.progonos.com/furuti/MapProj/Dither/ProjConf/projConf.html

Gail, M. H., & Benichou, J. (2000). Encyclopedia of epidemiologic methods. John Wiley & Sons.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. Geographical Analysis, 24(3), 189–206.

Goovaerts, P., & Jacquez, G. M. (2004). Accounting for regional background and population size in the detection of spatial clusters and outliers using geostatistical filtering and spatial neutral models: the case of lung cancer in Long Island, New York. International Journal of Health Geographics, 3(1), 14.

Graf, R., Grüntzig, K., Boo, G., Hässig, M., Axhausen, K. W., Fabrikant, S. I., … Folkers, G. (2016). Swiss Feline Cancer Registry 1965–2008: the Influence of Sex, Breed and Age on Tumour Types and Tumour Locations. Journal of Comparative Pathology, 154(2), 195–210.

Graf, R., Grüntzig, K., Hässig, M., Axhausen, K. W., Fabrikant, S. I., Welle, M., … Otto, V. (2015). Swiss Feline Cancer Registry: A Retrospective Study of the Occurrence of Tumours in Cats in Switzerland from 1965 to 2008. Journal of Comparative Pathology, 153(4), 266–277.

Grüntzig, K., Graf, R., Hässig, M., Welle, M., Meier, D., Lott, G., … Boo, G. (2015). The Swiss Canine Cancer Registry: a retrospective study on the occurrence of tumours in dogs in Switzerland from 1955 to 2008. Journal of Comparative Pathology, 152(2), 161–171.

Heseltine, E., Day, N. E., & Breslow, N. E. (1987). Statistical Methods in Cancer Research Volume II - The Design and Analysis of Cohort Studies. IARC Scientific Publication No. 82. Lyon: International Agency for Research on Cancer.

Higginson, J., Muir, C. S., & Munoz, N. (1992). Human cancer: epidemiology and environmental causes. Cambridge University Press.

Hirschfield, A., Brown, P. J. B., & Bundred, P. (1995). The spatial analysis of community health services on Wirral using geographic information systems. Journal of the Operational Research Society, 147–159.

Hjalmars, U., Kulldorff, M., Gustafsson, G., & Nagarwalla, N. (1996). Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection. Statistics in medicine (Vol. 15). Wiley Online Library.

IACR. (2017). International Classification of Diseases for Oncology. Retrieved March 24, 2017, from http://codes.iarc.fr/

Jacquez, G. M. (2008). Spatial cluster analysis. The Handbook of Geographic Information Science, 395–416.

Jarup, L. (2004). Health and environment information systems for exposure and disease mapping, and risk assessment. Environmental Health Perspectives, 995–997.

Jerrett, M., Burnett, R., Goldberg, M., Sears, M., Krewski, D., Catalan, R., … Finkelstein, N. (2003). Spatial analysis for environmental health research: concepts, methods, and examples. Journal of Toxicology and Environmental Health Part A, 66(16–19), 1783–1810.

Jones, Whs. (1923). Hippocrates, Volume I. London: Loeb Classical Library, 298–301.

Keim, D. A., & Herrmann, A. (1998). The gridfit algorithm: An efficient and effective approach to visualizing large amounts of spatial data. In Visualization'98. Proceedings (pp. 181–188). IEEE.

Kelsey, J. L., Moore, A. S., & Glickman, L. T. (1998). Epidemiologic studies of risk factors for cancer in pet dogs. Epidemiologic reviews (Vol. 20). Baltimore, Maryland: Johns Hopkins University Press, 1979-.

Kinoshita, S., Wagatsuma, Y., & Okada, M. (2007). Geographical distribution for malignant neoplasm of the pancreas in relation to selected climatic factors in Japan. International Journal of Health Geographics, 6, 34.

Knox, E. G. (1989). Detection of clusters BT - Methodology of enquiries into disease clustering; London. In P. Elliott (Ed.), . London: Small Area Health Statistics Unit.

Kulldorff, M. (2015). SaTScanTM User Guide Version 9.4. Retrieved from http://www.satscan.org/techdoc.html

Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R., & Mostashari, F. (2005). A Space–Time Permutation Scan Statistic for Disease Outbreak Detection. PLOS Medicine, 2(3), e59.

Kulldorff, M., Huang, L., & Konty, K. (2009). A scan statistic for continuous data based on the normal probability model. International Journal of Health Geographics, 8(1), 58.

KUSA. (2016). KUSA - Kennel Union of Southern Africa. Retrieved March 27, 2017, from https://www.kusa.co.za/index.php

Lam, N. S. (2012). Geospatial Methods for Reducing Uncertainties in Environmental Health Risk Assessment: Challenges and Opportunities. Annals of the Association of American Geographers (Vol. 102). Taylor & Francis Ltd.

Last, J. M. (2007). A dictionary of public health.

Lawson, A. (2006a). Disease cluster detection: a critique and a Bayesian proposal. Statistics in Medicine, 25(5), 897–916. http://doi.org/10.1002/sim.2417

Lawson, A. (2006b). Statistical Methods in Spatial Epidemiology. Chichester [u.a.]: Wiley.

Lawson, A., & Williams, F. (2001). An introductory guide to disease mapping. John Wiley.

Lehari, G. (2004). Ulmers grosses Lexikon der Hunderassen : 345 Rassen in Wort und Bild. Stuttgart (Hohenheim) : Ulmer.

Levine, N. (2004). CrimeStat III: a spatial statistics program for the analysis of crime incident locations (version 3.0). Houston (TX): Ned Levine & Associates/Washington, DC: National Institute of Justice.

Liebisch, N., Goovaerts, P., & Kaufmann, A. (2002). New methods to generate neutral images for spatial pattern recognition. In International Conference on Geographic Information Science (pp. 181–195). Springer.

Lloyd, C. (2010). Spatial data analysis: an introduction for GIS users. Oxford university press.

Maguire, D. J. (1991). An overview and definition of GIS. Geographical Information Systems: Principles and Applications, 1, 9–20.

Marshall, R. J. (1991). A review of methods for the statistical analysis of spatial patterns of disease. Journal of the Royal Statistical Society. Series A (Statistics in Society), 154(3), 421–441.

Mayer, J. D., & Meade, M. S. (1994). A reformed medical geography reconsidered. The Professional Geographer, 46(1), 103–106.

McPherson, G. (2013). Applying and interpreting statistics: a comprehensive guide. Springer Science & Business Media.

Meade, M. S. (2014). Medical Geography. The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society. John Wiley & Sons, Ltd. Retrieved from http://dx.doi.org/10.1002/9781118410868.wbehibs204

Morgenstern, H. (1982). Uses of ecologic analysis in epidemiologic research. American Journal of Public Health, 72(12), 1336–1344.

Moseley, W. G., Lanegran, D. A., & Pandit, K. (2007). The introductory reader in human geography: contemporary debates and classic writings. Blackwell.

National Research Council. (1991). Animals as sentinels of environmental health hazards. National Academies Press.

National Research Council. (2012). Analysis of cancer Risks in populations near nuclear facilities: Phase I. National Academies Press.

NIH. (n.d.). Understanding Cancer - National Cancer Institute. Retrieved February 15, 2017, from https://www.cancer.gov/about-cancer/understanding

NKP. (2014). Nationale Strategie gegen Krebs 2014-2017. Bern: Dialog Nationale Gesundheitspolitik. Retrieved from http://www.oncosuisse.ch/file/10 Kommunikation/KLS_Nationale Strategie gegen Krebs_Bericht_d_130703_1.pdf

Nødtvedt, A., Berke, O., Bonnett, B. N., & Brønden, L. (2012). Current status of canine cancer registration–report from an international workshop. Veterinary and Comparative Oncology, 10(2), 95–101.

OECD, & WHO. (2011). OECD Reviews of Health Systems: Switzerland 2011.

Openshaw, S. (1984). The modifiable areal unit problem (Vol. 38). Norwich : Geo Books.

Openshaw, S., Charlton, M., & Craft, A. (1988). Searching for Leukaemia Clusters Using a Geographical Analysis Machine. Regional Science Association, 64, 95–106.

Pickle, L. W., Szczur, M., Lewis, D. R., & Stinchcomb, D. G. (2006). The crossroads of GIS and health information: a workshop on developing a research agenda to improve cancer control. International Journal of Health Geographics, 5(1), 51.

Pinho, S., Carvalho, S., Cabral, J., Reis, C. A., & Gärtner, F. (2012). Canine tumors: A spontaneous animal model of human carcinogenesis. Translational Research, 159(3), 165–172.

Porta, M., Greenland, S., Hernán, M., Dos Santos Silva, I., & Last, J. M. (2014). A dictionary of epidemiology. (M. Porta, Ed.) (6. Ed). Oxford university press.

Pospischil, A., Grüntzig, K., Graf, R., Boo, G., Folkers, G., Otto, V., & Fabrikant, S. I. (2015). One Medicine-One Oncology – Incidence and Geographical Distribution of Tumors in Dogs and Cats in Switzerland from 1955-2008. 3rd GRF One Health Summit 2015, October(One Health One Planet One Future), 108–111. Retrieved from www.grforum.org

Raisz, E. (1938). General cartography.

Reif, J. S. (2011). Animal sentinels for environmental and public health. Public Health Reports, 50–57.

Rezaeian, M., Dunn, G., Leger, S. S., & Appleby, L. (2007). Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. Journal of Epidemiology and Community Health, 61(2), 98–102.

Richardson, D. B., Volkow, N. D., Kwan, M., Kaplan, R. M., Goodchild, M. F., & Croyle, R. T. (2013). Spatial turn in health research. Science (New York, NY) (Vol. 339). NIH Public Access.

Richardson, S., & Monfort, C. (2000). Ecological correlation studies. Spatial Epidemiology: methods and applications. Oxford University Press Oxford, UK.

Roth, R. E., Woodruff, A. W., & Johnson, Z. F. (2010). Value-by-alpha maps: An alternative technique to the cartogram. The Cartographic Journal, 47(2), 130–140.

