What can a GIS bring to an epidemiological study in Ireland: A case study of nitrates in drinking water and cancer risk.

Margaret Forrest

A dissertation submitted to the University of Dublin, in partial fulfilment of the requirements for the degree of Master of Science in Health Informatics

2007

Declaration

I declare that the work described in this dissertation is, except where otherwise stated, entirely my own work, and has not been submitted as an exercise for a degree at this or any other university.

Signed: ______________________________

Margaret Forrest

11th September 2007

Permission to lend and/or copy

I agree that the Trinity College Library may lend or copy this dissertation upon request.

Signed: ______________________________

Margaret Forrest

11th September 2007

Acknowledgments

I would like to thank the staff at Trinity College for their encouragement and support, in particular Gaye Stephens, Lucy Hederman and Mary Sharp. I would also like the thank Dr. Harry Comber of the National Cancer Registry and Fiona Lawlor of the Environmental Protection Agency for their help with acquiring data. Finally many thanks to family and friends.

Summary

Excluding heart disease and stroke, cancer kills more people in Ireland every year than any other cause. The aetiology of many cancers is to a large degree unknown. National and international differences in incidences of cancer signify that geographic variation may provide etiological clues. The purpose of this study is to understand the value of using a Geographical Information System (GIS) for an epidemiological study of cancer incidence and nitrate levels in drinking water in Ireland. Data was sourced from Ordnance Survey Ireland (OSi), the Environmental Protection Agency (EPA) and the National Cancer Registry in Ireland (NCRI). The cancers reviewed were oesophagus, stomach, colon, rectum, head and neck and colorectal. A GIS was used to calculate a nitrate risk rate for each Electoral Division (ED) based on chemical data from national monitoring points. ED risk rates were analysed against cancer rates by calculating Pearson's product moment correlation coefficient (r). Overall there were no significant correlations between cancer incidence and nitrate levels. However, a path from nitrate levels to cancer aetiology is biologically plausible. The results do not mean that there is no link. They simply mean that in Ireland nitrate levels are not high and as a consequence associations with cancer incidence were not found. Therefore nitrate is not likely to be a cancer risk factor in an Irish context. The study illustrates the added value a GIS can bring to such studies. These include visualisation techniques to illustrate patterns and gaps in data, linkage of disparate datasets to undertake analysis, and its use as a tool to foster collaboration between multi-disciplinary teams. The study also highlights issues in terms of data acquisition and quality, technology and ecological study limitations.

Table of Contents

Table of Contents I
List of Figures II
List of Tables II
List of Abbreviations III
1. Introduction 1
2. Review of the literature 4
2.1 Literature Search Methodology 4
2.2 GIS Techniques and Concepts 5
2.2.1 Overlay Analysis 6
2.2.2 Disease Maps 7
2.2.3 Smooth Maps and Spatial Filters 11
2.2.4 Disease Clusters 12
2.2.5 Geographical Correlation Analysis 13
2.2.6 Environmental Exposure Assessment 14
2.3 GIS data 15
2.3.1 Geocoding 15
2.3.2 Data Types and Sources 15
2.4 GIS and Health in Ireland 16
2.4.1 Examples of Health Applications of GIS in Ireland 16
2.4.2 Limits of Health Applications of GIS in Ireland 17
2.5 Nitrates in Drinking Water and Cancer Risk 18
2.6 Literature Review Conclusion 20
3. Materials and Methods 21
3.1 Datasets 21
3.1.1 Electoral Divisions 21
3.1.2 Ground and Surface Water Chemical Data 23
3.1.3 Cancer Data 23
3.2 Technologies 24
3.3 Methodology 24
3.3.1 Data Preparation 24
3.3.2 Spatial Analysis 25
3.3.3 Statistical Analysis 27
4. Results 29
4.1 Drinking Water Quality and Nitrate Levels in Ireland 29
4.2 Nitrate Levels and Cancer Aetiology 31
4.3 Nitrate Levels and Cancer Incidence in Ireland 32
5. Discussion 34

5.1 Benefits of GIS 34
5.1.1 Visualisation and Hypothesis Generation 34
5.1.2 Collation and Analysis of Disparate Data 36
5.1.3 Multidisciplinary Teams 37
5.2 Limitations of GIS 38
5.2.1 Data Acquisition and Quality..38
5.2.2 Ecological Studies 40
6. Conclusion 42
7. References 45

List of Figures

Figure 1 Cholera Deaths in the Broad Street area 1854 (Snow 1855). 8
Figure 2 Choropleth map of male lung cancer incidence by ED for Dublin, Ireland (Kelly and Rybaczuk 2002)..9
Figure 3 Isopleth map of Sudden Infant Death Syndrome mortality rates per 1000 live births in North Carolina from 1974-1984 (Berke 2004) 12
Figure 4 Dublin region showing the four Local Authority Areas and EDs 22
Figure 5 Choropleth Map of Nitrate Risk levels for EDs in Ireland. 26

List of Tables

Table 1 Initial Key Word Literature Search...4
Table 2 Nitrate Risk Level Values 25
Table 3 Number of EDs with each Nitrate Risk Level 27
Table 4 Number of EDs (with data) with each nitrate risk level...32
Table 5 Correlation coefficient between nitrate risk levels and cancer SIRs in Ireland 33

