

# Mapping and Analysing cancer incidence in South Africa

**SJ Jansen van Rensburg**  
**21104417**

Dissertation submitted in fulfilment of the requirements for the degree *Magister Scientiae* in **Geography and Environmental Management** at the Potchefstroom Campus of the North-West University

Supervisor: Prof IJ van der Walt

Co-supervisor: Mr DP Cilliers

May 2014

# TABLE OF CONTENTS

<b>ABSTRACT .....</b>	<b>1</b>
<b>CHAPTER 1</b>	
<b>INTRODUCTION.....</b>	<b>2</b>
1.1 Research background .....	5
1.2 Objective .....	5
1.3 Study layout .....	5
<b>CHAPTER 2: LITERATURE REVIEW OF KNOWN ENVIRONMENTAL AND OTHER CARCINOGENS THAT ARE POSSIBLY INFLUENCING THE DEVELOPMENT OF THE 14 CANCER TYPES ANALYSED IN THIS DISSERTATION</b>	
<b>Introduction to literature review .....</b>	<b>6</b>
<b>2.1 Leukaemia's</b>	
2.1.1 Acute Lymphoblastic Leukaemia.....	7
2.1.2 Acute Myeloid Leukaemia .....	8
2.1.3 Environmental and other possible carcinogens linked to Leukaemia .....	8
<b>2.2 Lymphomas</b>	
2.2.1 Hodgkin's Lymphoma and Non-Hodgkin's Lymphoma.....	11
2.2.2 Burkitt's Lymphoma.....	11
2.2.3 Environmental and other possible carcinogens linked to Lymphomas .....	11
<b>2.3 Brain Tumours</b>	
2.3.1 Astrocytoma .....	13
2.3.2 Glioma.....	14
2.3.3 Medulloblastoma.....	14

2.3.4 Environmental and other possible carcinogens linked to brain tumours.....	14
<b>2.4 Hepatoblastoma</b> .....	17
2.4.1 Environmental and other possible carcinogens linked to Hepatoblastoma.....	17
<b>2.5 Kaposi Sarcoma</b> .....	18
2.5.1 Environmental and other possible carcinogens linked to Kaposi Sarcoma .....	18
<b>2.6 Neuroblastoma</b> .....	21
2.6.1 Environmental and other possible carcinogens linked to Neuroblastoma .....	21
<b>2.7 Nephroblastoma (Wilm's Tumour)</b> .....	23
2.7.1 Environmental and other possible carcinogens linked to Nephroblastoma .....	23
<b>2.8 Osteosarcoma</b> .....	25
2.8.1 Environmental and other possible carcinogens linked to Osteosarcoma .....	25
<b>2.9 Retinoblastoma</b> .....	27
2.9.1 Environmental and other possible carcinogens linked to Retinoblastoma.....	28

### **CHAPTER 3: MATERIALS AND METHODS**

3.1 Data quality .....	30
3.2 The software used in this study .....	34
3.3 Data preparation.....	38
3.4 Data Analysis .....	46

### **CHAPTER 4: RESULTS**

<b>4.1 Successfulness of GIS methodology</b> .....	67
<b>4.2 Result of data quality review</b> .....	68
<b>4.3 Results of the Hotspot Analysis (Getis-Ord Gi*) Tool</b>	
4.3.1 Difference between general results and accurate results produced by the Hotspot Analysis (Getis-Ord Gi*) Tool.....	69

4.3.2 General and accurate results discussed per cancer type	
4.3.2.1 Leukaemia.....	71
4.3.2.2 Lymphoma.....	73
4.3.2.3 Brain Tumours.....	77
4.3.2.4 Hepatoblastoma .....	81
4.3.2.5 Kaposi Sarcoma .....	82
4.3.2.6 Neuroblastoma .....	83
4.3.2.7 Nephroblastoma (Wilm’s Tumour).....	84
4.3.2.8 Osteosarcoma .....	85
4.3.2.9 Retinoblastoma .....	86
<b>CHAPTER 5: CONCLUSION AND RECOMMENDATIONS</b>	
5.1 Conclusion .....	88
5.2 Recommendations .....	90
<b>REFERENCES.....</b>	<b>92</b>

## **ABSTRACT**

The primary aim of this dissertation was to develop and validate a methodology for identifying spatial clusters (hotspots) of various paediatric cancers within South Africa by using GIS software. The Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool was used for this purpose. A series of spatial clusters (hotspots) were identified by the tool for each cancer type and these clusters were compared with the exiting literature regarding known environmental and other carcinogens. The quality of the cancer data used in the dissertation was however found to be questionable and significantly underreported. This caused the results of the tool to also be questionable. The dissertation therefore concluded that the tool could be successfully used to identify spatial clusters of cancer in principle. It was however found that the results of the tool needed to be viewed without caution in this dissertation due to the low quality of the cancer data used.

**KEY WORDS: CANCER HOTSPOTS, GIS, HOTSPOT ANALYSIS, SPATIAL CLUSTERS**

## CHAPTER 1: INTRODUCTION

In the last 30 years, the global burden of cancer has more than doubled and it is believed that it will continue to rise each year (Boyle & Levin, 2008; Jemal *et al.*, 2011). In the 1960s it was believed that cancer is a disease that only occurred in western industrialised countries with high resources. Today however, the majority of the global cancer burden is found in low- and medium-resource countries (Thom, 2008). According to the 2008 World Cancer Report (Boyle & Levin, 2008) cancer deaths doubled between 1975 and 2000 and are expected to double again in 2020 and nearly triple in 2030, leading to approximately 27 million cases of cancer and 17 million annual cancer related deaths by 2030. These increases in cancer incidences will arguably hit low- and medium-resource countries the hardest mainly due to the issue of limited health-care budgets (Mulcahy, 2008).

However, prevention of cancer is possible in some cases by means of identifying risk factors (Wang & Chen, 2001). Risk factors refer to those factors, or a combination of factors, that may be responsible for causing cancer (HHS, 2003; Sanderson *et al.*, 2009). Identification of risk factors is not a simple task, and delivering effective prevention can be even more difficult. Furthermore, factors of high priority must be identified to prioritise intervention strategies. Cancer has many causes, with factors outside (environmental factors), as well as inside the body, contributing to the development thereof (Delpomme *et al.*, 2007; HHS, 2003; Irigaray *et al.*, 2007). According to the U.S. Department of Health and Human Services (HHS, 2003) two thirds of all cases of cancer in the U.S. are related to human exposure to a wide variety of natural and man-made substances in the environment. Environmental factors can be divided into two groups (HHS, 2003; Irigaray *et al.*, 2007; Wang & Chen, 2001):

- Lifestyle choices (smoking, excessive alcohol consumption, poor diet, lack of exercise and excessive exposure to sunlight); and
- External factors (exposure to certain medical drugs, radiation, viruses, bacteria and chemicals present in the air, water, food and workplace).

In some cases specific environmental exposures are linked to specific kinds of cancer, such as asbestos that is linked primarily to lung cancer. In other cases factors can be linked to various types of cancer, such as smoking that can be linked to at least thirteen types of cancer (HHS, 2003). However, the chance that an individual will develop cancer in response to a particular environmental agent depends on several interacting factors such as length of exposure, frequency of exposure, exposure to other agents, genetic factors, diet, lifestyle, overall health, age and gender (HHS, 2003). Due to the complex interplay of these factors it is difficult to predict whether specific environmental exposure will cause a particular person to develop cancer. Nevertheless, it is known that certain environmental factors such as, tobacco, alcohol, ultraviolet radiation, pesticides, dioxins, metals and solvents may increase the risk (Delpomme *et al.*, 2007; HHS, 2003; Irigaray *et al.*, 2007).

By geographically mapping cancer cases (by type), certain patterns with regard to the interplay between cancer cases and environmental factors could be revealed (Boyle & Smans, 2008). The process of mapping a disease to identify its possible origin is not a new concept and has been used in the past. Probably the most renowned example is that of British physician John Snow who used the geographical method to identify the source of a cholera outbreak in London by mapping and analysing cholera incidences (McLeod, 2000). The first map of cancer however, was produced by Dr Alfrid Haviland in 1891, and communicated the influence of clays and limestone on cancer, indicated by the geographical distribution of cancer among women in England and Wales (Boyle & Smans, 2008). Currently, detailed maps showing the geographic distribution of cancer incidence in South Africa don't exist.

By using a Geographical Information System (GIS) to present cancer incidence on maps, the cancer scenery for South Africa could be obtained (Frenzel-Beyme *et al.*, 1979), which in turn could enable researchers to investigate spatial patterns and quantify the relationship between cancer and other health, socioeconomic and environmental variables (Brewer, 2006). Once detailed datasets indicating cancer incidence in South Africa have been derived, further analysis can be done to geostatistically examine the relationship between cancer incidence and certain land-uses and environmental factors.

GIS can thus serve as an additional tool in the exploration, analysis and communication of cancer and cancer related data (Brewer, 2006).

The reasoning behind the mapping of cancer is not only to produce a set of maps merely suitable for informative viewing, but a publication that will draw the attention of medical practitioners, scientists, public health officials and politicians to important features of cancer distribution in South Africa. Access to such information will hopefully stimulate further steps to combat the disease (Boyle & Smans, 2008). The mapping of cancer incidence can thus be an important tool in the fight against cancer, but before cancer maps can be created, the methodology for creating such maps needs to be established and tested to ensure that the maps that are created are trustworthy and statistically significant. The main purpose of this dissertation was therefore to identify and test a methodology that ensures that the process used to create these maps yields statistically significant results.

The cancer incidence data used in this analysis was studied and the 14 cancer types with the most cases were chosen for the analysis. The data consisted of paediatric cancer cases of children between the ages of 0 and 15 years old. The data was also used in conjunction with population statistics of children between the ages of 0 and 15 years old. Due to the limitations of the population statistics, that will be discussed in more detail later in this dissertation, the number of cancer cases used in the study was limited, although it should be remembered that even though the number of cancer cases used in the analysis was limited, it was still the best available data to use at the time and still a sample that represented the location of cancer incidence in South Africa. The geographic information linked to each cancer case was also not sufficient to do the analysis on the precise detailed scale first desired, namely the 2001 census enumeration area boundaries. The study was therefore conducted on a larger scale, namely on a local municipal scale, since many of the cases only had the town or suburb name as geographical reference point. The coordinates of the middle of these towns or suburbs were then used instead of the precise location where every patient lived (house address) to plot the data.



## **1.1 Research background**

The primary aim of this dissertation was to develop and validate a methodology on how to present approximately 7177 cancer cases in children across South Africa in a series of statistically significant maps by using GIS software. The created maps were presented according to the 14 cancer types chosen. An attempt was made to test whether the methodology used to create the maps resulted in the identification of statistically significant geographical clusters.

## **1.2 Objective**

A geographical presentation of cancer incidence can provide invaluable insight into the relationship between cancer and environmental factors within South Africa, but the process used to identify these cancer clusters needs to be scientific and therefore needs to be tested first.

## **1.3 Study layout**

Chapter 2 of this dissertation will review the existing literature on the known environmental carcinogens of the various cancer types analysed. Chapter 3 will then focus on the cancer data used in the study. This chapter will describe the quality of the cancer data used as well as how the data was prepared for analysis. This chapter also discusses the GIS process that was used to conduct the analysis. Chapter 4 of this dissertation will review the effectiveness of using GIS software to map cancer incidence in South Africa, as well as the results of the GIS analysis itself. This chapter will also highlight the limitations of implementing a GIS methodology for mapping cancer in South Africa, based on data quality concerns and limitations. Finally chapter 5 will focus on the conclusions that were drawn from the quality of the data, the preparation of the data, the GIS analysis itself and the interpretation of the results. This chapter will also provide recommendations for future studies similar to this one based on lessons learned during the study.

## **CHAPTER 2: LITERATURE REVIEW OF KNOWN ENVIRONMENTAL AND OTHER CARCINOGENS THAT ARE POSSIBLY INFLUENCING THE DEVELOPMENT OF THE 14 CANCER TYPES ANALYSED IN THIS DISSERTATION.**

Once it is known (geographically) where a specific cancer occurs more frequently, it is good to have a literature background on what the possible environmental and other carcinogens are for that specific cancer so one can know which environmental factors to investigate first within that area. By knowing which factors are possibly influencing the development of a specific cancer, one can focus one's efforts from the beginning on those factors that are most likely to have an influence on the development of the cancer. This is important since the resources for fighting cancer are always limited and cannot be wasted. Each of the 14 cancer types is therefore supported in this chapter by literature background regarding that specific cancer. The purpose of this section is therefore to know what to look for once a hotspot has been identified.

If suspected environmental carcinogens are present in the hotspot areas and not present in other areas where there are no hotspots then it can be concluded that those environmental carcinogens could indeed be influencing the development of that cancer. It is therefore necessary to understand the specific cancer type in order to truly understand the possible environmental carcinogens that might be present in the hotspot area.

The data that was used in this dissertation was derived from the National Paediatric Cancer Registry and was made available by Prof. Cristina Stefan, a paediatric oncologist at the Faculty of Health Sciences at the Tygerberg Campus of the Stellenbosch University. The data consists of approximately 7177 cancer cases diagnosed in the time period of 1992 to 2008 and is representative of cancer cases diagnosed all over South Africa.

The 14 cancer types analysed include:

- Acute Lymphoblastic Leukaemia (*Cancer of blood and bone marrow*)
- Acute Myeloid Leukaemia (*Cancer of blood and bone marrow*)
- Astrocytoma (*Brain Cancer*)
- Burkitt's Lymphoma (*Cancer commonly found in the jawbones*)
- Glioma (*Brain Cancer*)
- Hepatoblastoma (*Cancer of the liver*)
- Hodgkin's Lymphoma (*Cancer of lymph tissue, spleen, liver and other sites*)
- Kaposi's Sarcoma (*Cancer usually found in the skin*)
- Medulloblastoma (*Brain Cancer*)
- Nephroblastoma (*Cancer of the kidney*)
- Neuroblastoma (*Cancer of the adrenal glands, abdomen, thorax ,and neck*)
- Non-Hodgkin's Lymphoma (*Cancer of lymph tissue, spleen, and liver*)
- Osteosarcoma (*Cancer that starts in the bones*)
- Retinoblastoma (*Cancer of the eye*)

## **2.1. Leukaemia's**

### **2.1.1. Acute Lymphoblastic Leukaemia**

Acute Lymphoblastic Leukaemia (ALL) is a type of Leukaemia that starts from white blood cells in the bone marrow, the soft inner part of bones. It develops from cells called lymphocytes, a type of white blood cell central to the immune system, or from lymphoblasts, an immature type of lymphocyte. Acute Lymphoblastic Leukaemia invades the blood and can spread throughout the body to other organs, such as the liver, spleen, and lymph nodes. But it does not normally produce tumours as do many types of cancer. It is an acute type of Leukaemia, which means it can progress quickly. Without treatment, it can be fatal within a few months (Faderl & Kantarjian, 2011; WebMD, 2013a).

### **2.1.2 Acute Myeloid Leukaemia**

Acute Myeloid Leukaemia (AML) starts in the bone marrow. This is the soft inner parts of bones. With acute types of Leukaemia such as AML, bone marrow cells don't mature the way they're supposed to. These immature cells, often called blast cells, just keep building up. Without treatment, AML can quickly be fatal. This type of leukaemia can spread quickly to the blood and to other parts of the body (Proytcheva, 2011; WebMD, 2013b).

### **2.1.3 Environmental and other possible carcinogens linked to Leukaemia.**

Paternal exposure to pesticides has been linked to Leukaemia by Ma *et al.* (2002) and Shu *et al.* (1988). Various other studies also found an association with paternal pesticide use and Acute Lymphoblastic Leukaemia as well as household pesticide use and Acute Lymphoblastic Leukaemia (Alderton *et al.*, 2006; Belson *et al.*, 2007; Buffler *et al.*, 2005; Soldin *et al.*, 2009; Infante-Rivard & Weichenthal, 2007; Van Maele-Fabry *et al.*, 2011; Jurewicz & Hanke, 2006; Ma *et al.*, 2002; Meinert *et al.*, 2000; Menegaux *et al.*, 2006; Monge *et al.*, 2007). Maternal exposure to pesticides has also been linked to ALL. In a study conducted by Shu *et al.* (1988), a significant association was found between maternal exposure to pesticides during pregnancy and the risk of developing ALL in children younger than 15. In a study conducted by Rull *et al.* (2009) elevated ALL risk was associated with lifetime moderate exposure, but not high exposure, to certain physicochemical categories of pesticides, including Organophosphates, Chlorinated Phenols, and Triazines, and with pesticides classified as insecticides or fumigants. A similar pattern was observed for several toxicological groups of pesticides.

Pesticide exposure has also been linked to childhood Acute Myeloid Leukaemia and adult Acute Myeloid Leukaemia (Buckley *et al.*, 1989; Ries *et al.*, 1999 & Shu *et al.*, 1988). A study conducted by Krain (1991) found that high altitude exposure correlated less significantly with Leukaemia. Although this study only found a weak correlation between high altitude exposure and Leukaemia, it could indicate the

effect that increased ultraviolet radiation possible has on Leukaemia. The effect of other forms of radiation on Leukaemia has been observed by Ahlbom *et al.* (2001); Calvente *et al.* (2010); Wunsch-Filho *et al.* (2011) and the International Agency for Research on Cancer (IARC, 2002). All of these studies found a correlation between electromagnetic fields and childhood acute Leukaemia. There have been certain authors who put immense importance on the effect of radiation on Leukaemia such as Stewart *et al.* (1956) and Ries *et al.* (1999) who stated that the only known non-genetic risk factor for ALL is ionizing radiation, either from utero exposure to diagnostic X-rays or from postnatal exposure to therapeutic doses. According to the American Cancer Society, Leukaemia is the most radiation-induced cancer (American Cancer Society, 2010). There have also been links between other factors and Leukaemia, such as high birth weight and socioeconomic status. Although these factors are not direct environmental factors they might be indirectly caused by certain environmental factors.

In a study conducted by Yeazel *et al.* (1997) a statistically significant association between high birth weight and ALL as well as AML was found in children whose disease was diagnosed before the age of 2. Various other studies have also found an association between high birth weight and Leukaemia (MacIvlahon & Newill, 1962; Fasal *et al.*, 1971; Gold *et al.*, 1979; Roman *et al.*, 2013; Smith *et al.*, 2009). Internationally, the incidence of ALL is generally high in economically advantaged countries and low in disadvantaged countries (Kroll *et al.*, 2011). In Asia, Vietnam has a very low rate of 9.2 per million per year and Hong Kong has a high rate of 40.6 per million per year (Parkin *et al.*, 1998). In the United States of America, it was found that black children were only half as likely as white children to develop ALL. Similarly, in New Zealand, the incidences of ALL in Maori children were found to be half of that found in other children (Parkin *et al.*, 1998).

The Greaves' hypothesis and the Kinlen hypothesis are two possible explanations of how and why socioeconomic status might have an effect on the development of Leukaemia, with specific reference to the possible role of infections. The Greaves' hypothesis stipulates that common ALL results from two mutations that occur during periods of rapid proliferation of B-cells or their precursors. The second mutation

occurs postnatally during the proliferation of antibody producing cells after the infant's first exposure to multiple infections. If exposure to infection is delayed until after infancy, the likelihood of the second mutation occurring is increased thereby increasing the likelihood of developing ALL (Greaves, 1988).

The Kinlen hypothesis states that the spread of a viral infection that occurs when infected and susceptible individuals come in contact with each other leads to childhood Leukaemia. This "population mixing" occurs when, for example, large numbers of people move into a rural and previously sparsely populated area (Kinlen, 1988). By the time the Leukaemia has developed and been diagnosed, the infectious agent may no longer be present, making direct investigation of these hypotheses difficult. No infectious agents have been identified, although single studies have suggested Varicella (Vianna & Polan, 1976), Influenza (Fedrick & Alberman, 1972), and the Epstein–Barr Virus (Lehtinen *et al.*, 2003) as possible etiologic factors.

The possibility of better sanitation and less crowding that might occur with higher socioeconomic status could possibly result in a delayed exposure to infectious agents, thus supporting the Greaves' hypothesis. It is therefore interesting to note that the incidence of Acute Lymphoblastic Leukaemia is higher in industrialised countries than in developing countries (Parkin *et al.*, 1998). It is also interesting to note that Fasal *et al.* (1971); Greenberg and Shuster (1985); and McWhirter (1982) all found that the risk of Acute Lymphoblastic Leukaemia was higher for people of higher socioeconomic status. Fraumeni and Miller (1967) and Hrusak *et al.* (2002) observed a rapid growth of Acute Lymphoblastic Leukaemia in several countries during a time period of economic growth.

## **2.2 Lymphomas**

### **2.2.1 Hodgkin's Lymphoma and Non-Hodgkin's Lymphoma.**

Hodgkin's disease is a type of Lymphoma, a cancer that starts in white blood cells called lymphocytes. Lymphocytes are part of the body's immune system. There are two types of lymphomas, namely, Hodgkin's disease (Hodgkin's Lymphoma) and Non-Hodgkin's Lymphoma. These two main types of lymphomas differ in how they behave, spread and respond to treatment. Doctors can most often tell the difference between them by looking at the cancer cells under a microscope. In some cases laboratory tests may be needed to tell them apart. Lymph tissue is found in many parts of the body like lymph nodes; the spleen; the bone marrow, and the digestive tract. Lymphoma can therefore start in any part of the body (American Cancer Society, 2012).

### **2.2.2 Burkitt's Lymphoma**

Burkitt's Lymphoma is a form of Non-Hodgkin's Lymphoma, which occurs commonly in the jawbones among African children. This tumour is believed to be caused by the Epstein-Barr Virus and nowadays, this lesion is being reported from other parts of the world as well (Purkait, 2003).

### **2.2.3 Environmental and other possible carcinogens linked to Lymphomas.**

According to Mueller *et al.* (1996); Ferry (2006) and Molyneux *et al.* (2012), Burkitt's Lymphoma in children occurs endemically in hot, wet, rural lowlands and it is therefore thought that repeated malarial infections play a role in its etiology, possibly by its mitogenic effects. Climates associated with higher altitudes may also play a role in the development of Lymphomas as Krain (1991) found a weak association between Hodgkin's Lymphoma and high altitude exposure.

A possible reason why climate could have an effect on the development of Lymphomas might be because of the viruses that exist in certain climates. According to Montesano and Hall (2001); Talbot and Crawford (2004); and Carozza *et al.* (2009) there are various viruses that have been linked to Lymphomas such as the Epstein-Barr Virus that has been linked to Burkitt's Lymphoma and Hodgkin's Lymphoma and the Helicobacter Pylori Virus that has been linked to Gastric Lymphoma (Molyneux *et al.*, 2012; Biegging *et al.*, 2010).

The use of pesticides might also help these viruses infect children. This might indeed be what is happening since Hayes *et al.* (2006); Repetto (2013); Krieger (2010); and Satoh and Gupta (2010) agree that many pesticides are strong immunosuppressors. This argument is strengthened by the work of Sandlund *et al.* (1996) and Heise (2010), who found that immunodeficiency is known to increase the risk of Non-Hodgkin's Lymphoma. Other authors such as Menegaux *et al.* (2006); Daniels *et al.* (1997); Zahm and Ward (1998); George and Shukla (2011); Boccolini *et al.* (2013); Bolognesi and Merlo (2011) and the United States Environmental Protection Agency (US EPA, 2003) all agree that there is an overall increase in the relative risk of developing Non-Hodgkin's Lymphoma when parents or children have been exposed to pesticides. This argument is also supported by the work of Buckley *et al.*, (2000) who found a significant association between the risk of Non-Hodgkin's Lymphoma and increased frequency of reported pesticide use in the home.

It is not only the chemicals used in pesticides that have been linked to Lymphomas. Strong associations between Trichloroethylene and Non-Hodgkin's Lymphoma have been found in various studies (Radican *et al.*, 2008; Lipworth *et al.*, 2011; Purdue *et al.*, 2011). Trichloroethylene is a volatile organic chemical used primarily as an industrial solvent. The most common use therefore is to remove grease from fabricated metal parts and some textiles. It is also an ingredient in adhesives, paint removers, type writer correction fluid, rug-cleaning fluids and pepper sprays (ECSA, 2011; US EPA, 2006; US EPA, 2000).



This chemical is not only found within the factory it is used. According to the United States Environmental Protection Agency (US EPA, 2001) and Bellar (1974) it has been found in ambient air, surface water and ground water. Background levels of Trichloroethylene have also been found in industrial settings, homes undergoing renovations and homes using private wells located near Trichloroethylene disposal or contaminated sites.

According to Wallace (1985) and Mahle *et al.* (2007) the most likely Trichloroethylene exposure route for children is through pulmonary, oral or dermal routes. Indirect factors that might be facilitating the development of Lymphomas are socioeconomic status and diet. Although these are not direct environmental factors they may be indirectly caused by certain environmental factors. According to Scherr and Mueller *et al.* (1996); Soliman *et al.* (2013); and the American Cancer Society (American Cancer Society, 2013a), the incidence of Non-Hodgkin's Lymphoma is generally higher in developed countries. This could possibly indicate an unknown effect that socioeconomic status might have on the development of Non-Hodgkin's Lymphoma.

## **2.3 Brain Tumours**

### **2.3.1 Astrocytoma**

An Astrocytoma is a tumour that arises from the star-shaped cells (astrocytes) that form the supportive tissue of the brain. Other supportive cells of the brain include oligodendrocytes and ependymal cells. Collectively, these cells are known as glial cells and the tissue they form is known as glial tissue. Tumours that arise from the glial tissue, including Astrocytomas, are collectively referred to as Gliomas (Ferri, 2011; WebMD, 2013c).

### **2.3.2 Glioma**

A Glioma is a type of tumour that starts in the brain or spine. It is called a Glioma because it arises from glial cells. It is composed of neuroglia in any of its stages of development, sometimes extended to include all intrinsic neoplasms of the brain and spinal cord, as Astrocytomas and Ependymomas (Farlex, 2012a).

### **2.3.3 Medulloblastoma**

Medulloblastomas are malignant tumours formed from poorly developed cells at a very early stage of their life. They develop in the cerebellum, in a part of the skull called the posterior fossa, but may spread to other parts of the brain. Very rarely, Medulloblastomas may spread to other parts of the body. If they do spread to other parts of the brain or to the spinal cord, this is usually through the cerebrospinal fluid (CSF). CSF is the fluid that surrounds and protects the brain and the spinal cord (Halperin *et al.*, 2010; Macmillan Cancer Support, 2013).

### **2.3.4 Environmental and other possible carcinogens linked to brain tumours.**

The use of pesticides by parents and the exposure to pesticides by children have been linked to the relative risk for developing brain tumours in various studies (Menegaux *et al.*, 2006; Daniels *et al.*, 1997; Zahm & Ward, 1998; Van Maele-Fabry *et al.*, 2013; Shim *et al.*, 2009; Greenop *et al.*, 2013). It is possible that pesticides might not be the only agricultural factor that may be associated with brain tumours. The exposure to farm animals has also been linked to elevated levels of brain tumours (Bunin *et al.*, 1994; Efird *et al.*, 2003; Holly *et al.*, 1998). A more recent study conducted by Christensen *et al.* (2012) however found no association between farm animal exposure and the development of brain cancer, thus indicating that the possible influence of farm animal exposure needs further investigation.

Pesticides are not only used on farms but also in urban environments. A study conducted by Rosso *et al.* (2008) found a significant association between Medulloblastoma and lawn care pesticides. The result from this study strengthens the argument that pesticides are facilitating the development of brain tumours as well as other cancers. There are also other farming activities that could be carcinogenic such as the treatment of wooden fence droppers or wooden farming equipment with wooden handles. An association between Astocytoma and wood preservatives has been found by Schuz *et al.* (2001a).

There is evidence from various studies that vineyards and citrus farms seem to be associated with the development of brain cancers. A study conducted by Musicco *et al.* (1988) suggests that exposure to alkyl ureas (components of fungicides used extensively in vineyards) may explain the significant positive association between farming and Glioma observed in their case control study in Italy. A death certificate study in France found significantly increased standardised mortality ratios for vineyard farmers in regions with higher pesticide use compared with farmers in areas with low pesticide use (Viel *et al.*, 1998).

In south western France in the Bordeaux area, which is known for its vineyards, brain tumours have been extensively registered since 1999 and the incidence is among the highest reported in the world (Elia-Pasquet *et al.*, 2004). Propargite is an insecticide that is primarily used in orchards and vineyards. It was the highest ranked among individual pesticides for potential cancer hazard based on reported use in California weighted by exposure and carcinogenic potential (Grunier *et al.*, 2001).

In a study conducted by Jensen and Nordberg (1980) twenty seven patients who died from cancer were found to have significantly higher levels of pesticide in their fat tissues in comparison with forty four people who died of other illnesses. The pesticide Dichloro-Diphenyl-Trichloroethane (DDT), which was used in orange and grapefruit production, was found in these fat tissues. Studies conducted by Carreon *et al.* (2005) and Lee *et al.* (2005) also found possible associations between winegrowing and the development of brain cancer. In an ecological study in the

province of Trento, Italy, that was conducted by Ferrari and Lovaste (1986) it was found that the highest incidences of brain tumours were found in regions of intensive fruit and wine cultivation.

There are also natural sources that have been linked to brain tumours such as exposure to lead or mercury (Hayes, 1997; Wu *et al.*, 2012). Electromagnetic fields of extremely low frequency have been linked to the elevated risk of developing brain tumours in a study conducted by Mack *et al.* (1991). The possible effect of electromagnetic fields on the development of brain cancer however still needs further investigation since various studies have not found a significant association between electromagnetic fields and the development of brain cancer (Tynes & Haldorsen, 2003; Labreche *et al.*, 2003; Kheifets *et al.*, 2008). A study conducted by Bidoli *et al.* (1993) found a weak association between brain tumours and male residents living above 200 metres in north-eastern Italy.

Birth weight seems to be influencing the development of brain cancers and though it is not an environmental factor in itself, it might be indirectly influenced by the environment. High birth weight has been linked to the increased development of brain tumours in various studies (MacIvlahan & Newill, 1962; Fasal *et al.*, 1971; Gold *et al.*, 1979; Ross, 2006). Harder *et al.* (2008) found a strong association between high birth weight and Medulloblastoma specifically. Studies conducted by Seidman *et al.* (1982) and Reid *et al.* (2013) both found a positive association between asbestos exposure and the development of brain cancer. The results of these studies are supported by Bunderson-Schelvan *et al.* (2011) and Reid (2012), who agree that asbestos exposure could be contributing to the development of brain cancers.

## **2.4 Hepatoblastoma**

Hepatoblastoma is the most common form of liver cancer in children, although it is a comparatively uncommon paediatric solid tumour. The disease usually affects children younger than 3 years (Farlex, 2012b; Medscape, 2013).

### **2.4.1 Environmental and other possible carcinogens linked to Hepatoblastoma.**

There are certain sources in the natural environment which have been linked to Hepatoblastoma. The inhalation of arsenic oxides can cause liver cancer if it is swallowed according to Szymanska-Chabowska *et al.* (2002). The mechanisms of the action of metals and metalloids are not clear yet. They could act as co-carcinogens by activating procarcinogens in the liver according to Cantor *et al.* (2006) and Hayes (1997).

They could also act by replacing the natural enzyme- associated metal, thus inactivating the metabolic pathway of key enzymes. Berkel and Bako (1992) as well as Petrovski *et al.* (2012) found that in general areas with low levels of selenium tend to produce more cancers of the liver. The protective effects of selenium against cancer have been highlighted by the work of Bayoumy (2001); Abdulah *et al.* (2005); and Sarg and Gross (2007). The Hepatitis B and Hepatitis C Viruses have also been linked to Hepatoblastoma (Talbot & Crawford, 2004). Latini *et al.* (2004) found a link between Phthalate exposure, which is used in Polyvinylchloride in medical devices, and Hepatoblastoma.

It was found by Birch (2011) and Mclaughlin *et al.* (2006) that maternal and paternal preconceptional and gestational tobacco smoking was a risk factor for Hepatoblastoma, especially when both parents were regular smokers. In 2009 the International Agency for Research on Cancer classified tobacco smoke (via the parents) as a carcinogen for Hepatoblastoma (Birch, 2011). The work of Sorahan

and Lancashire (2004) found that the risk of developing Hepatoblastoma doubled if both parents smoked. Parental alcohol consumption and obesity have also been linked to the development of Hepatoblastoma (English *et al.*, 1995; World Cancer Research Fund, 1997).

## **2.5 Kaposi Sarcoma**

Kaposi Sarcoma is a cancer that develops from the cells that line lymph or blood vessels. The abnormal cells of Kaposi Sarcoma form purple, red, or brown blotches or tumours on the skin. These affected areas are called lesions. The skin lesions of Kaposi Sarcoma may look bad, but in many cases, the lesions cause no symptoms. In other cases, the disease causes painful swelling, especially in the legs, groin area, or skin around the eyes. Kaposi Sarcoma can cause serious problems (or even become life threatening) when the lesions are in the lungs, liver, or digestive tract. Kaposi Sarcoma in the digestive tract, for example, can cause bleeding, while tumours in the lungs may cause difficulty breathing (Purkait, 2003).

### **2.5.1 Environmental and other possible carcinogens linked to Kaposi Sarcoma.**

Kaposi Sarcoma has been found to be higher in areas where there are widespread swamps and wetlands (Dal Maso *et al.*, 2005). These are all areas where bloodsucking insects such as mosquitoes, biting midges and black flies are abundant and where malaria is endemic (Geddes *et al.*, 1995; Ascoli *et al.*, 2003). Fertile reddish-brown volcanic clay soils seem to be present where high levels of Kaposi Sarcoma are found. The Great Rift Valley in Eastern Zaire, western Uganda and Tanzania as well as further south in Malawi is an example of where there are volcanic mountains. There are also volcanic mountains in the mountainous areas of Cameroon, where Kaposi Sarcoma seems to be elevated (Cook-Mozaffari *et al.*, 1998). The volcanic soils found in these areas contain uniform colloidal particles of kaolinite, which forms preferentially over parent rock of alkalic basalt under tropical conditions of high altitude and seasonal rainfall. According to Price (1990) a clay

emulsion disaggregates into small kaolin particles in water, which can gain entry to the skin through sweat glands. Their high negative charge permits adsorption of cations including iron oxides. Peasants living in these areas are chronically exposed to the wet sticky clay and it can therefore enter the sweat glands and pores of their feet. This process is possibly aided by micro-abrasions caused by the high quartzite content of these soils. According to Price (1990) it is known that the penetration of ultrafine particles of clay into the skin of the feet during barefoot walking can lead to dermal lymphatic damage and impaired local immunity.

According to Ziegler (1993) and Pelsler *et al.* (2009), the chronic exposure to volcanic clays facilitates the development of Kaposi Sarcoma in Africa and Sicily, because it is common among rural peasants, has a predilection for the lower legs, is thought to originate from lymphovascular cells, and finally, because it is enhanced by impaired immunity. According to Ziegler *et al.* (1997) and Ziegler *et al.* (2003), barefoot walking as well as clay soil exposure increases the likelihood of developing Kaposi Sarcoma. In a study conducted by Krauskopf (1979), a magmatic substrate similar to that of the East African Rift system was found in Iceland and the Faroe Islands as well as in Mediterranean volcanic regions such as Sardinia and Sicily. According to Simonart *et al.* (1999) these areas are all known to exhibit high incidence rates of classic Kaposi Sarcoma.