Sawada, M. (2004). Global spatial autocorrelation indices-Moran's I, Geary's C and the general cross-product statistic. Research paper from the Laboratory for Paleoclimatology and Climatology at the University of Ottawa.

Schneider, D., Greenberg, M. R., Donaldson, M. H., & Choi, D. (1993). Cancer clusters: The importance of monitoring multiple geographic scales. Social Science & Medicine, 37(6), 753–759.

Schwabe, C. W. (1984). Animal monitors of the environment. Veterinary Medicine and Human Health, 562–578.

Schweikart, J., & Kistemann, T. (2004). Geoinformationssysteme im Gesundheitswesen. Heidelberg: Wichmann.

Selvin, S. (1996). Statistical analysis of epidemiologic data. Oxford: Oxford University Press.

Shi, W. (2009). Principles of modeling uncertainties in spatial data and spatial analyses. CRC Press.

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2005). Thematic Cartography and Geographic Visualization. Upper Saddle River, NJ, USA: Pearson Prentice Hall.

Snow, J. (1855). On the mode of communication of cholera. John Churchill.

Stahl Jr, R. G. (1997). Can mammalian and non-mammalian "sentinel species"; data be used to evaluate the human health implications of environmental contaminants?

Studdert, V. P., Gay, C. C., & Blood, D. C. (2011). Saunders comprehensive veterinary dictionary (4th Editio). Elsevier Health Sciences.

Swiss Confederation. (2015). Registrierung und lückenlose Rückverfolgbarkeit zum Schutz von Hund und Mensch. Medienmitteilung 17. Dezember 2015.

Twigg, L. (1990). Health based geographical information systems: their potential examined in the light of existing data sources. Social Science & Medicine, 30(1), 143–155.

Upton, G., & Cook, I. (1996). Understanding statistics. Oxford University Press.

Van der Schalie, W. H., Gardner, H. S., Bantle, J. A., De Rosa, C. T., Finch, R. A., Reif, J. S., … Folmar, L. C. (1999). Animals as sentinels of human health hazards of environmental chemicals. Environmental Health Perspectives, 107(4), 309.

Verma, M. (2015). Cancer epigenetics: risk assessment, diagnosis, treatment, and prognosis. Humana Press.

Waller, L. A., & Jacquez, G. M. (1995). Disease models implicit in statistical tests of disease clustering. Epidemiology, 584–590.

Wartenberg, D. (2001). Investigating disease clusters: why, when and how? Journal of the Royal Statistical Society: Series A (Statistics in Society), 164(1), 13–22.

WCRF. (n.d.). Data for cancer frequency by country. Retrieved March 24, 2017, from http://www.wcrf.org/int/cancer-facts-figures/data-cancer-frequency-country

WHO. (1999). Future trends in veterinary public health. Teramo: WHO. Retrieved from http://www.who.int/zoonoses/vph/en/

WHO. (2004). A glossary of terms for community health care and services for older persons. WHO.

WHO. (2015). Public Health. Retrieved March 30, 2015, from http://www.who.int/trade/glossary/story076/en/

WHO. (2017). WHO - Cancer Fact Sheets. Retrieved February 15, 2017, from http://www.who.int/mediacentre/factsheets/fs297/en/

Wilkinson, P., Thakrar, B., Shaddick, G., Stevenson, S., Pattenden, S., Landon, M., … Elliott, P. (1997). Cancer incidence and mortality around the Pan Britannica Industries pesticide factory, Waltham Abbey. Occupational and Environmental Medicine (Vol. 54). BMJ Publishing Group Ltd.

Winslow, C. E. (1920). The untilled fields of public health. Science, 51(1306), 23–33.

Zenk, S. N., Schulz, A. J., Matthews, S. A., Odoms-Young, A., Wilbur, J., Wegrzyn, L., … Stokes, C. (2011). Activity Space Environment and Dietary and Physical Activity Behaviors: A Pilot Study. Health & Place, 17(5), 1150–1161.

Zhang, C., Luo, L., Xu, W., & Ledwith, V. (2008). Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. Science of the Total Environment, 398(1), 212–221.

# Appendix

**A. Disease maps: All cancer types II**

**B. Disease maps: Adenoma/ Adenocarcinoma III**

**C. Disease maps: Adenocarcinoma of Anal Glands IV**

**D. Disease maps: Dermatofibroma/ Dermatofibrosarcoma, NOS V**

**E. Disease maps: Hemangioma/ Hemangiosarcoma VI**

**F. Disease maps: Lipoma/ Liposarcoma VII**

**G. Disease maps: Mastocytoma/ Mast Cell Sarcoma VIII**

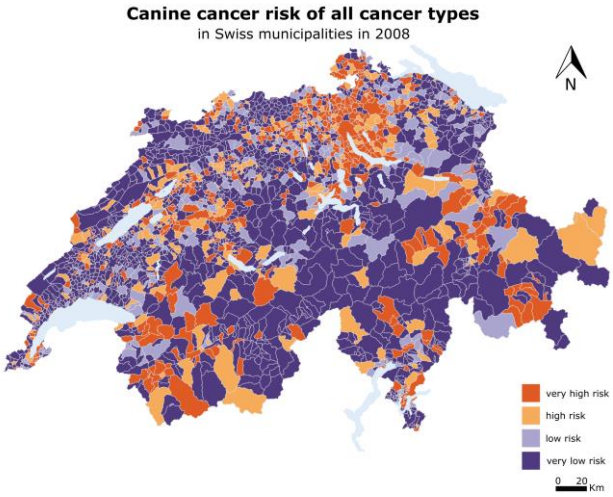
**H. Disease maps: Melanoma IX**

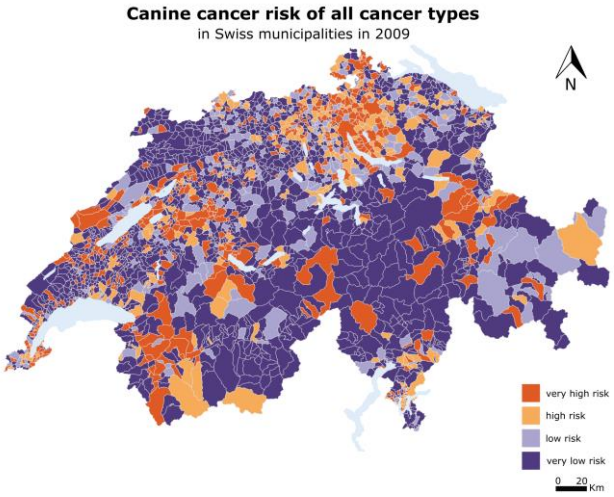
**I. Disease maps: Mixed tumours X**

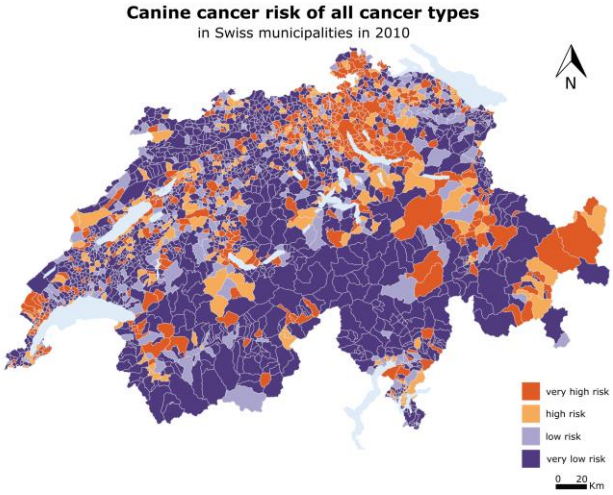
**J. Disease maps: Non-Hodgkin Lymphoma, NOS XI**

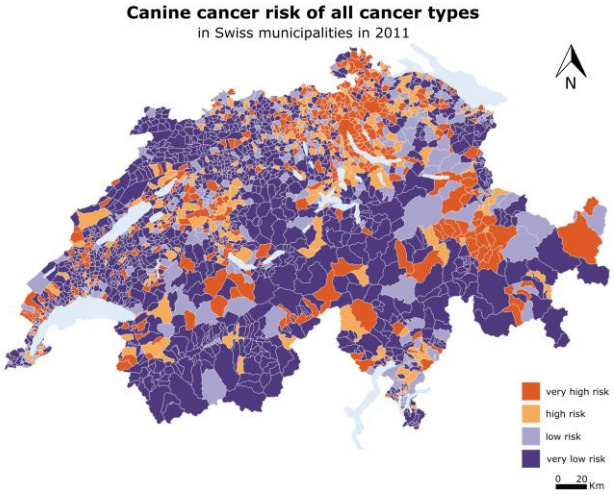
**K. Disease maps: Trichoepithelioma XII**