List of Abbreviations

A&E	Accident & Emergency
BOD	Biochemical Oxygen Demand
CDC	Centre for Disease Control and Prevention
CSO	Central Statistics Office
DOEHCL	Department of the Environment, Heritage and Local Government
DOHC	Department of Health and Children
ED	Electoral Division
EPA	Environmental Protection Agency
FSAI	Food Safety Authority of Ireland
GIS	Geographic Information System
GPS	Global Positioning System
HIPE	Hospital In-Patient Enquiry Scheme
HSE	Health Service Executive
IARC	International Agency for Research on Cancer
ICD-9-CM	International Classification of Diseases, Ninth Revision, Clinical Modification
MeSH	Medical Subject Headings
Mg/L	Milligrams Per Litre
NCI	National Cancer Institute United States
NCRI	National Cancer Registry of Ireland
NO_2	Nitrite
NO_3	Nitrate
NPPB	National Postcode Project Board
OSi	Ordnance Survey Ireland
SAPS	Small Area Population Statistics
SIR	Standardised Incidence Ratio
SMR	Standardised Mortality Ratio
US FGDC	United States Federal Geographic Data Committee
WHO	World Health Organisation

1. Introduction

After heart disease and stroke, cancer kills more people on the island of Ireland every year than any other cause. One in three Irish people are likely to contract some form of cancer by the age of 74 (Walsh, Comber et al. 2001). The cause of many cancers is to a large degree unknown. National and international differences in incidences of cancer signify that geographic variation may provide etiological clues (Aase and Bentham 1998). A Geographical Information System (GIS) can facilitate in the investigation and interpretation of geographical variations in diseases in relation to different factors (Elliott and Wartenberg 2004). The term GIS was added to Medical Subject Headings (MeSH) in 2003. This step is indicative of the increased prevalence and importance of GIS in health and healthcare research (Boulos 2004).

‘Medical geography' and ‘spatial epidemiology' are in effect equivalent expressions for the examination of disease in the context of its geographical distribution (Bithell 2000). All through the history of medicine associations between disease and the chemical environment have been discovered. Examples of well know causal correlations include dental health and fluorine and between goitre and iodine. By relating the geographical distribution of chemical elements to rates of endemic diseases possible links may be established or eliminated helping with the formation of fresh etiological hypotheses (Bølviken 2001). The area of spatial epidemiology has gained momentum in recent years. The reasons for this includes a cognisance of the impact environment and location can have on human health and an increased availability of data. Epidemiological studies are using more and more complex statistical methods, but maps prevail as a vital descriptive means of representing information visually which can illustrate features that could escape regular statistical analysis (Bithell 2000).

The purpose of this study is to understand the value of using a GIS for an epidemiological study in Ireland. A case study will be undertaken to review nitrates in drinking water and possible links with cancer. Nitrates exist naturally in soils, waters, plants, and in meats. They are part of the universal nitrogen cycle and are ubiquitous in the natural world. Humans have altered the nitrogen cycle resulting in steady build up of nitrate in water supplies. Nitrate is now an extensive pollutant of drinking water; however it’s possible health effects are unclear. In the body nitrate (NO_3) is reduced to nitrite (NO_2), which can form N-nitroso compounds. N-nitroso compounds are recognised animal carcinogens (Gulis, Czompolyova et al. 2002; Ward, deKok et al. 2006; WHO 2006). A World Health Organisation (WHO) standard of 50 mg/L for Nitrate (NO_3) or 3 mg/L for Nitrite (NO_2) in drinking water exists. This was originally set to protect bottle fed babies from developing methemoglobinemia or blue-baby syndrome

(Gulis, Czompolyova et al. 2002; Ward, deKok et al. 2006; WHO 2006). Some research has indicated potentially chronic health effects linked with high rates of nitrate in drinking water in particular in relation to cancer incidence. Conversely other research has found no negative health outcome in relation to nitrate levels. Results and findings are inconsistent across studies (Grinsven, Ward et al. 2006).

The case study will attempt to establish if correlations exist between drinking water nitrate rates and cancer incidence in Ireland. To facilitate the case study data availability will be assessed and data will be sourced from different relevant organisations. The cancers reviewed will include oesophagus, stomach, colon, rectum, head and neck and colorectal. GIS techniques and applicability will be considered and the most appropriate for Irish data will be selected in consultation with GIS expert and environment engineers. These techniques will be utilised to determine nitrate risk rates by appropriate administrative areas in Ireland. Nitrate risk rates will then be compared with cancer rates to establish if associations, negative or positive, exist. Throughout the process the relevance, benefits, drawbacks and usefulness of using a GIS for a study in Ireland will be assessed.

The dissertation contains six chapters in total. Chapter 2 of this thesis includes a review of the literature published in relation to GIS, epidemiology and health. Initially it describes the literature search methodology adopted. It then presents GIS methods and concepts theoretically. Once the theory in introduced concrete applications of the theories in a health context are outlined. It discusses the types of data available and likely sources of data. It also describes data related issues. It looks at the use of GIS in healthcare in Ireland as well as reviewing articles relating to nitrates and ill health correlations.

The methodology for the case study is outlined in Chapter 3. It describes the datasets sourced and used including the spatial and attribution information included in each dataset. Also it outlines to technologies utilised in order to perform the GIS and statistical analysis. It explains the methodological approach adopted and why this approach was used.