In a study conducted by Montella *et al.* (1996), a nearly twofold increase in Kaposi Sarcoma was observed for people living near volcanic soils. According to Ollier (1984), volcanic soils are highly weatherable, which allows it to release significant quantities of iron compounds into the environment. The high incidence rates of Kaposi Sarcoma in regions where volcanic soils are present may therefore point to the prolonged exposure to indigenous iron oxide rich volcanic soils as a common aetiological risk factor. Cellular iron content has also been linked to the development of cancers by the work of Weinberg (1996) and Huang (2003). According to the American Cancer Society, skin cancer is strongly linked to radiation exposure (American Cancer Society, 2010).

The effect of radiation on skin cancer is also highlighted by the work of Krain (1991) who found that high altitude exposure, and therefore increased ultraviolet radiation exposure, correlated significantly with skin cancer. An interesting finding was that tobacco use was found to be associated with a decreased risk of developing of Kaposi Sarcoma (Goedert *et al.*, 2002; Nawar *et al.*, 2005; Ford *et al.*, 2005). Goedert *et al.* (2002) found that patients with Kaposi Sarcoma were only 25% as likely to be current or former smokers as compared to Non-Kaposi Sarcoma controls. Mineral oils and lubricants have also been found to be associated with skin cancers (Kane *et al.*, 1984; Mackerer *et al.*, 2003). According to Rabkin and Yellin (1994), HIV is known to increase the incidence of Kaposi Sarcoma. The relationship between HIV and Kaposi Sarcoma is of special interest in South Africa because of the HIV endemic that exists in the country. According to the American Cancer Society (2013b), when HIV damages the immune system, people who are also infected with the Kaposi Sarcoma Herpes Virus are more likely to develop Kaposi sarcoma. The identification of HIV hotspots in South Africa could possibly increase the treatment and prevention of Kaposi Sarcoma.

If a similar study to this dissertation could be carried out to find hotspots of HIV then researchers would be able to predict where the resources for fighting Kaposi Sarcoma should go. In a study conducted by Jones *et al.* (2000) it was found that the incidence of Kaposi Sarcoma decreased as new effective antiretroviral agents developed. It is thus clear that resources such as antiretroviral agents can effectively combat Kaposi Sarcoma, but it would first be necessary to identify in which geographic areas HIV has the highest incidence, in order to identify where Kaposi Sarcoma has the highest probability of developing.



## **2.6 Neuroblastoma**

Neuroblastoma is a cancer that forms in the nerve tissue. It usually begins in the adrenal glands, which sit on top of one's kidneys. It may also begin in the neck, chest or spinal cord. The cancer often begins in early childhood. Sometimes it begins before a child is born. By the time doctors find the cancer, it has usually spread to other parts of the body. The most common symptoms are a lump in the abdomen, neck or chest, bulging eyes, dark circles around the eyes, bone pain, a swollen stomach and trouble breathing in babies, painless bluish lumps under the skin in babies, and the inability to move a body part (Copel & Han, 2012; Medline Plus, 2013a).

### **2.6.1 Environmental and other possible carcinogens linked to Neuroblastoma.**

The effects of pesticides on Neuroblastoma have been highlighted by the work of Michalek *et al.* (1996); Kramer *et al.* (1987); Schwartzbaum (1992); Bunin *et al.* (1990); Spitz and Johnson (1985); and Carozza *et al.* (2008), who all found that parental occupational exposure to pesticide lead to increased incidences of Neuroblastoma in children. This argument is further supported by the work of Nasterlack (2007) who found an elevated risk for Neuroblastoma in children whose mothers or fathers were occupationally exposed to herbicides, insecticides or pesticides. Kerr *et al.* (2000) found that mothers and fathers who were significantly more exposed to insecticides during the mother's pregnancy had more children diagnosed with Neuroblastoma than control parents who were not exposed to insecticides during pregnancy.

Daniels *et al.* (2001) found modestly elevated levels of Neuroblastoma in children when both parents reported using pesticides in the home and garden. A strong association between children diagnosed after one year of age and Neuroblastoma was found specifically for pesticides used in the home and the garden. Many studies have found an association between Neuroblastoma in children and paternal

occupational exposure to pesticides. It is however clear from the abovementioned work that the role of maternal exposure to pesticides should not be ignored, since it is the mother who is the most likely to touch the child and thereby transfer pesticide chemicals onto the child. A seven-fold risk increase for high-stage Neuroblastoma was observed with maternal occupational exposure to pesticides by the work of Schuz *et al.* (2001c). It is clear from the abovementioned work that there might be chemicals in agricultural pesticides that facilitate the development of Neuroblastoma.

It is however not only in the agricultural setting where chemicals are used. Many chemicals used in the industrial setting have also been linked to various cancers and it is possible that Neuroblastoma might be no exception. The carcinogenic effect of Trichloroethylene has already been discussed in this chapter. The argument that industrial chemicals might also play a role in the development of Neuroblastoma is strengthened when the global pattern of the disease is considered.

Neuroblastoma incidences appear to be low in Africa, Central America, and South America and high in North America, Australia, and New Zealand (Parkin *et al.*, 1998). Rates are high in Europe with intermediate rates in some Eastern European countries. Rates are mixed in Asia with low rates in China, India, and Thailand and high rates in Japan (Parkin *et al.*, 1998). Generally, the rates are higher in industrialised nations and low in developing nations. The fact that Neuroblastoma is generally higher in industrialised nations could point to the fact that industrial chemicals might be playing a role in the development of Neuroblastoma.

It is possible that people living in developing or less industrialised countries will have lower socioeconomic status. It is interesting to note that Menegaux *et al.* (2004) and Daniels *et al.* (2002) found that day-care attendance, childhood infections and breast feeding were related with lower levels of Neuroblastoma. The protective effect of breast feeding has also been highlighted by the work of Hamos *et al.* (1996) and Hamburger (1988).

The work done by Macvlahon and Newill (1962); Daling *et al.* (1984); Fasal *et al.* (1971); Gold *et al.* (1979); Uravama *et al.* (2007) and Harder *et al.* (2010), all found a positive association between high birth weight and Neuroblastoma. Contradictory to these findings however is the findings of Schuz *et al.* (2001c), who found that Neuroblastoma was strongly associated with a shorter gestational period and a birth weight of less than 2.5 kg.

## **2.7 Nephroblastoma (Wilm's Tumour)**

Nephroblastoma (also known as Wilm's Tumour) is a rare type of kidney cancer. It causes a tumour on one or both kidneys. It usually affects children, but can happen in adults. Having certain genetic conditions or birth defects can increase the risk of developing it. Children that are at risk should be screened every three months until they turn eight. Symptoms include a lump in the abdomen, blood in the urine, and a fever for no reason. Tests that examine the kidneys and blood are used to find the tumour (Eble *et al.*, 2004; Medline Plus, 2013b).

### **2.7.1 Environmental and other possible carcinogens linked to Nephroblastoma**

The work done by Menegaux *et al.* (2006); Daniels *et al.* (1997); Cooney *et al.* (2007); Tsai *et al.* (2006); and Zahm and Ward (1998) all found an association between pesticides and Nephroblastoma. This argument is strengthened by the work of Sharpe *et al.* (1995); Kristensen *et al.* (1996); Clapp *et al.* (2008); Corozza *et al.* (2008) and Fear *et al.* (1998) who also found positive associations between Nephroblastoma and pesticides. The potential risk of fathers contaminating their children with pesticide chemicals was tested by Fear *et al.* (1998), who found an association between paternal occupational exposure to pesticides and Nephroblastoma. Once again this does not seem to be related to fathers only since Schuz *et al.* (2001b) found a non-significantly increased risk for children to develop Nephroblastoma when their mothers were exposed to pesticides.

Several studies have found a positive association between Nephroblastoma and paternal employment as a welder or mechanic (Bunin *et al.*, 1989b; Kantor *et al.*, 1979; Olshan *et al.*, 1990; IARC, 2008; United States Department of Labour, 2013). Trichloroethylene (previously discussed in this chapter) has also been strongly associated with Nephroblastoma (Christensen *et al.*, 2013; Moore *et al.*, 2010; Karami *et al.*, 2012; Scott & Jinot, 2011). The incidence of Nephroblastoma is high in North America, Europe, Australia, and New-Zealand, but generally lower in Asia and South America (Parkin *et al.* 1998). One can again speculate that there is a reason why Nephroblastoma is higher in industrialised areas such as North America and Europe. This global trend seems to support the argument that industrial chemicals are related to increased levels of Nephroblastoma. It might be possible that children in industrialised countries are more exposed to industrial chemicals.

Szymanska-chabowska *et al.* (2002) found that the inhalation of arsenic oxide can cause lung cancer but if it is swallowed, can cause Nephroblastoma. Berkel and Bako (1992) found that elevated incidences of Nephroblastoma occurred in low selenium regions. The exposure to lead and cadmium has been associated with the increased risk of developing Nephroblastoma in two studies (Hayes, 1997; Kantor *et al.*, 1979). It is not only the exposure to lead and cadmium that seems to be having an effect on the development of Nephroblastoma. The exposure to hydrocarbons and boron has also been associated with increased incidences of Nephroblastoma (Wilkins & Sinks, 1984). The possible association between hydrocarbons and the development of Nephroblastoma still needs further investigation however as stated by the work of Bosetti *et al.* (2006), who found that the results of various other studies were not consistent enough to prove that there is a significant relationship between hydrocarbon exposure and the development of Nephroblastoma.

Maternal hypertension (high blood pressure) during pregnancy was linked to the increased occurrence of Nephroblastoma in a study conducted by Lindblad *et al.* (1996). Other environmental factors that have been linked to Nephroblastoma include, radiation (Hicks *et al.*, 1984); tobacco smoking (Montesano & Hall, 2001); and maternal consumption of coffee and tea during pregnancy (Bunin *et al.*, 1987; LeMasters & Bove, 1980; Olshan *et al.*, 1993; Schuz *et al.*, 2001b). In a study

conducted by Bidoli *et al* (1993) it was found that woman living in locations above 200 metres of sea level seemed to be protected against Nephroblastoma. Although the exact reason for this is unclear it does indicate that elevation might have an influence on the development of Nephroblastoma. High birth weight has also been associated with increased incidences of Nephroblastoma in children in various studies (Daling *et al.*, 1984; Heuch *et al.*, 1996; Leisenring *et al.*, 1994; Schuz *et al.*, 2001b; Smulevich *et al.*, 1999; Yeazel *et al.*, 1997; Chu *et al.*, 2010; Rangel *et al.*, 2010).

## **2.8 Osteosarcoma**

Osteosarcoma is the most common bone cancer in children. The age of diagnosis is around 15. Boys and girls are equally likely to develop this tumour until the late teen years, when it occurs more often in boys. Osteosarcoma is also common in people over age 60. Osteosarcoma tends to occur in the bones of the:

- Shin (near the knee)
- Thigh (near the knee)
- Upper arm (near the shoulder)

Osteosarcoma can develop in any bone, but occurs most commonly in large bones and in the area of bone with the fastest growth rate (Link & Eilber, 1997; Medline Plus, 2013c).

### **2.8.1 Environmental and other possible carcinogens linked to Osteosarcoma.**

There seems to be a variation in the incidence of Osteosarcoma internationally with no obvious pattern (Parkin *et al.*,1998). Incidences seem to be similar across the Americas, Australia, and New Zealand, with somewhat lower rates in Asia. The rates in Europe range from the very low in Slovakia to the highest of any country, in Portugal. The incidence rates of Osteosarcoma were lower in 1975–1978 than in later years, for unknown reasons (Ries *et al.*, 1999). Treatment with radiation or

alkylating agents for a previous childhood cancer is known to increase the risk of Osteosarcoma (Hawkins *et al.*, 1996; Newton *et al.*, 1991; Tucker *et al.*, 1987; The Ohio State University Wexner Medical Centre, 2013; American Cancer Society, 2013c).

The exposure to radium has been found to be associated with the development of Osteosarcoma (Guse *et al.*, 2002; American Cancer Society, 2013c; US EPA, 2013). A study conducted by Polednak (1978) found that workers who applied radium containing paint to watch faces experienced an increased risk for developing Osteosarcoma. According to Finkelstein and Kreiger (1996) it is difficult to calculate whether the levels of radium in the drinking water of some locations are high enough to cause an increased risk to children or adults to develop Osteosarcoma. The exposure to pesticides again seems to have an effect on the development of Osteosarcoma. Single studies have found associations between Osteosarcoma and parental employment in any type of farming, and exposure to herbicides and pesticides (Krieger, 1992; Schwartzbaum *et al.*, 1991; Carozza *et al.*, 2008; Merletti *et al.*, 2006). Associations with paternal employment in agriculture have also been observed in other studies (Holly *et al.*, 1992; Hum *et al.*, 1998; Valery *et al.*, 2002; Winn *et al.*, 1992).

A genetic factor that might be influencing the development of Osteosarcoma is height. Large breeds of dogs experienced high rates of Osteosarcoma (Tjalma, 1966) and because of this, researchers investigated whether greater height was associated with increased risk in humans. In two studies, Osteosarcoma patients were observed to be taller at diagnosis than controls (Fraumeni, 1967; Gelberg *et al.*, 1997). Two other studies however observed no such association (Buckley *et al.*, 1998; Operskalski *et al.*, 1987).

Some studies have suggested that fluoride intake may be linked to the development of Osteosarcoma in children and adolescents. As a result of this argument, epidemiological evidence on the relationship between fluoride exposure and Osteosarcoma has been reviewed by various scientific organizations (Liteplo *et al.*,

2002; Coleman *et al.*, 2007). Animal and human studies seem to be inconclusive and contradictory, showing a positive association in some and a negative association in others, while some showed no association (McDonagh *et al.*, 2000). In various geographical areas bone cancer incidence rates and time trends in fluoridated and non-fluoridated water have been examined by several studies (Hoover *et al.*, 1991; Hrudey *et al.*, 1990; Freni *et al.*, 1992).

In general, these studies highlight the scarcity of information regarding the relationship between fluoride and Osteosarcoma during childhood and adolescence, and some did not differentiate between other types of bone cancers and Osteosarcoma (Takahashi *et al.*, 2001; Yang *et al.*, 2000). In a more recent study the age specific incidence rates of Osteosarcoma between the Republic of Ireland (where approximately 70% of the population receives fluoridated water) and Northern Ireland (where water fluoridation is not implemented) was compared. The study however did not observe any significant difference between the two areas (Comber *et al.*, 2011). In a study conducted by Levy and Leclerc (2012) it was found that males with Osteosarcoma were significantly more likely to have been exposed to fluorinated drinking water when compared to matched controls. It is clear that the possible carcinogenic effect of fluorinated drinking water has not been conclusively proven.

## **2.9 Retinoblastoma**

Retinoblastoma is a cancer that forms in the tissues of the retina (the light-sensitive layers of nerve tissue at the back of the eye). It usually occurs in children younger than 5 years. It is caused by a mutation in a gene controlling cell division, causing cells to grow out of control and become cancerous. It may be hereditary or non-hereditary. In approximately half of the cases, this mutation develops in a child whose family has never had eye cancer.

In other cases, the mutation is present in several family members. If the mutation runs in the family, there is a 50% chance that an affected person's children will also

have the mutation. They will therefore have a high risk of developing Retinoblastoma themselves. The cancer generally affects children under the age of 6. It is most commonly diagnosed in children aged 1 to 2 years (US National Cancer Institute, 2012; Medline Plus, 2013d).

### **2.9.1 Environmental and other possible carcinogens linked to Retinoblastoma.**

Retinoblastoma is known for its strong hereditary characteristics, and yet 60% of Retinoblastoma incidences occur non-hereditary and unilateral (Bonaiti-Pellie & Briard-Guillemot, 1981). Although the molecular changes leading to Retinoblastoma are well understood, the role of exposures has rarely been studied. Some studies found an association between Retinoblastoma and certain paternal occupations and in vitro fertilisation as well as maternal use of multivitamin supplements and barrier contraception (Bunin *et al.*, 1989a; Moll *et al.*, 2003). The effects of viruses on the development of Retinoblastoma have also been studied. In a study conducted by Orjuela *et al.* (2000) it was found that Human Papilloma Virus (HPV) sequences were detected in about one-third of Retinoblastoma cases in Mexico, suggesting the possible role of HPV infection.

It is interesting to note that Retinoblastoma was found to be higher in developing countries in a few studies conducted in the 1970s. These studies however included relatively small numbers of cases and used crude population estimates (Albert *et al.*, 1974; Benezra & Chirambo, 1976; Freedman & Goldberg, 1976). The small numbers of cases used in these studies could have possibly lead to inaccurately high estimates of incidence. This trend was found again in a study conducted in the 1990s by Parkin *et al.* (1998) who found annual incidence rates of 2–7 per million per year in North America and Europe, somewhat higher rates of 6–8 per million per year in Central and South America, a wider range of rates (3–9 per million per year) in Asia with the highest in Vietnam and India, as well as generally high rates of 10–24 per million per year in Africa. This pattern supports the possibility of higher rates occurring in developing countries, as it can be seen that the highest rates are clearly in Africa with a range of rates in other developing countries. The possible influence



of an environmental factor could be indicated in these trends as it is possible that people living in developing countries might be exposed (or not exposed) to certain environmental factors that people in developed countries are not exposed to (or exposed to). The challenge is to identify these possible environmental factors.

Table 1: Summary of possible environmental and other carcinogens.

<b>Cancer Type</b>	<b>Possible Environmental Carcinogens</b>
Leukaemia	Parental exposure to pesticides High altitudes Electromagnetic fields and ionizing radiation High birth weight Living in an economically advanced country Infections such as Influenza and the Epstein-Barr Virus
Lymphoma	Epstein-Barr Virus as well as Helicobacter Pylori Virus Hot, wet climates Malarial infections High altitudes Parental exposure to pesticides Exposure to Trichloroethylene Living in an economically advanced country
Brain Tumours	Parental exposure to pesticides Farm animal exposure Wood preservative exposures Vineyards and wine cultivation Lead exposure Mercury exposure Electromagnetic fields High birth weight Asbestos exposure
Hepatoblastoma	Inhalation of Arsenic Oxides Areas where there are low levels of selenium Hepatitis B Virus as well as Hepatitis C Virus Exposure to Trichloroethylene Exposure to Phthalate Preconceptional tobacco smoking Obesity Preconceptional alcohol consumption
Kaposi Sarcoma	Areas that consists of swamps and wetlands Areas where malaria is endemic High altitude Areas know to have fertile volcanic clay soils Radiation exposure
Neuroblastoma	Parental exposure to pesticides Exposure to Trichloroethylene Living in an industrialised country High birth weight
Nephroblastoma	Parental exposure to pesticides Parental employment as welder or mechanic Exposure to Trichloroethylene Living in an industrialised country Inhalation of Arsenic Oxides Areas where there are low levels of selenium Lead exposure Cadmium exposure Hydrocarbon exposure Maternal hypertention during pregnancy Radiation exposure Maternal tobacco smoking Maternal consumption of coffee or tea during pregnancy High birth weight
Osteosarcoma	Treatment of previous cancers with radiation Radium exposure Pesticide exposure Flouride intake
Retinoblastoma	In vitro fertilisation Maternal use of multivitamin supplements Human Papilloma Virus Living an a developing country

## CHAPTER 3: MATERIALS AND METHODS

### 3.1 Data quality

The data used in this study cannot be considered completely accurate. According to Stefan (2012), the custodian of the data, the amount of new cancer cases for children under the age of 15 diagnosed in South Africa is much less than is expected when compared to the number of cases diagnosed in developed first world countries. According to Stefan (2012) the global trend is to expect approximately 140 to 150 new cancer cases for every one million children under the age of 15. It is therefore expected that 2000 to 2500 new cancer cases should be reported in South Africa each year for children under the age of 15. The reality in South Africa is that only about 500 new cancer cases are reported or diagnosed each year for children under the age of 15, which leads to the conclusion that cancer is significantly underreported in South Africa. The suspicion that cancer is underreported in South Africa is also shared by Dr Georgia Demetriou, who is the spokesperson for the SA Oncology Consortium and an oncologist at the Donald Gordon Oncology Centre (Medical Chronicle, 2011). According to Dr Demetriou there is a very good indication that cancer is probably underreported in South Africa. This view is also shared by Dr Magda Heunis, who is the head of radiation oncology at the University of Stellenbosch. According to Dr Heunis there is a lack of necessary infrastructure to record accurate cancer statistics in South Africa (Medical Chronicle, 2011).

Dr Carl Albrecht (CANSA's head of research) also supports the statement that the data contained in the South African Paediatric Cancer Registry is probably significantly underreported, especially in terms of the black population group of South Africa (Albrecht, 2006). According to Dr Demetriou, a possible reason for the underreported data in the South African Paediatric Tumour Registry is the fact that South Africa's registry is pathology based, resulting in patients being missed if they demise without a diagnosis (Medical Chronicle, 2011). In order to understand why cancer is underreported it is first necessary to determine where most of these cases are expected to be diagnosed.

According to the General Household Survey that was conducted by Statistics South Africa in 2010 (where 25653 households were interviewed across South Africa), 60.8 % of the households interviewed indicated that they go to public clinics when they have health related concerns, while 24.3% went to private doctors, and 9.4% went to public hospitals (Statistics South Africa, 2011). It is therefore reasonable to expect that most cancer incidences are first diagnosed in public clinics in South Africa. It is therefore worthwhile to take a closer look at the country's public clinics in order to try and understand the possible reasons for South Africa's underreported cancer problem.

The chaotic state of South Africa's public clinics may be contributing to the underreporting of cancer incidences. According to the Consolidated Report of Inspections of Primary Health Care Delivery Sites, conducted by the Public Service Commission in 2010 (where various public clinics across South Africa were inspected), the majority of public clinics in South Africa were under staffed and the staff at most clinics were overworked (Public Service Commission, 2010). This report also indicated that many of the clinics visited did not have sufficient medical equipment. The first concern that arises here is that there is overworked medical staff conducting diagnoses with questionable medical equipment. With the exception of most clinics visited in Kwa-Zulu Natal, Gauteng and the Western Cape, most of the other public clinics visited had insufficient medical equipment. In the case of the Taylors Clinic that was visited for example, it was observed that the clinic had a serious shortage of nurses, which together with non-existent administrative support, resulted in each nurse being involved in various duties at once such as conducting registration, testing patients for blood pressure and conducting consultations.

It is therefore reasonable to suspect that there could be mistakes made when patients are diagnosed. Some cancer incidences could therefore be missed as they could be falsely diagnosed as something else. The other concern that was identified when the Consolidated Report of Inspections of Primary Health Care Delivery Sites (Public Service Commission, 2010) was studied was that most of the clinics visited did not have sufficient computers. The clinics that did have computers did not use them because the staff had no computer skills. The result of this is that most patient records are paper based and manually stored in filing cabinets. The storage of

patient records in filing cabinets raises concern when the chaotic unorganised state of the clinics visited in this report is considered.

To illustrate the concern, consider the state of the following clinics that were visited in the Consolidated Report of Inspections of Primary Health Care Delivery Sites (Public Service Commission, 2010). At one clinic in Gauteng it was found that patient files; equipment; vaccine and medical fridges were stored in unlocked steel cabinets in the waiting room. At another clinic in Gauteng it was observed that the clinic was very small and in a very untidy state. A clinic in Limpopo lacked office space and was in a state of disrepair. This clinic had no filing cabinets or shelves, as a result patient files with crucial information were placed in trolleys. Another clinic in Limpopo also lacked office space. Most of the clinics visited in the Northern Cape did not have sufficient computer equipment and were coping with the bare minimum. At a clinic in Northwest only two consulting rooms were observed. One of these rooms was used as a nursing sister's office, while the other room was used as the counsellor's office. In another clinic in Northwest, counselling was conducted in the kitchen where the health promoter also used the same kitchen to write reports. It was observed at other clinics that the reception areas were used to store patient files due to a lack of space. It was observed that one clinic did not have lockable cupboards which could result in files being misplaced or easily lost. In Northwest five out of the six clinics visited had no functional computers, fax machines, or photo copier machines and patient records were handled manually.

The abovementioned examples clearly indicate the unorganised state that seems to be the trend in most public clinics in South Africa. The lack of office space is also highlighted by these examples. It is reasonable to suspect that patient files could be lost or misplaced in these unorganised filing cabinets that seem to be stored in any space available. The fact that most of these filing cabinets are unlocked also increases the possibility of these records being lost or misplaced. A possible reason for the underreported nature of cancer in South Africa could therefore be because these records are lost due to the chaotic state of the clinics that diagnose and store the records. The majority of these records are not computer-based and once a file is lost it is lost indefinitely.

The underreported nature of cancer in South Africa does not seem to be because of patients going to traditional healers, spiritual healers or pharmacies, where it is likely that formal patient records won't be kept. The General Household Survey conducted by Statistics South Africa in 2010 confirmed that this does not seem to be the case since only 0.2% of the 25653 households interviewed indicated that they go to traditional healers when they have health related concerns. Another 0.2% indicated that they go to spiritual healers, while only 0.3% indicated that they go to pharmacies (Statistics South Africa, 2011).

A limitation within the data according to Stefan (2012) is that a significant amount of people give false information with regard to their addresses in order to gain access to specific medical centres. This casts suspicion over the geographic information linked to each cancer case. According to the General Household Survey that was conducted by Statistics South Africa in 2010 (Statistics South Africa, 2011), as well as the Consolidated Report on Inspections of Primary Health Care Delivery Sites that was conducted by the Public Service Commission in 2010 (Public Service Commission, 2010), the primary concerns of people visiting public clinics are the long waiting periods; the unavailability of medicine; clinic staff being rude and in some cases the unhygienic state of some of the clinics. The Consolidated Report on Inspections of Primary Health Care Delivery Sites, for example, found that at a clinic in Northwest, citizens reported that it is a known practice that one must arrive at the clinic at 5 o'clock in the morning if one has realistic hopes of being helped that day. The report also found that patients complained of staff being rude and showing them disrespect (Public Service Commission, 2010).

These findings support the possibility that patients could be lying about their addresses to gain access to specific clinics that have better service than the clinics in their area. This could have serious effects on the quality of the geographic data that was used to plot the cancer cases. Although it has been indicated that this might be what is happening, one must ask the question of how serious of a problem it really is and whether it has a significant impact on the data.

According to the patients that were interviewed in both the General Household Survey, as well as the Consolidated Report on Inspections of Primary Health Care Delivery Sites, the distance they need to travel to get to their nearest clinic is not a problem (Public Service Commission, 2010; Statistics South Africa, 2011). The General Household Survey found that 91.1 % of the 25653 households interviewed indicated that they go to their nearest public clinic when they have health related concerns. This finding implies that most people visiting clinics are honest about where they live. This however needs to be verified and warrants further investigation.

The data used in this study can therefore not be considered completely reliable and although there are justified concerns about the quality of the data, it is still the best available data to use at this time. As stated before, the main purpose of this study was to develop and validate a methodology for creating cancer maps using GIS software. The result of whether the methodology was successful should therefore be viewed as more important than the analysis results itself. These results should however not be completely rejected as they may indicate factors worth investigating. The maps that were created from the data might indicate actual spatial clusters of cancers or it might indicate false clusters of cancer. The quality of the data used makes it impossible to say for sure, but the process used to get to those maps can be thoroughly verified, and the reliability and validity of that process can be determined.

### **3.2 The software used in this study.**

A Geographic Information System (GIS) is a computer system that is capable of storing virtually any information that can be found on a paper map. It can however be of much greater use than a traditional map. A GIS can display maps on a computer screen, where it can provide detailed information about their features, such as roads, buildings or streams. A GIS can also search and analyse these map features and their attributes in ways that paper maps cannot (George & Korte, 2001). The difference between a paper map and a GIS map is that the latter displays intelligence (Kennedy, 2013). A spatial query can be conducted in a GIS map for example, where the GIS can be asked to show only certain features with certain

attributes. A GIS can be defined as a computer based technology and methodology for collecting, managing, analysing, modelling and representing geographic data for a wide range of applications.

A GIS consists of computer software and hardware and can be viewed in the following ways according to Bernhardsen (2002):

- A data processing system designed for map production and visualization.
- A data analysis system for examining conflicts over plans or optimizing the design of transport systems.
- An information system to be used in responding to queries about land ownership or soil type.
- A management system to support the operations of a utility company, helping it to maintain its distribution network of pipes or cables.
- A planning system to aid the design of road systems, excavations, or forest harvest operations.
- An electronic navigation system for land or sea transport.

A GIS can consist of vector data models or raster data models. A vector data model is a representation of the world using points, lines or polygons. Vector models are useful for objects that have discrete boundaries, such as a country's border, a land parcel, or a street (Esri, 2013a). A raster model is a representation of the world as a surface divided into a regular grid of cells. Raster models are useful for storing data that varies continuously such as in an aerial photograph, a satellite image, a surface of chemical concentrations, or an elevation surface (Esri, 2013b).

The basic difference between vector models and raster models according to Gregory and Eil (2007) is that vector models store information about discrete objects and gives them precise locations which, in the case of polygons will have unambiguous boundaries. The raster model stores a continuous representation of the Earth's surface which does not have such precise locations, nor requires unambiguous boundaries. According to Gregory and Eil (2007) vector data models are more

suitable for human features, while raster data models are more suitable for physical or environmental features.

According to Buckey (2013) the following can be seen as advantages and disadvantages of vector data models:

#### Advantages

- Data can be represented in its original resolution.
- Graphic input is usually more aesthetically pleasing (traditional cartographic representations).
- Accurate geographic location in the data is maintained.

#### Disadvantages

- The location of each vertex needs to be stored explicitly.
- For effective analysis vector data must be converted into a topological structure, which is processing intensive and usually requires extensive data cleaning.
- Algorithms for manipulative and analysis functions are complex and may be processing intensive, which could limit the functionality for large datasets.

According to Buckey (2013) the following can be seen as advantages and disadvantages of raster data models:

#### Advantages

- The geographic location of each cell is implied by its position in the cell matrix.
- Due to the nature of the data storage technique, data analysis is usually easy to program and quick to perform.
- The inherent nature of raster maps is ideally suited for mathematical modelling and quantitative analysis.
- Grid-cell systems are very compatible with raster-based output devices such as electrostatic plotters or graphic terminals.



## Disadvantages

- The cell size determines the resolution at which the data is represented.
- It is difficult to adequately represent linear features depending on the cell resolution and it is therefore accordingly difficult to establish network linkages.
- Processing of associated attribute data may be cumbersome if large amounts of data exist.
- Most output maps from grid-cell systems do not conform to high-quality cartographic needs.

According to Florenza (2012) and GIS and Science (2009), a GIS has the following additional benefits:

- It improves the decisions made by government officials.
- It improves the allocation of resources and planning.
- It improves transparency for citizen engagement.
- It improved communications during a crisis.
- It can cope with large amounts of data.
- It can cover a large study area.
- It can conveniently select any sub-study area.
- It results in cost savings due to greater efficiency.

It is clear from the above that a GIS has many advantages and perhaps its greatest advantage is that it is not limited to a specific sector. According to George and Korte (2001) and Longley *et al.* (2004), GIS technology can be used in a wide range of applications.

During this dissertation the following GIS applications were used:

***ArcEditor 10 (data mapping)***

ArcEditor is used for the editing of spatial data. It provides tools for the creation of maps. It was chosen because of its capability to create, edit and present maps.

***ArcCatalog 10 (data management and storing)***

ArcCatalog is used to organise and manage various types of geographic data. It was chosen because of its data storing and management capabilities.

***ArcGIS Spatial Analyst (spatial analysis)***

ArcGIS Spatial Analyst provides tools for helping one understand spatial relationships within one's data. It also helps one to perform statistical calculations by using geographic data. It was chosen because of its ability to perform a statistical analyses based on geographic data.

In addition to the GIS software used, Google Earth was also used to find the latitude and longitude coordinates of each of the cancer cases. Google Earth is a virtual globe that can be used as a map. It allows one to view satellite imagery of any part of the world. It can be used to search for any location on planet earth. A satellite image as well as the coordinates of that area can then be seen.

**3.3 Data preparation**

The cancer data used in this dissertation was contained in Microsoft Excel format and the geographic data linked to every individual cancer case was used to identify the coordinates (geographic location) of each case. Two new fields were added to

the excel document for both the latitude and longitude coordinates of each case. The coordinates were determined using Google Earth as well as the street address or town name linked to each cancer case. Of the 7177 cancer cases available within the data for the chosen 14 cancer types, approximately 900 cases could not be plotted because they contained insufficient geographic information. Within the remaining 6268 cases, approximately 2000 had more specific geographic information linked to them, such as the physical street address of the cancer case. The remaining portion of the 6268 cases contained either the name of the town in which the patient lived or the name of the suburb in which the patient lived or only the postal code of the city or town in which the patient lived.

The first 2000 cases could be plotted very accurately because of the specific geographic information linked to them but the remaining 4268 cases had to be plotted according to the coordinates of the town in which they lived. The town in which they lived was determined by either the name of the town, the suburb or the postal code linked to each cancer case. The latitude and longitude coordinates of the centre of each town or suburb was then linked to each corresponding cancer case. The coordinates were then used to plot points that represent the cancer cases as can be seen in figure 3.4.1 (which represents a portion of the Acute Lymphoblastic Leukaemia cases that were plotted). The picture of the plotted points tempts one in to identifying where clusters are right away, but one cannot statistically conclude that there are clusters by just looking at the dots. It is also important to realise that where there are more people, such as the case in Gauteng for example, there will be more cancer cases or dots because there are more people living in that area.

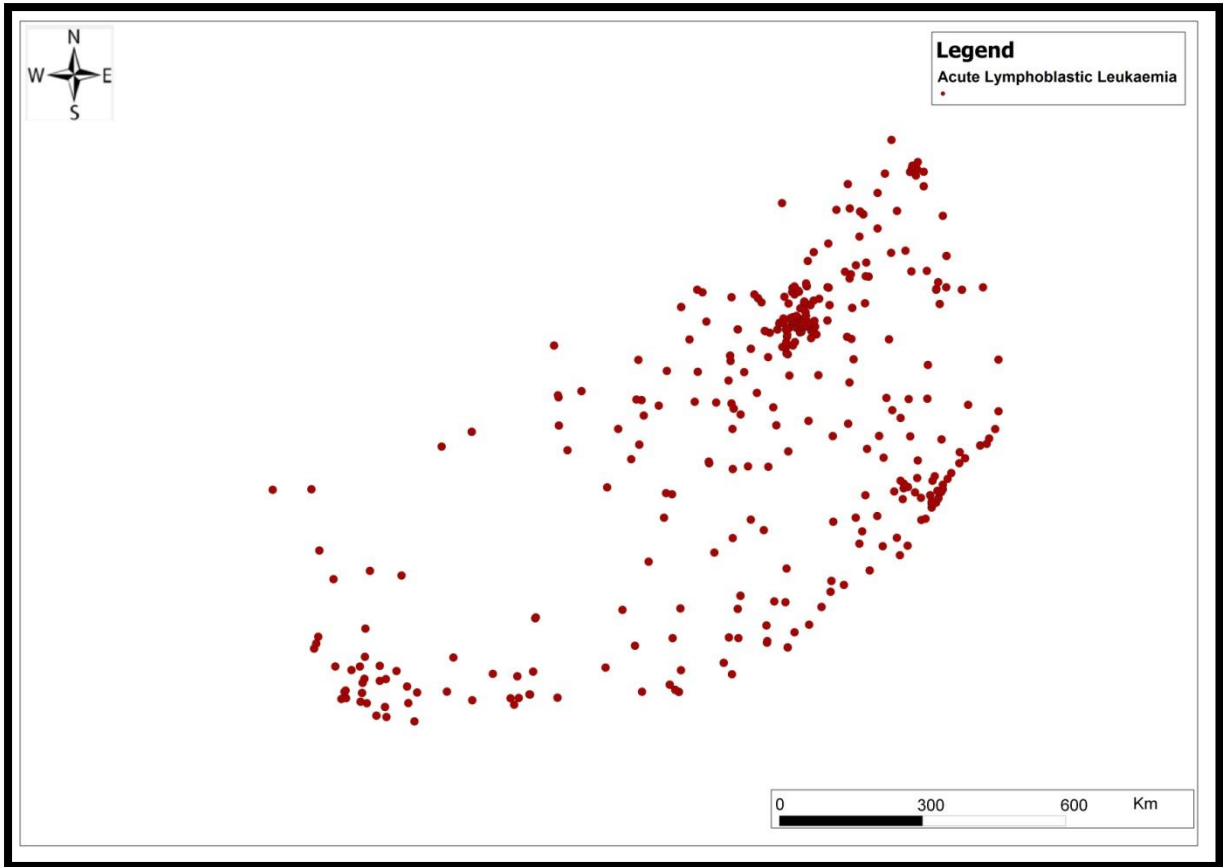


Figure 3.4.1: A portion of Acute Lymphoblastic Leukaemia cases plotted as dots.