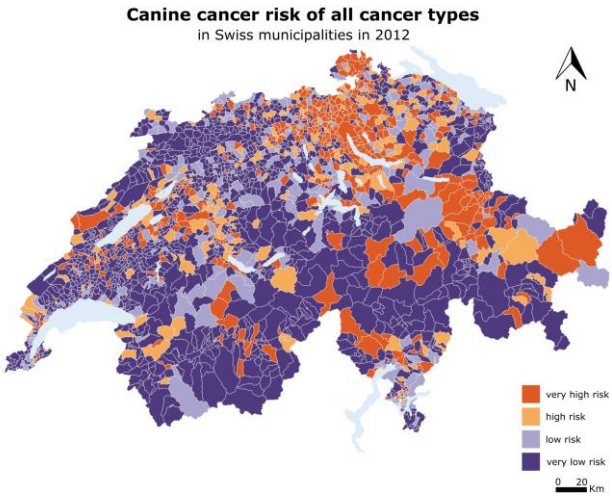
# A. Disease maps: All cancer types

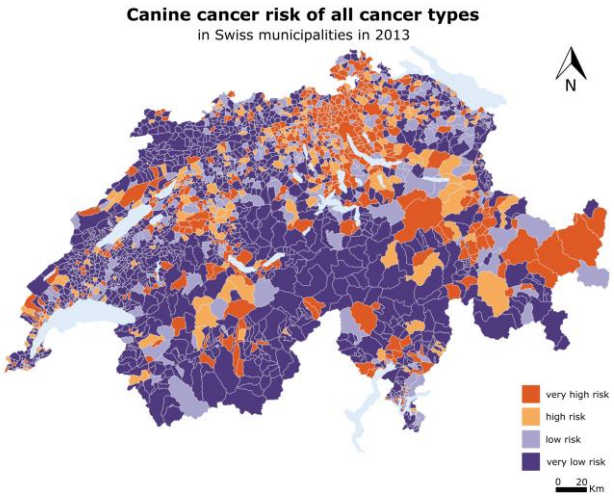


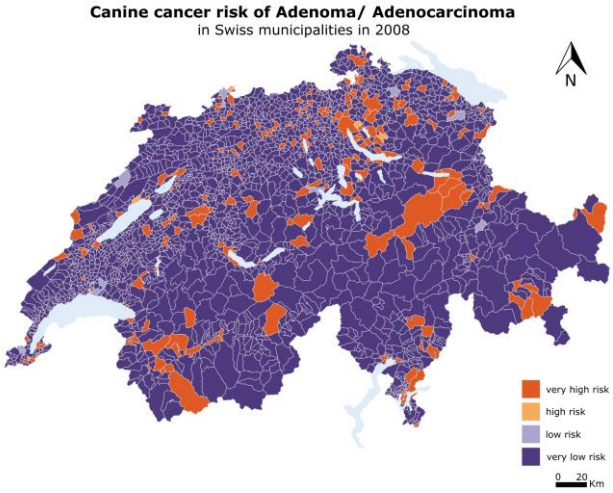



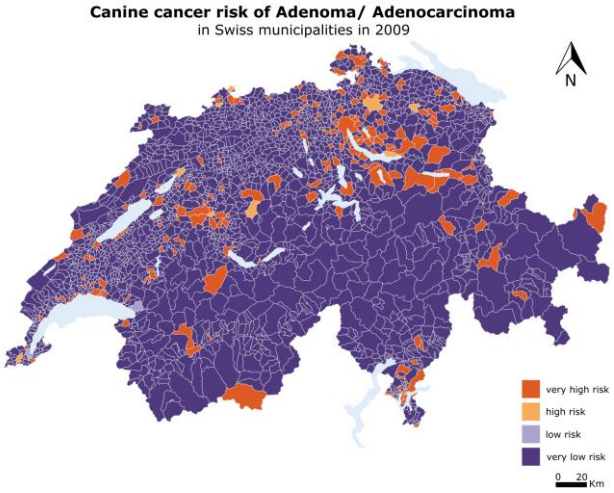



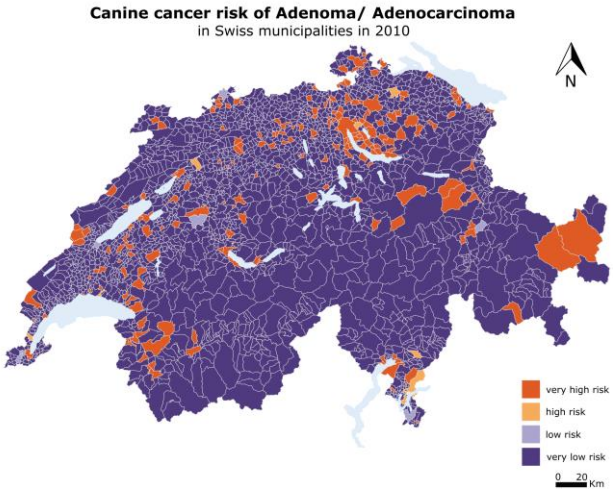



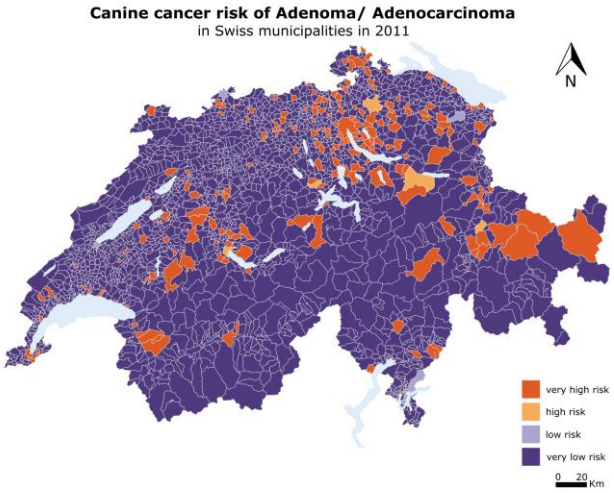



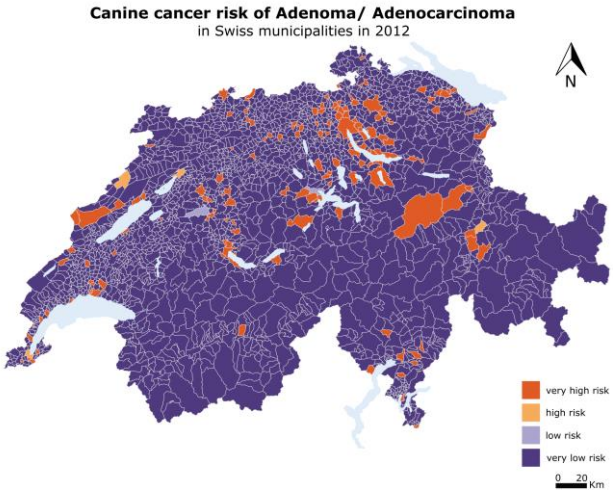
# B. Disease maps: Adenoma/ Adenocarcinoma

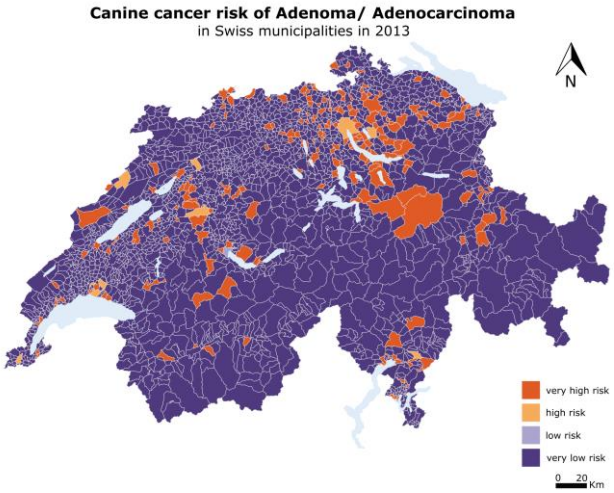


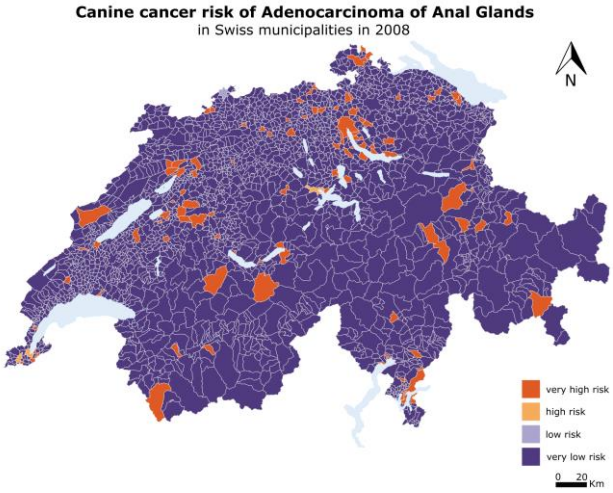



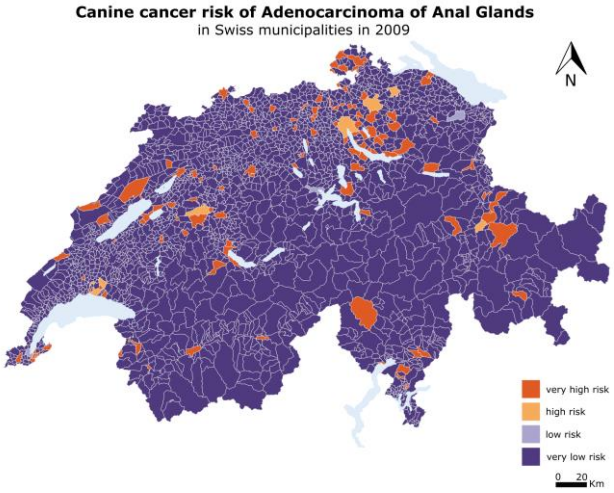



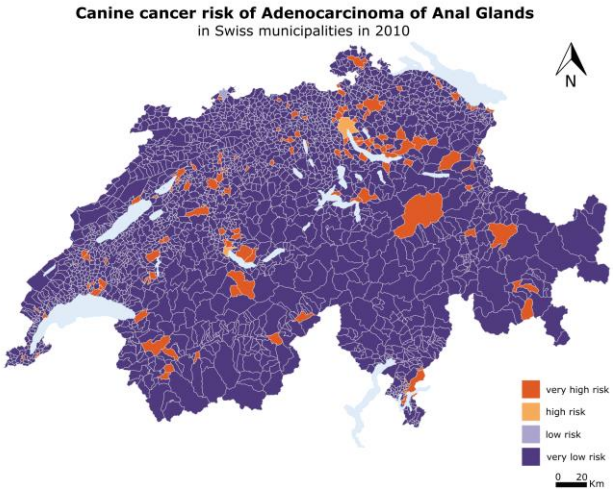



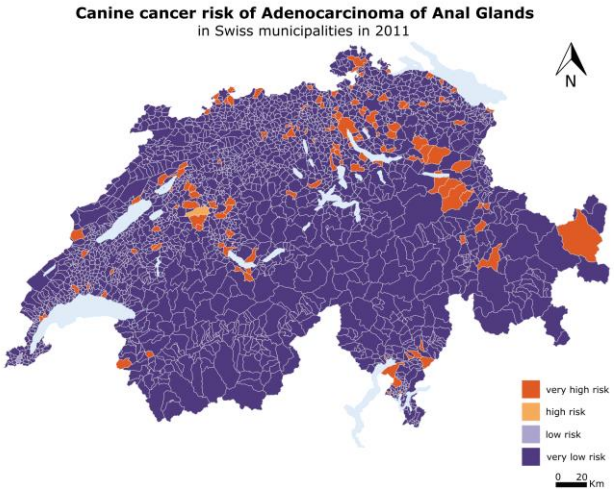



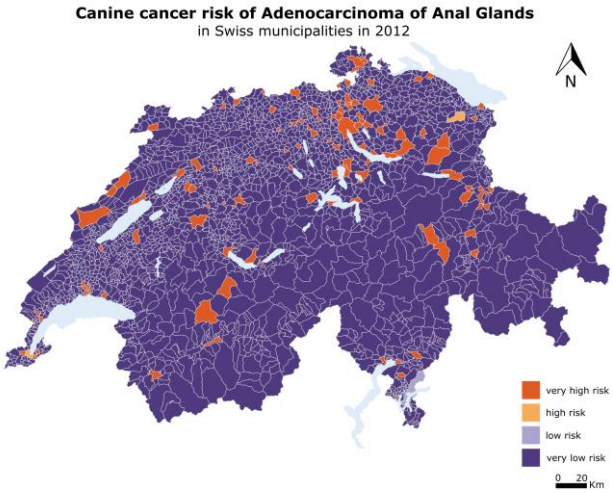
# C. Disease maps: Adenocarcinoma of Anal Glands