Chapter 4 discusses the results. Initially it describes drinking water quality in Ireland generally and specifically in relation to nitrate levels in ground and surface water. It illustrates the possible etiological pathway for nitrates and cancer from nitrate exposure to the generation of possibly carcinogenic N-nitroso compounds. This chapter also outlines the results of the geographical and statistical analysis of an ED’s nitrate risk rate and any association with cancer incidence in Ireland

The value of a GIS for an epidemiological study in Ireland is discussed in Chapter 5. Benefits are described including visualisation techniques to illustrate patterns and gaps in data. The power of using geography to link disparate datasets to undertake analysis is outlined. The use of GIS as a tool to foster collaboration between multi-disciplinary teams is indicated. This chapter also highlights issues in regard to the use of GIS in Irish health research as well as discussing problems relating to data acquisition and quality, technology and ecological studies.

Chapter 6 contains the conclusions of this research. It provides a brief review of the main results. It includes suggestions for further research and discusses the contributions of the study.

In conclusion cancer is a prevalent and endemic disease and the second biggest cause of death in Ireland after arteriosclerosis, which causes heart disease and strokes. This research uses a GIS to undertake a case study examining nitrate levels and their possible associations with cancer. In doing so the research discovers the likely impact of nitrate levels in relation to cancer incidence in Ireland. The research also uncovers the potential benefits and pitfalls of utilising a GIS for an epidemiological study in Ireland. The level and distribution of diseases and their causes are essential inputs into improving population health. By understanding causes and risks diseases can be prevented. The focus can then shift from palliative and curative services to interventions and appropriate resource allocations in areas with at risk populations (Murray, Ezzati et al. 2003). When it comes to cancer prevention is most certainly better than cure.

2. Review of the literature

In this chapter the literature relating to GIS, health and epidemiology is reviewed. Initially the literature review methodology is explained. GIS techniques and concepts are then introduced from a theoretical point of view. Once the concept has been presented practical applications of the theory, in a health care context, are outlined. In the next section data related issues are discussed. The use of GIS in Irish healthcare is then summarised. Finally articles relating to nitrates and ill health correlations are reviewed.

2.1 Literature Search Methodology

Electronic databases that contain journals relating to Geography, Health and Epidemiology, were identified using the Information Resources section of the Trinity College library web site. Initially a key word search was performed to establish a base set of articles. Table 1 provides a synopsis of the key word search. The search terms are listed in the first column. The electronic databases searched and the number of results found is indicated in the subsequent columns.

	Literature Database								
Search Terms	Bio Med	Blackwell & Synergy	Cambridge Journals	GeoScience World	Oxford Journals	Pub Med	Science Direct	Springer Link	Taylor & Francis Journals
GIS AND Health	237	202	8	71	16	489	181	71	50
GIS AND Health AND Epidemiology	137	11	5	5	3	189	21	6	17
GIS AND Health AND Epidemiology AND Spatial	136	11	7	5	1	86	11	4	12
GIS AND Cancer	115	129	2	7	33	91	17	6	24
GIS AND Ireland	14	3	0	79	0	22	20	7	28
Nitrates AND Drinking AND Water AND Cancer	2	0	0	0	7	78	10	12	12

Table 1 Initial Key Word Literature Search

Originally nine electronic databases were chosen. Some journals are held in more than one database therefore some of the articles were retrieved more than once. However the extension of the search across several databases ensured the results were not narrowed by a limited initial journal set (Levy and Ellis 2006). Where possible the full text of articles was searched. An exception to full text searching was made for the Blackwell and Synergy, Oxford Journals and Cambridge Journals databases. These database searches do not have the ability to search by "whole word" only. Therefore a full text search for "GIS AND Health" returns 13191, 3151 and 750 articles respectively. The reason for this is because any word containing the letters "gis" constitutes a match. To refine the results only the title and abstract were searched.

Once the articles were retrieved a concept centric approach was adopted (Webster and Watson 2002). Articles were grouped based on different emerging concepts. Articles that related to GIS in so far as the term referred to Geographical Information Systems were included. Articles that were written in English were also included as were articles relating to the Republic of Ireland. Finally due the large volume of literature retrieved, for the most part, articles relating the GIS and health but not epidemiology are not incorporated. These articles primarily related to service provision and equity of access to health care. An exception to this rule was used for Ireland. Articles relating to GIS, health and Ireland are limited. Therefore all of these articles were reviewed.

Where appropriate, backward searching took place whereby articles and authors referenced in the articles were retrieved and evaluated. For standout articles a forward search took place to identify additional articles that have cited the article. In addition to the database searches two Irish journals were individually searched as results on the use of GIS in health in Ireland proved difficult to find. These journals were Irish Geography and the Irish Medical Journal. Finally the book catalogue in Trinity College was searched to find relevant books on the subject. This search methodology was adopted based on the recommendations of Levy and Ellis (2006) and Webster and Watson (2002).

2.2 GIS Techniques and Concepts

A GIS can be described as a system for the collection, storage, analysis and visualisation of spatially referenced data (Gatrell and Loytonen 1998). Out-of-the-box the GIS toolbox is rich. Many techniques exist such as data retrieval, buffering, spatial querying, map generation. GIS techniques combined provide a powerful tool to the user. The emergence of web-based GIS applications help the public locate facilities that meet their requirements and display health atlases of different areas. These sites encourage the public to become more involved with public health decision making (Higgs 2004).

Spatial analysis is the interpretation and investigation of geographical variations in disease in relation to various factors e.g. environmental, socio-economic, and infectious risk factors (Elliott and Wartenberg 2004). A GIS allows an epidemiologist, or researcher, inspect spatial patterns in their data and comprehend the relationship between a disease and other factors. In addition high quality maps provide a facility for an epidemiologist to present a persuasive case for further studies, interventions and service provision (Brewer 2006).