To compensate for this, it was decided to look at the percentage of children under the age of 15 who had cancer in various areas, therefore creating the need for census data and area boundaries. Statistics South Africa provided the census data for 1990, 1996, and 2001 as well as the community survey data for 2007. They also provided a shapefile containing the 2001 census enumeration area boundaries as can be seen in figure 3.4.2. Enumeration areas are the smallest geographical units (pieces of land) into which the country can be divided for census or survey enumeration. It is an area small enough for one census fieldworker to enumerate alone in the allocated census period and it typically contains between 100 and 250 households (Statistics South Africa, 2001). Each one of the polygons shown in figure 3.4.2 contained population statistics for that specific polygon. It was decided to spatially join the plotted cancer dots to the 2001 enumeration area polygons, using the Spatial Join Tool found in ArcGIS, so that the computer could register how many dots fell in each polygon. It would then show this number in the join count field of the attribute table. This field together with the field containing the total population under the age of 15 could then be used to calculate the percentage of children under the

age of 15 who had cancer for each individual polygon. This percentage could then be used in the analysis to determine where there might be clusters of a specific cancer.

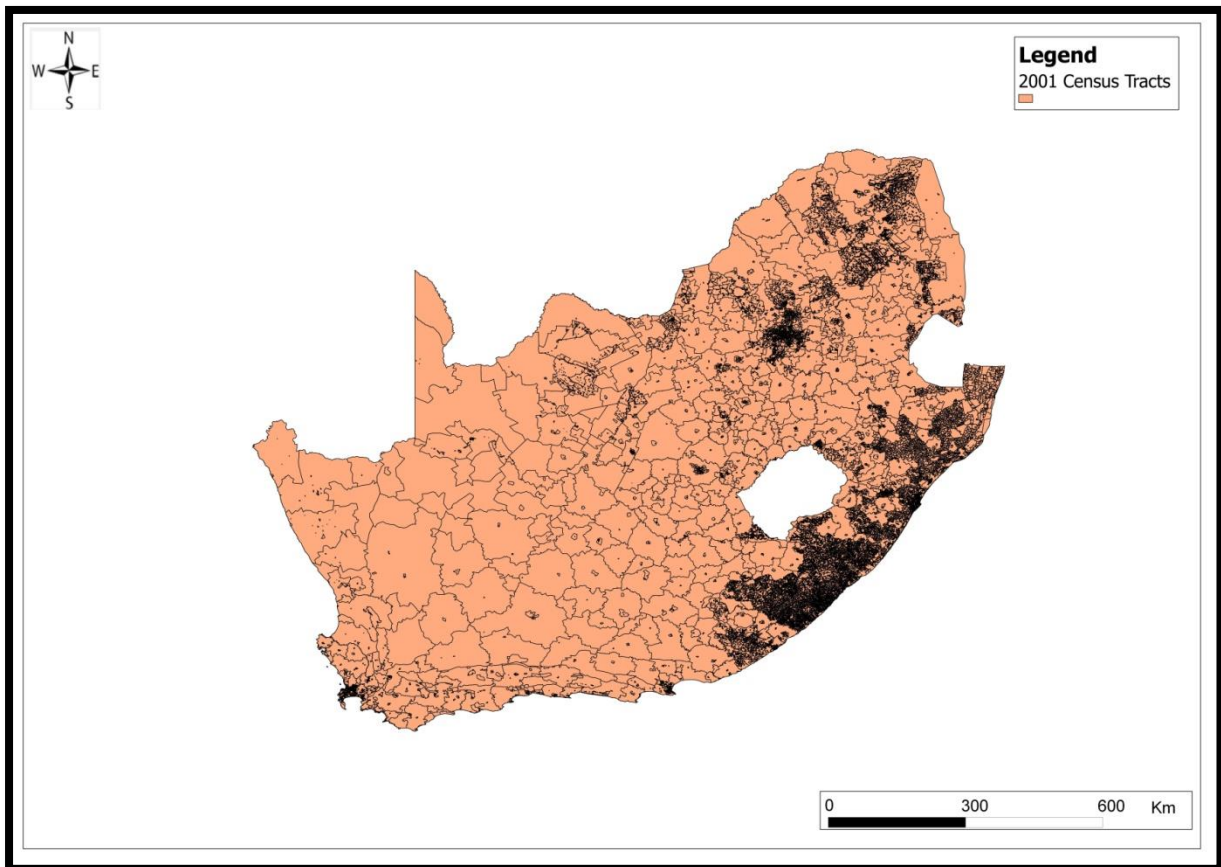


Figure 3.4.2: 2001 Census enumeration areas as provided by Statistics South Africa.

This approach was hampered with various limitations. The first thing to consider was that the cancer data was from 1992 to 2008, so using polygons that contain only the 2001 census population statistics would be a mistake since the total population under the age of 15 for a specific polygon would not have stayed the same for this entire period. Calculating a cancer ratio for cancer incidences from 1992 to 2008 based on only the population statistics of 2001 would result in a suspect ratio. It was therefore decided to use the average population under the age of 15 for each polygon for the entire period of 1992 to 2007. To do this the census data needed to be converted to identical polygon boundaries so the population statistics from the various census counts could be compared, since the enumeration area boundaries did not stay the same for the entire period of 1992 to 2007. The 1996 enumeration

boundaries for example weren't the same as the 2001 enumeration boundaries (Dube, 2005). The ideal would then be to have the census statistics from each census count (1990, 1996, 2001 and 2007) converted to the same set of boundaries (the 2001 census enumeration area boundaries as seen in figure 3.4.2). The 1990, 1996, 2001 and 2007 census data could then be linked to identical boundaries so that each polygon had census data for 1990, 1996, 2001 and 2007. The 1990, 1996, 2001 and 2007 census data for each polygon could then be used to calculate the average population under the age of 15 for each polygon. Once the average population was calculated the number of cancer cases plotted in each polygon could then be determined as a percentage of the average population.

This process was however hampered by the lack of specific geographic information within the cancer data. Only the 2000 cases that had specific geographic information linked to them could be used at this scale (2001 enumeration boundaries), since they could be plotted at their exact location. The remaining 4268 cases could only be plotted according to a much coarser resolution, because they only contained information about the postal code, town or suburb in which the patient lived. These dots could not be plotted at the exact location where the patient lived and this caused a problem since the enumeration areas become smaller in densely populated areas. In a city like Potchefstroom for example, the entire city is divided into 18 enumeration areas. A cancer case in Potchefstroom with specific geographic information linked to it, such as the physical address could be correctly plotted in one of those 18 enumeration areas. The problem arises when the geographic information linked to a cancer case in Potchefstroom is limited to only "Potchefstroom". In these situations the latitude and longitude coordinates of the centre of Potchefstroom were linked to those specific cases. The lack of specific geographic information in the cancer data therefore limits the scale on which the analysis can be conducted. If these dots were to be linked to the 2001 census enumeration areas, the dots would end up in the middle of Potchefstroom which would cause the computer to register them in an enumeration area that is in the centre of the city. In reality those patients might live in an enumeration area at the edge of city.

It was therefore decided to use the local municipal boundaries instead, because these boundaries were bigger. These boundaries insured that each cancer case

could be plotted in the correct polygon. In the case of Potchefstroom, the entire city fell within the municipal boundary. The fact that there were a large number of cancer cases that only had postal codes as geographic information was also considered when the local municipal boundaries were chosen. It is reasonable to expect that many farmers who live on the outskirts of a city or town like Potchefstroom might have their mail posted to the town post office and therefore would have Potchefstroom's postal code but might actually live outside the city.

The data from the 1996, 2001 and 2007 census counts were converted to the local municipal boundaries as they were in 2005 (figure 3.4.4) with the help of Statistics South Africa. Unfortunately Statistics South Africa could not convert the census data prior to 1996 because the enumeration areas used from 1996 onwards did not exist prior to 1994 (Dube, 2005). The census data after 2007 could also not be converted because it was not yet available. The data from the 1996, 2001 and 2007 census counts were therefore used to determine the average population under the age of 15 for each local municipal boundary as they were in 2005. Each 2005 local municipal boundary therefore contained census data for 1996, 2001 and 2007. The combined census data of each polygon was then used to determine the average population of children under the age of 15 (for the period 1996 to 2007) for each of these polygons. The calculated average population was then used together with the plotted cancer cases to determine a cancer ratio for each of the 2005 local municipal polygons (for the period of 1996 to 2007).

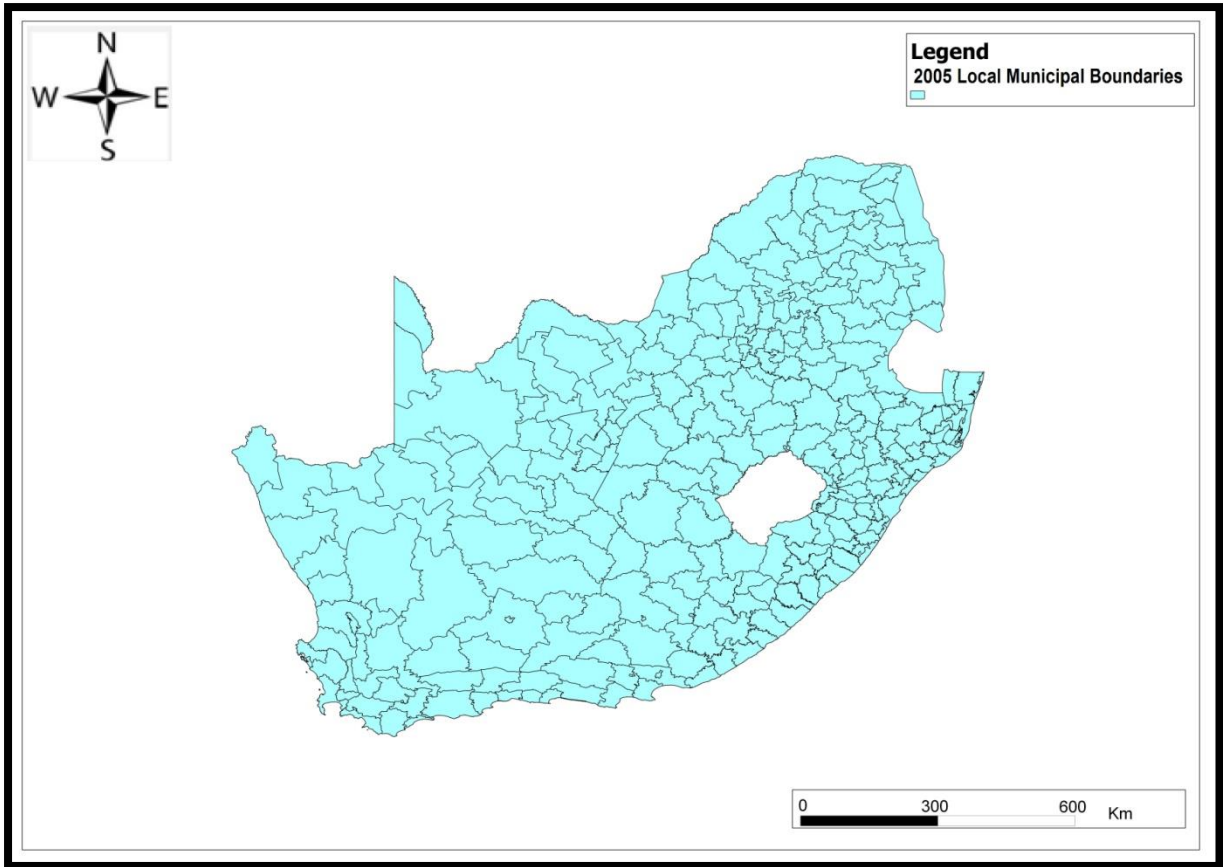


Figure 3.4.4: 2005 Local municipal boundaries

This limited the quantity of cancer cases that could be used in the analysis, since the cases prior to 1996 and after 2007 had to be left out of the analysis to ensure that the calculated ratios were specific to the time period of 1996 to 2007.

The following table (Table 2) shows the number of cancer cases that could be plotted for the time period of 1992 to 2008 (overall 92-08). It also shows the number of cases that could be plotted that fall within the time period of 1996 to 2007 (96-07). Lastly it also shows the big picture of how many of the total number of cases that could be plotted was actually used in this analysis (big picture). In other words, up to approximately 87% of the total available cancer cases (for the entire period of 1992 to 2008) were plotted successfully on the local municipal scale, but needed to be used in conjunction with population statistics that were limited. The result was that only the cases that could be linked to the population statistics could be used. This resulted in only the cases between 1996 and 2007 being used. As a result only 67% of the total available cases (1992 to 2008) were used. The remaining 33% of cases



were left out either because they were diagnoses before 1996 or after 2007, or did not have sufficient geographic information linked to them. The issue of missing data in research is a common phenomenon and according to Greenland and Finkle (1995) and Raghunathan (2004) the most common approach to missing data is to simply omit the missing data and do the analysis with what remains. Omitting missing data has the advantage of ensuring simplicity and comparability across an analysis but it also reduces the statistical power of the analysis because it doesn't use all the available data according to Enders (2010) and Little and Rubin (2002). The limitations that were found within the data used in this study significantly reduced the statistical power of the analysis and should be kept in mind when viewing the final results. The omission of the unusable data was however a necessity because if the effects of the missing data were not taken into account the results of the study would be biased and misleading (Bennet, 2001).

Table 2: Number of cancer cases that could be plotted.

<p><b>Acute Lymphoblastic Leukaemia</b> Overall 92-08 (1461/1674) 87% 96-07 (1112/1251) 89% Big picture (1112/1674) 66%</p>	<p><b>Hodgkin's Lymphoma</b> Overall 92-08 (413/467) 88% 96-07 (307/337) 91% Big picture (307/467) 65%</p>	<p><b>Osteosarcoma</b> Overall 92-08 (258/299) 86% 96-07 (207/233) 89% Big picture (207/299) 69%</p>
<p><b>Acute Myeloid Leukaemia</b> Overall 92-08 (485/556) 87% 96-07 (382/423) 90% Big picture (382/556) 68%</p>	<p><b>Kaposi's Sarcoma</b> Overall 92-08 (144/164) 88% 96-07 (188/138) 85% Big picture (118/164) 71%</p>	<p><b>Retinoblastoma</b> Overall 92-08 (581/689) 84% 96-07 (446/526) 85% Big picture (446/689) 64%</p>
<p><b>Astrocytoma</b> Overall 92-08 (252/285) 88% 96-07 (201/222) 90% Big picture (201/285) 70%</p>	<p><b>Medulloblastoma</b> Overall 92-08 (193/218) 88% 96-07 (144/156) 92% Big picture (144/218) 66%</p>	
<p><b>Burkitt's Lymphoma</b> Overall 92-08 (306/344) 88% 96-07 (241/267) 90% Big picture (241/344) 70%</p>	<p><b>Nephroblastoma</b> Overall 92-08 (1051/1214) 86% 96-07 (798/901) 88% Big picture (798/1214) 65%</p>	
<p><b>Glioma</b> Overall 92-08 (197/220) 89% 96-07 (167/187) 89% Big Picture (167/220) 75%</p>	<p><b>Neuroblastoma</b> Overall 92-08 (520/593) 87% 96-07 (372/412) 90% Big picture (372/593) 62%</p>	
<p><b>Hepatoblastoma</b> Overall 92-08 (118/132) 89% 96-07 (88/96) 92% Big picture (88/132) 66%</p>	<p><b>Non-Hodgkin's Lymphoma</b> Overall 92-08 (289/322) 90% 96-07 (218/236) 92% Big picture (218/322) 67%</p>	

### 3.4 Data Analysis

GIS software is a very powerful tool for visually representing data. The symbology of the maps one wishes to represent can be manipulated at the touch of a button. The symbology can be represented as categories, quantities, or charts. Data represented in quantities can for example be represented as equal intervals, quintiles, natural breaks, geometric intervals or standard deviations (Esri, 2012a). The problem with using ordinary GIS symbology to symbolise data is that the end result would be extremely subjective since each GIS user might favour a different style of symbology (Esri, 2012a). To illustrate the abovementioned points consider the case of Acute Lymphoblastic Leukaemia. In figure 3.5.1 the percentage (ratio) of children with Acute Lymphoblastic Leukaemia is represented as quantities with graduated colours, five classes and a natural breaks classification. In figure 3.5.2 the percentage of children with Acute Lymphoblastic Leukaemia is represented as quantities with graduated colours, five classes and an equal interval classification. When these two figures are compared, it is clear that there is a significant difference in the data presentation because of the two different types of classifications used. From this example it is clear that using the normal GIS symbology settings could be very subjective and therefore not trustworthy.

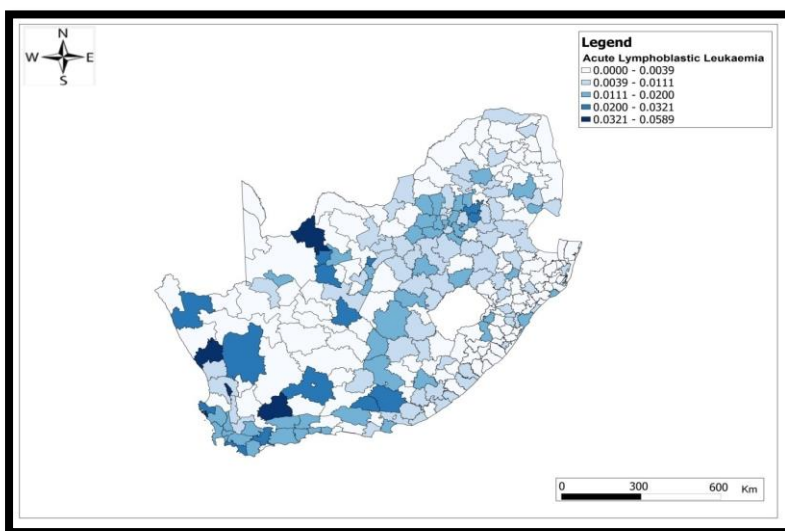


Figure 3.5.1: Percentage (ratio) of Acute Lymphoblastic Leukaemia presented as quantities with graduated colours, 5 classes and a natural breaks classification.

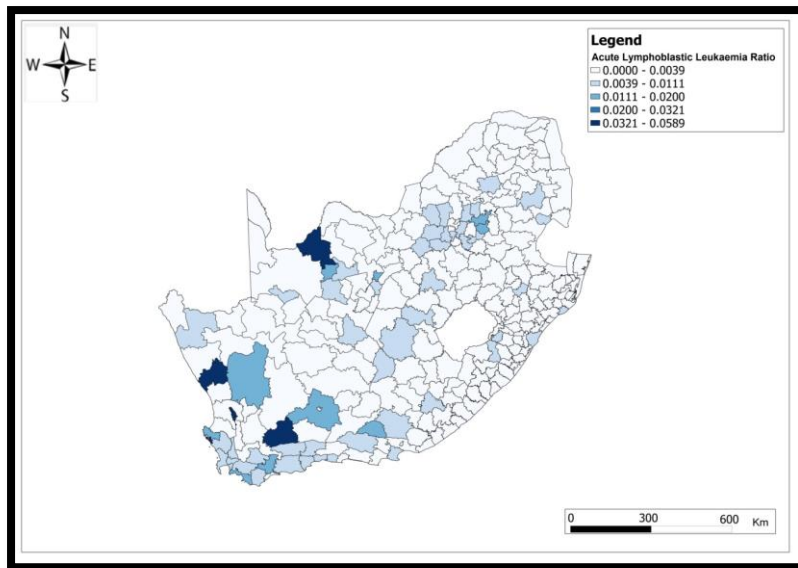


Figure 3.5.2: The percentage (ratio) of Acute Lymphoblastic Leukaemia presented as quantities with graduated colours, 5 classes and an equal interval classification

To overcome this problem the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool was used to statistically determine where possible clusters of cancer are and to present the results in a statistically unbiased objective manner.

The tool uses a G-statistic, or more specifically a  $G_i^*$  statistic. A G-statistic tells one whether hotspots (clusters of high values) or cold spots (clusters of low values) exist within the study area (Getis, 1991). The G-statistic takes each polygon in the study area and gives it a turn to be the target polygon. For each target polygon a radius is drawn around it based on a distance the analyst specifies. All the neighbouring polygons that fall within this distance are considered neighbours of the target polygon. To calculate the  $G_i^*$  value the GIS sums the ratios of all the neighbours that are located within this distance and divides this value by the sum of all the ratios of all the polygons in the entire study area (Lee & Wong, 2001). This value is then called the calculated  $G_i^*$  value and since a G-statistic is a relative measure, one does not really know what a large or small value means unless one compares it to the expected G-statistic for a random distribution at the distance the analyst specified.

Only then can it be determined whether the distribution of values (cancer ratios) is significantly different than a random distribution. The expected G-statistic at a given distance is what the value of G would be if there was no particular concentration of high or low values (ratios) in the study area (Getis, 1991). The expected  $G_i^*$  value for a random distribution is equal to the sum of all the spatial weights for a given distance, divided by the number of polygons in the study area, minus 1 (Lee & Wong, 2001). Spatial weights will be explained later in this chapter.

A group of polygons with high calculated  $G_i^*$  values indicates a cluster or concentration of polygons with high attribute values (high ratios) and a group of polygons with low calculated  $G_i^*$  values indicates a cold spot. A calculated  $G_i^*$  value near 0 indicates that there is no concentration of either high or low values surrounding the target polygon and this can occur when the surrounding values are near the mean, or when the target polygon is surrounded by a mix of high and low values (Getis & Ord, 1996). The GIS determines the statistical significance of the calculated  $G_i^*$  value by comparing it to the expected  $G_i^*$  value that would exist if there was a random distribution in the data where there would be no concentration of either high or low values (ratios). The GIS calculates the statistical significance of the calculated  $G_i^*$  value by calculating a Z-score. The GIS calculates the Z-score by subtracting the expected  $G_i^*$  value for the polygon (given a random distribution) from the calculated  $G_i^*$  value for the polygon. The difference is then divided by the square root of the variance for all the polygons in the entire study area (Lee & Wong, 2001).

The GIS calculates a Z-score for each polygon at the specified distance. As with the  $G_i^*$  value itself, a high Z-score for a polygon indicates that its neighbours have high attribute values, and vice versa. A Z-score near 0 indicates no apparent concentration of similar values. To determine whether a Z-score is statistically significant, it is compared to the range of values for a given confidence level, for example, at a confidence level of 95%. A Z-score would have to be less than -1.96 or greater than 1.96 to be statistically significant (Lee & Wong, 2001).

A Z-score indicates how certain one can be that one's observed pattern is not simply due to random chance (Longley & Batty, 2003). The further away a Z-score is from 0, the more one's results deviate from the mean and the smaller is the probability that the spatial pattern one is looking at is due to random chance. As can be seen in figure 3.5.3, if the Z-score is zero (indicated by black arrow), then the probability that the spatial pattern is due to chance is at its greatest (indicated by yellow bracket), but if the Z-score is 1 (indicated by red arrow), then the probability that the pattern is due to random chance decreases (indicated by green bracket), and finally if the Z-score is 2 (indicated with orange arrow), then the probability that the pattern one is looking at is due to random chance will decrease even further (Indicated with the red bracket). To conclude then, the further away the Z-score is from 0, the smaller the probability is that the spatial pattern one is looking at is due to random chance, and the smaller the probability is, the more statistically significant the spatial pattern is, which is the desired effect (Esri, 2009).

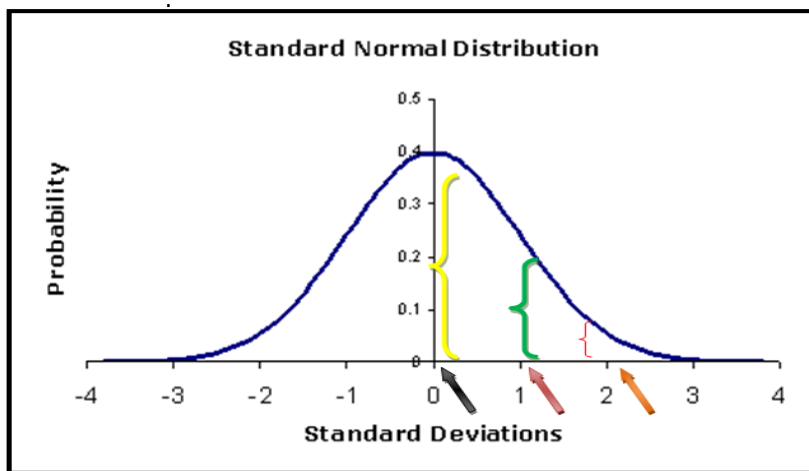


Figure 3.5.3: Standard normal distribution (Esri, 2009)

It is only human to favour a pattern or relationship one expects to see, so to maintain impartiality, one sets out to prove the opposite, the so called null hypothesis. One's initial hypothesis is called the alternative and by doing significance tests one can decide whether one should or should not reject the null hypothesis (Burt & Barber, 1996). In order to decide whether to reject the null hypothesis, one first needs to decide what the risks are that one is willing to accept for being wrong. In a study like this one, where one is trying to find hotspots of cancer, one needs to make

especially sure that the chances of identifying false hotspots are minimised since the resources for fighting cancer are always limited and needs to be allocated to the correct areas. It is therefore necessary that the risk of being wrong needs to be as small as possible.

This degree of risk, called the confidence level or significance level is expressed as a probability ranging from 0.0 to 1.0., and in order to decide whether or not to reject the null hypothesis the desired level of confidence needs to be compared with the observed level of confidence. The observed level of confidence, known as the p-value, is calculated using sample data (in this case, neighbourhoods). It is therefore important to note that the characteristics of a sample (its size and whether it is a random sample) affect its observed level of confidence (Earickson&Harlin, 1994).

The most common confidence levels for statistical tests are 0.10, 0.05 and 0.01. If a study is repeated with a confidence interval of 0.1 or 0.05, then 99/100 or 95/100 times you would get the same results if your study is designed the same with a 1% or a 5% error each time, respectively. The results that would be obtained if the study is repeated many times and on each occasion a 95% confidence interval is calculated, then 95% of these intervals would contain the true effect (Ebdon, 1985).

Each statistical tool in the GIS software has an appropriate test, and these tests provide a statistic that represents the p-value and the desired confidence level has a corresponding critical value (Lee & Wong, 2001). If the value of the test statistic exceeds the critical value, one rejects the null hypothesis. Most spatial statistics tools calculate a test statistic at the same time they calculate the initial statistic and report both. Many of the tools (including the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool,) calculate a Z-score, which is a reference measure for the standard normal distribution, with a mean of zero and standard deviation of 1 (Lee & Wong, 2001).

The critical values of -1.96 and 1.96 are standard deviations from the mean. This means that 95% of the area underneath the standard normal curve falls between plus and minus 1.96 standard deviations from the mean. The other five percent of

the area is called the rejection region. If a Z-score falls within the rejection region, there is only a 5 percent chance that one would be wrong to reject the null hypothesis (Bailey & Gatrell, 1995). In the case of the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool which uses a confidence level of 0.01, 99% of the area under the curve falls between -2.58 and 2.58 standard deviations of the mean (Lee & Wong, 2001).

These tests were developed within classic, non-spatial, statistics. Spatial data, however contradicts some of the assumptions of inferential statistics, and the assumptions the tests make, producing questionable results (Lee & Wong, 2001). Spatial data often violates one of the main assumptions of statistics, namely the independence of observation within a sample. In a hypothetical random spatial distribution, every feature or observation would have an equal probability of occurring at any given location, and the location of any given feature or observation would have no influence on any other feature or observation within the dataset. Spatial data violates this assumption for example when the selling price of a particular house, influences nearby housing values or when commercial burglaries only occur where there are businesses (Burt & Barber, 1996). The outcome of spatial statistics is therefore somewhat predetermined.

An increase in computing power is allowing researchers to explore methods that rely on computer-generated distributions derived from millions of permutations over a dataset, such as Monte Carlo simulation and Bootstrapping. These methods may eventually replace theoretical distribution models and their accompanying mathematical equations, and avoid the issues associated with statistical tests based on sampling, such as assumptions of normality and spatial independence (Lee & Wong, 2001). One never truly knows whether the null hypothesis is true or false, one decides to either reject it or not, with some level of confidence.

The only thing one can do is to try and reduce the risk that one will draw the wrong conclusion from the significance test by minimizing the likelihood that one will make an error in rejecting the null hypothesis (Bailey & Gatrell, 1995). If the null hypothesis is rejected and it is actually false, or if it is not rejected and it is actually

true, then the conclusion that is drawn is correct, but if the null hypothesis is rejected and it is actually true, or if it is not rejected and it is actually false, an error is made. Statisticians call the first a *Type 1* error and the second a *Type 2* error (Mitchell, 2005).

The risk is usually less with a *Type 1* error, for example in this case where potential environmental factors may be contributing to the development of a particular type of cancer. In this case the null hypothesis states that the cancer incidences are distributed randomly throughout the study area. If a *Type 1* error is committed the null hypothesis will be rejected, thereby falsely concluding that the cancer incidences are clustered. It will then be needed to move to a more specific level of analysis, such as examining the relationships between the cancer clusters and particular environmental factors. By doing this, the error will likely be uncovered. On the other hand, if a *Type 2* error is committed and the null hypothesis is not rejected when it should be, it will be falsely concluded that the cancer incidences occur randomly (Lee & Wong, 2001). This could lead to the analyst abandoning his/her research prematurely. The bigger risk in this study is therefore to commit a *Type 2* error.

According to Mitchel (2005) the best way to favour a *Type 1* error is to specify a less stringent confidence level, since the null hypothesis is already established in spatial statistics. If a confidence level of 0.1 is used instead of a confidence level of 0.05, it is more likely to conclude that there is no significant clustering. If it is perceived that there is no clustering (the null hypothesis is true). If a higher confidence level of 0.01 is set, it is less likely that it will be concluded that there is significant clustering, in this case it is more likely that the null hypothesis will be accepted. If there is no clustering then a *Type 2* error will be committed (Burt & Barber, 1996).

To conclude then, the Z-score provides a basis on which the desired confidence level can be determined, for example, if a confidence level of 0.01 or 99% is desired then all the polygons in the study area with a Z-score greater than 2.58 or less than -2.58 are considered statistically significant. The result is that there is a less than 1% chance that rejecting the null hypothesis will be a mistake. This means that there is a



less than 1% chance that the cancer clusters are due to random chance and it can therefore be concluded that there must be environmental factors in those areas which cause them to have more incidences.

The Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool requires a specified distance. The tool uses this distance to determine a radius (neighbourhood) around each target polygon. All of the neighbouring polygons that fall within this distance are considered neighbours of the target feature and the values of these neighbouring polygons are therefore included in the calculation of that specific target polygon's Z-score.

There are two types of distances to choose from. The *Euclidean* distance is the straight line distance between two points (as the crow flies) and the *Manhattan* distance is the distance between two points measured along the axis at right angles, calculated by summing the difference between the x and y coordinates (Esri, 2012b). Calculations based on both the *Euclidean* and *Manhattan* distances require projected data. Projected coordinate systems are defined as a flat, two dimensional surfaces. The earth is seen as a sphere or spheroid but this three-dimensional surface needs to be transformed to a flat surface in order for the GIS to use it (Grafarend & Krumm, 2010). This mathematical transformation is called a map projection. One easy way to understand how map projections alter spatial properties is to visualise shining a light through the centre of the earth onto a surface (called a projection surface). Imagine the earth's surface is clear with the grid of intersecting lines of latitude and longitude (also known as the graticule) drawn on it. Wrap a piece of paper around the earth. A light at the centre of the earth will cast the shadows of the graticule onto the piece of paper. One can now unwrap the piece of paper and lay it flat. The shape of the graticule on the flat paper will be different from that on the earth (Snyder, 1997). There are different types of projections that preserve different properties, such as conformal projections that preserve local shape, equal area projections that preserve area, and equidistant projections that preserve distance (Esri, 2012b).

In this analysis it was decided to use the Africa-Equidistant-Conic projection with the central meridian set to the +25 longitude line and the standard parallel lines set to the -21 and -35 latitude lines respectively. This ensured that the distances used by Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool were accurate (Esri, 2009). The +25 longitude line was chosen as the central meridian because it cuts through the middle of South Africa. South Africa is located between the -21 and -35 latitude lines. Selecting these parameters therefore ensured that the distance properties of the study area were not distorted (Esri, 2009).

The properties of Equidistant Conic projections are that they preserve local shapes along the standard parallel lines. The distortions of local shapes however increase as the distance from the standard parallel lines increase. The distortion of area is constant along any given parallel but increases with distance from the standard parallel lines. Directions are true along the standard parallels. The most important factor however is that both distance and scale characteristics are true along the meridian and standard parallel lines (Lliffe & Lott, 2008).

The distance from one polygon to another is calculated by determining the distance between the centroids (the middle point of each polygon) (Esri, 2009). The distance that is chosen should reflect the inherent relationships among the features analysed. The more realistically one can model how features interact with each other in space, the more accurate one's results will be. According to a communication with Dr Lauren Scott (Scott, 2012), a project engineer on Esri's geoprocessing team and an expert in the use of statistics in a geospatial context, the best distance method to use is the Fixed Distance Band.

The Fixed Distance Band method works well for polygon data where there is a large variation in polygon size. The Fixed Distance Band method is also recommended when running the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool according to Esri (2009). One can think of the Fixed Distance Band as a moving window that momentarily settles on top of each polygon. It then looks at that polygon within the context of its neighbours. It is important to use a distance band that reflects maximum spatial

autocorrelation. Spatial clustering on the landscape indicates that there is evidence of underlying spatial processes at work. The distance that exhibits maximum spatial clustering, as measured by the Spatial Autocorrelation (Global Moran's I) Tool, is the distance where those spatial processes are most "active" or most pronounced Esri (2009).