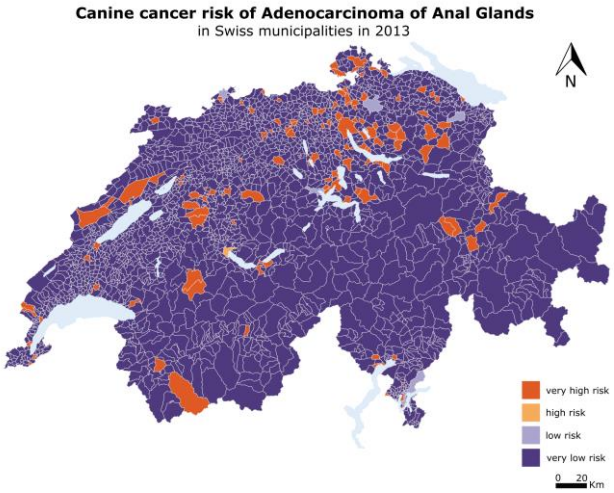


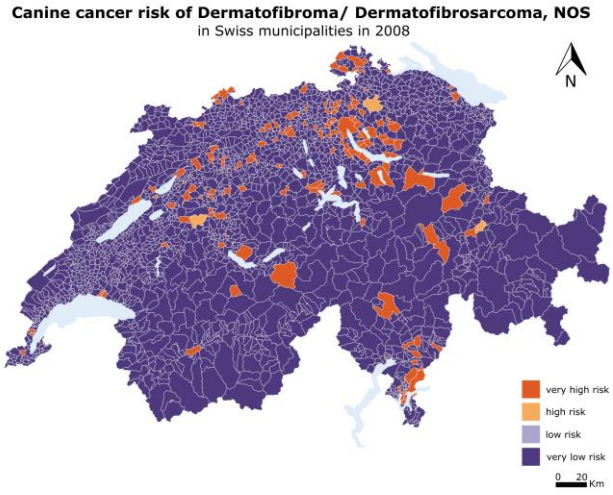



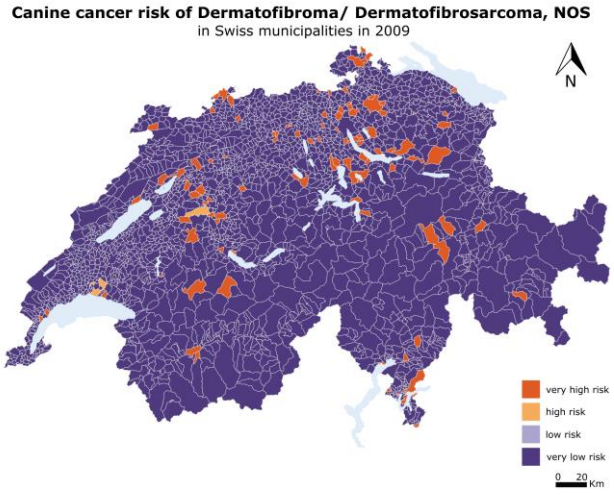



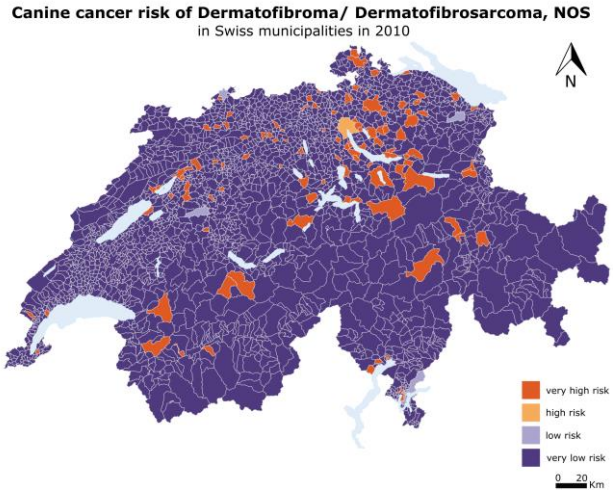



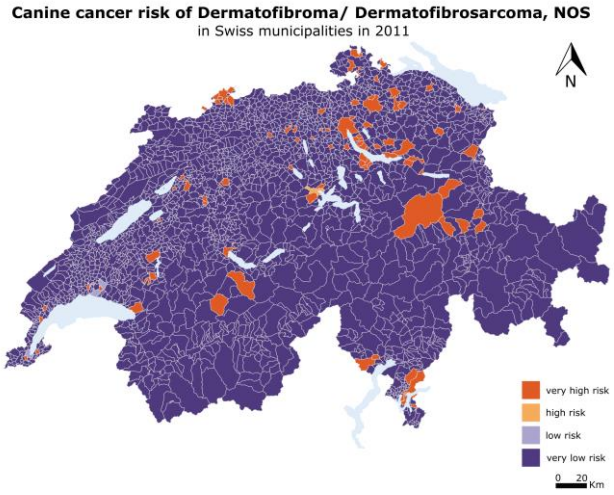



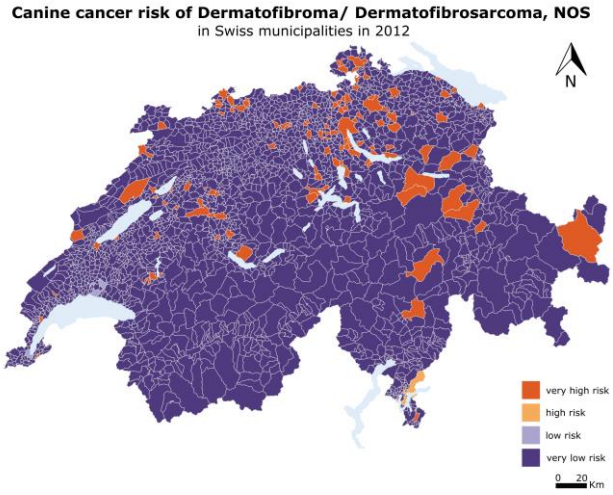
# D. Disease maps: Dermatofibroma/ Dermatofibrosarcoma, NOS