A range of GIS analysis techniques are described in this section. These areas are not strict and tend to overlap. A study will typically use a combination of these techniques based on the nature of the problem and the available data (Berke 2004).

2.2.1 Overlay Analysis

GIS can merge and evaluate complex data from multiple sources. Each data source can contain different variables that can be added as a layer on a map. These maps layers can be displayed using various data combinations so that patterns emerge (Phillips, Kinman et al. 2000). This technique is known as overlay analysis. Overlay analysis is the ability to collate data from disparate sources and link the data together using geography as the key. This facilitates the combination of different geographical datasets e.g. patients attending a hospital can be mapped in relation to a buffer zone around that hospital.

Overlay analysis is one of the value adds of a GIS. It facilitates the ability to view geographical variations in data (Chung, Yang et al. 2004). Variations are important for numerous reasons. Areas with gaps in access can be identified with a view to relocating resources or planning new facilities. Service utilisation trends can be established to understand usage patterns and preferences. From a surveillance point of view areas with high or low rates of unfavourable outcomes related to health can be established with a view to identifying and if possible eliminating causative factors. At risk populations can be identified enabling a more targeted provision of services and resources (Clarke, McLafferty et al. 1996).

Glass, Schwartz et al. (1995), overlaid land data from six different databases that contained 53 different environment variables using a GIS. They assessed environmental risk factors for Lyme disease for residents of Baltimore County, Maryland. Lyme disease is transmitted to humans through ticks. Using GIS techniques and case-control methods they were able to detect residents at high risk. Guidry and Margolis (2005) found that certain schools, in North Carolina (USA), whose pupils were classified as high risk in terms of developing respiratory infections, were disproportionately afflicted by the flooding from Hurricane Floyd. To determine this they used a GIS to overlay satellite images of flooded areas with school and demographic data. They concluded that a GIS can be used to discover and prioritise schools or other public facilities most impacted by natural disasters. Miranda, Dolinoy et al. (2002) defined a model to predict lead exposure risk in children in order to shift programs away from being mitigative towards being preventive. They amalgamated tax assessor data, U.S.

Census Demographic data, and North Carolina blood lead screening data into a spatial overlay theme. Although the datasets were unique and from different sources they could do this because they shared a mutual geographic backbone. Using overlay analysis the data was integrated and they were able to statistically analyse all data layers combined. They concluded that although GIS has limitations it holds enormous possibilities for improving how environmental health agencies plan and implement both existing programs and those linked with new emerging problems.

2.2.2 Disease Maps

When evaluating data it is good practice to generate and examine some initial graphs. This process helps the analyst get an early sense of the data and unusual characteristics can be identified. This is called disease mapping in spatial epidemiology (Berke 2004). Disease maps are considered a basic tool in the analysis of public health data. Their use dates back to 1854 when John Snow produced maps of cholera victim's addresses in relation to polluted water pumps (Lawson 2001).

A section of the original map created by Dr. John Snow can be seen in Figure 1. Snow suspected that cholera was spread through contaminated drinking water. He marked cholera deaths by placing a line parallel to the front of building where the person had died. In doing so Snow was able to identify a cluster of cases and trace the spread in this area to a water pump on the corner of Cambridge and Broad Street (Snow 1855).

Figure 1 Cholera Deaths in the Broad Street area 1854 (Snow 1855).

Disease maps play an important descriptive role in spatial analysis. A valid disease map should be based on rates. A rate is the relationship between the number of individuals displaying a particular characteristic and the total population. Populations grow and shrink over time and vary from one place to another. In general, the absolute number of cases of a disease, for comparison purposes, has limited importance. The use of rates is recommended. A disease map should also be cleaned of noise and adjusted for geographical variations in the underlying population for example age, sex and other known risk factors (Bailey 2001). The resulting disease map should provide insight into possible causes, effects and trends in the vast amount of data. This will provide an invaluable starting point for epidemiological enquiry (Sankoh, Berke et al. 2002).

Traditional disease maps look at disease rates at a particular spatial scale encompassing administrative or other areas. Patterns of disease rates are often depicted using choropleth maps. Events are portrayed on these maps as shaded polygons whereby each polygon represents a shaded area and areas with the same shade have the same rate of disease occurrence (Moore and Carpenter 1999; Berke 2002; Krivoruchko, Gotway et al. 2003). An example of a choropleth map can be seen in Figure 2. This map represents lung cancer rates in Dublin. The darker areas have higher rates (Kelly and Rybaczuk 2002).

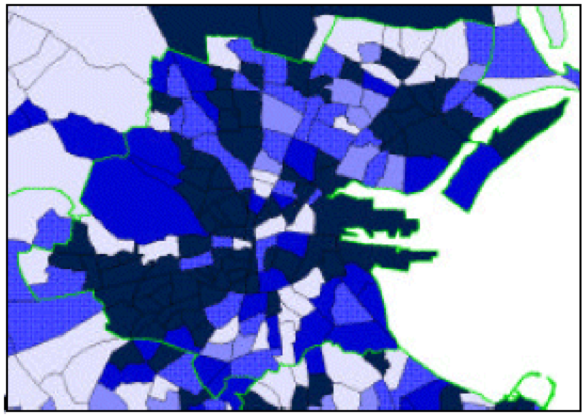

Figure 2 Choropleth map of male lung cancer incidence by ED for Dublin, Ireland (Kelly and Rybaczuk 2002)