The Spatial Autocorrelation (Global Moran's I) Tool calculates a Z-score for each distance specified. The tool needs to be run at different distances to determine where peaks in Z-scores are. Every peak represents a distance where the processes promoting spatial clustering are pronounced. Multiple peaks are common. Generally the peaks associated with larger distances reflect broad trends (a broad east to west trend, for example, where the west is a giant hotspot and the east a giant cold spot). In general one will be most interested in peaks associated with smaller distances. An inconspicuous peak often means there are many different spatial processes operating at a variety of spatial scales. One probably wants to look for other criteria as well in order to determine which fixed distance method is best for one's analysis (perhaps the most effective distance for remediation). If the Z-score never peaks (in other words, it just keeps increasing) and if one is using aggregated data (for example, counties), it usually means that the aggregation scheme is too coarse. This means that the spatial processes of interest are operating at a scale that is smaller than the scale of one's aggregation units. If one can move to a smaller scale of analysis (moving from counties to tracts, for example), this may help find a peak distance Esri (2009).

The Moran's I statistic measures feature similarity. It determines whether there are polygons with similar values clustered together or not. These values can either be low values or high values. The important thing is that they should be similar (Longley & Batty, 2003). A group of polygons next to each other with the values of; 2, 3, 1, 2, and 4 respectively would be considered as a group of polygons with similar values. A group of polygons next to each other with the values of; 33, 34, 32, 36 and 35 respectively would also be considered as a group of polygons with similar values. A

group of polygons next to each other with the values of; 22, 3, 67, 200, 2 and 14 would however not be considered as a group of polygons with similar values.

The first thing the Moran's I statistic does is to add all the values of all the polygons in the study area together. It then divides this value by the number of polygons in the study area to get the mean value of all the polygons (Mitchell, 2005). If the polygons in figure 3.5.4 were representative of the entire study area, then the Spatial Autocorrelation (Global Moran's I) Tool would add all the values together ( $2+3+3+10+12=30$ ) and divide this value (30) by the number of polygons in the study area, which in this case would be 5 polygons. The mean value would therefore be 6 (30 divided by 5).

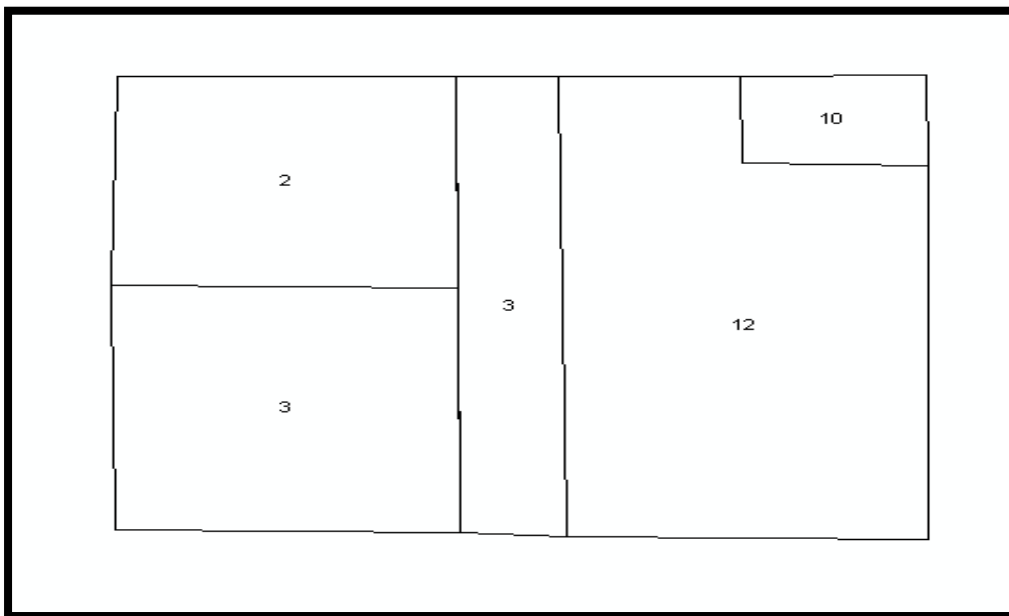


Figure 3.5.4: Example of neighbourhood (Mitchell, 2005)

The tool then takes each pair of polygons in the study area and works out how each polygon in that pair differs from the mean. Each polygon's difference from the mean is then multiplied to calculate that pair's cross-product (Longley & Batty, 2003). A high (positive) cross-product indicates that nearby polygons have similar values, while a low (negative) cross-product indicates that nearby polygons have dissimilar values (Mitchell, 2005).

A pair of polygons with similar values such as the values of 10 and 12 (as can be seen in figure 3.5.5) will have a positive cross-product. The tool takes the value of each polygon in the pair and subtracts the mean (6) from it. The result of each polygon is then multiplied to calculate the cross-product, for example  $(12-6) \times (10-6) = 6 \times 4 = 24$  (Mitchell, 2005). The cross-product is therefore 24, indicating that two similar values (12 and 10) are clustered together.

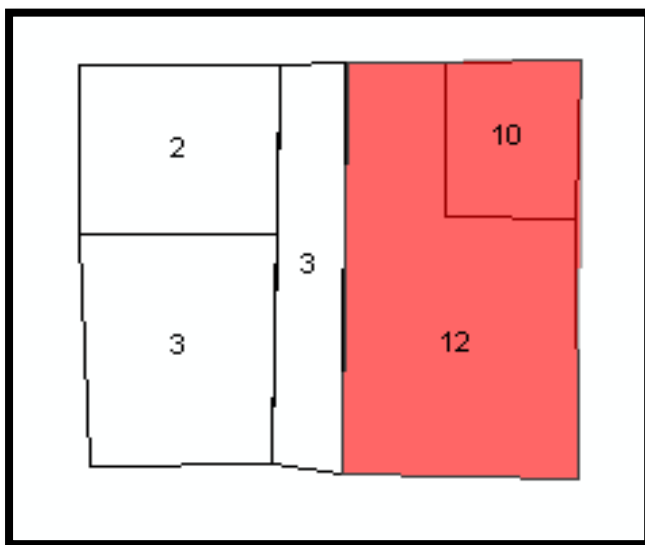


Figure 3.5.5: A pair of polygons with similar values (Mitchell, 2005)

If the same calculation is done for a pair of polygons with dissimilar values such as the values of 3 and 12 (figure 3.5.6), a negative cross-product will be calculated. The mean will be subtracted from the value of each polygon in the pair. The result for each polygon will then be multiplied to calculate a low cross-product, for example  $(12-6) \times (3-6) = 6 \times -3 = -18$  (Mitchell, 2005). The cross-product for this pair will therefore be -18, indicating that two dissimilar values are next to each other.

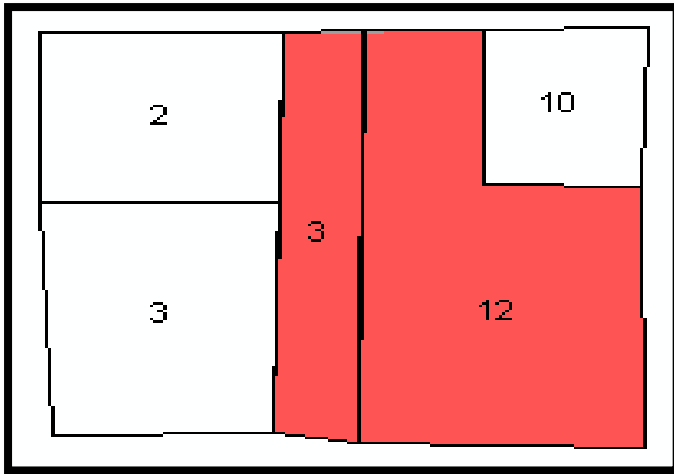


Figure 3.5.6: A pair of polygons with dissimilar values (Mitchell, 2005)

After the cross-product for each pair of polygons in the study area is calculated, they are summed together. If there were roughly as many pairs with positive cross-products as there were with negative cross-products, the result of adding the cross-products together would be close to 0, indicating a random distribution where some neighbours have similar values and some don't (Longley & Batty, 2003). If there were more positive cross-products than negative cross-products in the study area, the result of adding the cross-products together would be bigger than 0 indicating a clustered pattern where most of the polygons in the study area are surrounded by other polygons with similar values (Bailey & Gatrell, 1995).

In order to calculate a Z-score the statistic creates a scenario where it determines what the expected result would be if the cross-products of all the polygons in the study area were added together, assuming that there were no significant similarities between the polygons (Mitchell, 2005). The tool then subtracts this result from the actual observed result that was calculated when all the cross-products in the study area were added together. The result is the Z-score. According to Scott (2012) it is necessary to ensure that the distances one chooses to use when running the Spatial Autocorrelation (Global Moran's I) Tool, will result in each feature having at least one neighbour. In order to do this one can use the Calculate Distance Band from Neighbourhood Count Tool.

The Calculate Distance Band from Neighbourhood Count Tool returns the minimum, maximum, and average distance to a specified Nth nearest neighbour (N is an input parameter) for a set of features (Esri, 2009). The value of N can be changed to find the minimum, maximum, or average distance to the nearest quantity of neighbours desired. For example if N is set as 1, then the tool will calculate the minimum, maximum and average distance that ensures that each polygon has at least one neighbour. If N is set to 2 then the tool will calculate the minimum, maximum and average distance that ensures that each polygon has at least 2 neighbours. The Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool needs each polygon to have at least one neighbour. It is therefore useful to use the Calculate Distance Band From Neighbourhood Count Tool because it will ensure that the chosen distances result in each polygon having at least one neighbour.

According to Scott (2012) it is also good practice to use the average distance produced by the Calculate Distance Band From Neighbour Count tool as the interval between the distances one uses when running the Spatial Autocorrelation (Global Moran's I) Tool. It was observed in this study that the average distance produced by the Calculate Distance Band From Neighbour Count Tool had limitations when it was used as an interval between distances. The study area contained areas where there were large polygons, surrounded by other large polygons. The result of this was that the centroids (middle points of each polygon) were on average further apart than they would have been if the study area had smaller polygons. It was therefore found that the calculated average distance was too large, which caused possible Z-scores peaks to be missed. To illustrate this consider figure 3.5.7. In figure 3.5.7 the Spatial Autocorrelation (Global Moran's I) Tool was run using distances with an interval of 12724 metres, starting with a distance of 147300 metres, which is the maximum distance that will ensure that each polygon has at least one neighbour. The peaks that were identified were at 182000 metres, 275000 metres and 352000 metres. Figure 3.5.7 indicates what the results of the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool looked like with the first peak distance of 182000 metres.

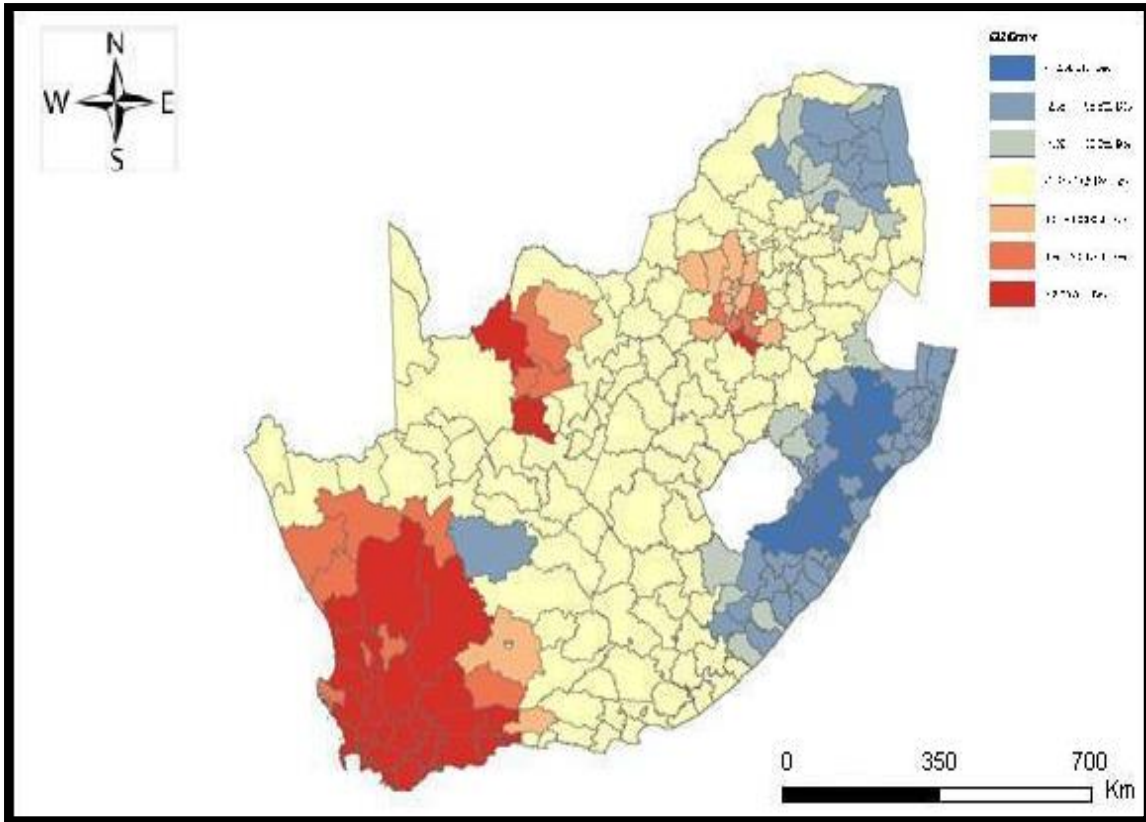


Figure 3.5.7: Spatial Autocorrelation (Global Moran's I) Tool run with an interval of 12724 metres and peak distance of 182000 metres.

Now consider figure 3.5.8. In figure 3.5.8 the Spatial Autocorrelation (Global Moran's I) Tool was run with distances that had a 49000 metre interval (the average distance calculated by the Calculate Distance Band from Neighbourhood Tool). The first distance used was again 147300 metres, which is the maximum distance that will ensure that each polygon has at least one neighbour according to the Calculate Distance Band From Neighbourhood Tool. The peaks that were identified by using these distances, were at 400000 metres, 745000 metres and 845000 metres. Figure 3.5.8 indicates what the results of the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool looked like when the first peak distance of 40000 metres was used.



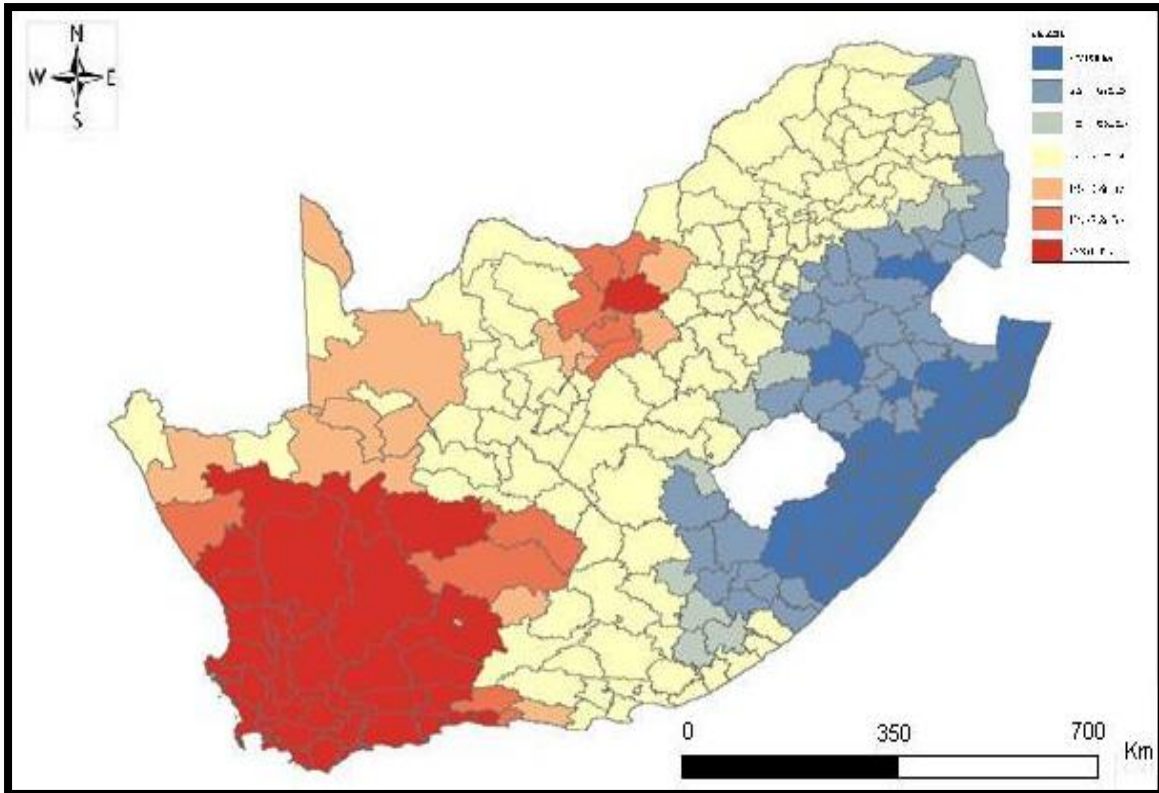


Figure 3.5.8: Spatial Autocorrelation (Global Moran's I) Tool run with an interval of 49000 metres and a peak distance of 40000 metres.

It is clear that figure 3.5.7 gives a more detailed representation than figure 3.5.8, because of the smaller peak distance used. In this case the differences between the two figures were small but in other cases this could cause significant differences. It is therefore recommended that future studies following this procedure should consider the different sizes of the polygons in the study area. The distance calculated by the Calculate Distance Band From Neighbour Count Tool to ensure that each polygon has at least one neighbour, is a good distance to use as an interval, but then the size of the polygons in the study area should be roughly the same. Important Z-score peaks might therefore be missed if the Spatial Autocorrelation (Global Moran's I) Tool is run with an inadequate interval distance. Once the distances have been determined where there are Z-score peaks, it is necessary to insure that each target polygon has on average at least 8 neighbours. This is done by running the Generate Spatial Weights Matrix Tool.

This tool calculates the average number of neighbours there are in the study area for each target polygon at the specific distance (Esri, 2009). If a peak Z-score distance of 182000 metres is run with this tool for example, and the tool calculates an average number of neighbours of 22, then it means that on average each polygon in the study area has 22 neighbours if a radius of 182000 metres is drawn around each target polygon. It is therefore important to insure that the average number of neighbours is never less than 8 since this will cause some target polygons not to have enough neighbours, causing the tool to calculate false hotspots. Polygons near the edge of the study area generally have fewer neighbours. The results of such polygons can thus be skewed (Getis&Ord, 1996).

If fewer than 30 polygons are analysed, the results might be skewed because of the effect of any outliers (Lee & Wong, 2001). Outliers are exceptionally high or low values, beyond what one would expect even with a skewed distribution of data. Since outliers can throw off the results of one's analysis, one needs to know whether they are present. Outliers often represent data errors. Values can be entered incorrectly into the database or associated with the wrong polygon. Once outliers have been identified, it is necessary to check the original data source to make sure that the values for those polygons are correct (Burt & Barber, 1996). Missing values often show up as outliers, and if a valid value for the polygon cannot be obtained, it might be necessary to remove it from the dataset before performing one's analysis. It is important to note that outliers are not always data errors.

According to Earickson and Harlin (1994) they may in fact represent valid but unexpected values, for example, a mansion in an otherwise modest neighbourhood, which has a much higher value than its neighbours. An outlier might also reflect a previously unknown condition which could alter the assumptions of the analysis performed or change the approach followed (Ebdon, 1985). A block group which has an unexpectedly high number of crimes, for example, might point to a previously unknown drug-trafficking hotspot. When outliers represent valid data values, it is necessary to measure their influence on one's analytical results. One way to do this

according to Burt and Barber (1996) is to run the analysis without the outliers. If the results are very close to each other, the outliers are not having a strong impact on the results, but if the results deviate strongly, it might be necessary to find another analysis method that will not be sensitive to the presence of any outliers.

Outliers were identified for Acute Lymphoblastic Leukaemia; Astrocytoma; Burkitt's Lymphoma; Glioma; Hepatoblastoma; Hodgkin's Lymphoma; Medulloblastoma and Non-Hodgkin's Lymphoma. The Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool was run with the outliers present and again with the outliers removed for each cancer type. It was observed that the outliers did not have a prominent effect on the results. Due to it being good practice and the fact that the main focus of this dissertation was to determine a best practice methodology for creating cancer maps, the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool was run with the outliers removed. In figure 3.5.9 the effect of an outlier can be seen. In this instance the outlier caused the first Z-score peak to occur earlier, thereby causing a different result. As can be seen in figure 3.5.9 the first Z-score peak was found at 182000 metres (when the outlier was still present), but dropped to 175000 metres when the outlier was removed. The R1 indicated in figure 3.5.9 represents the cancer ratio for the first cancer type (Acute Lymphoblastic Leukaemia).

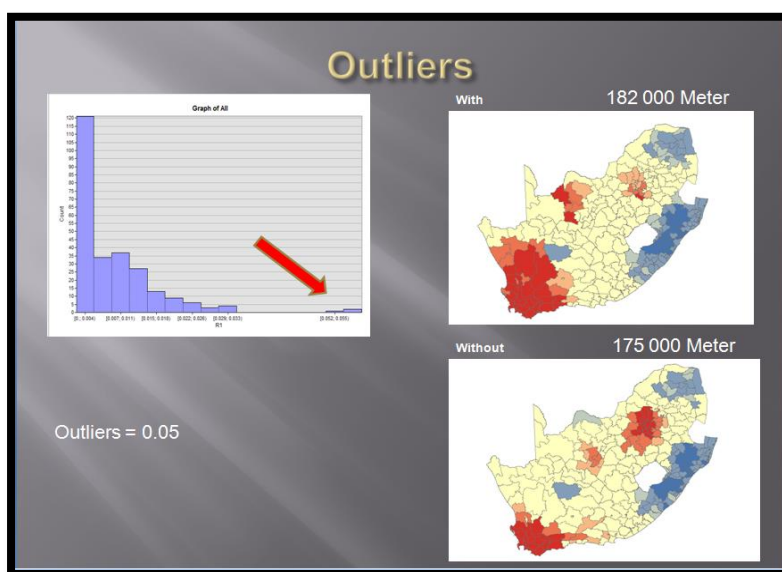


Figure 3.5.9: The effect of an outlier on the results of Acute Lymphoblastic Leukaemia.

The Z-scores calculated by the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool or other local statistic tools can produce misleading results. This happens because the tests assume independence between the polygons. The GIS runs the tests to calculate a Z-score for each polygon. These tests however end up using many of the same neighbours for adjacent polygons (Getis & Ord, 1996). This violates the independence of the test, especially since the values of adjacent polygons are more likely to be similar in any case. Researchers Art Getis and Keith Ord have suggested an alternative test to overcome this issue, called the Bonferroni Type Test. This test uses a more stringent criterion to ensure that results are significant at the confidence level specified (Getis & Ord, 1996).

By changing the critical value, the Bonferroni Correction makes it more difficult for any one test to be statistically significant. In the correction's simplest form, one divides the original confidence level by the number of tests (number of polygons in the study area) to get an adjusted confidence level (Getis & Ord, 1996). With a confidence level of 0.10 and 30 polygons in the study area, the adjusted confidence level would be 0.003 ( $0.10/30$ ). One would then compare the calculated Z-score for each feature to the critical value obtained using this adjusted confidence level.

In the case of Acute Lymphoblastic Leukaemia, for example, the Bonferroni Correction was used to find a new critical value. At a 99% confidence level where there is a less than 1% chance (0.01) that the hotspots observed are due to random chance, the confidence level of 0.01 was divided by the number of polygons in the study area, 257. The result of this calculation ( $0.01/257$ ) was 0.000039. What this means is that all the polygons with a p-value smaller than 0.000039 can be considered statistically significant at a 99% confidence interval (Getis & Ord, 1996). This can also be done for a 95% confidence interval (0.05), where the calculation would be ( $0.05/257$ ). The result of this calculation would then be 0.000195, which means that all the polygons with a p-value smaller than 0.000195 can be considered accurate on a 95% confidence interval.

To illustrate these results please consider figure 3.5.10. In this instance all the polygons with p-values smaller than 0.000039 for Acute Lymphoblastic Leukaemia were selected in blue. This means that all the polygons selected in blue can be considered statistically significant at a 99% confidence interval. The hotspots that are not selected in blue can therefore not be considered statistically significant at a 99% confidence interval, but some of them might be statistically significant at a 95% confidence interval.

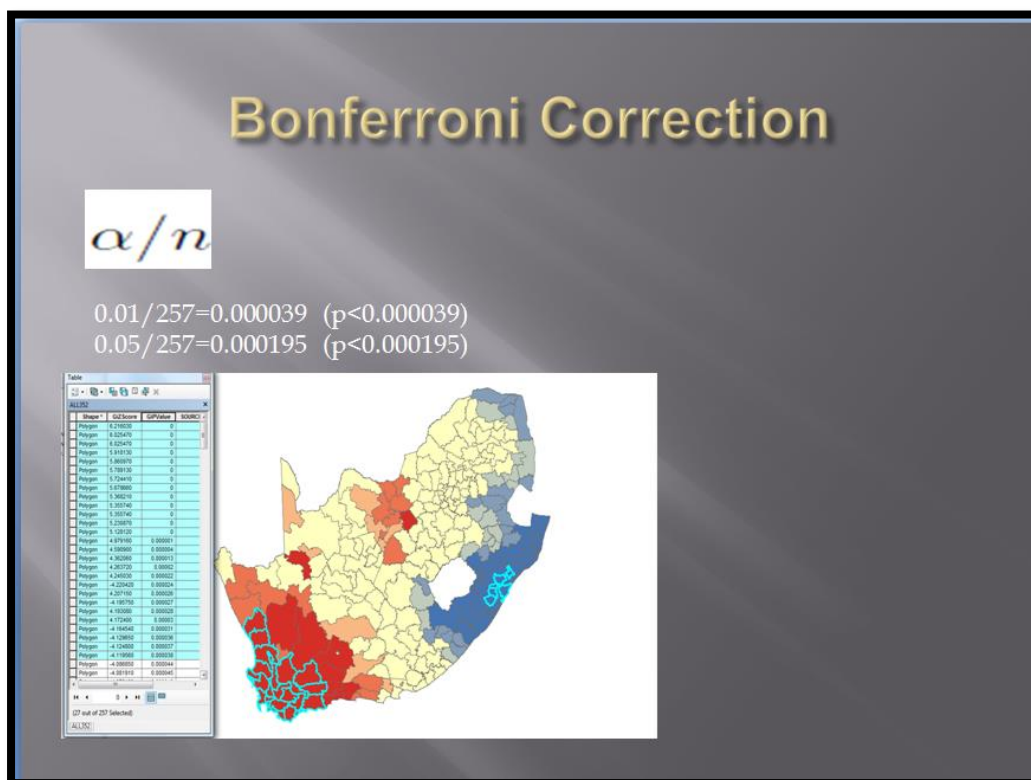


Figure 3.5.10: Example of the Bonferroni Correction performed on a 99% confidence interval where all the p-values smaller than 0.000039 were selected in blue.

By considering the limitations of the Hotspot Analysis (Getis-Ord Gi\*) Tool and by doing the Bonferroni Correction, statistically significant hotspots on a 99% confidence interval can be identified. GIS software can therefore be used to identify spatial clusters of cancers that are statistically significant. It is clear that the Hotspot Analysis (Getis-Ord Gi\*) Tool has certain limitations, but it is also clear that there are

ways to compensate for these limitations to still arrive at statistically significant results that can be invaluable in the fight against cancer.

## CHAPTER 4: RESULTS

### 4.1 Successfulness of GIS methodology.

It was found in this study that GIS software can be successfully applied to map cancer incidences. The use of GIS is particularly valuable because of its capability to perform statistical calculations in a spatial context. It was found that normal GIS symbology could be used to represent cancer incidences but that this would result in biased representations. It was found that the use of spatial statistics within the GIS software could overcome this limitation.

The study found that the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool could be successfully implemented to map cancer incidences since the tool uses a G-statistic which calculates a Z-score that can determine whether a cluster of high or low values are statistically significant or just due to random chance. It was found that the tool could only be used successfully if the various factors influencing the tool were understood and accounted for. These factors included the specified distance that needed to be used to ensure that each target polygon had enough neighbours; the extent of the spatial autocorrelation that was associated with that distance; the possible effects of polygons at the edge of the study area as well as the possible effects of outliers within the data.

It was also found that spatial statistics often violate one of the main assumptions of statistics, namely the independence of observation in a sample. Spatial data often violates this assumption because one spatial variable often influences other variables. Commercial burglaries for example tend to occur where there are businesses, and businesses tend to cluster. The outcome of an analysis performed with spatial statistics could therefore be somewhat predetermined. It was however found that the problem could be overcome by performing the Bonferroni Correction.

This study found that there are definite best practice guidelines that needs to be followed when using the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool. It was found that once the various elements influencing the Tool were understood and accounted for, the Tool could be successfully used to map the cancer incidences.

#### **4.2 Result of data quality review.**

The cancer incidences diagnosed in South Africa (for children under the age of 15) were found to be significantly less than the incidences diagnosed in first world countries. It was found that in South Africa only approximately 500 new cases are reported each year, while approximately 2500 new cases are reported in first world countries each year.

The most probable reason for the underreported nature of South Africa's cancer incidences (for children younger than 15) was attributed to the fact that South Africa's registry is pathology based, thus missing cases where patients die before they are diagnosed. It was also found that the disorganised state of the country's clinics could play a role in cancer cases being missed. It was found that there is a possibility that people visiting public clinics lie about their addresses in order to gain access to better clinics which are not the areas where they actually live, thus impairing the quality of the geographic data that is linked to each cancer case.

The geographic data linked to the cancer cases limited the scale on which the analysis could be done. The majority of the geographic data consisted of only the town, city or suburb name in which the patient lived. The data was therefore insufficient to perform the analysis on the preferred (finer) enumeration area scale and had to be done on a (coarser) local municipal scale. Approximately 87% of the total cancer data (1992-2008) could be plotted on the local municipal scale. The data however needed to be used in conjunction with population statistics in order to calculate the ratios of cancer incidences. The population statistics limited the quantity of cancer cases that could be used in the analysis since only the census



data from 1996, 2001 and 2007 could be converted to the same set of boundaries (2005 local municipal boundaries) by Statistics South Africa.

The biggest limitation of this study is therefore that approximately 33% of the total cancer data provided by the South African Paediatric Tumour Registry could not be used (20% could not be linked to the population statistics because the cases were diagnosed before 1996 or after 2007, and 13% could not be plotted because they had no geographical information linked to them at all). The remaining 67% of the total data could be used on a (coarser) local municipal scale, but not on a (finer) enumeration area scale (which would have resulted in only 28% of the total cancer data being used). This study was therefore conducted with 33% of the total available data for analysis missing or unusable.

### **4.3 Results of the Hotspot Analysis (Getis-Ord Gi\*) Tool**

#### **4.3.1 Difference between general results and accurate results produced by the Hotspot Analysis (Getis-Ord Gi\*) Tool.**

When interpreting these results, it should be remembered that there are concerns about the quality of the data that was used. The results may therefore not be accurate, not due to the GIS process but due to the data itself. There might however be some hotspots worth investigating, since the analysis was still based on actual cancer data albeit suspect data. There might be certain trends that could be worth investigating. The reality however is that the quality of the data cannot be guaranteed and the accuracy of the final maps can therefore also not be guaranteed. In terms of using the correct GIS methodology to identify significant cancer clusters it is necessary to understand that once the Hotspot Analysis (Getis-Ord Gi\*) Tool has been run the process is not finished. It is vital that the Bonferroni Correction is performed. Spatial data often violates one of the main assumptions of statistics, namely the independence of observation in a sample. The Bonferroni Correction is therefore necessary to compensate for this violation.

The general results therefore indicate hotspots that have not been corrected by the Bonferroni Correction and although these results might indicate possible trends, they do not take the limitations of spatial data into consideration and should therefore be viewed with caution but not completely ignored as they might indicate possible trends. The Bonferroni Correction yields accurate results on a 99% or 95% confidence interval and thus mitigates the limitations of the spatial data. In other words, the results can only truly be trusted once the Bonferroni Correction has been performed, since only then will the limitations of spatial data be accounted for.

The general results are discussed with the accurate results since these results might still indicate possible trends worth investigating, but when conclusions have to be made on results that have taken the violation of independence of observation within spatial data into account, the accurate results found after the Bonferroni Correction was performed needs to be used. These hotspots are indicated by blue margins in the results.

The Bonferroni Correction was done for each cancer type. It should be noted that only Acute Lymphoblastic Leukaemia; Hodgkin's Lymphoma; Burkitt's Lymphoma; Astrocytoma; Medulloblastoma and Retinoblastoma showed p-values smaller than 0.000039 (99% confidence interval). These are thus the only cancer types that showed hotspots on a 99% confidence interval after the Bonferroni Correction was performed. In the case of Glioma no p-values smaller than 0.000039 (99% confidence interval) were observed after the Bonferroni Correction, however p-values smaller than 0.000195 (95% confidence interval) were observed. The Bonferroni Correction therefore indicated hotspots for Glioma on a 95% confidence interval but not on a 99% confidence interval. The remaining cancer types namely, Acute Myeloid Leukaemia; Non-Hodgkin's Lymphoma; Hepatoblastoma; Kaposi Sarcoma; Neuroblastoma; Nephroblastoma and Osteosarcoma did not have any p-values smaller than 0.000039 (99% confidence interval) or 0.000195 (95% confidence interval) after the Bonferroni Correction was performed. These cancer types did not show any hotspots on a 99% or 95% confidence interval and therefore only had general results.

There were also coldspots indicated within the results. These coldspots indicated areas where there were significantly less cancer incidences than one would expect. These areas should however be viewed with caution since none of them were found to be significant on a 99% or 95% confidence interval after the Bonferroni Correction was performed. These areas should not be ignored as they might indicate areas where possible protective environmental factors against cancer could exist. It should however be noted that the most probable cause for these coldspots are the amount of underreported cancer incidences within these areas. Most of these coldspots were found in rural areas where it is likely that they were not reported.

### **4.3.2 General and accurate results discussed per cancer type.**

#### **4.3.2.1 Leukaemia**

When the general result for Acute Lymphoblastic Leukaemia (in figure 4.1) is considered it is clear that there seems to be hotspots of Acute Lymphoblastic Leukaemia in the Western Cape, Gauteng and the western part of Northwest. There are also coldspots indicated in Kwa-Zulu Natal, the Eastern Cape and Limpopo. It is important to note that hotspots based on a 99% confidence interval were found for Acute Lymphoblastic Leukaemia in the Western Cape (figure 4.2). This should therefore be seen as a significant area on which research efforts can be focussed. According to the literature review the possible environmental and other carcinogens linked to Leukaemia include; pesticide exposure; high altitudes; electromagnetic fields; ionizing radiation; high birth weight; and certain infections such as Varicella, Infuenza and the Epstein-Bar Virus. It is therefore worth investigating whether any of these variables are present in this area and to what extent they might be influencing the development of Leukaemia The general result for Acute Myeloid Leukaemia indicated hotspots in the central part of the country, as well as coldspots in Kwa-Zulu Natal, the Eastern Cape and Northern Cape (as seen in figure 4.3). No hotspots were identified on a 99% or 95% confidence interval after the Bonferroni Correction was performed for this cancer type.