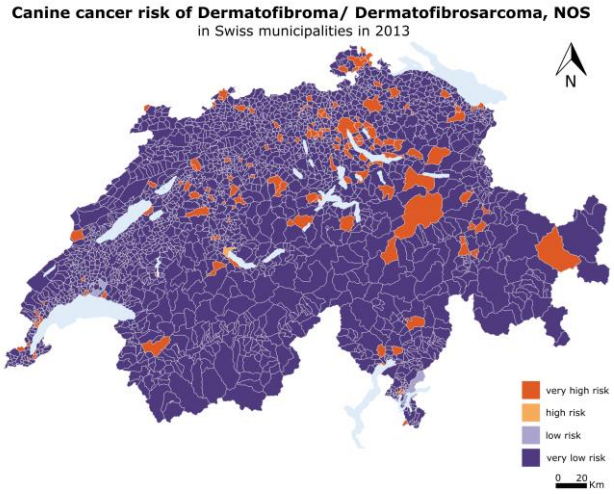


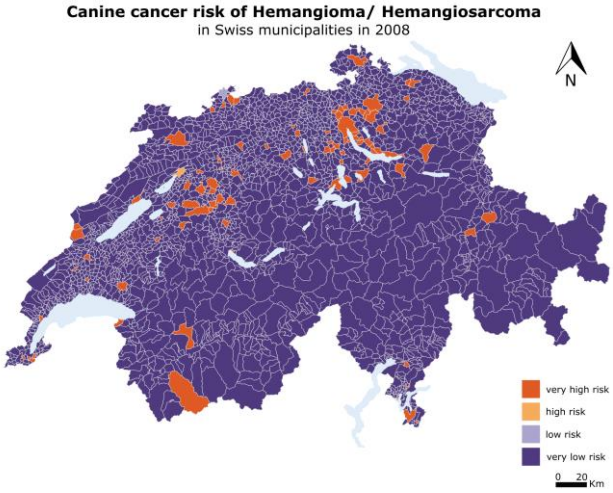



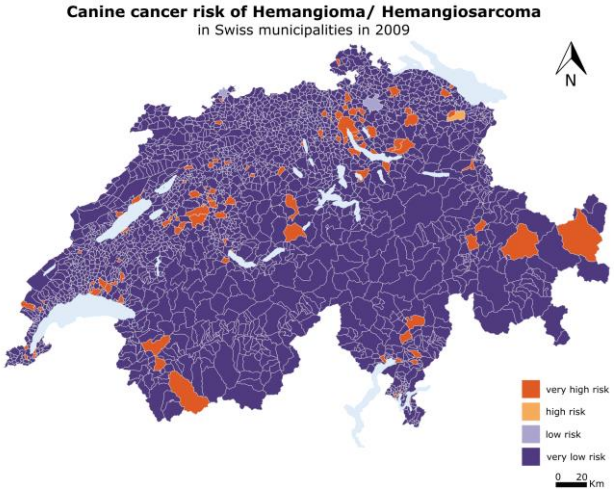



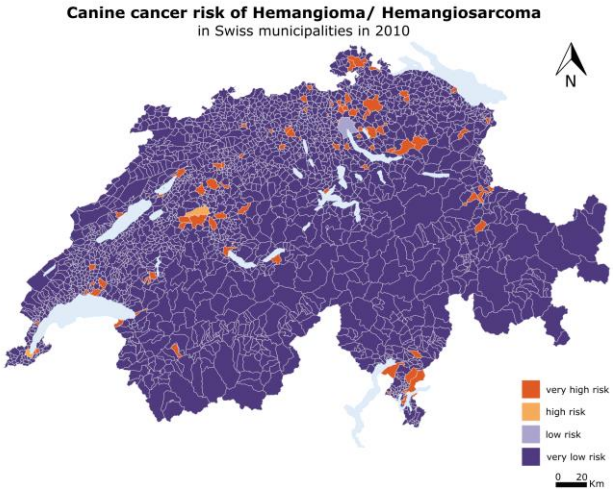



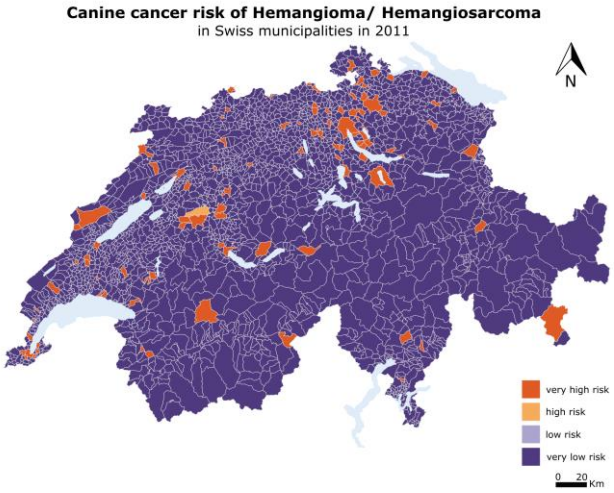



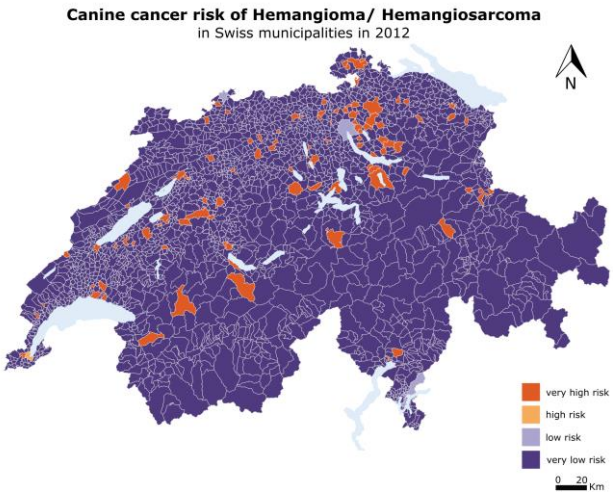
# E. Disease maps: Hemangioma/ Hemangiosarcoma