Choropleth disease maps are often referred to as disease atlases or thematic maps. These atlases were traditionally paper based but increased use of technology has led to the availability of these atlases on the Internet either as PDF files or as interactive maps. Bell, Hoskins et al. (2006) discuss a number of applications that generate choropleth maps to help inform policy makers and educate the public. In the United States the National Cancer Institute (NCI) collaborated with the Centres for Disease Control and Prevention (CDC) and developed a web site which profiles incidences of cancer by state. This site links cancer rates, screening information, prevalence of risk factors and demographics. This helps planners focus interventions on areas of most need (NCI and CDC 2006). Epidemiological Query and Mapping System (EpiQMS) was developed by the State of Washington and also implemented by the State of Pennsylvania. It combines maps, graphs and tables to provide health related data to the public (Washington State Department of Health Health 2007). The Health Service Executive (HSE) in Ireland commissioned a health atlas application in February 2006 (E-Tenders 2006). This atlas hopes to exploit the possibility of integrating GIS, health datasets and statistics in a web and open source environment. The proof of concept was due to complete in 2006 (HSE 2006). An all Ireland health observatory is also available online which provides map, chart and tabular data. The observatory aims to 'support those working to improve health and reduce health inequalities by producing and disseminating health intelligence, and strengthening the research and information infrastructure on the island of Ireland' (IPH 2006).

Choropleth maps are an uncomplicated method of portraying incidence of disease but can be an over simplification. Firstly there appears to be an even spread of a rate across a polygon or administrative area. An area with a dense population could have significant variability within it. Secondly rates in adjacent or bordering areas can vary dramatically. In reality this is not the case. Tobler (1970) published a paper in Economic Geography in 1970 which stated for the first time the first law of geography: ''everything is related to everything else, but near things are more related than distant things.'' This simple but effective statement became known as Tobler's first law (TFL) of geography. Tobler tried to capture a fundamental principle of geography by calling it the first law (Sui 2004). Choropleth maps can break this law by displaying adjacent areas with very different rates. Finally when areas are divided along administrative, census or political boundaries physical area size tends to be determined based on population density. This results in small geographic areas with large populations and large geographic areas with small populations. These large geographic areas can visually dominate the map although they represent a small proportion of the population (Rushton 1998; Bithell 2000; Berke 2004).

In a classic book "How to Lie with Maps" Monmonier illustrates many ways one can introduce nuances and tailor a map to communicate specific meanings (Monmonier 1996; Houghton 2005). When a glossy colourful map is produced and presented it is tempting to generate a hypothesis based on a noticeable pattern. Jacquez (1998) refers to his phenomenon as the 'gee whiz' effect.

Notwithstanding these limitations choropleth maps are widely used. They are relatively easy to construct in comparison to more complicated smoothed maps. Additionally people tend to be familiar with administrative areas and therefore request information based on these areas. Administrative areas also protect the anonymity of individuals (Boulos 2004). Disease maps offer a quick visual synopsis of multifaceted geographic information and can highlight subtle patterns that would be lost in tabular presentations. They are used for a number of reasons. These include for illustrative purposes, to create aetiological hypotheses, to aid surveillance by identifying areas at seemingly high risk, and to help policy creation and allocation of resources (Elliott and Wartenberg 2004).

While a disease map is a helpful descriptive method, displaying areas with high and low rates, it falls short in terms of telling us if these rates are due to random spatial variation or if they actually expose an uneven risk for the disease in the different geographical locations. Statistical analysis, of the map and corresponding data, must be performed to determine if the perceived map pattern is statistically significant. Only then can a hypotheses be formed (Jacquez 1998; Kulldorff 1998; Krivoruchko and Gotway 2002). This analysis is performed through spatial statistical analysis, which includes smoothing techniques, spatial filtering, disease clustering techniques and geographical correlation analysis. These analytical techniques are described in the next section.

2.2.3 Smooth Maps and Spatial Filters

The previous section outlined some limitations of choropleth maps. These related primarily to the nature of aggregated data and the use of administrative areas. The "small number problem" also manifests when using data that is aggregated geographically. If the population in an area is very small, rates cannot be precise. If the number of cases (e.g., bone cancer incidences) is also very small, a change in a rate from one year to the next can produce a dramatic result. In these scenarios the rates are viewed as unstable (Krivoruchko, Gotway et al. 2003). A method of overcoming the shortcomings of choropleth maps and mitigating against the small number problem is to produce smoothed maps. Smoothed maps represent disease rates assuming spatial continuity. They maintain statistical stability of the rates (Rushton 1998). Although smoothing produces increased rate reliability, one has to be careful one doesn't smooth away genuine clusters or mask linear characteristics where rates change dramatically (Rushton 2003).

Smoothed rates are rates that have been spatially adjusted by the rates in their neighbouring areas. Therefore the rate reflects the rate of the area plus it's neighbour (Talbot, Kulldorff et al. 2000). The technique is similar to that of using moving averages over time except that the smoothing occurs spatially rather than chronologically.

One smoothing method is empirical Bayes. A number of studies using empirical Bayes estimates can be found in the research publications. Yiannakoulias, Rowe et al (2003) applied the technique in order to understand the geography of injuries relating to falls in the elderly in Alberta, Canada. They also applied a cluster detection technique. They claim that a description of the geography of a disease or incidence can improve the success of prevention strategies by detecting high-risk areas even when the individual and background factors that explain the patterns are unknown. Maheswaran, Strachan et al. (2002) undertook a study examining if there was a correlation between early life incidences on the geographic variation in adult mortality for cancer and stoke in England and Wales. They smoothed rates before assigning scores to counties in recognition of imprecision in unsmoothed rates. Sankoh, Berke et al. (2002) completed a study of childhood mortality in rural Bukina Faso. They used mortality rates from 39 villages. Initially they mapped crude mortality rate but found they could not identify any spatial patterns. They then smoothed the data using empirical Bayes estimation. The resulting maps did not show a clear spatial trend but they did highlight that villages in the north-eastern part of the study area had higher incidences of mortality. They recommend this technique for exploratory mapping of disease or mortality.