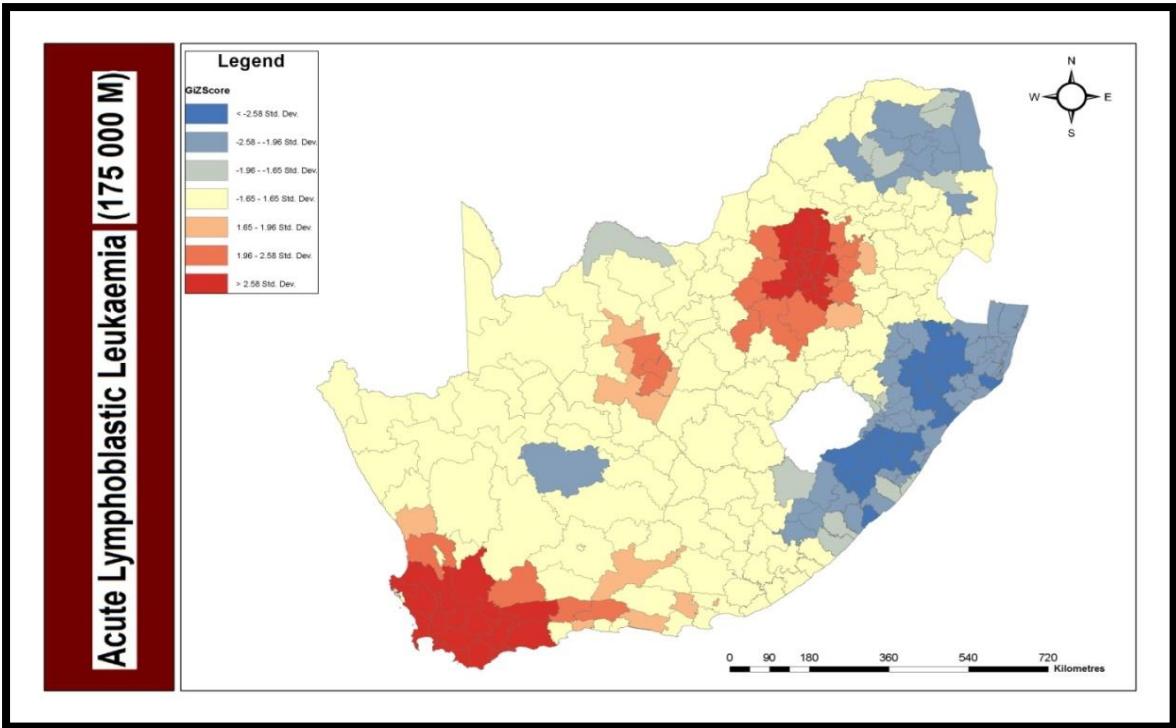


Figure 4.1: General result for Acute Lymphoblastic Leukaemia before the Bonferroni Correction was performed.

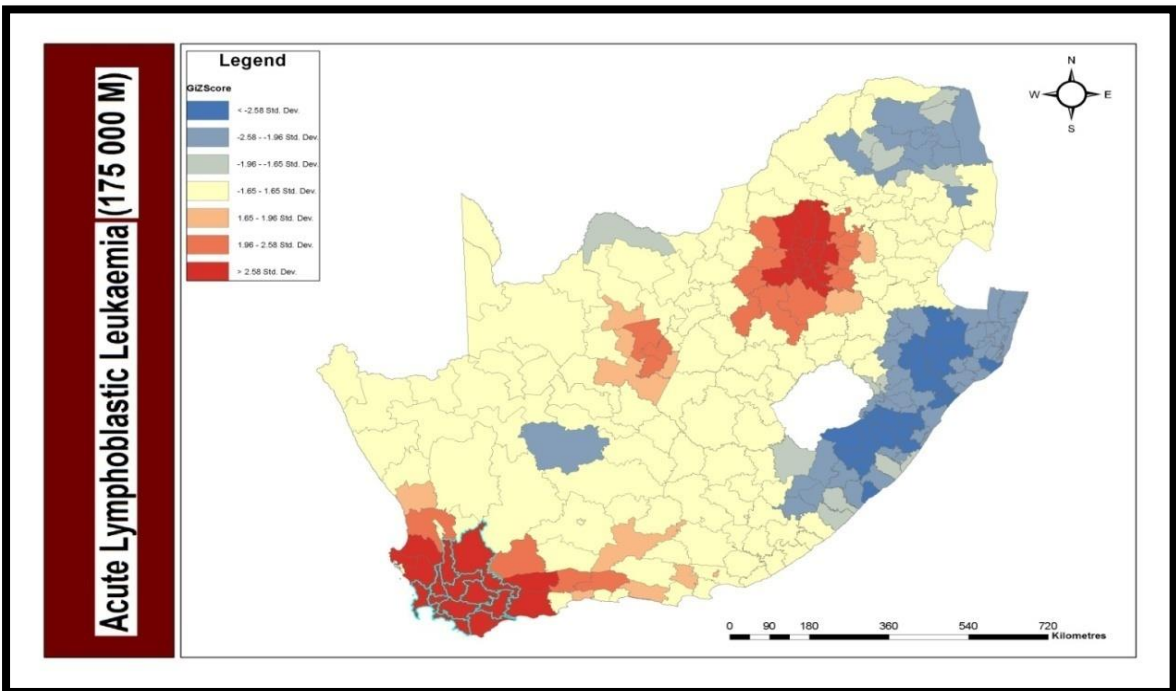


Figure 4.2: Accurate result on a 99% confidence interval for Acute Lymphoblastic Leukaemia after the Bonferroni Correction was performed.

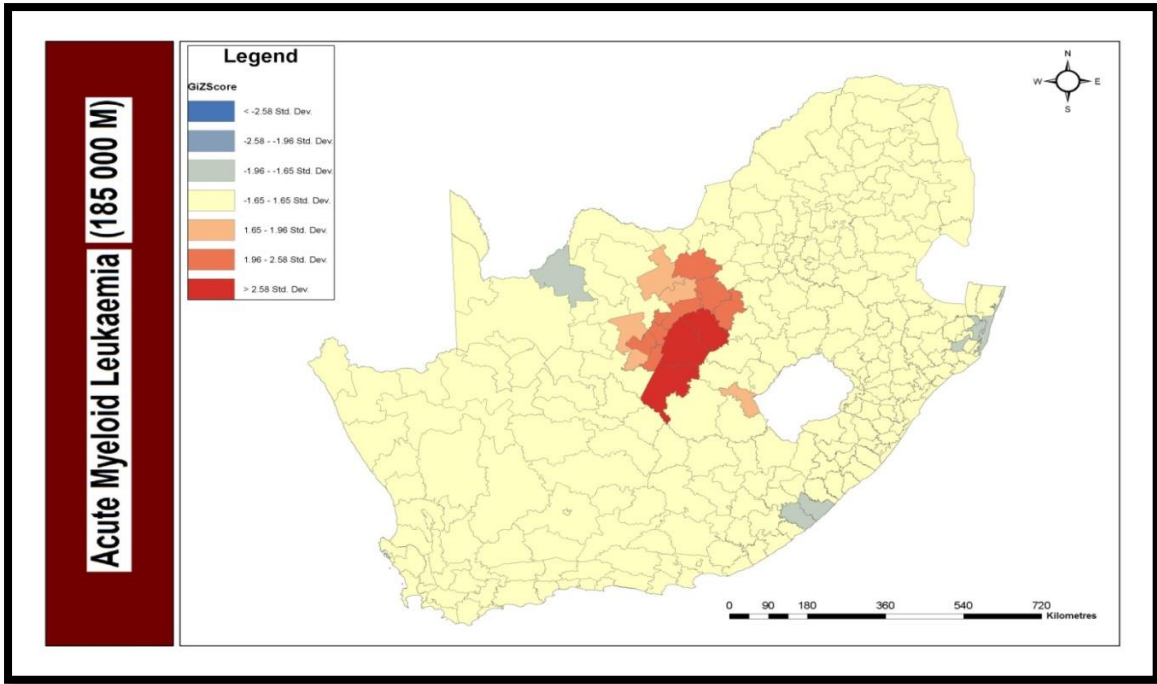


Figure 4.3: General result for Acute Myeloid Leukaemia. (No hotspots were found on a 99% or 95% confidence interval after the Bonferroni Correction was performed for this cancer type).

#### 4.3.2.2 Lymphoma

The general result for Hodgkin's Lymphoma (figure 4.4) indicated hotspots in the southern part of the Northern Cape, the southern part of the Free State and the majority of Mpumalanga as well as coldspots in the Eastern Cape. The general result for Burkitt's Lymphoma indicated hotspots in the Western Cape and some areas in the Eastern Cape as well as coldspots in Kwa-Zulu Natal and Limpopo (as seen in figure 4.7). It is important to note that hotspots based on a 99% confidence interval were observed for Hodgkin's Lymphoma in the Mpumalanga province, as well as for Burkitt's Lymphoma in the Western Cape province (as seen in figures 4.5 and 4.8). These areas should therefore be seen as the most significant areas. The literature review revealed that the possible environmental carcinogens linked to Lymphoma include hot and wet areas; pesticide exposure; exposure to Trichloroethylene; climates related to higher altitudes; malarial infections and possible associations with the Epstein-Barr Virus and Helicobacter Pylori Virus.

It is therefore worth investigating whether any of these variables are present in the identified areas. The general result for Non-Hodgkin's Lymphoma shows hotspots in the western part of the Northern Cape as well as coldspots in the Eastern Cape (as seen in figure 4.6). There was however no hotspots found on a 99% or 95% confidence interval for this cancer type after the Bonferroni Correction was performed.

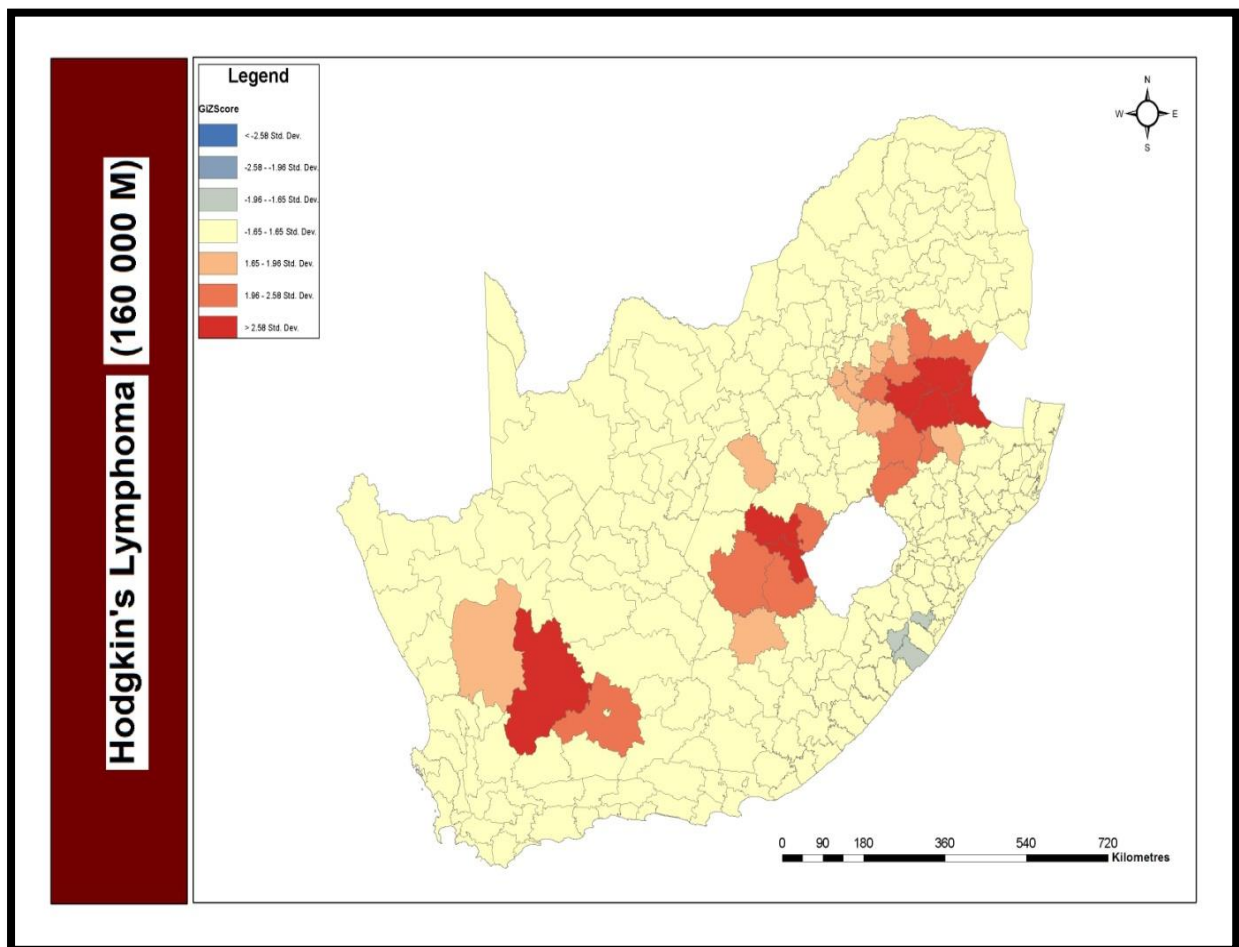


Figure 4.4: General result for Hodgkin's Lymphoma before the Bonferroni Correction was performed.

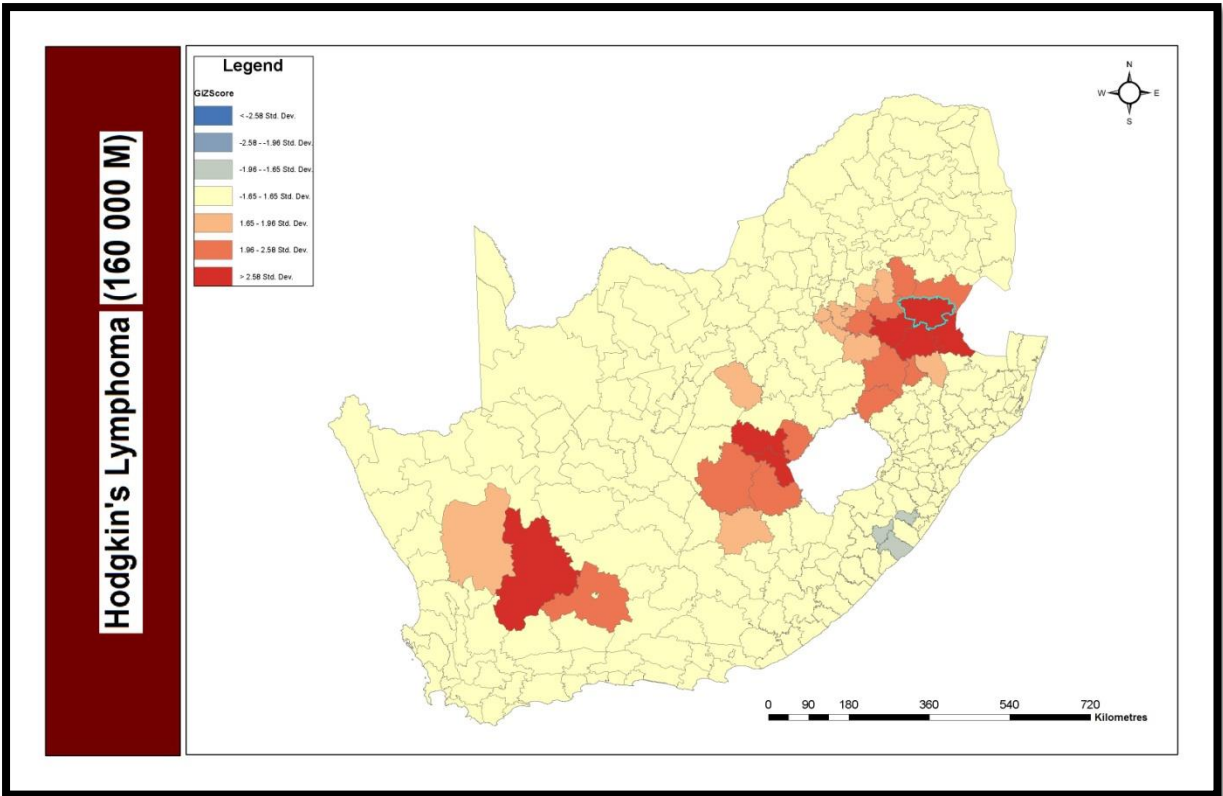


Figure 4.5: Accurate result for Hodgkin's Lymphoma after the Bonferroni Correction was performed on a 99% confidence interval.

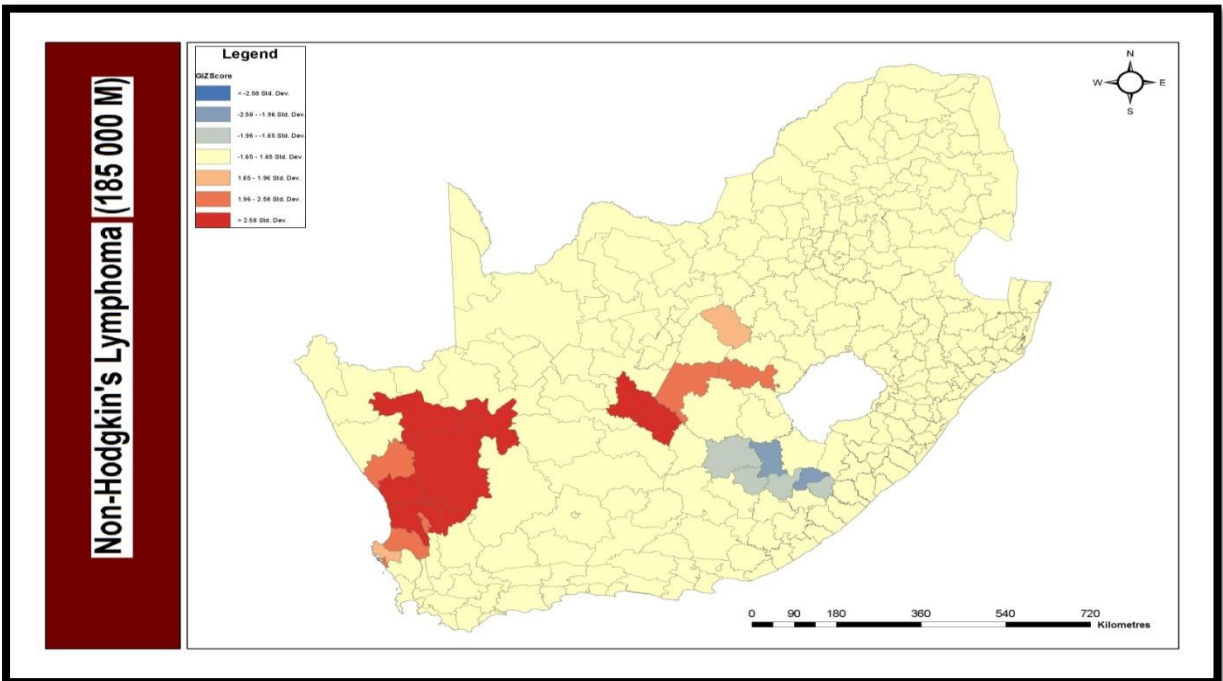


Figure 4.6: General result for Non-Hodgkin's Lymphoma (No hotspots were found on a 99% or 95% confidence interval for this cancer type).



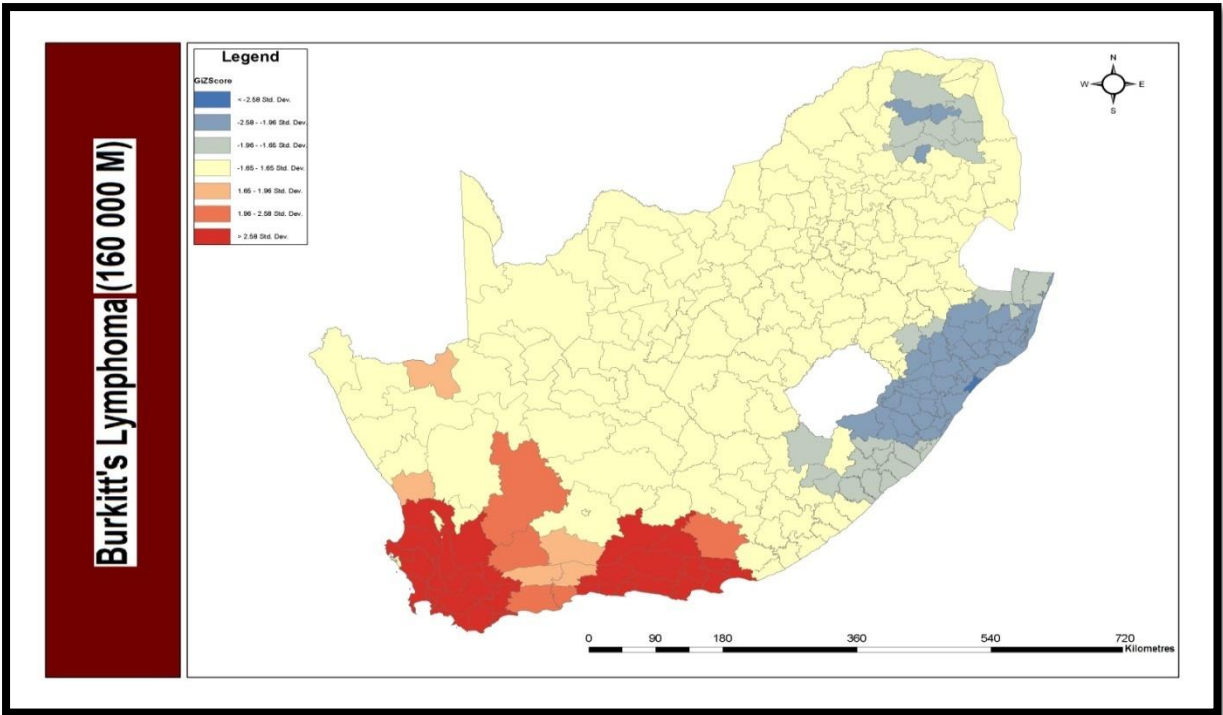


Figure 4.7: General result for Burkitt's Lymphoma before the Bonferroni Correction was performed.

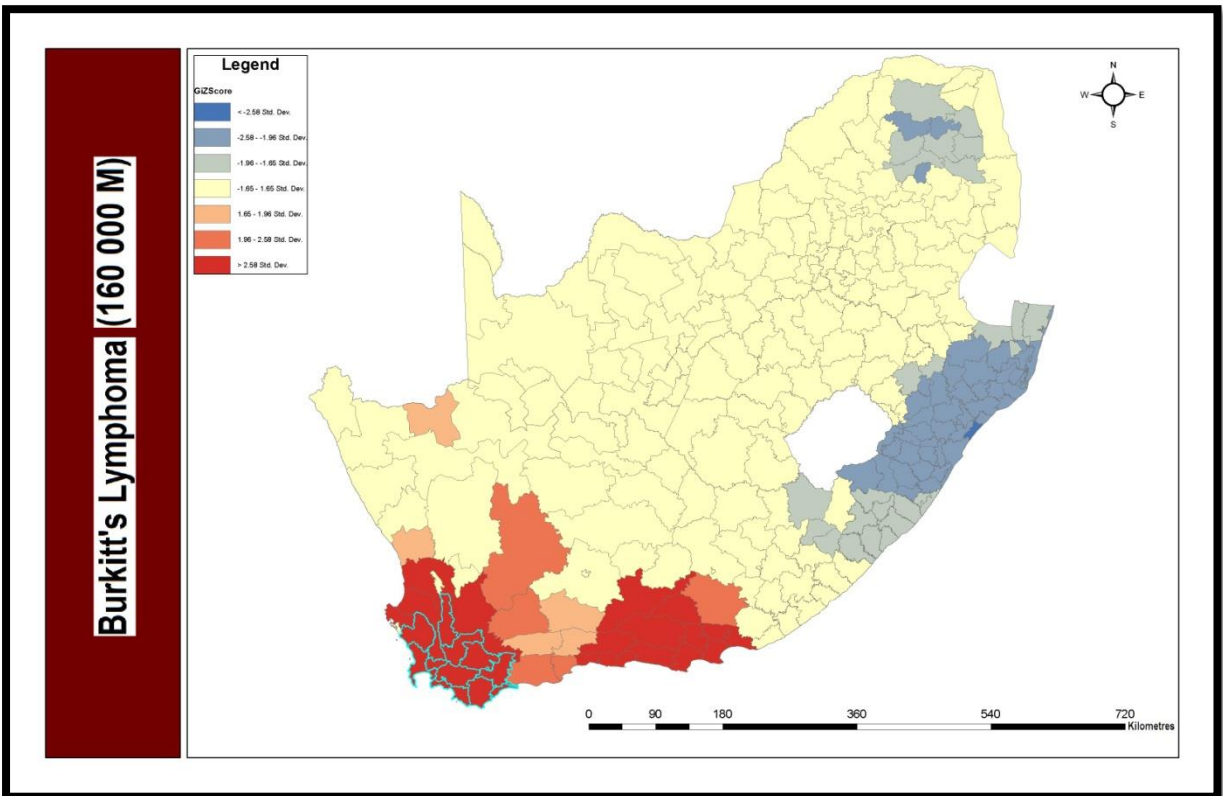


Figure 4.8: Accurate result for Burkitt's Lymphoma after the Bonferroni Correction was performed on a 99% confidence interval.



### 4.3.2.3 Brain Tumours

The general result for Astrocytoma indicated hotspots in the Western Cape as well as coldspots in Kwa-Zulu Natal and Mpumalanga (figure 4.9). The general result for Medulloblastoma indicated hotspots in the Western Cape, with coldspots in the Eastern Cape and Limpopo (figure 4.11). The general result for Glioma also indicated hotspots in the Western Cape as well as in the eastern parts of the Northern Cape, the western parts of Northwest and the western parts of the Free State. The result indicated coldspots in parts of Mpumalanga and Limpopo (figure 4.13). Hotspots based on a 99% confidence interval were observed for Astrocytoma and Medulloblastoma in the Western Cape (as seen in figure 4.10 and figure 4.12). These hotspots are of particular interest since the literature review revealed a strong relationship between brain cancer and vineyards. The Western Cape is well known for its vineyards and the fact that hotspots for both Astrocytoma and Medulloblastoma were found in this area justifies further investigation.

In the case of Glioma there were no hotspots found on a 99% confidence interval after the Bonferroni Correction was performed. There were however hotspots found on a 95% confidence interval after the Bonferroni Correction was performed. These hotspots were found in the northern parts of the Northern Cape Province. These hotspots are again of particular interest when the literature review regarding brain cancer is considered. The literature review revealed that there is a possible relationship between asbestos exposure and the development of brain cancer. These hotspots are located in close proximity to the town of Kuruman which was previously known as an asbestos mining town. The area around Kuruman was extensively mined for asbestos before the use of asbestos was banned. These hotspots therefore indicate a possible link between asbestos exposure and the development of brain cancer within this area. According to the literature review brain cancer can also be facilitated by the exposure to pesticides; farm animals; wood preservatives; lead; mercury and electromagnetic fields. High birth weight was also found to significantly increase the chances of developing brain cancer. It is therefore

worth investigating whether any of these variables are present within the identified hotspot areas.

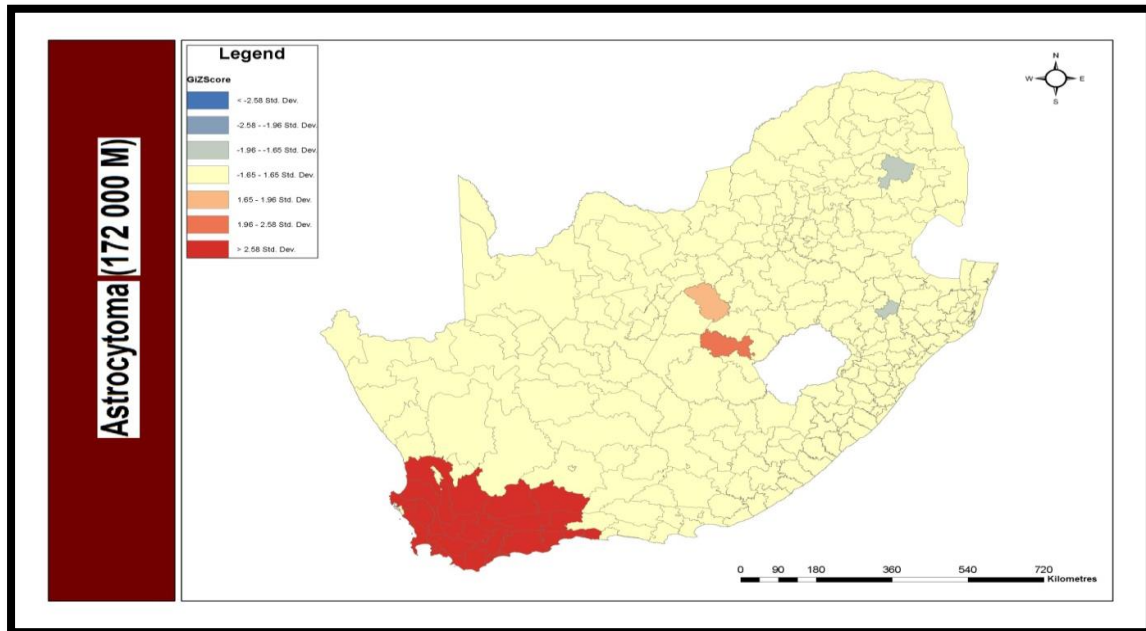


Figure 4.9: General result for Astrocytoma before the Bonferroni Correction was performed.

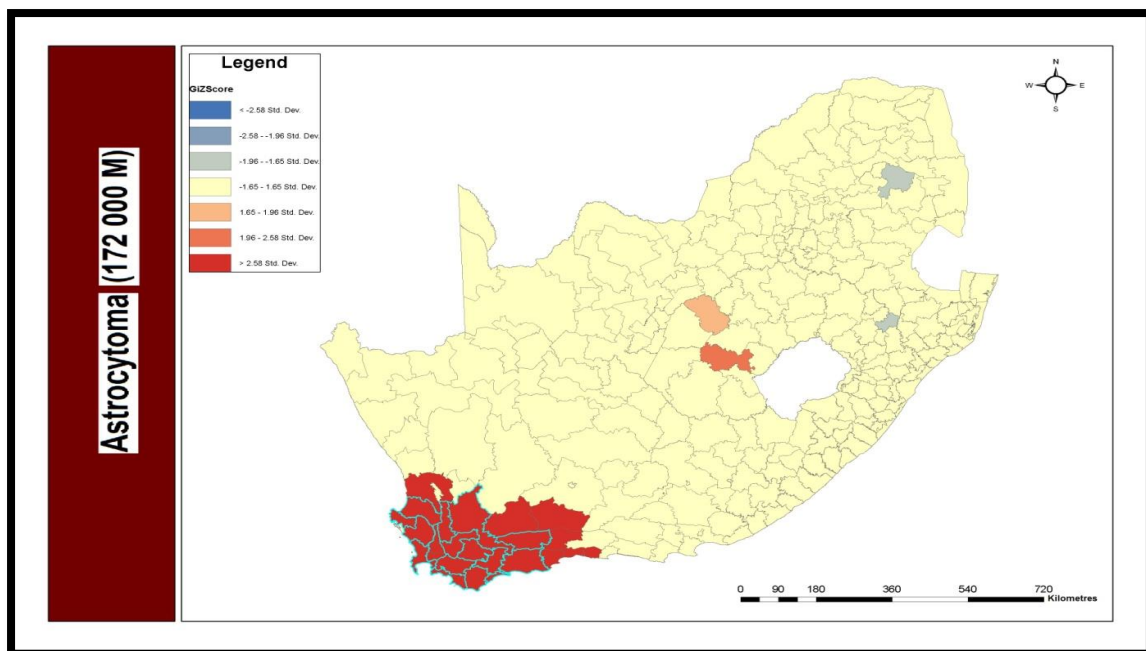


Figure 4.10: Accurate result for Astrocytoma on a 99% confidence interval after the Bonferroni Correction was performed.

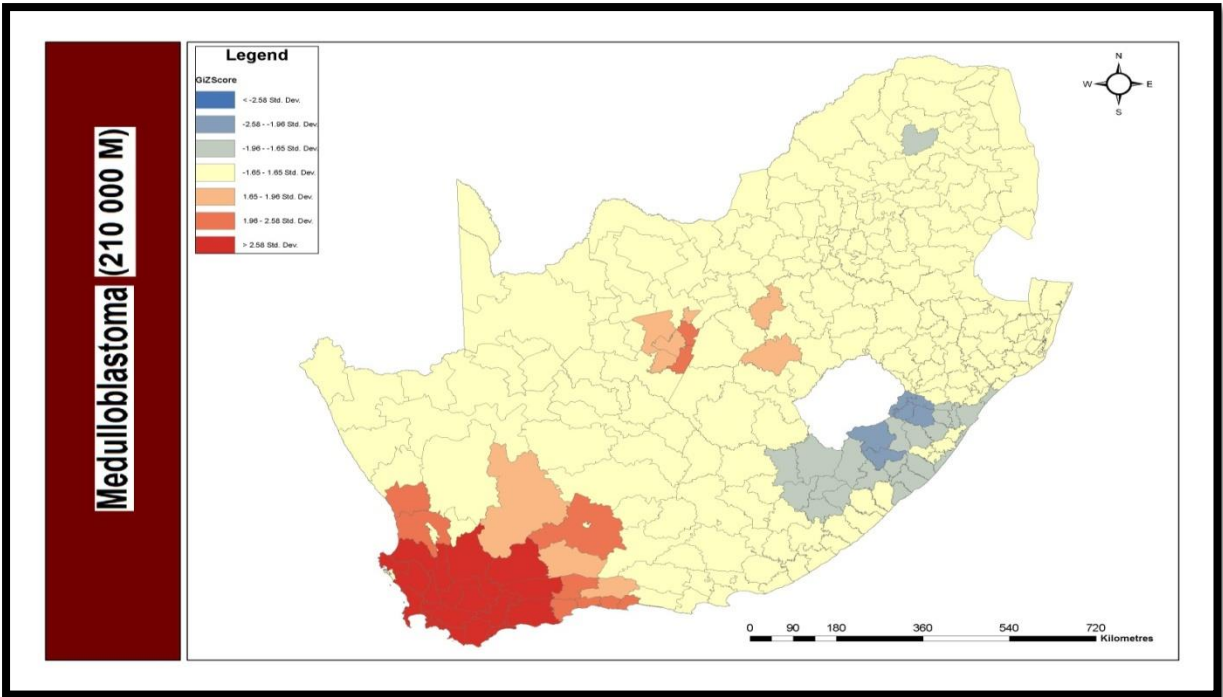


Figure 4.11: General result for Medulloblastoma before the Bonferroni Correction was performed.