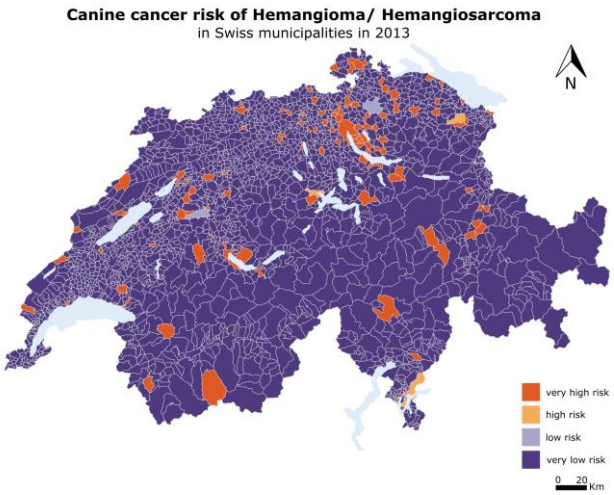


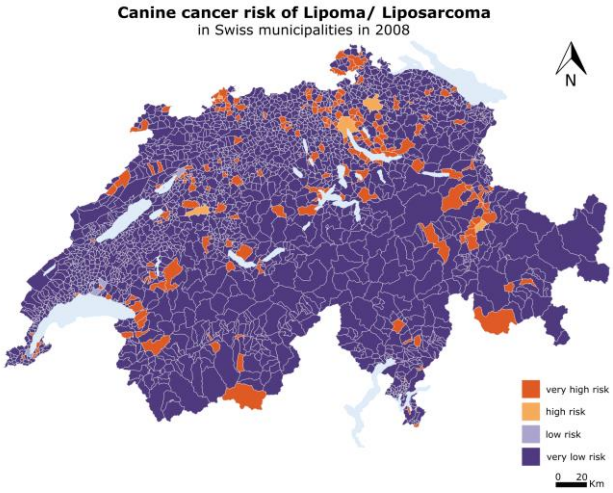



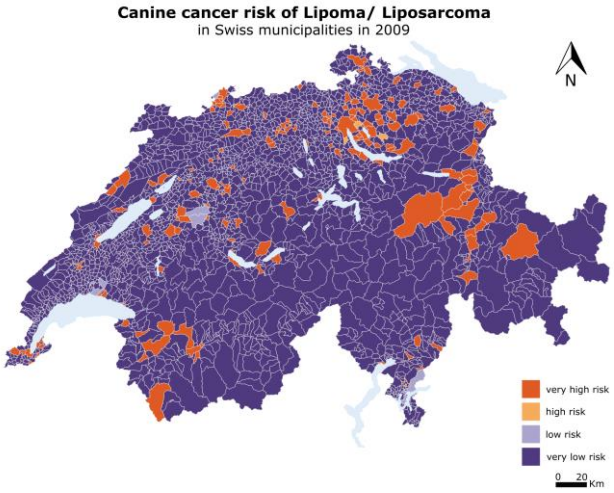



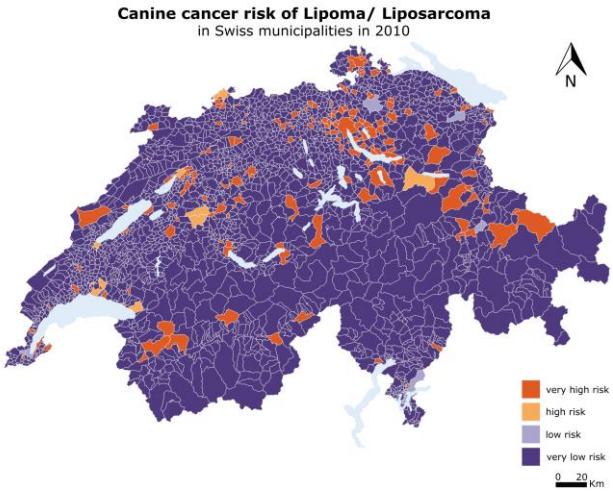



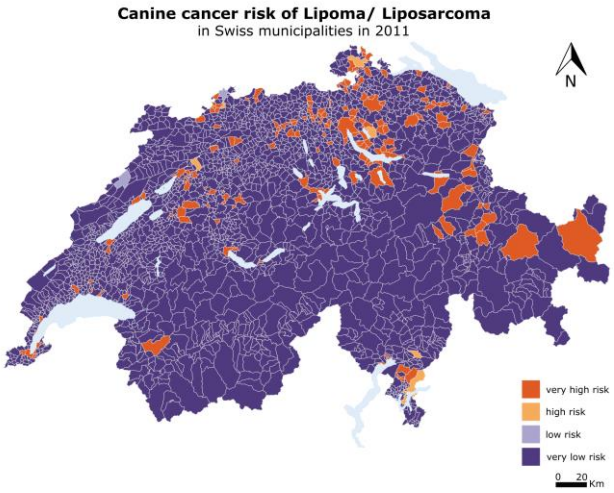



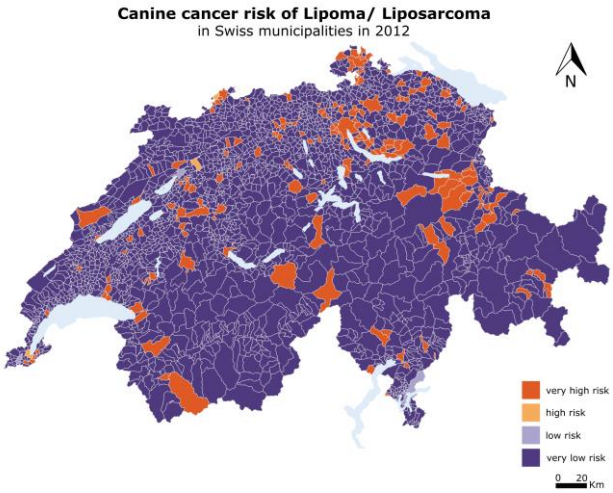
# F. Disease maps: Lipoma/ Liposarcoma

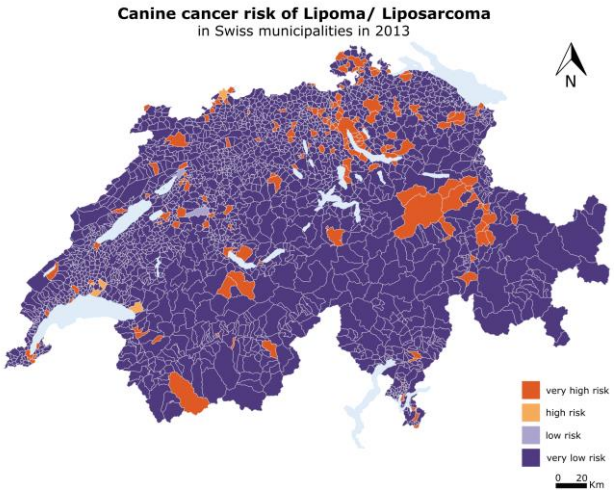


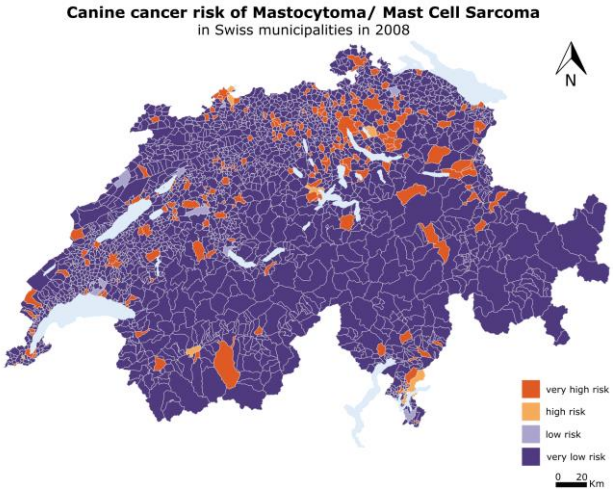



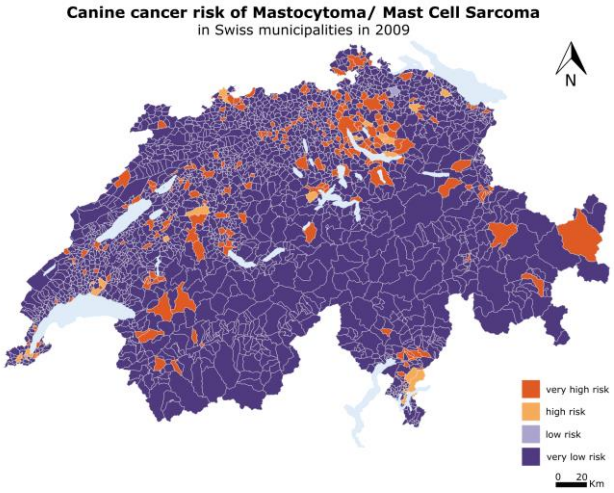



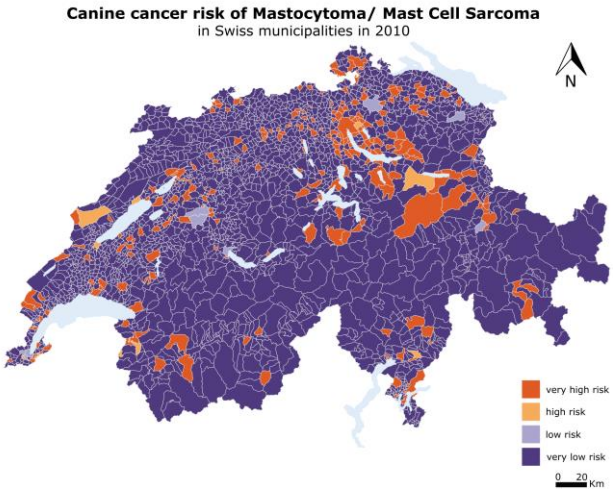



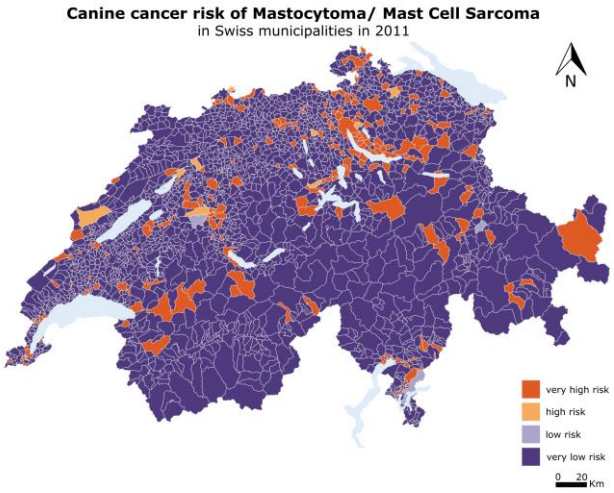



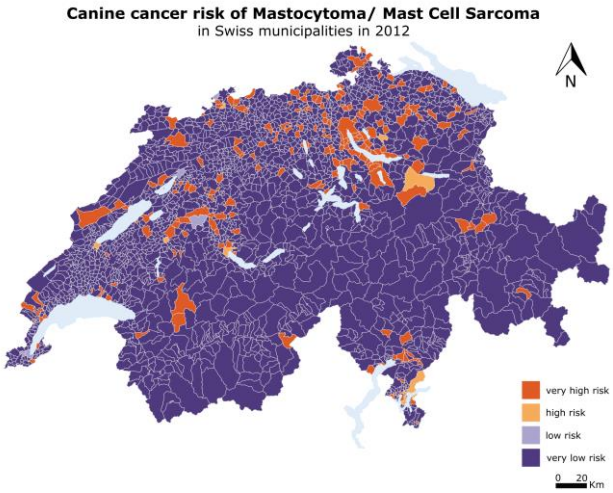
# G. Disease maps: Mastocytoma/ Mast Cell Sarcoma