Spatial filtering is another smoothing technique. It can be used with both point based and area based data. The estimated rate at a specific location is defined as the observed rate within a specific distance of that location. Circles or buffer zones are generated around locations or points. These circles can overlap enabling shared observations between adjacent or neighbouring points. Once the estimated rates are calculated, contouring software can generate isarithmic maps. Isopleth, also known as isarithmic, maps represent a continuous smoothed map of data where regions with a constant rate or range of rates can be recognised (Talbot, Kulldorff et al. 2000). An advantage of spatial filtering is the ability to investigate using different filter sizes and therefore different spatial resolutions. This results in estimated rates that are not dependent on the boundaries of the areas used to aggregate the data (Boulos 2004; Ozdenerol, Williams et al. 2005). An example of an isopleth map can be seen in Figure 3.

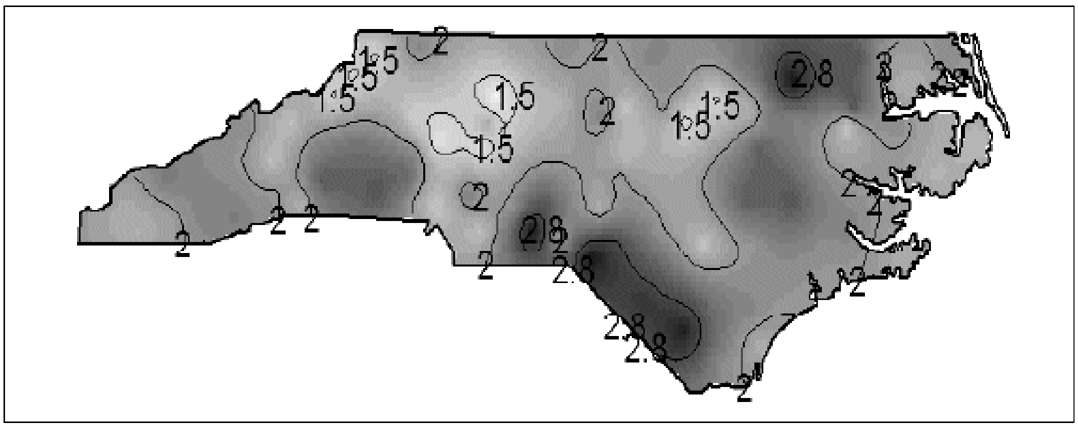

Figure 3 Isopleth map of Sudden Infant Death Syndrome mortality rates per 1000 live births in North Carolina from 1974-1984 (Berke 2004)

Ozdenerol, Williams et al. (2005) compared the spatial filtering technique and a cluster detection technique spatial scan statistics (SaTScan) in examining clusters of low birth weight. They found the filtering technique has advantages over SaTScan as it treats rates as a continuous spatial distribution and doesn't conceal spatial patterns.

2.2.4 Disease Clusters

A disease cluster is classified as a defined area within an overall study region with a prominent rise in the occurrence of a disease (a hot-spot cluster). The detection of a cluster of disease can help epidemiologists understand risk factors and lead to improved comprehension of aetiology. Additionally, locating disease clusters facilitates more complete analysis using targeted case-control studies in the field (Aamodt, Samuelsen et al. 2006). The typical null hypothesis in spatial analysis of disease is that the number of cases in an area is proportional to the number at risk. If there are no clusters then cases are either randomly or evenly spread (Moore and Carpenter 1999).

Smoothed rates and spatial filtering have been discussed in the previous section. These methods are a form of cluster detection as they highlight statistical patterns on a map. Another method is SatScan which estimates the probability that an event at a point is greater than that occurring by chance and displays a probability map (Ozdenerol, Williams et al. 2005). Numerous other cluster detection techniques exist. Reader (2001) used a method based on kernel density estimation to detect clusters of low-birth weight in two regions in Florida, USA. Wang (2004) examined spatial clusters of four top types of cancer in Illinois. These cancer types were breast, lung, colorectal, and prostate cancers.

2.2.5 Geographical Correlation Analysis

Correlation analysis is focussed on establishing whether a relationship exists between variables e.g. X and Y (Walford 1995). Geographic correlation analysis focuses on geographic variations across population groups based on different attributes like environmental exposure (e.g. air pollution), demographics (e.g. age, gender, race) and risk factors (e.g. smoking and diet) (Elliott and Wartenberg 2004).

Various geographic correlation studies exist. For example in 1999 Staessen et al looked at the relationship between bone density and environmental exposure to cadium in ten Belgian districts. They found that even a small degree of exposure can lead to skeletal demineralisation, which may cause bone fragility and increase the risk of fractures (Staessen, Roels et al. 1999). Hales, Black et al. (2003) examined the correlation between social status and public health risk due to drinking water quality. They used a GIS to compare the grade of community supply of drinking water with a deprivation index. They found that deprived communities were exposed to more risk from community water supplies and were probably shouldering an uneven load in relation to adverse health effects. Aragonés, Ramis et al. (2007) conducted a spatial analysis of oesophageal cancer mortality in Spain. They found significant differences in the risk of dying from oesophageal cancer based on area of residence. They discovered that some variations were due to the existence of known modifiable risk factors like smoking and diet. They also found excess risk that could be attributed to environmental or socio-cultural influences. They concluded that small area geographic studies have the potential to prove extremely useful in determining locations where interventions and further research should be focussed.