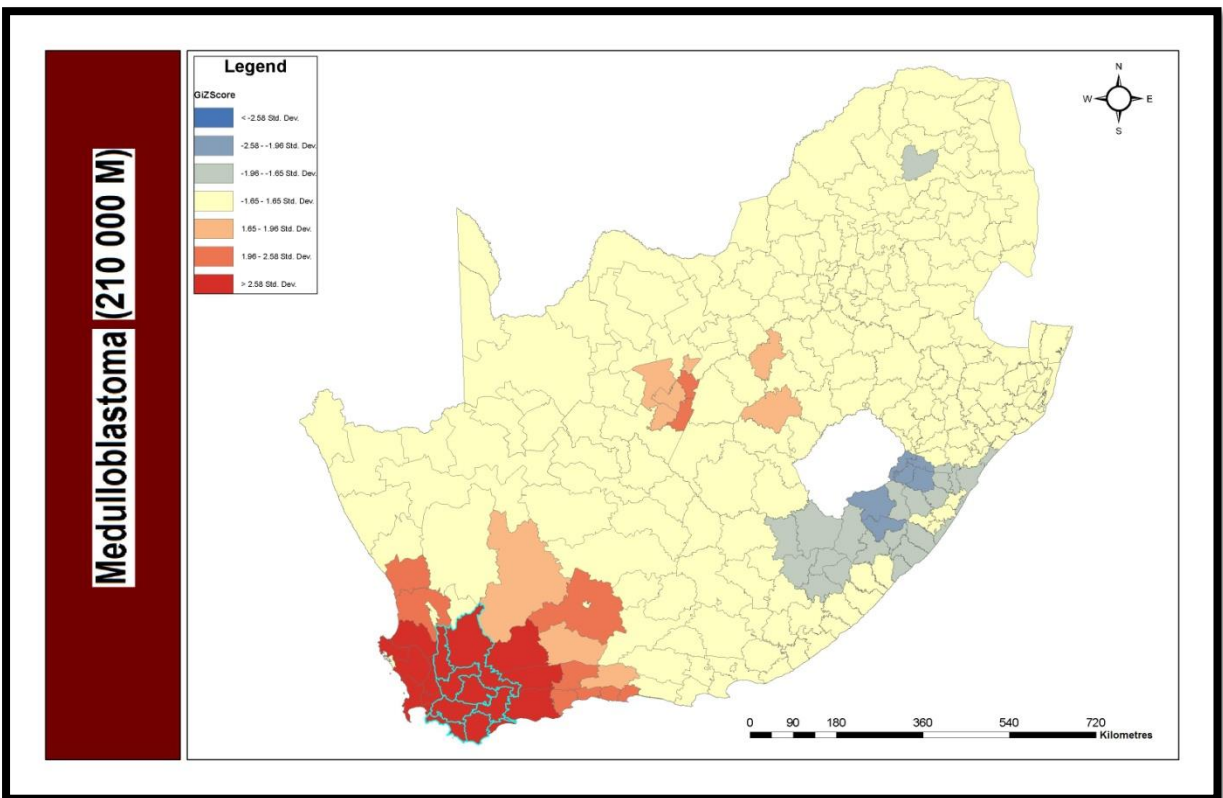


Figure 4.12: Accurate result for Medulloblastoma on a 99% confidence interval after the Bonferroni Correction was performed.

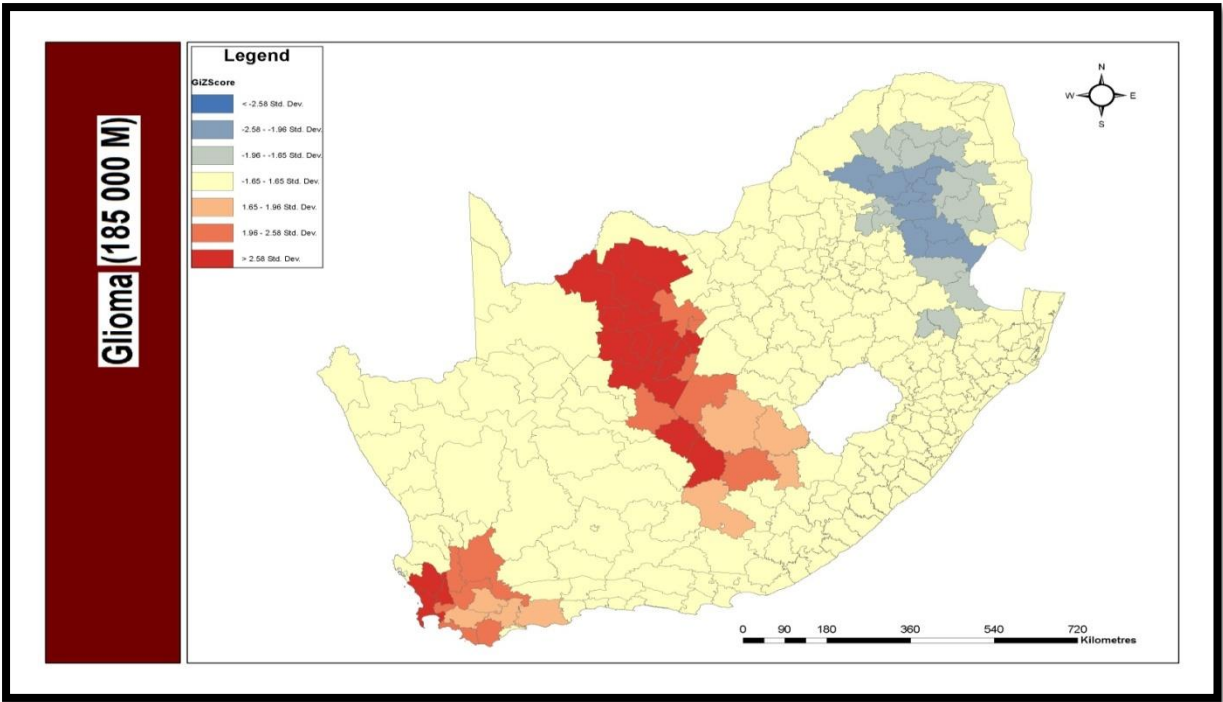


Figure 4.13: General result for Glioma before the Bonferroni Correction was performed.

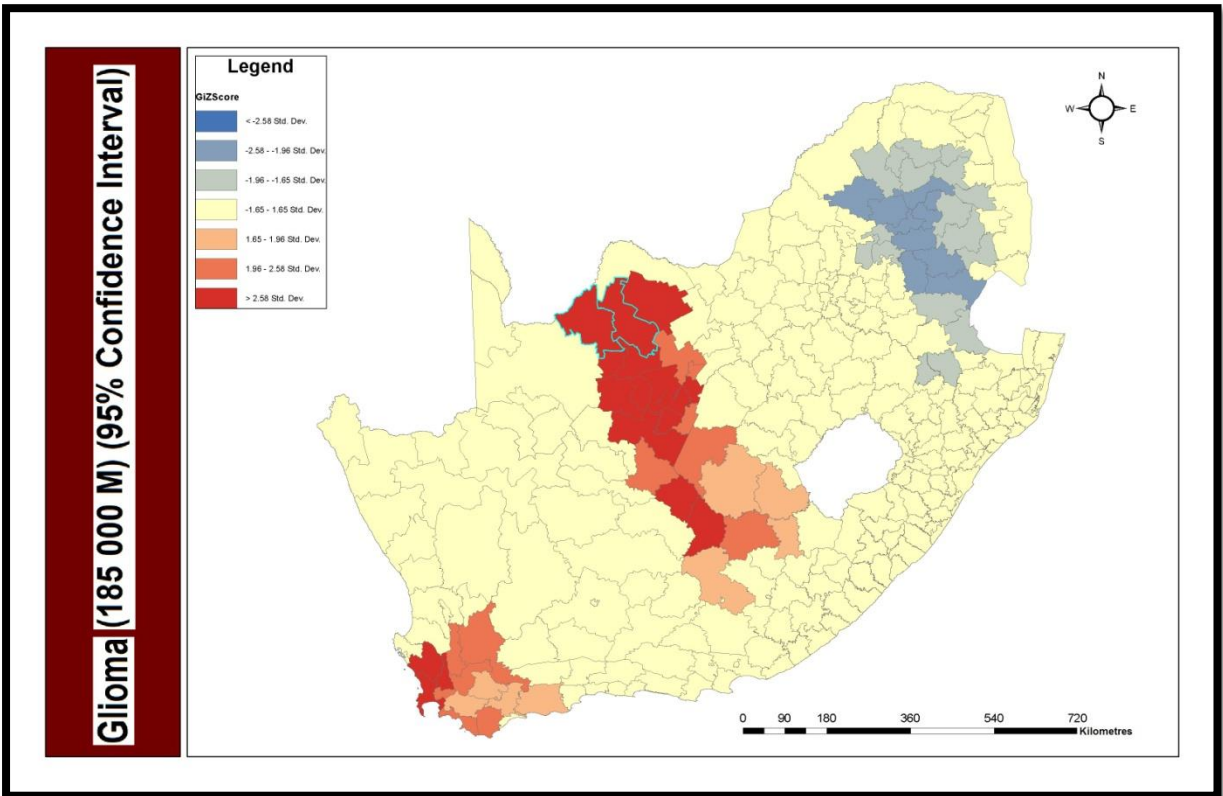


Figure 4.14: Accurate result for Glioma on a 95% confidence interval after the Bonferroni Correction was performed. (No hotspots were found on a 99% confidence interval for Glioma after the Bonferroni Correction was performed.)

#### 4.3.2.4 Hepatoblastoma

The general result for Hepatoblastoma indicated hotspots in the western part of Kwa-Zulu Natal, the eastern part of the Northern Cape and the central part of the Western Cape. According to the literature review the possible environmental factors that could contribute to the development of Hepatoblastoma include the inhalation of arsenic oxides; the exposure to Phthalate; prenatal tobacco smoking; alcohol consumption; obesity and the Hepatitis B Virus and Hepatitis C Virus. It is therefore worth investigating whether any of these variables are present within the identified hotspot areas. There were no hotspots found on a 99% or 95% confidence interval for Hepatoblastoma after the Bonferroni Correction was performed.

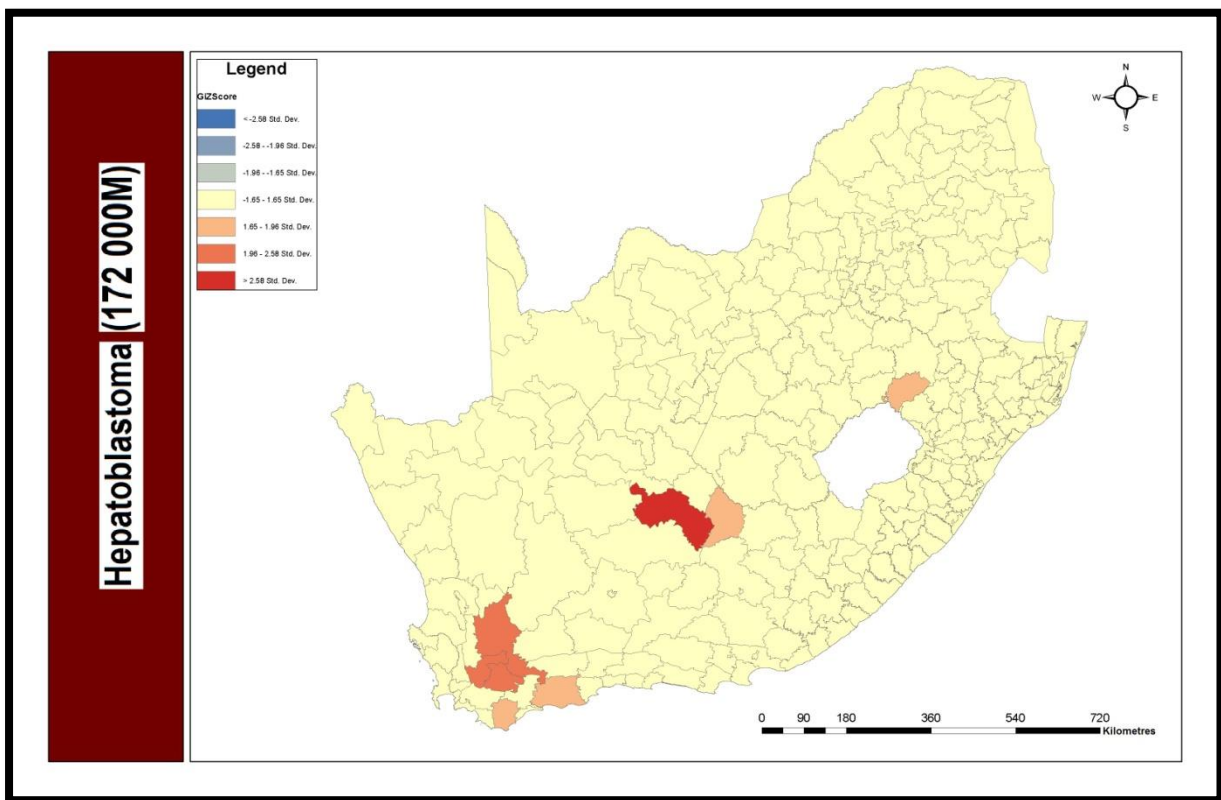


Figure 4.15: General result for Hepatoblastoma (No hotspots were found on a 99% or 95% confidence interval for Hepatoblastoma.)



### 4.3.2.5 Kaposi Sarcoma

The general result for Kaposi Sarcoma indicated hotspots in the Free State, Kwa-Zulu Natal and the Western Cape, as well as one coldspot in the northern part of Kwa-Zulu Natal (as seen in figure 4.16). There were no hotspots found on a 99% or 95% confidence interval after to Bonferroni Correction was performed for this cancer type. According to the literature review Kaposi Sarcoma seems to be higher in areas with widespread wetlands and fertile volcanic clay soils. Kaposi Sarcoma also seems to be present in areas where malaria is endemic. It is therefore worth investigating whether any of these factors are present within the identified hotspot areas.

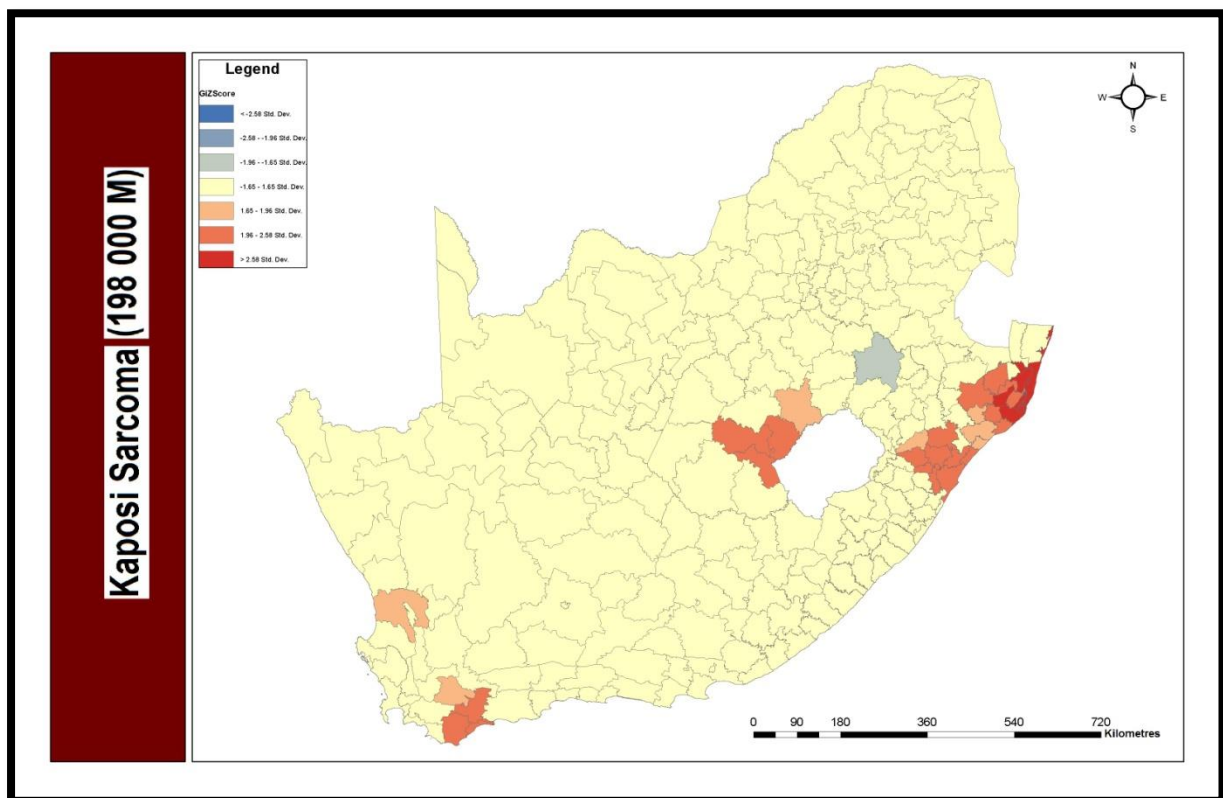


Figure 4.16: General result for Kaposi Sarcoma (No hotspots were found on a 99% or 95% confidence interval for Kaposi Sarcoma.)

### 4.3.2.6 Neuroblastoma

The general result for Neuroblastoma indicated that there were hotspots in the central part of the Eastern Cape as well as the eastern part of the Northern Cape. The result also indicated coldspots in Limpopo, Kwa-Zulu Natal and the Eastern Cape (as seen in figure 4.17). There were no hotspots found on a 99% or 95% confidence interval for Neuroblastoma after the Bonferroni Correction was performed. According to the literature review the exposure to pesticides and Trichloroethylene significantly increases the likelihood of developing Neuroblastoma. The literature review also revealed that high birth weight significantly increases the chances of developing Neuroblastoma. It is therefore worth investigation whether any of these variables are present within the identified hotspot areas.

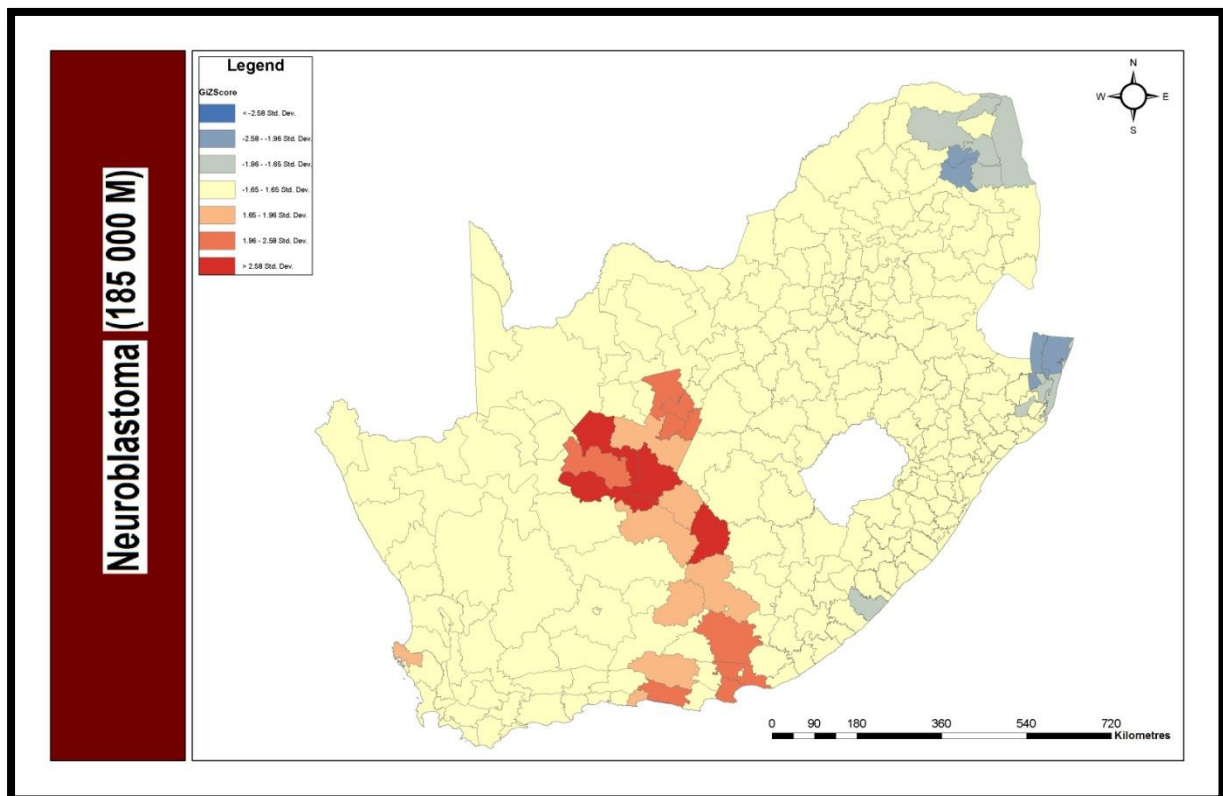


Figure 4.17: General result for Neuroblastoma (No hotspots were found on a 99% or 95% confidence interval for Neuroblastoma after the Bonferroni correction was performed).

#### 4.3.2.7 Nephroblastoma (Wilm's Tumour)

The general result for Nephroblastoma revealed hotspots in the southern parts of Northwest, the northern parts of the Free State and the north western parts of the Northern Cape. The result also indicated coldspots in Limpopo and Kwa-Zulu Natal (as seen in figure 4.18). There were no hotspots found on a 99% or 95% confidence interval for Nephroblastoma after the Bonferroni Correction was performed. The literature review revealed that both pesticide exposure and Trichloroethylene exposure were associated with the development of Nephroblastoma. Nephroblastoma was also associated with the exposure to lead; cadmium and hydrocarbons. The literature review also revealed a particularly strong association between Nephroblastoma and paternal employment as a welder or mechanic. High birth weight was also found to be a significant contributor to the development of Nephroblastoma. It is therefore worth investigating whether any of these variables are present within the identified hotspot areas.

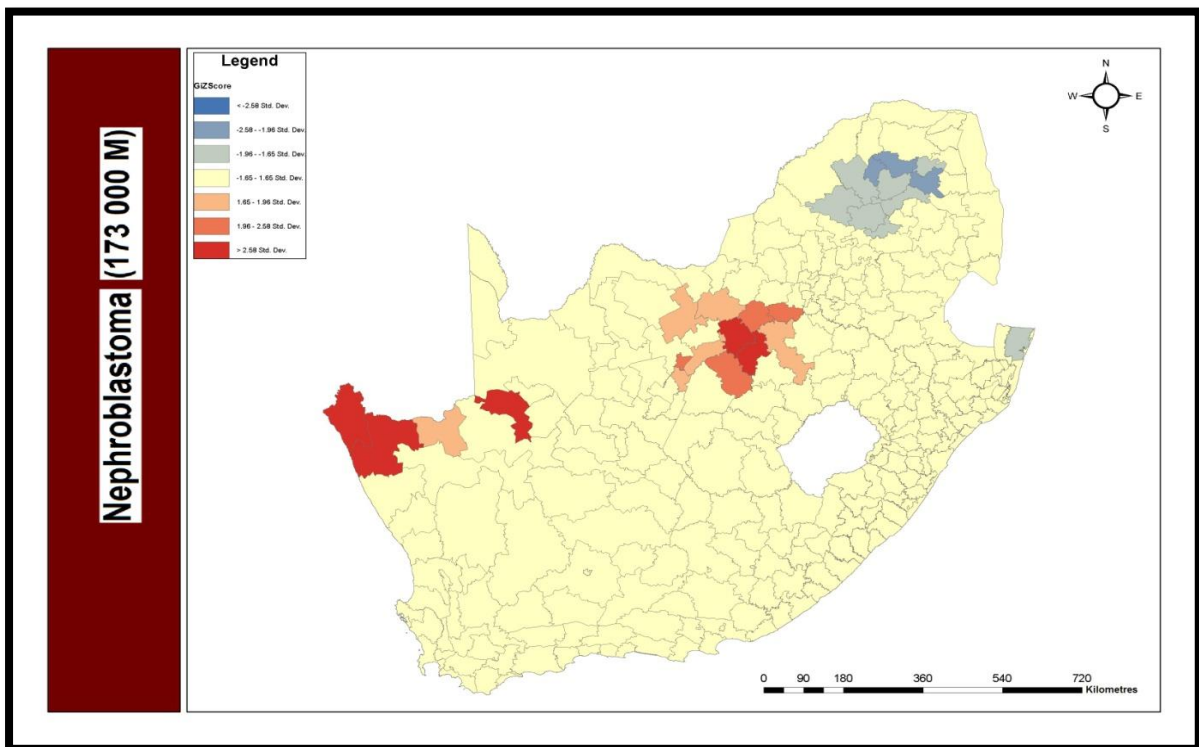


Figure 4.18: General result for Nephroblastoma (No hotspots were found on a 99% or 95% confidence interval for Nephroblastoma after the Bonferroni Correction was performed).



#### 4.3.2.8 Osteosarcoma

The general result for Osteosarcoma indicated hotspots in the western parts of Northwest, as well as the western parts of the Western Cape. The result also indicated coldspots in the eastern parts of the Northern Cape as well as the central parts of Limpopo (as seen in figure 4.19). There were no hotspots identified on a 99% or 95% confidence interval for Osteosarcoma after the Bonferroni Correction was performed. The literature review revealed that there was a strong association between the development of Osteosarcoma and the treatment of previous cancers with radiation. It was also found that the exposure to pesticides and radium as well as the intake of fluoride increased the likelihood of developing Osteosarcoma. It is therefore worth investigating whether any of these factors are present within the identified hotspot areas.

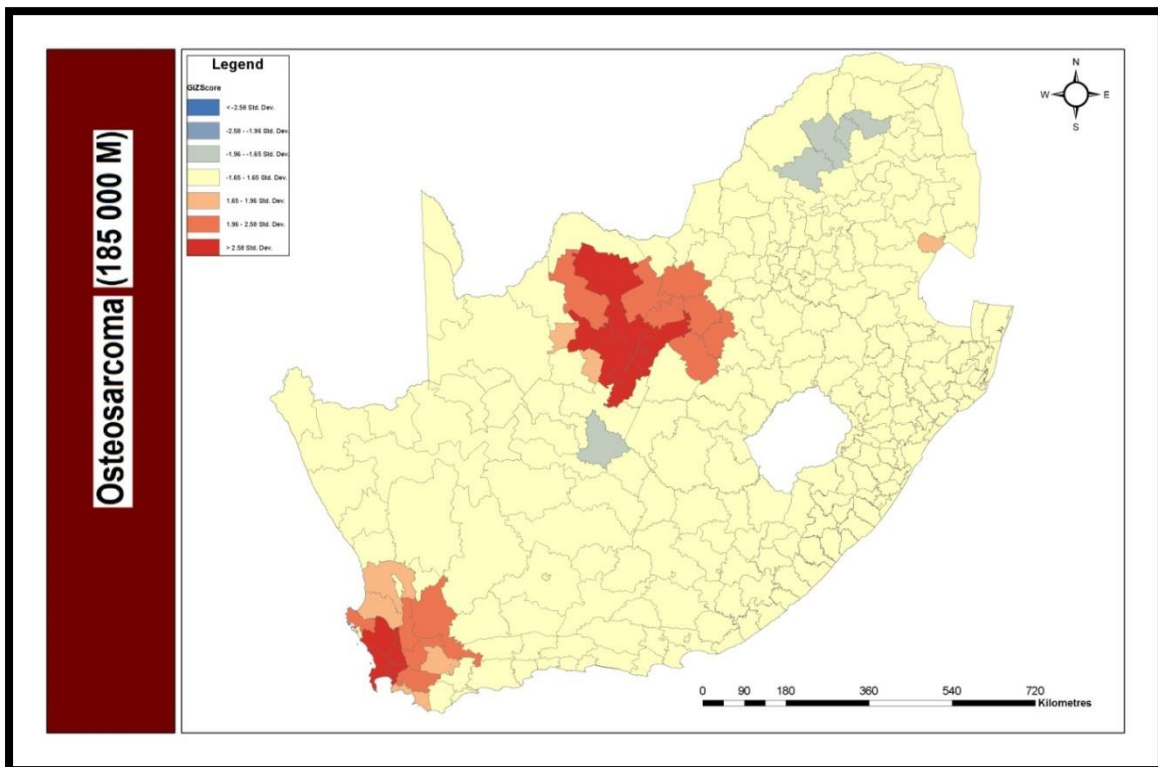


Figure 4.19: General result for Osteosarcoma (No hotspots were found on a 99% or 95% confidence interval for Osteosarcoma after the Bonferroni Correction was performed.)

### 4.3.2.9 Retinoblastoma

The general result for Retinoblastoma indicated hotspots in the northern parts of the Free State and in the central and eastern parts of the Northern Cape. The result also indicated coldspots in Kwa-Zulu Natal and the Eastern Cape (as seen in figure 4.20). The Bonferroni Correction revealed a hotspot based on a 99% confidence interval in the central part of the Northern Cape (as seen in figure 4.2). The literature review revealed that there was limited evidence of environmental factors contributing to the development of Retinoblastoma. Two of the possible factors that were associated with the development of Retinoblastoma included maternal use of multivitamin supplements and the Human Papilloma Virus.

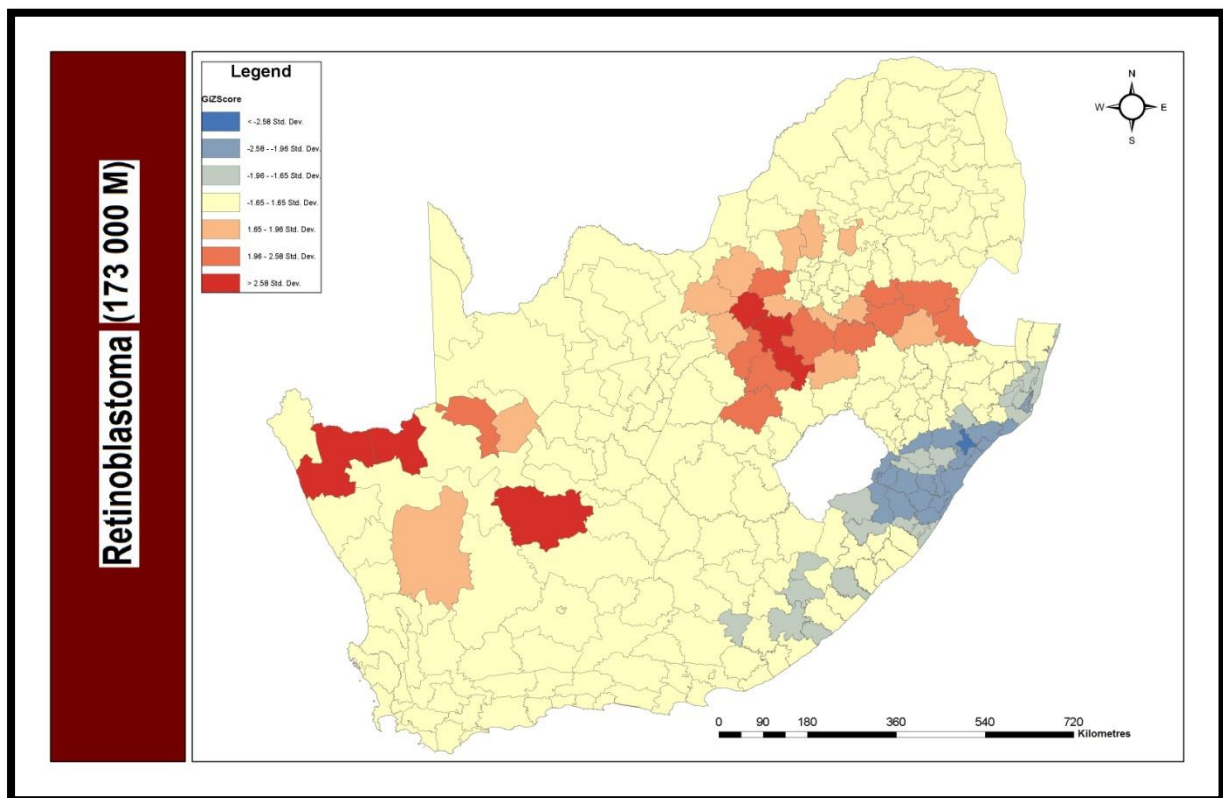


Figure 4.20: General result for Retinoblastoma before the Bonferroni Correction was performed.

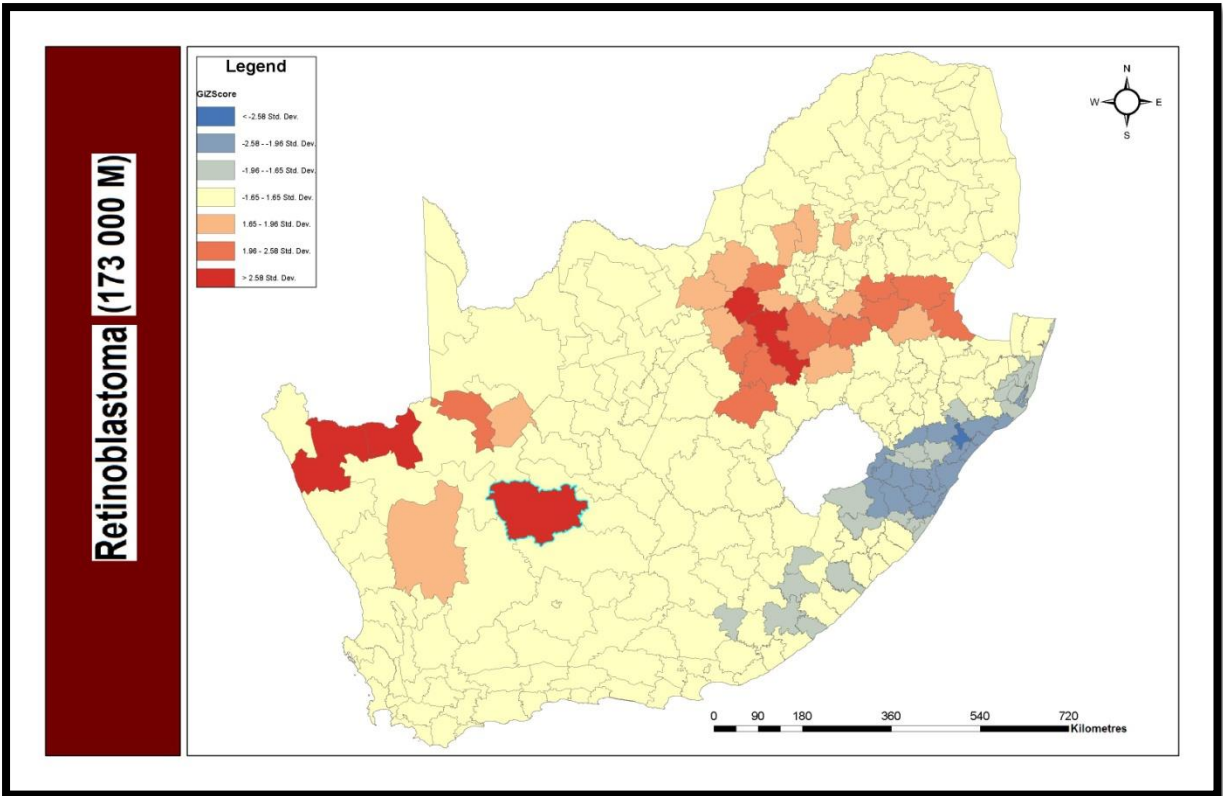


Figure 4.21: Accurate result for Retinoblastoma on a 99% confidence interval after the Bonferroni Correction was performed.

## **CHAPTER 5: CONCLUSION AND RECOMMENDATIONS**

### **5.1 Conclusion**

The main purpose of this dissertation was to develop and validate a methodology to scientifically identify spatial clusters (hotspots) of various cancer types within South Africa by using Esri GIS software. This purpose was achieved by identifying several spatial clusters of various cancer types within South Africa. The process was validated and its limitations identified and mitigated.