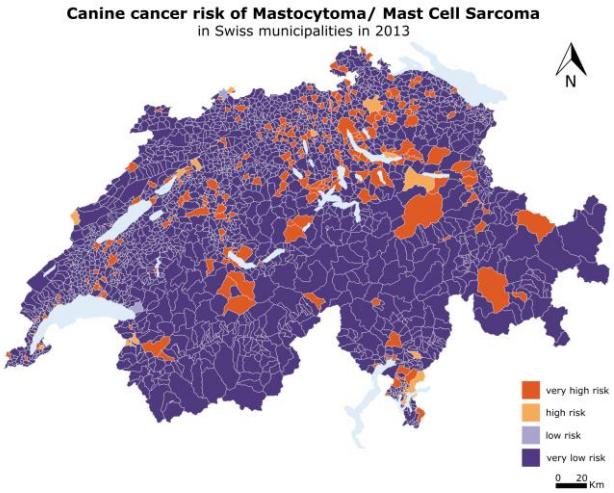


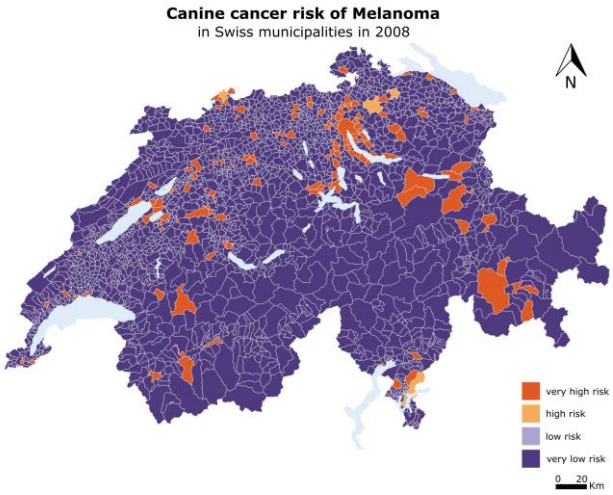



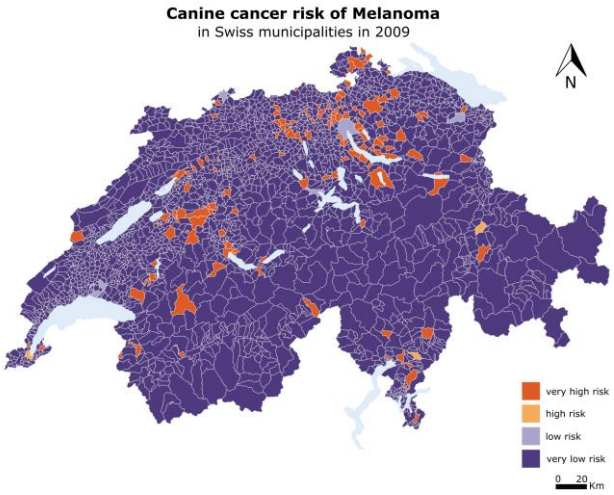



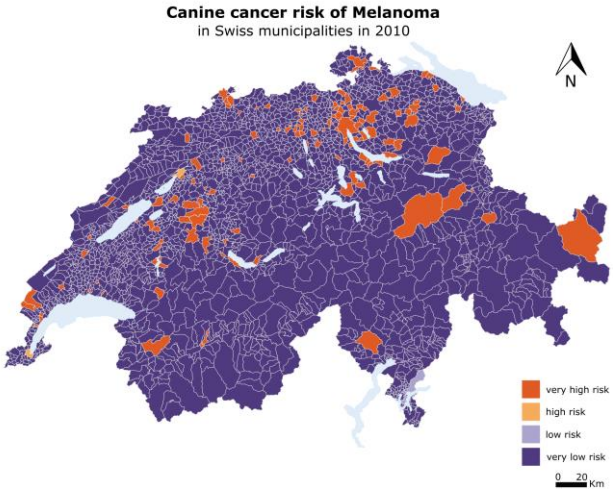



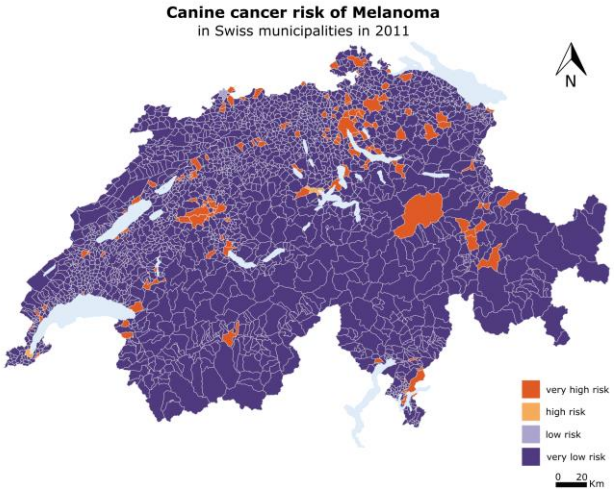



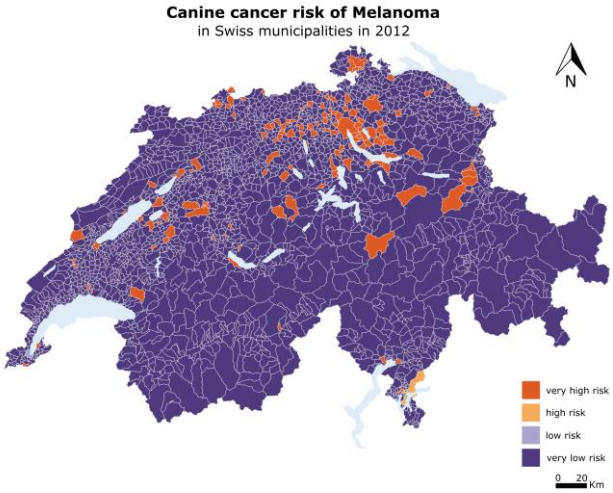
# H. Disease maps: Melanoma

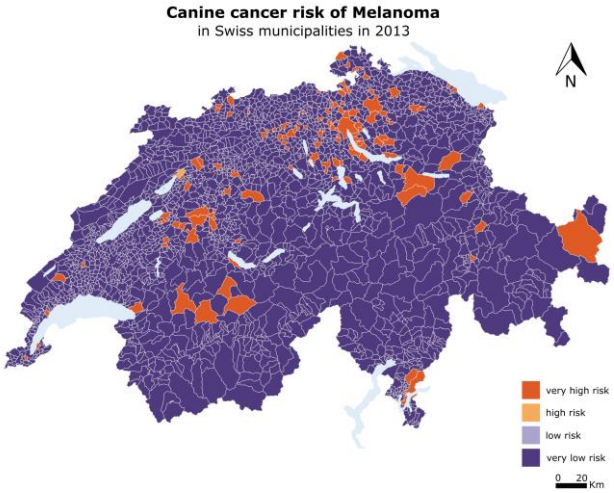


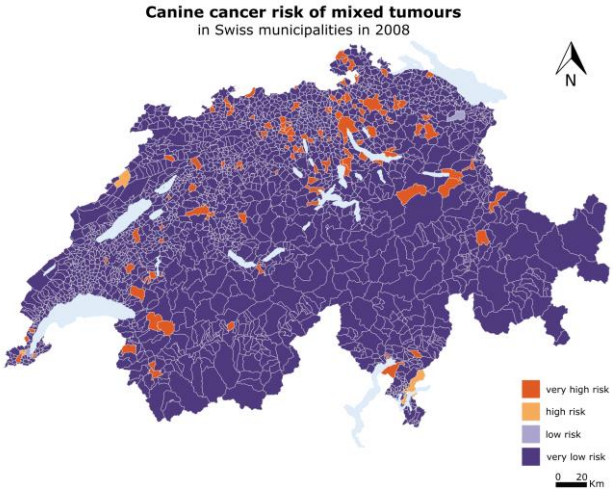



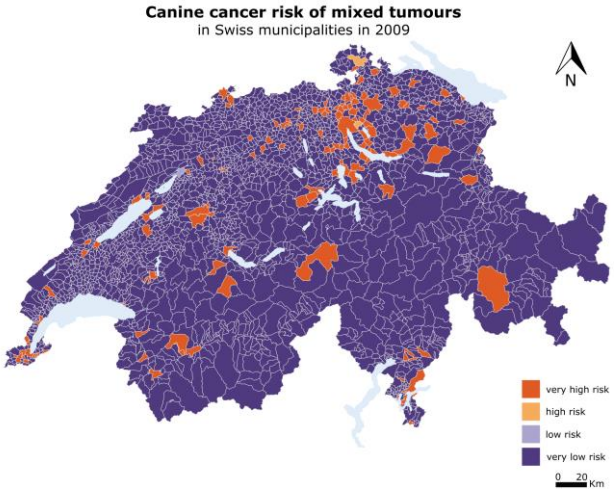



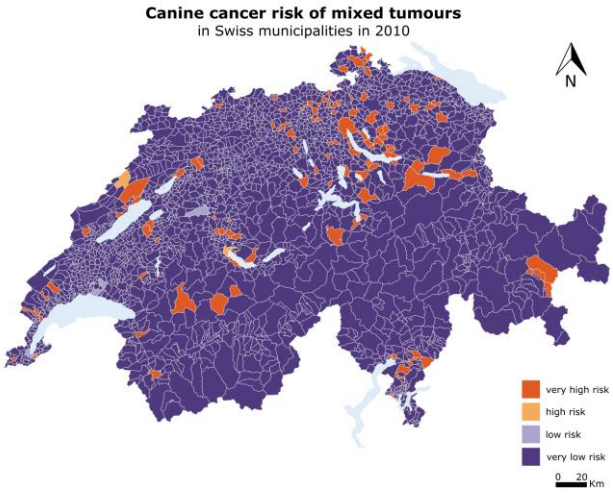



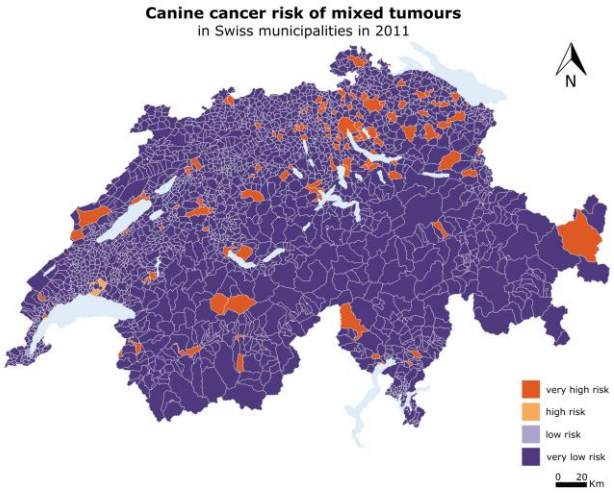



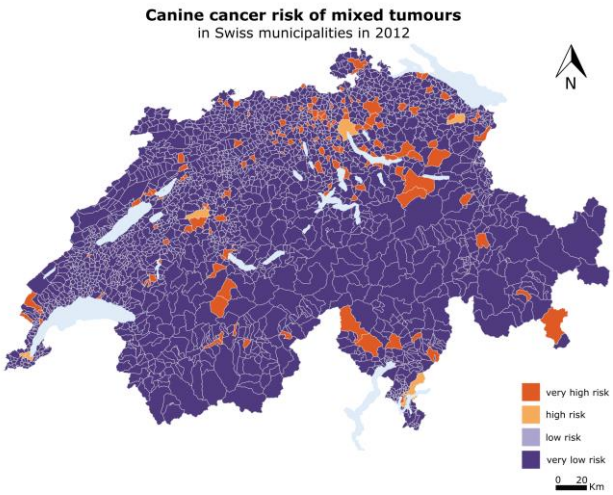
# I. Disease maps: Mixed tumours

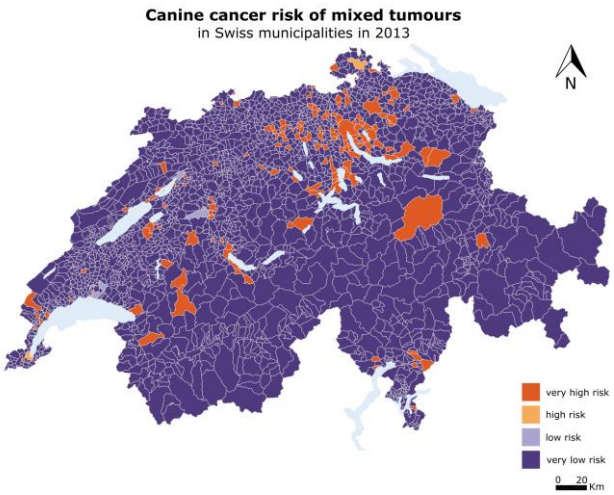


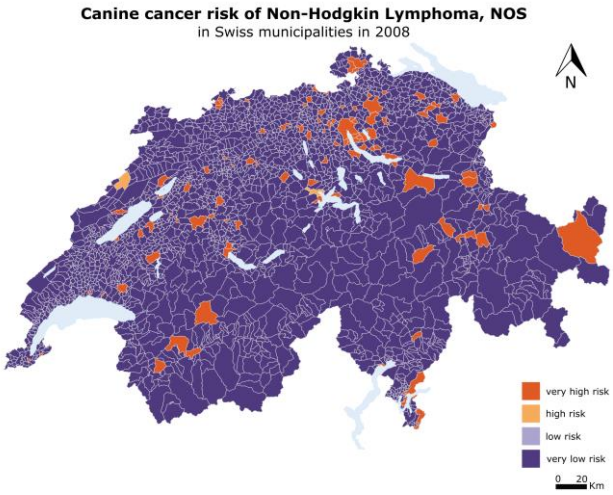