2.2.6 Environmental Exposure Assessment

Environmental health research examines risk factors relating to a person's surroundings. These include air quality, water quality, heat and sun exposure and also exposure to disease carrying organisms (Puolstrup and Hansen 2004). GIS technology provides many opportunities for environmental epidemiologists to look at geographic disease distribution and links to corresponding environmental exposures. The most costly and difficult aspect of an environmental health study is obtaining data relating to exposure. A GIS can merge information from existing databases and generate estimates of exposure. These exposure estimates combined with population and disease data can be utilised to propose and support hypotheses concerning environmental causes of disease (Vine, Degnan et al. 1997).

GIS can aid the exposure assessment process in many ways. This includes identifying the study population e.g. identifying all people living within a distance of a landfill site or along a transmission line. In environmental research the polluting agent can be physical or chemical. Once it's known a GIS can help identify the medium that carries the agent e.g. air, water, animals or insects A GIS can also help define exposure routes. A GIS can also aid the measurement of levels of contamination. Ideally to assess the levels of a contaminating agent one would measure the level during the critical time the contamination occurred. This is rarely possible or practical. A GIS can help model contamination levels to predict exposure. GIS has great potential to aid understanding and awareness of the spatial link between health and pollution (Maantay 2002; Nuckols, Ward et al. 2004).

Numerous and varied examples of the use of GIS in environmental exposure assessment can be found in the literature. Nuckols, Ward et al. (2004) found 15 studies since 1998 that illustrated the successful use of GIS in environmental epidemiology. Brody, Aschengrau et al. (2004) conducted a study of the relationship between breast cancer risk and exposure to pesticides. They used a GIS in a case-controlled study in Cape Cod, Massachusetts, USA. The study group included 1,165 women diagnosed with breast cancer and 1,006 controls. They examined exposures to a variety of pesticides. They did not find any general connection between pesticides and breast cancer incidence. They note however that their study was hampered somewhat by incomplete data. In 2000, Kohli, Brage et al. examined the correlation between leukaemia in children and ground radon levels. This was an ecological study set in Sweden. Population data was mapped. Radon maps were then overlaid and risk of exposure for each year was defined as low, medium or high. This was then assessed against the tumour registry using standardised mortality ratios (SMR) of children with acute lymphatic leukaemia. They found that children living in areas that were classified as low risk in relation to ground radon were less likely to develop acute lymphatic leukaemia. Nerriere, Zmirou-Navier et al. (2005) used a GIS to help assess the number of lung cancer deaths that could be attributed to chronic exposure to fine particles (pm2.5). These fine particles can reach the airways and alveoli in the lungs. The study looked at adult populations in Grenoble, Rouen, Paris and Strasbourg and they found that traffic emissions still play a significant part in the lung cancer problem in France.

2.3 GIS data

Spatial data is data that has a geographical component. Typically this geographical component will be a set of co-ordinates e.g. Latitude, Longitude. These co-ordinates represent a point, line or polygon of the earth's surface (Boscoe, Ward et al. 2004) . An example of a point location would be a person's address. A line could be a transmission line, a road, a rail line or river. A polygon is an enclosed area. Examples of polygons are counties and countries.

2.3.1 Geocoding

In order to associate an incident or address to a set of map co-ordinates a process called geocoding takes place. Geocoding takes a location or address and returns a set of co-ordinates. This process is a prerequisite for spatial analysis of the data. The development of GIS in the last two decades has meant increased geocoding of data where by data is stored with a geographical dimension (Talbot, Kulldorff et al. 2000). The US Federal Geographic Data Committee (FGDC) states that location is a key feature of 80-90% of government data (US Federal Geographic Data Committee 2003). The same can be said for other countries.

The level of accuracy required by this geocoding process is determined by the nature of the study. In some instances geocoding to an administrative area might be sufficient, in other cases an exact co-ordinate might be needed (Boulos 2004). A geocoding engine can automatically process an entire database of addresses or incidents. The quality and cleanliness of the data will dictate the match and success rate. Once the automated process has taken place a user can interactively match remaining addresses (Vine, Degnan et al. 1997).

2.3.2 Data Types and Sources

In spatial studies data is usually aggregated to administrative areas to help protect the privacy of individuals (Boscoe, Ward et al. 2004). Disease maps in general depict data in the form of areas. One reason is due to confidentiality requirements. Another reason is attributed to the nature of other datasets that might be sourced and used for a study. These would use counts for administrative areas (Bithell 2000). A cancer study for example might include cancer registry data, population or census data and perhaps environmental data. These datasets are sourced from different agencies and generally summarised to an area level (Boscoe, Ward et al. 2004).

The selection of areal units used for mapping can radically affect the map produced (Webster, Vieira et al. 2006). This problem has been given considerable attention by quantitative geographers, it is referred to as the modifiable areal unit problem. This problem occurs when the correlation coefficient between two variables changes at different scales i.e. for different aggregation units (Walford 1995). Studies that show a positive correlation at one scale can show a negative correlation at another. In general it is accepted that the more aggregated the data the less reliable it will be. For example, in the United States data aggregated to city or state is likely to be less reliable than blocks or census block groups (Maantay 2002).