The central problem identified in this dissertation was the lack of quality within the cancer data used. It was determined that the South African cancer incidences for children under the age of 15 were probably significantly underreported. Various possible reasons for this problem were discussed and it was determined that the data was possibly not representative of the actual situation within South Africa. Cancer ratios per population were identified as the best method to use in order to determine where cancer clusters exist. It was found that this would be more scientific than simply using the amount of cancer incidences in a specific area, since this would cause biased results as one would expect to find more cancer incidences in areas where more people lived.

It was determined that census data would be needed in order to calculate these ratios according to the population of certain areas. The problem experienced in this regard was that the cancer data contained incidences from 1992 to 2008. Statistics South Africa was however only able to provide the census data for 1996, 2001 and the community survey of 2007. This limited the amount of incidences that could be used in the study to only those that occurred between 1996 and 2007. The incidences prior to 1996 and after 2007 were therefore not included in the study. It was determined that the census data prior to 1994 could not be converted to the 2005 local municipal boundaries (used in the study) because it was obtained in such a manner that prevented Statistics South Africa from converting it.

The census data that was obtained after 2007 was also not used since it was not yet available at the time of the study. It was determined that the local municipal boundaries were the best classification to use in order to determine the locations of the hotspots. A significant percentage of the data lacked sufficient geographic information which made plotting it on a smaller scale than the local municipal boundaries impossible since the exact location of each cancer incidence would then be needed. This study found that the local municipal boundaries tend to shift every few years between census counts but with the help of Statistics South Africa the 1996 and 2001 census data as well as the 2007 community survey data were converted to the local municipal boundaries as they were in 2005.

This made it possible to compare the population statistics of 1996, 2001 and 2007 because they all had identical geographical boundaries (2005 local municipal boundaries). The average population between the ages of 0 and 15 years for each municipal polygon (for the time period between 1996 and 2007) was then calculated. This average population statistic was then used together with the cancer incidence data of each municipal polygon (for the time period of 1996 to 2007) to determine the cancer ratio per population (0-15 years) for each cancer type and for each municipal polygon. This ratio was then used as the input for the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool to determine where cancer clusters exist.

The GIS technology was found to be capable of determining scientifically significant clusters (hotspots) of cancer within South Africa. It was found that the GIS technology was not the problem but rather the quality of the cancer data within the South African Paediatric Tumour Registry. Before the quality of South Africa's cancer data is not significantly improved a study like this one cannot truly be done since the results of such a study will always be based on questionable data no matter how scientific the methodology is. A study similar to this one based on high quality data could be truly invaluable in the fight against cancer. If one can identify where cancer occurs more often and one considers the known or suspected environmental and other factors that could facilitate the development of that cancer, one can truly start to determine or even predict where that cancer is likely to occur in the future. These environmental and other factors can then be mitigated and lives can be saved. In

order to save lives we first need to know what is causing the problem before we can fix it. The geographical location of cancers clusters (hotspots) can focus efforts to where it matters.

Although the quality of the cancer data used in this study is questionable it should be noted that it is still the best available data to use at this time and that it is still a sample representing cancer incidences in South Africa even if it is underreported. The results of the Hotspot Analysis (Getis-Ord  $G_i^*$ ) Tool should therefore not be completely ignored.

## **5.2 Recommendations**

It is recommended that the low quality of South Africa's cancer data be further investigated. Only with high quality data will high quality results be obtained and only with high quality results will one truly be able to make a difference.

It is recommended that the geographic location of each child diagnosed with cancer should always be noted in as much detail as possible. If future studies like this one is to succeed, this aspect of South Africa's cancer data will have to be vastly improved since only 27% of the data used in this study had precise geographic information linked to the diagnosed patient. It is recommended that the latitude and longitude coordinates of each cancer patient's residential location be noted rather than the physical address of the patient. This is particularly highly recommended in South Africa because there are many rural villages where no formal street addresses exist. Cancer incidences in these areas might therefore be missed. By using the latitude and longitude coordinates it will ensure that the data is highly accurate geographically. This will make it possible to plot each cancer case at its exact location.

This will enable a future study to conduct an analysis on a much smaller scale. The more accurate one can determine where hotspots are the more focussed one's

efforts can be and the better the chances are that carcinogenic environmental factor will be identified. It is also recommended that the limitations of the GIS Tools used be understood and mitigated to ensure that the results produced are scientifically significant.

## REFERENCES

ABDULAH, R., MIYAZAKI, K., NAKAZAWA, M. & KOYAMA, H. 2005. Chemical forms of selenium for cancer prevention. *Journal of Trace Elements in Medicine and Biology*, 19:141-150.

AHLBOM, A., CARDIS, E., GREEN, A., LINET, M., SAVITZ, D. & SWERDLO W. 2001. Review of epidemiological literature on EMF and health. *Environ Health Perspect*, 109:911-33.

ALBERT, D.M., LAHAR, M., LESSER, R. & CRAFT, J. 1974. Recent observations regarding retinoblastoma: I. Ultrastructure tissue, culture growth, incidence, and animal models. *Trans. Ophthalmol. Soc. U.K*, 94:909–927.

ALBRECHT, C. 2006. Overview of the South African Cancer Research Environment as basis for discussions concerning the activation of Carisa (Cancer Research Initiative of South Africa) <http://www.sahealthinfo.org/cancer/overviewdocument.pdf> (Date of access: 17 Oct. 2012.).

ALDERTON, L.E., SPECTOR, L.G., BLAIR, C.K., ROESLER, M., OLSHAN, A.F., ROBISON, L.L., ROSS, J.A. 2006. Child and maternal household chemical exposure and the risk of acute leukaemia in children with Down's syndrome: a report from the Children's Oncology Group. *Am. J. Epidemiol*, 164:212–221.

AMERICAN CANCER SOCIETY. 2010. Radiation Exposure and Cancer. <http://www.cancer.org/Cancer/CancerCauses/OtherCarcinogens/MedicalTreatments/radiation-exposure-and-cancer> (Date of access: 30 Junly 2012).

AMERICAN CANCER SOCIETY. 2012. Hodgkin Disease Overview: What is Hodgkin Disease. <http://www.cancer.org/cancer/hodgkindisease/overviewguide/hodgkin-disease-overview-what-is-hodgkin-disease> (Date of access: 10 Nov. 2012).

AMERICAN CANCER SOCIETY. 2013a. Non-Hodgkin's Lymphoma. <http://www.cancer.org/acs/groups/cid/documents/webcontent/003126-pdf.pdf> (Date of access: 10 November 2013).

AMERICAN CANCER SOCIETY. 2013b. Epidemic (AIDS-related) Kaposi sarcoma. <Http://www.cancer.org/cancer/kaposisarcoma/detailedguide/kaposi-sarcoma-what-is-kaposi-sarcoma> (Date accessed: 15 July 2013).



AMERICAN CANCER SOCIETY. 2013c. What are the risk factors for Osteosarcoma? <http://www.cancer.org/cancer/osteosarcoma/detailedguide/osteosarcoma-risk-factors> (Date of access: 4 November 2013)

ASCOLI, V., ZAMBON, P., MANNO, D., GUZZINATI, S., ZORZI, M. & ARCA, B. 2003. Variability in the incidence of classic Kaposi Sarcoma in the Veneto region, Northern Italy. *Tumouri*, 89:122–124.

BAILEY, T.C. & GATRELL, A.C. 1995. *Interactive Spatial Data Analysis*. London: Longman

BAYOUMY, K.E. 2001. The protective role of selenium on genetic damage and on cancer. *Mutation Research*, 475:123-139.

BELLAR, T.A. 1974. The Occurrence of Organohalides in Chlorinated Drinking Waters. *Journal of the American Water Works Association*, 66(703):706.

BELSON, M., KINGSLEY, B. & HOLMES, A. 2007. Risk factors for acute leukaemia in children: a review. *Environ. Health Perspect*, 115:138–145.

BENEZRA, D. & CHIRAMBO, M.C. 1976. Incidence of retinoblastoma in Malawi. *J. Pediatr. Ophthalmol*, 13:340– 343.

BENNET, D.A. 2001. How can I deal with missing data in my study? *Aust N Z J Public Health*, 25(5):464-9.

BERKEL, J. & BAKO, G. 1992. Selenium and Cancer: Overview and Data from Alberta. *Western Geographic Series*, 27: 78.

BERNHARDSEN, T. 2002. *Geographic Information Systems: An Introduction*. New York: John Wiley & Sons, Inc.

BIDOLI E, FRANCESCHI S, DAL MASO L, GUARNERI S, BARBONE F. 1993. Cancer mortality by urbanization and altitude in a limited area in North-eastern Italy. *Rev EpidemiolSantePublique*, 41:374–82.

BIEGING, K.T., SWANSON-MUNGERSON, M., AMICK, A.C. & LONGNECKER, R. 2010. Epstein-Barr Virus and Burkitt's Lymphoma: A role for Latent Membrane Protein 2A. *Cell Cycle*, 9(5):901-908.

BIRCH, J.M. 2011. Epidemiology of pediatric liver tumours, (In Zimmermann A, Perilongo G, Malogolowkin M, von Schweinitz D (eds): Pediatric Liver Tumours. Heidelberg: Springer.)

BOCCOLINI, P.D.M., BOCCOLINI, C.S., CHRISMA, J.R., MARKOWITZ, S.B., KOIFMAN, S., KOIFMAN, R.J. & MEYER, A. 2013. Pesticide use and Non-Hodgkin's Lymphoma mortality in Brazil. *International Journal of Hygiene and Environmental Health*, 216:461-466.

BOLOGNESI, C. & MERLO, F.D. 2011. Pesticides: Human Health Effects. *Encyclopaedia of Environmental Health*, 438-453.

BONAITI-PELLIE, C., BRIARD-GUILLEMOT, M.L. 1981. Segregation analysis in hereditary retinoblastoma. *Hum Genet*, 57:411-9.

BOSETTI, C., BOFFETTA, P. & LA VECCHIA, C. 2006. Occupational Exposures to Polycyclic Aromatic Hydrocarbons and Respiratory and Urinary Tract Cancers: a Quantitative Review to 2005. *Annals of Oncology*, 18:431-446.

BOYLE, P. & LEVIN, B. 2008. World cancer report 2008. Lyon: IARC Press.

BOYLE, P. & SMANS, M. 2008. Atlas of cancer mortality in the European Union and the European economic area. France, Lyon: International Agency for Research on Cancer.

BREWER, C.A. 2006. Basic mapping principles for visualizing cancer data using geographic information systems (GIS). *American journal of preventive medicine*, 30(2s):25-36.

BUCKEY, D.J. 2013. The GIS Primer: An introduction to Geographic Information Systems. [http://bgis.sanbi.org/gis-primer/page\\_19.htm](http://bgis.sanbi.org/gis-primer/page_19.htm) (Date of access: 6 October 2013).

BUCKLEY, J.D., MEADOWS, A.T., KADIN, M.E., LE BEAU, M.M., SIEGEL, S., ROBISON, L.L. 2000. Pesticide exposures in children with non-Hodgkin lymphoma. *Cancer*, 89:2315-2321.

BUCKLEY, J.D., PENDERGRASS, T.W., BUCKLEY, C.M., PRITCHARD, D.J., NESBIT, M.E., PROVVISOR, A.J. & ROBISON, L.L. 1998. Epidemiology of Osteosarcoma and Ewing's sarcoma in childhood: a study of 305 cases by the Children's Cancer Group. *Cancer*, 83:1440–1448.

BUCKLEY, J.D., ROBISON, L.L., SWOTINSKY, R., GARABRANT, D.H., LEBEAU, M., MANCHESTER, P., NESBIT, M.E., ODOM, L., PETERS, J.M., WOODS, W.G., HAMMOND, G.D., 1989. Occupational exposures of parents of children with acute nonlymphocytic leukaemia: a report from the Children's Cancer Study Group. *Cancer Res.* 49, 4030–4037.

BUFFLER, P.A., KWAN, M.L., REYNOLDS, P. & URAYAMA, K.Y., 2005. Environmental and genetic risk factors for childhood leukaemia: appraising the evidence. *Cancer Invest*, 23:60–75.

BUNDERSON-SCHELVAN, M., PFAU, J., CROUCH, R. & HOLIAN, A. 2011. Nonpulmonary Outcomes of Asbestos Exposure. *Journal of Toxicology and Environmental Health, Part B, Critical Reviews*, 14(1-14): 122-152.

BUNIN, G.R., BUCKLEY, J.D., BOESEL, C.P., RORKE, L.B., MEADOWS, A.T., 1994. Risk factors for astrocytic glioma and primitive neuroectodermal tumour of brain in young children: a report from the Children's Cancer Group. *Cancer Epidemiol., Biomarkers Prev*, 3:197–204.

BUNIN, G.R., KRAMER, S., MARRERO, O. & MEADOWS, A.T. 1987. Gestational risk factors for Wilms' tumour: results of a case-control study. *Cancer Res*, 47:2972–2977.

BUNIN, G.R., MEADOWS, A.T., EMANUEL, B.S., BUCKLEY, J.D., WOODS, W.G. & HAMMOND, G.D. 1989a. Pre- and post-conception factors associated with heritable and non-heritable retinoblastoma. *Cancer Res*, 49:5730–5735.

BUNIN, G.R., NASS, C.C., KRAMER, S. & MEADOWS, A.T. 1989b. Parental occupation and Wilms' tumour: results of a case-control study. *Cancer Res*: 49, 725–729.

BUNIN, G.R., WARD, E., KRAMER, S., RHEE, C.A., MEADOWS, A.T. 1990. Neuroblastoma and parental occupation. *Am J Epidemiol*, 131:776-80.

BURT, J.E. & BARBER, G.M. 1996. *Elementary Statistics for Geographers*. New York: Guilford Press.

CALVENTE, I., FERMANDEZ, M.F., VILLALBA, J., OLEA, N. & NUNEZ, M.I. 2010. Exposure to electromagnetic fields (non-ionizing radiation) and its relationship with childhood Leukaemia : A systematic review. *Science of the Total Environment*, 408:3062-3069.

CANTOR, K.P., WARD, M.H., MOORE, L. & LUBIN, J. 2006. Water contaminants.part 20.In: Schottenfeld D, Fraumeni Jr F, editors. *Cancer epidemiology and prevention*. 3rd ed. New York: Oxford University Press.

CAROZZA, S.E., LI, B., ELGETHUN, K. & WHITWIRTH, R. 2008. Risk of childhood cancers associated with residence in agriculturally intense areas in the United States. *Environ Health Perspect*, 116(4):559-565.

CAROZZA, S.E., LI, B., WANG, Q., HOREL, S. & COOPER, S. 2009. Agricultural pesticides and risk of childhood cancers. *International Journal of Hygiene and Environmental Health*, 212:186-195.

CARREON, T., BUTLER, M.A., RUDER, A.M., WATERS, M.A., VIS-KING, K.E., CALVERT, G.M., SCHULTE, P.A., CONNALLY, B., WARD, E.M., SANDERSON, W.T., HEINEMAN, E.F., MANDEL, J.S., MORTON, R.F., REDING, D.J., ROSENMAN, K.D., TALASKA, G. & CANCER, B. 2005. Gliomas and farm pesticide exposure in women: the Upper Midwest Health Study. *Environ Health Perspect*, 113:546-51.

CHRISTENSEN, J.S., MORTENSEN, L.H., ROOSLI, M., FEYCHTING, M., TYNES, T., ANDERSEN, T.V., SCHMIDT, L.S., POULSEN, A.H., AVDIN, D., KUEHNI, C.E., PROCHAZKA, M., LANNERING, B., KLAEBOE, L., EGGEN, T. & SCHUZ, J. 2012. Brain tumors in childhood and adolescents and exposure to animals and farm life: a multicentre case-control study (CEFALO). *Cancer Causes Control*, 23(9):1463-73.

CHRISTENSEN, K.Y., VIZCAYA, D., RICHARDSON, H., LAVOUE, J. AROSEN, K. & SIEMIATYCKI, J. 2013. Risk of selected cancers due to occupational exposure to chlorinated solvents in a case-control study in Montreal. *J Occup Environ Med*, 55:198-208.

CHU, A., HECK, J.E., RIBEIRO, K.B., BRENNAN, P., BOFFETTA, P., BUFFLER, P., HUNG, R.J. 2010. Wilms' tumour: a systematic review of risk factors and meta-analysis. *Paediatric and Perinatal Epidemiology*, 24(5): 449-69.

CLAPP, R.W., JACOBS, M.M. & LOECHLER, E.L. 2008. Environmental and Occupational causes for cancer: New evidence 2005-2007. *Rev Environ Health*, 23:1-37.

COLEMAN, K., HARVEY, C. & WESTON, A. 2007. A systematic review of the efficacy and safety of fluoridation. Melbourne, Australia: Australian National Health and Medical Research Council.

COMBER, H., DEADY, S., MONTGOMERY, E., GAVIN, A. 2011. Drinking water fluoridation and Osteosarcoma incidence on the island of Ireland. *Cancer Causes Control*, 22(6):919–24.

COOK-MOZAFFARI, P., NEWTON, R., BERAL, V. & BUKITT, D.P. 1998. The geographical distribution of Kaposi Sarcoma and of lymphomas in Africa before the AIDS epidemic. *Br. J. Cancer*, 78:1521–1528.

COONEY, M.A., DANIELS, J.L., ROSS, J.A., BRESLOW, N.E., POLLOCK, B.H. & OLSHAN, A.F. 2007. Household pesticides and the risk of Wilms tumor. *Environ Health Perspect*, 115(1):134-7.

COPEL, J. & HAN, C.S. 2012. Obstetric Imaging. Philadelphia: Elsevier Saunders.

COROZZA, S.E., LI, B., ELGETHUN, K. & WHITWORTH, R. 2008. Risk of childhood cancers associated with residence in agriculturally intense areas in the United States. *Environ Health Perspect*, 116(4):559-65.

DAL MASO, L., POLESEL, J., ASCOLI, V., ZAMBON, P., BUDRONI, M. & FERRETTI, S. 2005. Classic Kaposi Sarcoma in Italy, 1985–1998. *Br. J. Cancer*, 92:188–193.

DALING, J.R., STARZYK, E., OLSHAN, A.F. & WEISS, N.S. 1984. Birthweight and the incidence of childhood cancer. *J Natl Cancer Inst*, 72:1039 -41.

DANIELS, J.L., OLSHAN, A.F. & SAVITZ, D.A. 1997. Pesticides and childhood cancers. *Environ. Health Perspect*, 105:1068–1077.

DANIELS, J.L., OLSHAN, A.F., POLLOCK, B.H., SHAH, N.R., STRAM, D.O. 2002. Breast-feeding and neuroblastoma, USA and Canada. *Cancer Causes Control*, 13:401–5.

DANIELS, J.L., OLSHAN, A.F., TESCHKE, K., HERTZ-PICCIOTTO, I., SAVITZ, D.A., BLATT, J. BONDY, M.L., NEGLIA, J.P., POLLOCK, B.H., COHN, S.L., LOOK, A.T., SEEGER, R.C.,

CASTLEBERRY, R.P. 2001. Residential pesticide exposure and neuroblastoma. *Epidemiology*, 12:20–27.

DELPOMME, D., IRIGARAY, P., SASCO, A.J., NEWBY, J.A., HOWARD, V., CLAPP, R. & HARDELL, L. 2007. The growing incidence of cancer: role of lifestyle and screening detection (Review). *International journal of oncology*, 30:1037-1049.

DUBE, C. 2005. Census Geography of South Africa. <http://www.statssa.gov.za> (Date of access: 17 October 2012.)

EARICKSON, R.J. & HARLIN, J.M. 1994. Geographic Measurements and Quantitative Analysis. London: Macmillan.

EBDON, D. 1985. Statistics in Geography. Oxford: Blackwell Publishing.

EBLE, J.N., SAUTER, G., EPSTEIN, J.I. & SESTERHENN, I.A. 2004. Pathology & Genetics: Tumours of the Urinary System and male Genital Organs. Lyon, France: IARC Press.

ECSA (EUROPEAN CHLORINATED SOLVENT ASSOCIATION). 2011. Health Profile of Trichloroethylene. <http://www.eurochlor.org/media/57105/health-profile-tri.pdf> (Date of access: 3 October 2013).

EFIRD, J.T., HOLLY, E.A., PRESTON-MARTIN, S., MUELLER, B.A., LUBIN, F., FILIPPINI, G., PERIS-BONET, R., MCCREDIE, M., CORDIER, S., ARSLAN, A., BRACCI, P.M., 2003. Farm-related exposures and childhood brain tumours in seven countries: results from the SEARCH International Brain Tumour Study. *Paediatr.Perinat. Epidemiol*, 17:201– 211.

ELIA-PASQUET, S., PROVOST, D. & JAFFRE, A. 2004. Incidence of central nervous system tumours in Gironde, France. *Neuroepidemiology*, 117:23110-117.

Enders, C.K. 2010. Applied Missing Data Analysis. New York: Guilford Press.

ENGLISH, D.R., HOLMAN, C.D.J., MILNE, E., WINTER, M.G., HULSE, G.K. & CODDE, J.P. 1995. The quantification of drug caused morbidity and mortality in Australia. Sydney: Commonwealth Department of Human Services and Health.

ESRI. 2009. Hot Spot Analysis (Getis-OrdGi\*: Spatial Statistics).<http://webhelp.esri.com/arcgisdesktop/9.9/> (Date of access: 9 Oct. 2012).

ESRI. 2012a. ArcGIS Resources: Mapping Centre. <http://mappingcentre.esri.com> (Date of access 27 Nov. 2012.)

ESRI. 2012b. ArcGIS Resources: What are map projections? <http://resources.arcgis.com> (Date of access: 12 Nov. 2012.)

ESRI. 2013a. GIS Dictionary: Vector data model. <http://support.esri.com/en/knowledgebase/GISDictionary/term/vector%20data%20model> (Date of access: 6 October 2013).

ESRI. 2013b. GIS Dictionary: Raster data model. <http://support.esri.com/en/knowledgebase/GISDictionary/term/raster%20data%20model> (Date of access: 6 October 2013).

FADERL, S. & KANTARJIAN, H. 2011. Leukaemias: Principles and Practice of theory. Chichester, West Sussex: Blackwell Publishing Ltd.

FARLEX. 2012a. The Free Dictionary: Glioma. <http://medical-dictionary.thefreedictionary.com/glioma> (Date of access: 12 Oct. 2012).

FARLEX. 2012b. The Free Dictionary: Hepatoblastoma. <http://medical-dictionary.thefreedictionary.com/hepatoblastoma> (Date of access: 16 Oct. 2012).

FASAL, E., JACKSON, E.W. & KLAUBER, M.R. 1971. Birth characteristics and leukaemia in childhood. *J Nat Cancer Inst*, 47:501-9.

FEAR, N.T., ROMAN, E., REEVES, G. & PANNETT, B. 1998. Childhood cancer and paternal employment in agriculture: the role of pesticides. *Br. J. Cancer*, 77 (5):825–829.

FEDRICK, J., ALBERMAN, E.D. 1972. Reported influenza in pregnancy and subsequent cancer in the child. . *Br. Med. J*, 2: 485–488.

FERRARI, G. & LOVASTE, M.G. 1986. Primary intracranial tumours in the province of Trento-Italy (1977–1984). Distribution by administrative units. *Neuroepidemiology*, 5:159-70.

FERRI, F.F. 2011. Practical Guide to the care of the medical patient. Philadelphia: Mosby Elsevir.

FERRY, J.A. 2006. Burkitt's Lymphoma: Clinicopathologic Features and Differential Diagnoses. *The Oncologist*, 11:375-383.

FINKELSTEIN, M.M., KREIGER, N. 1996. Radium in drinking water and risk of bone cancer in Ontario youths: a second study and combined analysis. *Occup. Environ. Med*, 53:305– 311.

FLORENZA, P. 2012. Top 10 Benefits of GIS Technology. <http://www.govloop.com/profiles/blogs/top-10-benefits-of-gis-technology> (Date of access: 6 October 2013).

FORD, P.W., HAMDEN, K.E., WHITMAN, A.G., BRYAN, B.A., CHINTALGATTY, V., MCCUBREY, J.A., DYSON, O.F., AKULA, S.M. 2005. Cigarette smoke concentrate inhibits Kaposi Sarcoma associated Herpesvirus infection. *Virus Research*, 114(2):172-176.

FRAUMENI, J.F. & MILLER, R.W. 1967. Epidemiology of human leukaemia: recent observations. *J. Natl. Cancer Inst*, 38:593– 605.

FRAUMENI, J.F. 1967. Stature and malignant tumours of bone in childhood and adolescence. *Cancer* 20, 967– 973.

FREEDMAN, J. & GOLDBERG, L. 1976. Incidence of retinoblastoma in the Bantu of South Africa. *Br. J. Ophthalmol*, 60:655– 656.

FRENI, S.C., GAYLOR, D.W. 1992. International trends in the incidence of bone cancer are not related to drinking water fluoridation. *Cancer*, 70(3):611–8.

FRENZEL-BEYME, R., LEUTER, R., WAGNER, G. & WIEBELT, H. 1979. *Krebatlas der Bundesrepublik Deutschland*. Berlin: Springer-Verlag

GEDDES, M., FRANCESCHI, S., BALZI, D., ARNIANI, S., GAFA, L. & ZANETTI, R. 1995. Birthplace and classic Kaposi Sarcoma in Italy, *J. Natl Cancer Inst*, 87:1015–1017.



GELBERG, K.H., FITZGERALD, E.F., HWANG, S.A. & DUBROW, R. 1997. Growth and development and other risk factors for Osteosarcoma in children and young adults. *Int. J. Epidemiol*, 26: 272– 278.

GEORGE, B. & KORTE, P.E. 2001. The GIS Book: How to implement, manage and assess the value of Geographic Information Systems. 5<sup>th</sup> Ed. New York: On Word Press.

GEORGE, J. & SHUKLA, Y. 2011. Pesticides and Cancer: Insights into toxicoproteomic-based findings. *Journal of Proteomics*, 74:2713-2722.

GETIS, A. & ORD, J.K. 1996. Local Spatial Statistics: An overview. New York: Wiley Publishing.

GETIS, A. 1991. Spatial Interaction and Spatial Autocorrelation: A Cross-Product Approach. *Environment and Planning*, 23:1269-77.

GIS AND SCIENCE. 2009. Top five benefits of GIS. <http://gisandscience.com/2009/09/14/top-five-benefits-of-gis> (Date of access: 6 October 2013).

GOEDERT, J.J., VITALE, F., LAURIA, C., SERRAINO, D. TAMBURINI, M. & MONTELLA, M. 2002. Classical kaposi's sarcoma working group, risk factors for classical Kaposi Sarcoma. *J. Natl Cancer Inst*, 94:1712–1718.

GOLD, E., GORDIS, L., TONASCIA, J. & SZKLO, M. 1979. Risk factors for brain tumours in children. *Am J Epidemiol*, 109:309-19.

GRAFAREND, E.W. & KRUMM, F.W. 2010. Map Projections: Cartographic Information Systems. New York: Springer.

GREAVES, M.F., 1988. Speculations on the cause of childhood acute lymphoblastic leukaemia. *Leukaemia*, 2:120–125.

GREENBERG, R.S., SHUSTER, J.L., 1985. Epidemiology of cancer in children. *Epidemiol. Rev*, 7: 22–48.

GREENLAND, S. & FINKLE, W.D. 1995. A Critical Look at Methods for Handling Missing Covariates in Epidemiologic Regression Analyses. *American Journal of Epidemiology*, 142 (12):1255-1264.

GREENOP, K.R., PETERS, S., BAILEY, H.D., FRITSCHI, L., ATTIA, J., SCOTT, R.J., GLASS, D.C., DE KLERK, N.H., ALVARO, F., ARMSTRONG, B.K. & MILNE, E. 2013. Exposure to pesticides and the risk of childhood brain tumors. *Cancer Causes Control*, 24(7):1269-78.

GREGORY, I. & ELL, P. 2007. Historical GIS: Technologies, Methodologies, and Scholarship. Cambridge: Cambridge University Press.

GRUNIER, R.B., HARNLY, M.E., REYNOLDS, P., HERTZ, A. & VON BEHREN, J. 2001. Agricultural Pesticide use in California: Pesticide prioritization, use densities, and population distributions for a childhood study. *Environ Health Perspect*, 109:1071-1078.

GUSE, C.E., MARBELLA, A.M., GEORGE, V. & LAYDE, P.M. 2002. Radium in Wisconsin drinking water: an analysis of Osteosarcoma risk. *Arch Environ Health*, 57(4):294-303.

HALPERIN, E.C., COSTINE, L.S., TARBELL, N.J. & KUN, L.E. 2010. Pediatric Radiation Oncology. Philadelphia: Lippincott Williams & Wilkins.

HAMBURGER, R.N. 1988. Early prevention of allergy. *JAMA*, 259:2468-2472.

HAMOS, M., ELLIS, L.A., PALLOCK, D.R. & HENDERSON, T.R. 1996. Breast feeding and the working mothers: effects of time and temperature of short time storage on proteolysis lipolysis and bacterial growth in milk. *Paediatrics*, 97:492-498.

HARDER, T., PLAGEMANN, A. & HARDER, A. 2008. Birth weight and subsequent risk of childhood primary brain tumours: a meta-analysis. *Am J Epidemiol*, 168:366-73.

HARDER, T., PLAGEMANN, A. & HARDER, A. 2010. Birth weight and risk of Neuroblastoma: a meta-analysis. *International Journal of Epidemiology*, 39(3):746-756.

HAWKINS, M.M., KINNIER WILSON, L.M., BURTON, H.S., POTOK, M.H.N., WINTER, D.L., MARSDEN, H.B. & STOVALL, M.A. 1996. Radiotherapy, alkylating agents, and risk of bone cancer after childhood cancer. *JNCI*, 88:270-277.

HAYES, R.B. 1997. The carcinogenicity of metals in humans. *Cancer Causes Control*, 8:371-85.

HAYES, T.B., CASE, P., CHUI, S., CHUNG, D., HAEFFELE, C. & HASTON, K. 2006. Pesticide mixtures, endocrine disruption, and amphibian declines: are we underestimating the impact? *Environ Health Perspect*, 114:40-50.

HEISE, W. 2010. GI-Lymphomas in immunosuppressed patients (organ transplantation; HIV). *Best Practice & Research Clinical Gastroenterology*, 24:57-69.

HEUCH, J.M., HEUCH, I. & KVALE, G. 1996. Birth characteristics and risk of Wilms' tumour: a nationwide prospective study in Norway. *Br. J. Cancer*, 74:1148–1151.

HHS (U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES). 2003. Cancer and the environment: what you need to know. Washington, D.C:HHS.

HICKS, N., ZACK, M. & GALDWELL, G. 1984. Childhood cancer and occupational radiation exposure in parents. *Cancer* 53 (8):1637–1643.

HOLLY, E.A., ASTON, D.A., AHN, D.K. & KRISTIANSEN, J.J. 1992. Ewing's bone sarcoma, paternal occupational exposure, and other factors. *Am. J. Epidemiol.*, 135: 122– 129.

HOLLY, E.A., BRACCI, P.M., MUELLER, B.A., PRESTON-MARTIN, S. 1998. Farm and animal exposures and pediatric brain tumours: results from the United States West Coast Childhood Brain Tumour Study. *Cancer Epidemiol., Biomarkers Prev.*, 7:797– 802.

HOOVER, R.N., DEVESA, S.S., CANTOR, K.P., LUBIN, J.H. & FRAUMENI, J.F. 1991. Fluoridation of drinking water and subsequent cancer incidence and mortality. *In: Review of fluoride benefits and risks. Report of the ad hoc subcommittee on fluoride of the committee to coordinate environmental health and related problems. Washington (DC): US Department of Health and Human Services.*

HRUDEY, S.E., SOSKOLNE, C.L., BERKEL, J. & FINCHAM, S. 1990. Drinking water fluoridation and Osteosarcoma. *Can J Public Health*, 81(6):415–6.

HRUSAK, O., TRKA, J., ZUNA, J., POLOUCKOVA, A., KALINA, T., STARY, J. 2002. Acute lymphoblastic leukaemia incidence during socioeconomic transition: selective increase in children from 1 to 4 years. *Leukaemia*, 16:720– 725.

HUANG, X. 2003. Iron overload and its association with cancer risk in humans: evidence for iron as a carcinogenic metal. *Mutat. Res*, 533:153–171.

HUM, L., KREIGER, N. & FINKELSTEIN, M.M. 1998. The relationship between parental occupation and bone cancer risk in offspring. *Int. J. Epidemiol*, 27:766– 771.

IARC (INTERNATIONAL AGENCY FOR RESEARCH ON CANCER). 2002. Non-ionizing radiation, part 1: static and extremely low-frequency (ELF) electric and magnetic fields. In: IARC monographs on the evaluation of carcinogenic risk to humans, vol. 80. Lyon: IARC Press.

IARC (INTERNATIONAL AGENCY FOR RESEARCH ON CANCER). 2008. Report of the Advisory Group to Recommend Priorities for IARC Monographs during 2010-2014. Lyon, France: IARC.

INFANTE-RIVARD, C. & WEICHENTHAL, S. 2007. Pesticides and childhood cancer: an update of Zahm and Ward's 1998 review. *J. Toxicol. Environ. Health B Crit. Rev*, 10:81–99.

IRIGARAY, P., NEWBY, J.A., CLAPP, R., HARDELL, L., HOWARD, V., MONTAGNIER, L., EPSTEIN, S. & BELPOMME, D. 2007. Lifestyle-related factors and environmental agents causing cancer: an overview. *Biomedicine & pharmacotherapy*, 61:640-658

JEMAL, A., BRAY, F., CENTER, M.M., FERLAY, J. WARD, E. & FORMAN, D. 2011. Global cancer statistics. *CA Cancer J Clin*, 61:69-90.

JENSEN, N.K. & NORDBERG, G. 1980. Cancer Patients found to have Higher Pesticide Residues in Body. *Environmental Research*, 23:257-263.

JONES, J.L., HANSEN, D.L., DWARKIN, M.S. & JAFFE, H.W. 2000. Incidence and trends in kaposi sarcoma in the era of effective antiretroviral therapy. *J Acquir Immune Defic Syndr*, 24(3):270-4.

JUREWICZ, J. & HANKE, W. 2006. Exposure to pesticides and childhood cancer risk: has there been any progress in epidemiological studies. *Int. J. Occup. Med. Environ. Health*, 19:152–169.

KANE, M.L., LADOV, E.N., HOLDSWORTH, C.E. & WEAVER, N.K. 1984. Toxicological characteristics of refinery streams used to manufacture lubricating oils. *Am J Ind Med*, 5:183-200.

KANTOR, A.F., CURNEN, M.G., MEIGS, J.W. & FLANNERY, J.T. 1979. Occupations of fathers of patients with Wilms's tumour. *J. Epidemiol. Community Health*, 33:253–256.

KARAMI, S., LAN, Q., ROTHMAN, N., STEWARD, P.A., LEE, K.M. & VERMEULEN, R. 2012. Occupational Trichloroethylene exposure and kidney cancer risk: a meta-analysis. *Occup Environ Med*, 69:858.