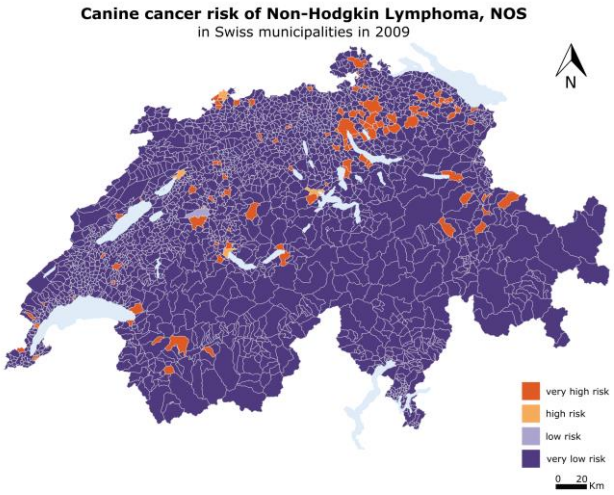



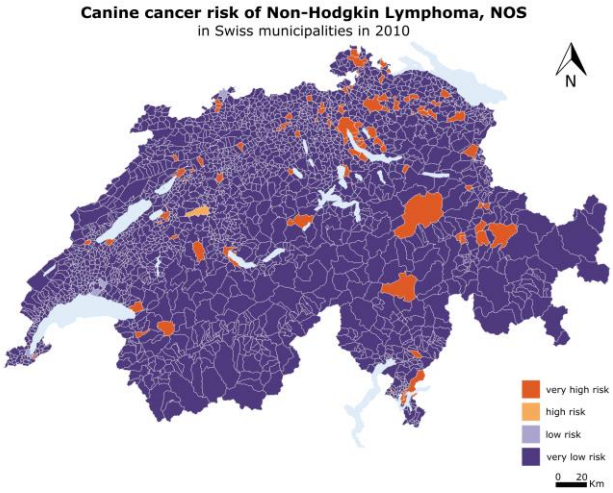



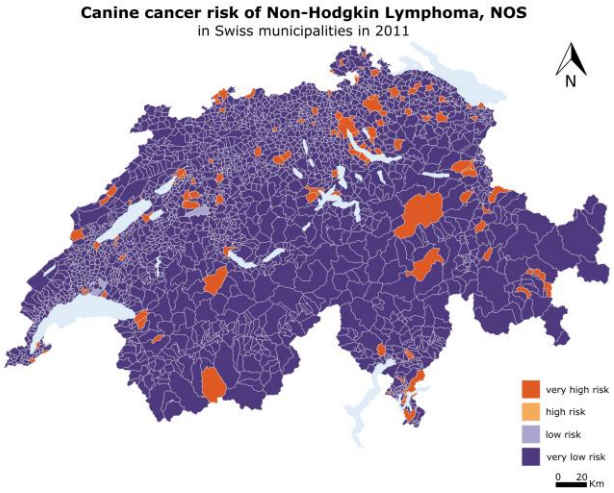



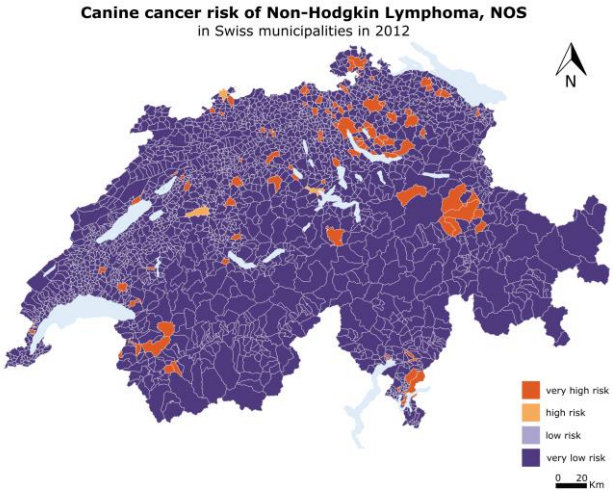
# J. Disease maps: Non-Hodgkin Lymphoma, NOS

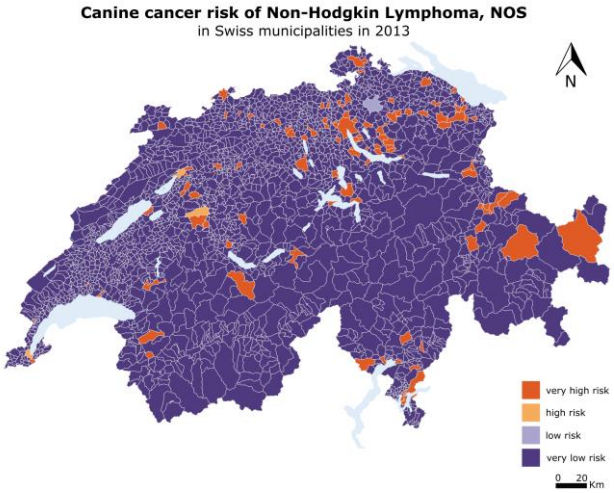


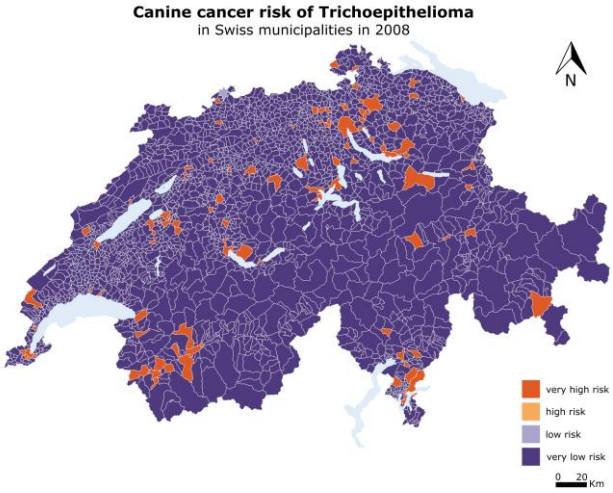



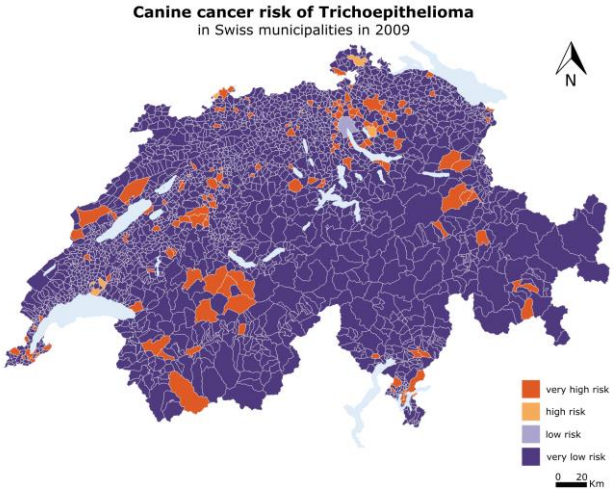



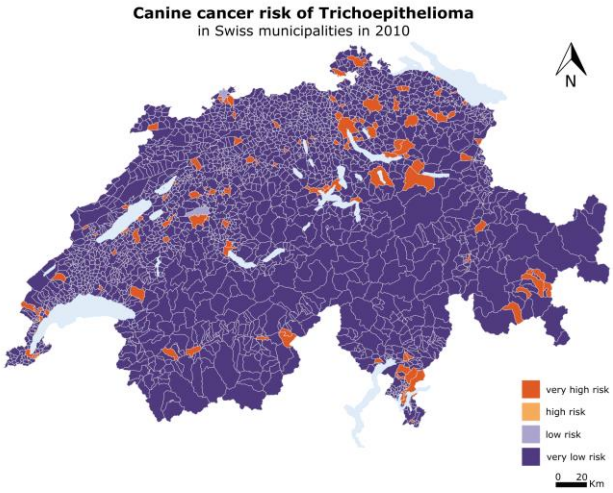



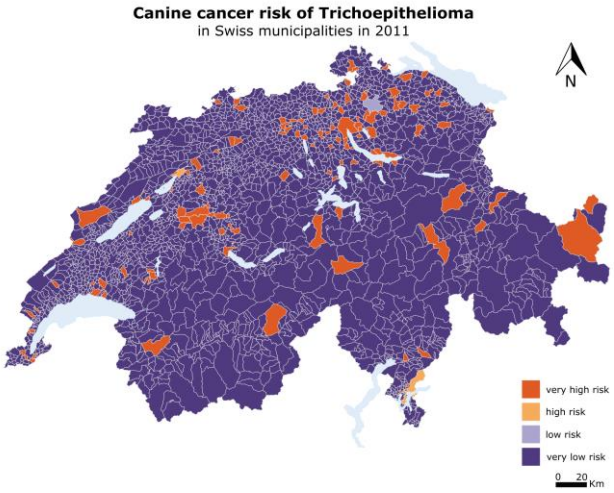



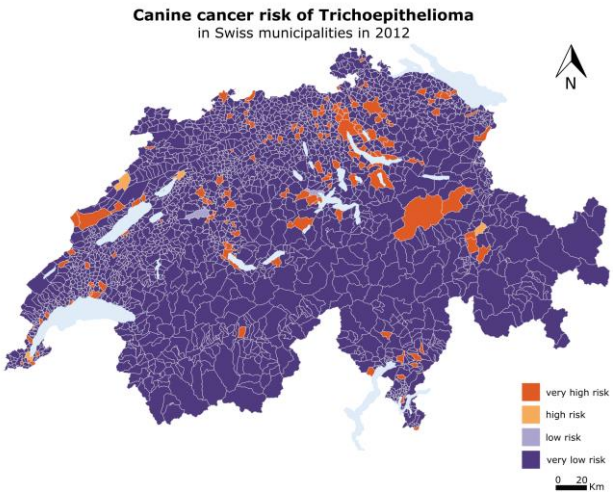
# K. Disease maps: Trichoepithelioma

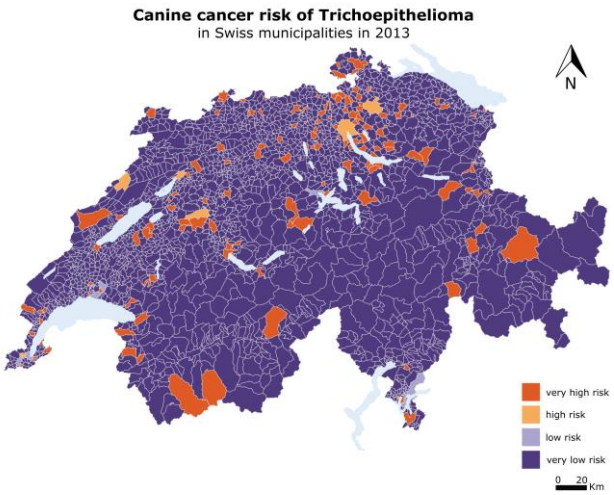


# Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Vanessa Sarah Guidetti

Friday, 21 April 2017