KENNEDY, M. 2013. *Introducing Geographic Information Systems with ArcGIS: A workbook approach to learning GIS*. 3<sup>rd</sup> Ed. Hoboken, New Jersey: John Wiley & Sons, Inc.

KERR, M.A., NASCA, P.C., MUNDT, K.A., MICHALEK, A.M., BAPTISTE, M.S., MAHONEY, M.C., 2000. Parental occupational exposures and risk of neuroblastoma: a case-control study (United States). *Cancer Causes Control*, 11:635–643.

KHEIFETS, L., MONROE, J., VERGARA, X., MEZEI, G. & AFIFI, A.A. 2008. Occupational electromagnetic fields and Leukaemia and brain cancer: an update to two meta-analysis. *J Occup Environ Med*, 50(6):677-88.

KINLEN, L., 1988. Evidence for an infective cause of childhood Leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain. *Lancet*, 1323–1327.

KRAIN, L.S. 1991. Aviation, high altitude, cumulative radiation exposure and their associations with cancer. *Med Hypotheses*, 34:33–40.

KRAMER, S., WARD, E., MEADOWS, A.T., MALONE, K.E. 1987. Medical and drug risk factors associated with neuroblastoma: a case-control study. *J Natl Cancer Inst*, 78:797-804.

KRAUSKOPF, K.B. 1979. *Introduction to Geochemistry*. London: McGraw-Hill, Kogahusha Ltd.

KRIEGER, N. 1992. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. *Am. J. Public Health*, 82:703–710.

KRIEGER, R. 2010. *Hayes's Handbook of Pesticide Toxicology: Volume 1*. 3<sup>rd</sup> Ed. London: Elsevier Inc.

KRISTENSEN, P., ANDERSEN, A. & IRGENS, L.M. 1996. Incidence and risk factors of cancer among men and women in Norwegian agriculture. *Scand J Work Environ Health*, 22:14–26.

KROLL, M.E., STILLER, C.A., MURPHY, M.F. & CARPENTER, L.M. 2011. Childhood leukaemia and socioeconomic status in England and Wales 1976-2005: Evidence of higher incidence in relatively affluent communities persists over time. *Br J Cancer*, 105(11):1783-7.

LABRECHE, F., GOLDBERG, M.S. & VALOIS, M.F. 2003. Occupational exposure to extremely low frequency magnetic fields and postmenopausal breast cancer. *American Journal of Industrial Medicine*, 44:643-652.

LATINI, G., GALLO, F. & DE FELICE, C. 2004. Birth characteristics and hepatoblastoma risk in young children. *Cancer*, 101:210.

LEE, J. & WONG, D.W.S. 2001. *Statistical Analysis with ArcView GIS*. New York: Wiley Publishing.

LEE, W.J., COLT, J.S., HEINEMAN, E.F., MCCOMB, R., WEISENBURGER, D.D., LIJINSKY, W. & WARD, M.H. 2005. Agricultural pesticide use and risk of glioma in Nebraska, United States. *Occup Environ Med*, 62:786-92.

LEHTINEN, M., KOSKELA, P., OGMUNSDOTTIR, H.M., BLOIGU, A., DILLNER, J., GUDNADOTTIR, M., HAKULINEN, T., KJARTANSDOTTIR, A., KVARNUNG, M., PUKKALA, E., TULINIUS, H., LEHTINEN, T. 2003. Maternal herpesvirus infections and risk of acute lymphoblastic leukaemia in the offspring. *Am. J. Epidemiol*, 158:207– 213.

LEISENRING, W.M., BRESLOW, N.E., EVANS, I.E., BECKWITH, J.B., COPPES, M.J. & GRUNDY, P. 1994. Increased birth weights of National Wilms' Tumour Study patients suggest a growth factor excess. *Cancer Res*, 54:4680–4683.

LEMASTERS, G.K. & BOVE, K.E. 1980. Genetic/environmental significance of multifocal nodular renal blastoma. *Am. J. Hematol. Oncol*, 2,:81– 87.

LEVY, M. & LECLERC, B.S. 2012. Fluoride in drinking water and Osteosarcoma incidence rates in the continental United States among children and adolescents. *Cancer Epidemiology*, 36:83-88.

LINDBLAD, P., ZACH, M., ADAMI, H.O. & ERICSON, A., 1996. Maternal and perinatal risk factors for Wilms' tumour: a nationwide nested case-control study in Sweden. *Int J Cancer*, 51 (1):38–41.

LINK, M.P., EILBER, F. 1997. Osteosarcoma. In: Pizzo PA, Poplack DG, eds. Principles and practice of pediatric oncology. Philadelphia: Lippincott-Raven.

LIPWORTH, L., SONDERMAN, J.S., MUMMA, M.T., TARONE, R.E., MARANO, D.E. & BOICE, J.D. 2011. Cancer mortality among aircraft manufacturing workers: an extended follow-up. *J Occup Environ Med*, 53:992-1007.

LITEPLO, R., GOMES, R., HOWE, P. & MALCOLM, H. 2002. Environmental health criteria 227. Fluorides. International Program on Chemical Safety. Geneva, Switzerland: World Health Organization, United Nations Environment Programme, International Labour Organization and Inter-Organization Programme for the Sound Management of Chemicals.

LITTLE, R.J.A. & RUBIN, D.B. 2002. Statistical Analysis with Missing Data. 2<sup>nd</sup> Ed. Hoboken: John Wiley & Sons, Inc.

LLIFFE, J. & LOTT, R. 2008. Datums and Map Projections for Remote Sensing, GIS, and Surveying. Dunbeath: Whittles Publishing.

LONGLEY, P.A. & BATTY, M. 2003. Advanced Spatial Analysis: The CASA book of GIS. Redlands, California: ESRI.

LONGLEY, P.A., GOODCHILD, M.F., MAQUIRE, D.J. & RHIND, D.W. 2004. Geographic Information Systems and Science. 2<sup>nd</sup> Ed. New York: Wiley Publishing.

MA, X., BUFFLER, P.A., GUNIER, R.B., DAHL, G., SMITH, M.T., REINIER, K. & REYNOLDS, P. 2002. Critical windows of exposure to household pesticides and risk of childhood leukaemia. *Environ. Health Perspect*, 110:955–960.

MACK, W., PRESTON-MARTIN, S. & PETERS, J.M. 1991. Astrocytoma risk related to job exposure to electric and magnetic fields. *Bioelectromagnetics*, 12:57-66.

MACKERER, C.R., GRIFFIS, L.C., GRABOWSKI, J.S. & REITMAN, F.A. 2003. Petroleum mineral oil refining and evaluation of cancer hazard. *Appl Occup Environ Hyg*, 18:890-901.

MACLVLAHON, B. & NEWILL, V.A. 1962. Birth characteristics of children dying of malignant neoplasms. *J Natl Cancer Inst*, 28:231-44.

MACMILLAN CANCER SUPPORT. 2013. Medulloblastoma. <http://www.macmillan.org.uk> (Date of access: 1 June 2013)

MAHLE, D.A., GEARHART, J.M., GRIGSBY, C.C., MATTIE, D.R., BARTON, H.A. & LIPSCOMB, J.C. 2007. Age-dependent partition coefficients for a mixture of volatile organic solvents in Sprague-Dawley rats and Humans. *J Toxicol Environ Health A*, 70:1745-1751.

MCDONAGH, M., WHITING, P., BRADLEY, M., COOPER, J., SUTTON, A. & CHESTNUTT, I. 2000. Fluoridation of drinking water: a systematic review of its efficacy and safety. Report No.: 18. York, United Kingdom: Center for Reviews and Disseminations, University of York.

MCLAUGHLIN, C.C., BAPTISTE, M.S. & SCHYMURA, M.J. 2006. Maternal and infant birth characteristics and hepatoblastoma. *Am J Epidemiol*, 163:818-28.

MCLEOD, K.S. 2000. Our sense of Snow: the myth of John Snow in medical geography. *Social science & medicine*, 50:923-935.

MCWHIRTER, W.R. 1982. The relationship of incidence of childhood lymphoblastic Leukaemia to social class. *Br. J. Cancer*, 46:640– 645.

MEDICAL CHRONICLE. 2011. New Registry to give Major Boost to Fight against Cancer. <http://www.medicalchronicle.co.za/new-registry-to-give-major-boost-to-fight-against-cancer> (Data of access: 16 Oct. 2012.).

MEDLINE PLUS. 2013a. Neuroblastoma. <http://www.nlm.nih.gov/medlineplus/neuroblastoma.html> (Date of access: 10 November 2013).

MEDLINE PLUS. 2013b. Nephroblastoma. <http://www.nlm.nih.gov/medlineplus/nephroblastoma.html> (Date of access: 10 November 2013).



MEDLINE PLUS. 2013c. Osteosarcoma. <http://www.nlm.nih.gov/medlineplus/ency/article/001650.htm> (Date of access: 10 November 2013).

MEDLINE PLUS. 2013d. Retinoblastoma. <http://www.nlm.nih.gov/medlineplus/ency/article/001030.htm> (Date of access: 10 November 2013).

MEDSCAPE. 2013. Pediatric Hepatoblastoma: Practice Essentials. <http://emedicine.medscape.com/article/986802-overview> (Date of access: 6 November 2013)

MEINERT, R., SCHUZ, J., KALETSCH, U., KAATSCH, P., MICHAELIS, J. 2000. Leukaemia and non-Hodgkin's lymphoma in childhood and exposure to pesticides: results of a register-based case-control study in Germany. *Am. J. Epidemiol*,151: 639–646.

MENEGAUX, F., BARUCHEL, A., BERTRAND, Y., LESCOEUR, B., LEVERGER, G., NELKEN, B., SOMMELET, D., H, EMON, D., CLAVEL, J. 2006. Household exposure to pesticides and risk of childhood acute Leukaemia. *Occup. Environ. Med*, 63:131-134.

MENEGAUX, F., OLSHAN, A.F., NEGLIA, J.P., POLLOCK, B.H. & BONDY, M.L. 2004. Day care, childhood infections, and risk of neuroblastoma. *Am J Epidemiol*, 159:843–51.

MERLETTI, F., RICHIARDI, L., BERTONI, F., AHRENS, W., BUEMI, A., COSTA-SANTOS, C., ERIKSSON, M., GUENEL, P., KAERLEV, L., JOCKEL, K.H., LLOPIS-GONZALEZ, A., MERLER, E., MIRANDA, A., MORALES-SUAREZ-VARELA, M., OLSSON, H., FLETCHER, T. & OLSEN, J. 2006. Occupational factors and risk of adult bone sarcomas: a multicentre case-control study in Europe. *Int. J. Cancer*, 118:721-727.

MICHALEK, A.M., BUCK, G.M., NASCA, P.C., FREEDMAN, A.N., BAPTISTE, M.S., MAHONEY, M.C. 1996. Health status, medication use, and risk of neuroblastoma. *Am J Epidemiol*, 1143:996-1001.

MITCHELL, A. 2005. The ESRI Guide to GIS Analysis: Spatial Measurements and Statistics. Vol. 2. Redlands, California: ESRI.

MOLL, A.C., IMHOF, S.M., CRUYSSBERG, J.R., SCHOUTEN-VAN MEETEREN, A.Y., BOERS, M. & VAN LEEUWEN, F.E. 2003. Incidence of retinoblastoma in children born after in-vitro fertilisation [comment]. *Lancet*, 361:309–310.

MOLYNEUX, E.M., ROCHFORD, R., GRIFFEN, B., NEWTON, R., JACKSON, G., MENON, G., HARRISON, C.J., ISRAELS, T. & BAILEY, S. 2012. Burkitt's Lymphoma. *Lancet*, 379:1234-44.

MONGE, P., WESSELING, C., GUARDADO, J., LUNDBERG, I., AHLBOM, A., CANTOR, K.P., WEIDERPASS, E. & PARTANEN, T. 2007. Parental occupational exposure to pesticides and the risk of childhood leukaemia in Costa Rica. *Scand. J. Work Environ. Health*, 33:293–303.

MONTELLA, M., FRANCESCHI, S., GEDDES, M., ARNIANI, S. & COCCHIARELLA, G. 1996. Classic Kaposi Sarcoma and volcanic soil in southern Italy, *Lancet*, 347:905.

MONTESANO, R. & HALL, J. 2001. Environmental causes of human cancers. *European Journal of Cancer*, 37:67-87.

MOORE, L.E., BOFFETTA, P., KARAMI, S., BRENNAN, P., STEWARD, P.S. & HUNG, R. 2010. Occupational Trichloroethylene exposure and renal carcinoma risk: evidence of genetic susceptibility by reductive metabolism gene variants. *Cancer Res*, 70:6527-6536.

MUELLER, N.E., EVANS, A.S., LONDON, W.T. 1996. Viruses. In: Schottenfeld, D., Fraumeni Jr., J.F. (Eds.). New York : Cancer Epidemiology and Prevention. Oxford Univ. Press

MULCAHY, N. 2008. Cancer to become leading cause of death worldwide by 2010. <http://www.medscape.com/viewarticle/585098> (Date of access: 21 Jan 2010)

MUSICCO, M., SANT, M. & MOLINARI, S. 1988. A case-control study of brain gliomas and occupational exposure to chemical carcinogens: the risk to farmers. *Am J Epidemiol*, 128:778–85.

MASTERLACK, M. 2007. Pesticides and Childhood Cancer: An Update. *International Journal of Hygiene and Environmental Health*, 210:645-657.

NAWAR, E., MBULAITEYE, S.M., GALLANT, J.E. WOHL, D.A., ARDINI, M. & HENDERSB, T. 2005. Risk factors for Kaposi Sarcoma among HHV-8 seropositive homosexual men with AIDS. *Int. J. Cancer*, 115:296–300.

NEWTON, W.A., MEADOWS, A.T., SHIMADA, H., BUNIN, G.R. & VAWTER, G.F. 1991. Bone sarcomas as second malignant neoplasms following childhood cancer. *Cancer*, 67:193– 201.

OLLIER, C. 1984. *Weathering*. Essex: Longman Group Ltd.

OLSHAN, A.F., BRESLOW, N.E., DALING, J.R., FALLETTA, J.M., GRUFFERMAN, S., ROBISON, L.L., WASKERWITZ, M. & HAMMOND, G.D. 1990. Wilmstumour and paternal occupation. *Cancer Res*, 50:1– 6.

OLSHAN, A.F., BRESLOW, N.E., FALLETTA, J.M., GRUFFERMAN, S., PENDERGRASS, T., ROBISON, L.L., WASKERWITZ, M., WOODS, W.G., VIETTI, T.J. & HAMMOND, G.D. 1993. Risk factors for Wilmstumour. Report from the National WilmsTumour Study. *Cancer*, 72:938– 944.

OPERSKALSKI, E.A., PRESTON-MARTIN, S., HENDERSON, B.E. & VISSCHER, B.R. 1987. A case-control study of Osteosarcoma in young persons. *Am. J. Epidemiol*, 126:118– 126.

ORJUELA, M., PONCE CASTANEDA, V., RIDAURA, C., LECONA, E., LEAL, C., ABRAMSON, D.H., ORLOW, I., GERALD, W. & CORDON-CARDO, C. 2000. Presence of human papilloma virus in tumour tissue from children with retinoblastoma: an alternative mechanism for tumour development. *Clin. Cancer Res*, 6:4010–4016.

PARKIN, D.M., KRAMAROVA, E., DRAPER, G.J., MASUYER, E., MICHAELIS, J., NEGLIA, J.P., QURESHI, S. & STILLER, C.A. 1998. *International Incidence of Childhood Cancer*. Lyon, France: International Agency for Research on Cancer.

PELSER, C., DAZZI, C., GRAUBARD, B.I., LAURIA, C., VITALE, F. & GOEDERT, J.J. 2009. Risk of classic Kaposi Sarcoma with residential exposure to volcanic and related soils in Sicily. *Ann Epidemiol*, 19(8):597-601.

PETROVSKI, B.E., PATAKI, V., JENEI, T., ADANY, R. & VOKO, Z. 2012. Selenium levels in men with liver disease in Hungary. *Journal of Trace Elements in Medicine and Biology*, 26:31-35.

POLEDNAK, A.P. 1978. Bone cancer among female radium dial workers. Latency periods and incidence rates by time after exposure: brief communication. *J. Natl. Cancer Inst*, 60:77–82.

PRICE, E.W. 1990. Podoconiosis: Non-Filarial Elephantiasis. Oxford: Oxford University Press.

PROYTICHEVA, M.A. 2011. Diagnostic Pediatric Hematopathology. Cambridge: Cambridge University Press.

PUBLIC SERVICE COMMISSION. 2010. Consolidated Report on Inspections of Primary Health Care Delivery Sites: Department of Health. Pretoria: Public Service Commission.

PURDUE, M.P., BAKKE, B., STEWART, P., DE ROOS, A.J., SCHENK, M., LYNCH, C.F., BERNSTEIN, L., MORTON, L.M., CERHAN, J.R., SEVERSON, R.K., COZEN, W., DAVIES, S., ROTHMAN, N., HARTGE, P. & COLT, J.S. 2011. A case-control study of occupational exposure to Trichloroethylene and Non-Hodgkin's Lymphoma. *Environ Health Perspect*, 119:232-8.

PURKAIT, S.K. 2003. Essentials of oral pathology. 2nd Ed. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd.

RABKIN, C.S. & YELLIN, F. 1994. Cancer incidence in a population with a high prevalence of infection with immunodeficiency virus type 1. *J Natl Cancer Inst*. 86(22):1711-6.

RADICAN, L., BLAIR, A., STEWART, P. & WARTENBERG, D. 2008. Mortality of aircraft maintenance workers exposed to Trichloroethylene and other hydrocarbons and chemicals: extended follow-up. *J Occup Environ Med*, 50:1306-1319.

RAGHUNATHAN, T.E. 2004. What to do with missing data: Some options for analysis of incomplete data. *Annu. Rev. Public Health*, 25:99-117.

RANGEL, M., CYPRIANO, M., DE MARTINO., LEE, M.L., LUISI, F.A., PETRILLI, A.S., STRUFALDI, M.W. & FRANCO, M.C. 2010. Leukemia, non-Hodgkin's lymphoma, and Wilms tumor in childhood: the role of birth weight. *Eur J Pediatr*, 169(7):875-81.

REID, A. 2012. Asbestos still haunts those exposed as kids in mining towns. *The Conversation*, 10 Oct. <http://theconversation.edu.au/asbestos-still-haunts-those-exposed-as-kids-in-mining-towns-9487> (Date of access: 17 Feb, 2013).

REID, A., FRANKLIN, P., OLSEN, N., SLEITH, J., SAMUEL, L., ABOAGYE-SARFO, P., DE KLERK, N. & MUSK, A.W. 2013. All-cause mortality and cancer incidence among adults exposed to blue asbestos during childhood. *American Journal of Industrial Medicine*, 56: 133-145.

REPETTO, R. 2013. Pesticides and the Immune System: An overview. <http://www.globalhealingcenter.com/effects-of-pesticides/pesticides-immune-system> (Date of access: 13 October 2013.)

REYNOLDS, P., URAYAMA, K.Y., VON BEHREN, J. & FEUSNER, J. 2004. Birth characteristics and hepatoblastoma risk in young children. *Cancer*, 100:1070–6.

RIES, L.A.G., SMITH, M.A., GURNEY, J.G., LINET, M., TAMRA, T., YOUNG, J.L. & BUNIN, G.R. 1999. Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975–1995. Bethesda: US National Cancer Institute.

ROMAN, E., LIGHTFOOT, T., SMITH, A.G., FORMAN, M.R., LINET, M.S., ROBINSON, L., SIMPSON, J., KAATSCH, P., GRELL, K., FREDERIKSEN, K. & SCHUZ, J. 2013. Childhood Acute Lymphoblastic Leukaemia and birth weight : Insights from a pooled analysis of case-control data from Germany, the United Kingdom and the United States. *European Journal of Cancer*, 49:1437-1447.

ROSS, J.A. 2006. High Birth Weight and Cancer: Evidence and Implications. *Cancer Epidemiol Biomarkers Prev*, 15:1.

ROSSO, A.L., HOVINGA, M.E., RORKE-ADAMS, L.B., SPECTOR, L.G., & BUNIN, G.R. 2008. A case-control study of childhood brain tumours and fathers' hobbies: a Children's Oncology Group study. *Cancer Causes Control*, 19:1201–7.

RULL, R.P., GUNIER, R., VON BEHREN, J., HERTZ, A., CROUSE, V., BUFFLER, P.A. & REYNOLDS, P. 2009. Residential proximity to agricultural pesticide applications and childhood acute lymphoblastic leukaemia. *Environmental Research*, 109:891-899.

SANDERSON, S.C., WALLER, J., JARVIS, M.J., HUMPHRIES, S.E. & WARDLE, J. 2009. Awareness of lifestyle risk factors for cancer and heart disease among adults in the UK. *Patient education and counselling*, 74:221-227.

SANDBLUND, J.T., DOWNING, J.R., CRIST, W.M. 1996. Non-Hodgkin's lymphoma in childhood. *N. Engl. J. Med.*, 334:1238–1248.

SARG, K.M.J. & GROSS, A.N. 2007. *The Cancer Dictionary*. 3<sup>rd</sup> Ed. New York: Infobase Publishing.

SATOH, T. & CUPTA, R.C. 2010. *Anticholinesterase Pesticides: Metabolism, Neurotoxicity and Epidemiology*. New Jersey: John Wiley & Sons, Inc.

SCHERR, P.A. & MUELLER, N.E. 1996. Non-Hodgkin's lymphomas, in: *Cancer Epidemiology and Prevention*. 2nd ed. Oxford: Oxford University Press.

SCHUZ, J., KALETSCH, U., KAATSCH, P., MEINERT, R. & MICHAELIS, J., 2001a. Risk factors for pediatric tumours of the central nervous system: results from a German population-based case-control study. *Med. Pediatr. Oncol*, 36:274–282.

SCHUZ, J., KALETSCH, U., MEINERT, R., KAATSCH, P. & MICHAELIS, J. 2001b. High-birth weight and other risk factors for Wilms tumour: results of a population-based case-control study. *Eur. J. Pediatr*, 160:333–338.

SCHUZ, J., KALETSCH, U., MEINERT, R., KAATSCH, P., SPIX, C. & MICHAELIS, J. 2001c. Risk factors for neuroblastoma at different stages of the disease. Results from a population based case control study in Germany. *Journal of Clinical Epidemiology*, 54:702-709.

SCHWARTZBAUM, J.A. 1992. Influence of mother's prenatal drug consumption on risk of neuroblastoma in the child. *Am J Epidemiol*, 135:1358-67.

SCHWARTZBAUM, J.A., GEORGE, S.L., PRATT, C.B. & DAVIS, B. 1991. An exploratory study of environmental and medical factors potentially related to childhood cancer. *Med. Pediatr. Oncol*, 19:115– 121.

SCOTT, C.S. & JINOT, J. 2011. Trichloroethylene and cancer: systematic and quantitative review of epidemiologic evidence for identifying hazards. *Int J Environ Res Public Health*, 8:4238-4272.

SCOTT, L. 2012. Verbal communication with the author. San Diego

SEIDMAN, H., SELIKOFF, I.J. & HAMMOND, E.C. 1982. Mortality of brain tumours among asbestos insulation workers in the United States and Canada. *Annals of the New York Academy of Science*, 381:160-171.

SHARPE, C.R., FRANCO, E.L. & DE CAMARGO, B. 1995. Parental exposures to pesticides and risk of Wilms' tumour in Brazil. *Am. J. Epidemiol*, 141 (3):210–217.

SHIM, Y.K., MLYNAREK, S.P. & VAN WIJNGAARDEN, E. 2009. Parental Exposure to Pesticides and Childhood Brain Cancer: U.S. Atlantic Coast Childhood Brain Cancer Study. *Environ Health Perspect*, 117(6):1002-1006.

SHU, X.O., GAO, Y.T. & BRINTON LA. 1988. A population-based case-control study of childhood leukaemia in Shanghai. *Cancer*, 62:635-644.

SIMONART, T., VAN VOOREN, J.P., HERBAUTS, J. & BOELAERT, J.R. 1999. High prevalence of Kaposi Sarcoma in Iceland and the Faroe Islands. *Br. J. Cancer*, 79:373.

SMITH, A., LIGHTFOOT, T., SIMPSON, J. & ROMAN, E. 2009. Birth weight, sex and childhood cancer: A report from the United Kingdom Childhood Cancer Study. *Cancer Epidemiology*, 33:363-367.

SMULEVICH, V.B., SOLIONOVA, L.G. & BELYAKOVA, S.V. 1999. Parental occupation and other factors and cancer risk in children: I. Study methodology and non-occupational factors. *Int. J. Cancer*, 83:712–717.

SNYDER, J.P. 1997. *Flattening the Earth: Two Thousand Years of Map Projections*. 2nd Ed. Chicago: University of Chicago Press.

SOLDIN, O.P., NSOULY-MAKTABI, H., GENKINGER, J.M., LOFFREDO, C.A., ORTEGA-GARCIA, J.A., COLANTINO, D., BARR, D.B., LUBAN, N.L., SHAD, A.T. & NELSON, D. 2009. Pediatric Acute Lymphoblastic Leukemia and Exposure to Pesticides. *The Drug Monit*, 31(4): 495-501.

SOLIMAN, A.M.R., SCHOTTENFELD, D. & BOFFETTA, P. 2013. *Cancer Epidemiology: Low and Middle income countries and special populations*. New York: Oxford University Press.

SORAHAN, T., LANCASHIRE, R.J. 2004. Parental cigarette smoking and childhood risks of hepatoblastoma: OSCC data. *Br J Cancer*,90:1016–8.

SPITZ, M.R., JOHNSON, C.C. 1985. Neuroblastoma and paternal occupation. *Am J Epidemiol*, 121:924-9.

STATISTICS SOUTH AFRICA. 2001. Census 2001: Concepts and Definitions. Pretoria: Stats SA.

STATISTICS SOUTH AFRICA. 2011. General Household Survey 2010. Pretoria: Stats SA.

STEFAN, C. 2012. Verbal communication with the author. Potchefstroom.

STEWART, A., WEBB, J., GILES, D. & HEWITT, D., 1956. Malignant disease in childhood and diagnostic irradiation in utero. *Lancet* 2, 447– 448.

SZYMANSKA-CHABOWSKA, A., ANTONOWICZ-JUCHNIEWICZ, J. & ANDRZEJAK, R. 2002. Some aspects of arsenic toxicity and carcinogenicity in living organism with special regard to its influence on cardiovascular system, blood and bone marrow. *Int J Occup Med Environ Health*, 15:101-16.

TAKAHASHI, K., AKINIWA, K. & NARITA, K. 2001. Regression analysis of cancer incidence rates and water fluoride in the U.S.A. based on IACR/IARC (WHO) data (1978–1992). International Agency for Research on Cancer. *J Epidemiol*;11(4):170-9.

TALBOT, S.J. & CRAWFORD, D.H. 2004. Viruses and tumours - an update. *European Journal of Cancer*, 40:1998-2005.

THE OHIO STATE UNIVERSITY WEXNER MEDICAL CENTRE. 2013. What is Osteosarcoma? [http://medicalcenter.osu.edu/patientcare/\\_services/bone\\_disorders/bone\\_cancer/osteosarcoma](http://medicalcenter.osu.edu/patientcare/_services/bone_disorders/bone_cancer/osteosarcoma) (Date of access: 15 October 2013)

THOM, A. 2008. Cancer will be the leading cause of death globally in 2010. [http://www.givegain.com/cgi-in/giga?cmd=cause\\_dir\\_news\\_Item&cause\\_id=105\\_6&news\\_id=57430](http://www.givegain.com/cgi-in/giga?cmd=cause_dir_news_Item&cause_id=105_6&news_id=57430) (Date of access: 21 Jan 2010).



TJALMA, R.A. 1966. Canine bone sarcoma: estimation of relative risk as a function of body size. *J. Natl. Cancer Inst*, 36:1137– 1150

TSAI, J., KAYE, W.E. & BOVE, F.J. 2006. Wilms' tumor and exposures to residential and occupational hazardous chemicals. *Int. J. Hyg. Environ.-Health*, 209:57–64.

TUCKER, M.A., D'ANGIO, G.J., BOICE, J.D., STRONG, L.C., LI, F.P., STOVALL, M., STONE, B.J., GREEN, D.M., LOMBARDI, F., NEWTON, W., HOOVER, R.N., FRAUMENI, J.F. & GROUP, L.E.S. 1987. Bone sarcomas linked to radiotherapy and chemotherapy in children. *N. Engl. J. Med*, 317:588–593.

TYNES, T., & HALDORSEN, T. 2003. Residential and occupational exposure to 50 Hz magnetic fields and haematological cancers in Norway. *Cancer Causes & Control*, 14:715-720.

UNITED STATES DEPARTMENT OF LABOUR. 2013. Welding Health Hazards. <http://www.osha.gov/doc/outreachtraining/htmlfiles/weldhlth.html>. (Date of access: 6 November 2013).

URAVAMA, K.Y., VON BEHREN, J., REYNOLDS, P. 2007. Birth characteristics and risk of Neuroblastoma in young children. *Am J Epidemiol*, 165(5):486-95.

US EPA (UNITED STATES ENVIRONMENTAL PROTECTION AGENCY). 2000. Trichloroethylene Hazard Summary. <http://www.epa.gov/ttn/atw/hlthef/tri-ethy.html> (Date of access: 16 Oct. 2012.).

US EPA (UNITED STATES ENVIRONMENTAL PROTECTION AGENCY). 2001. National Air Toxics Program: The Integrated Urban Strategy: Report to Congress. <http://www.epa.gov/ttnat01/urban/natprpt.pdf> (Date of Access: 16 Oct. 2012).

US EPA (UNITED STATES ENVIRONMENTAL PROTECTION AGENCY). 2003. America's children and the environment: measures of contaminants, body burdens and illness. *United States Environmental Protection Agency*, 76-81.

US EPA (UNITED STATES ENVIRONMENTAL PROTECTION AGENCY). 2006. Consumer Factsheet on: Trichloroethylene. <http://www.epa.gov/safewater/dwh/c-voc/trichlor.html> (Date of access: 16 Oct. 2012).

US EPA (UNITED STATES ENVIRONMENTAL PROTECTION AGENCY). 2013. Health effects of Radium. <http://www.epa.gov/radiation/radiationuclides/radium.html> (Date of access: 4 September 2013)

US NATIONAL CANCER INSTITUTE. 2012. Retinoblastoma. <http://www.cancer.gov/cancertopics/types/retinoblastoma> (Date of access: 16 November 2012).

VALERY, P.C., MCWHIRTER, W., SLEIGH, A., WILLIAMS, G., BAIN, C. 2002. Farm exposures, parental occupation, and risk of Ewing's sarcoma in Australia: a national case-control study. *Cancer Causes Control*, 13:263– 270.

VAN MAELE-FABRY, G., HOET, P. & LISON, D. 2013. Parental occupational exposure to pesticides as risk factor for brain tumors in children and young adults: A systematic review and meta-analysis. *Environment international*, 56:19-31.

VAN MAELE-FABRY, G., LANTIN, A.C. HOET, P. & LISON, D. 2011. Residential exposure to pesticides and childhood Leukaemia: A systematic review and meta-analysis. *Environment International*, 37:280-291.

VIANNA, N.J., POLAN, A.K., 1976. Childhood lymphatic leukaemia: prenatal seasonality and possible association with congenital varicella. *Am. J. Epidemiol*, 103:321– 332.

VIEL, J.F., CHALLIER, B. & PITARD, A. 1998. Brain cancer mortality among French farmers: the vineyard pesticide hypothesis. *Arch Environ Health*, 53:65–70.

VON SCHWEINITZ, D. 2012. Hepatoblastoma: Recent Developments in Research and Treatment. *Seminars in Pediatric Surgery*, 21:21-30.

WALLACE, L.A. 1985. Personal Exposures, Indoor-Outdoor Relationships, and Breath levels of Toxic Air Pollutants Measured for 355 Persons in New Jersey. *Atmospheric Environment*, 19(10):1651-1661.

WANG, L. & CHEN, S. 2001. Environment, genome and cancer. *Academie des Sciences (III) Sciences del la vie*, 324:1085-1091.

WEBMD. 2013a. Cancer Health Centre: Acute Lymphoblastic Leukaemia. <http://www.webmd.com/cancer/acute-lymphoblastic-leukemia> (Date of access: 6 September 2013).

WEBMD. 2013b. Cancer Health Centre: Acute Myeloid Leukaemia. <http://www.webmd.com/cancer/acute-myeloid-leukemia-symptoms-treatments> (Date of access: 6 September 2013).

WEBMD. 2013c. Brain Cancer Health Centre: Astrocytoma. <http://www.webmd.com/cancer/brain-cancer/astrocytoma> (Date of access: 16 September 2013).

WEINBERG, E.D. 1996. The role of iron in cancer. *Eur. J. Cancer Prev*, 5:19–36.

WILKINS, J.R. & SINKS, T.H. 1984. Occupational exposures among fathers of children with Wilms' tumour. *J. Occup. Med*, 26 (6):427–435.

WORLD CANCER RESEARCH FUND. 1997. Food, nutrition and the prevention of cancer: a global perspective. Washington, D.C.: American Institute for Cancer Research.

WU, W.T., LIN, Y.J., LIOU, S.H., YANG, C.H., CHENG, K.F., TSAI, P.J. & WU, T.N. 2012. Brain cancer associated with environmental lead exposure: Evidence from implementation of a National Petrol-Lead Phase-Out Program (PLPOP) in Taiwan between 1979 and 2007. *Environment International*, 40:97-101.

WUNSCH-FILHO, V., PELISSAN, D.M., BARBIEN, F.E., SANT'ANNA, L., DE OLIVEIRA, C.T., DE MATA, J.F., TONE, L.G., LEE, M.L., DE ANDREA, M.L., BRUNIERA, P., EPELMAN, S., FILHO, V.O. & KHEIFETS, L. 2011. Exposure to magnetic fields and childhood Acute Lymphoblastic Leukaemia in Sao Paulo, Brazil. *Cancer Epidemiology*, 35(6):534-639.

YANG, C.Y., CHENG, M.F., TSAI, S.S. & HUNG, C.F. 2000. Fluoride in drinking water and cancer mortality in Taiwan. *Environ Res*, 82(3):189–93.

YEAZEL, M.W., ROSS, J.A., BUCKLEY, J.D., WOODS, W.G., RUCCIONE, K. & ROBISON, L.L. 1997. High birth weight and risk of specific childhood cancers: A report from the Children's Cancer Group. *J. Pediatr*, 131:671– 677.

ZAHM, S.H. & WARD, M.H. 1998. Pesticides and childhood cancer. *Environ Health Perspect*, 106:893-908.

ZIEGLER, J.L. 1993. Endemic Kaposi Sarcoma in Africa and local volcanic soils. *Lancet*, 342:1348–1351.

ZIEGLER, J.L., NEWTON, R., BOURBOULIA, D., CASABONNE, D., BERAL, V. & MBIDDE, E. 2003. Risk factors for Kaposi Sarcoma: a case– control study of HIV-seronegative people in Uganda. *Int. J. Cancer*, 103:233–240.

ZIEGLER, J.L., NEWTON, R., KATONGOLE-MBIDDE, E., MBULATAIYE, S., DE COCK, K. & WABINGA, H. 1997. Risk factors of Kaposi Sarcoma in HIV-positive subjects in Uganda. *AIDS*, 11:1619–1626.