2.4 GIS and Health in Ireland

2.4.1 Examples of Health Applications of GIS in Ireland

GIS has been used in health applications in Ireland. The Small Area Research Unit (SAHRU) specialises in the application of statistical and GIS techniques to aid with spatial modelling, data mining, population projection and needs assessment and geographic accessibility measures (SAHRU 1996). As mentioned, previously in the review, the HSE in Ireland has commissioned a health atlas to make geographical health information available on-line. Kelly and Rybaczuk (2002) used GIS and GeoVisualisation techniques to investigate the spatial pattern of cancer, mortality and material deprivation. They discovered that the most appropriate scale for examining cancer rates in Ireland was by Electoral Division (ED) and that mapping small area cancer statistics has distinct advantages over coarser scales like county and health board. The National Cancer Registry of Ireland (NCRI) participated in a project to compare and map cancer incidence and mortality rates for various parts of Ireland and Britain from 1991-2000. A cancer atlas of the UK and Ireland was produced (NCRI 2005).

Two accessibility reports that use GIS have also been published in the past year. Teljeur, Barry et al. (2006) used a model to estimate travel times to Accident & Emergency (A&E) for road traffic accidents and acute myocardial infarction. This was prompted by the publication of the ‘Hanly’ report which proposed closing A&E services in some small acute hospitals and increasing capacity in larger hospitals (DOHC 2003). Three scenarios were modelled with the aid of a GIS. It was found that this approach to planning the provision of new services is useful in maintaining equity of access to services. Kalogirou and Foley (2006) used a similar GIS based modelling approach to asses the impact of the 'Hanly' report. In this study they used combined spatial data in the form of road networks, population demographics and hospital locations. They concentrated on people over 65. They modelled three scenarios and were able to see the likely increase or decrease in accessibility based on each one. They concluded that the method would be helpful in enabling informed policy decisions based on spatial outcomes. They also emphasise that the model could be extended or re-used in the future.

2.4.2 Limits of Health Applications of GIS in Ireland

Although the previous section highlighted the usefulness of GIS in helping to answer health questions in Ireland published articles in an Irish context are difficult to find. Kelly and Rybaczuk (2002) found that disease mapping in Ireland is not an area that is made use of . They found some spatial examinations of disease, specifically cancer but the majority of these were at the Health Board level – the largest administrative unit, made up of three or more counties. Teljeur, Barry et al. (2006) state that "there is a dearth of published data relating to travel times for emergency admissions". Houghton (2005) reviews the creation of disease and mortality maps and finds that although technology and GIS have advanced the production of maps in the last fifty years has declined. An example is the Annual Reports of Vital Statistics produced by the Central Statistics Office (CSO). In 1991 maps ceased to be included. Houghton argues that the reasons for this are twofold. Firstly he outlined the fact that there is an absence of a health-funding model in Ireland. Currently funds are allocated on a historical basis. An effort to introduce a spatial and equitable funding model would be likely to divert resources away from Dublin because of the younger age profile of the capital. This would have an impact in particular on the large hospitals. Secondly he emphasises the political implications of publicising health inequalities.

Another reason for the limited adoption of GIS can be attributed to the uneven and uncoordinated use of GIS in the Health Boards in Ireland. A survey conducted in 2003 revealed that 5 out of 8 health boards had a GIS. Some used the ArvView GIS package and others used a package from MapInfo but most used GIS purely for display rather than analysis purposes (Houghton 2001; Houghton 2004).

Perhaps the biggest challenge relating to GIS in Ireland is the issues relating to geocoding Irish addresses. The process of geocoding assigns geographic coordinates to addresses so that they can be mapped and analysed by a GIS. In Ireland a postcode system does not exist so the process of geocoding can be expensive and in many instances cost prohibitive. A dataset that is not geocoded cannot be used for analysis (Houghton 2004). An ability to accurately automatically geocode a dataset is crucial in GIS projects (Boulos 2004).

2.5 Nitrates in Drinking Water and Cancer Risk

Humans have altered the nitrogen cycle resulting in steady build up of nitrate in water supplies. Nitrate is now an extensive pollutant of drinking water; however it's possible health effects are ambiguous. In the body nitrate (NO_3) is reduced to nitrite (NO_2), which can form N-nitroso compounds. N-nitroso compounds are recognized animal carcinogens. The World Health Organisation (WHO) set a standard in drinking water at 50 mg/L for Nitrate (NO_3) or 3 mg/L for Nitrite (NO_2) (WHO 2006). This was set to protect bottle fed babies from developing methemoglobinemia or blue-baby syndrome (Gulis, Czompolyova et al. 2002; Ward, deKok et al. 2006; WHO 2006).

Research has indicated potentially chronic health effects linked with high rates of nitrate in drinking water. However results and findings are inconsistent across studies (Grinsven, Ward et al. 2006). A conference on drinking-water nitrate and health was held at the International Society for Environmental Epidemiology in August 2004. The purpose was to evaluate nitrate exposures and corresponding implication on health effects with regard to the current regulatory limit. The working group concluded that the risk factor for certain cancers, reproductive outcomes, and other chronic health effects and the part played by drinking-water nitrate exposure require further study before a change to the regulatory level for nitrate in drinking water can be contemplated (Ward, deKok et al. 2006).

The following set of studies found a positive correlation with drinking water and certain types of cancer. Volkmer, Ernst et al. (2006) undertook a study in Bocholt, Germany where a community had different drinking water nitrate levels. One group's rate was 10 mg/L while the other group's rate was 60 mg/L. They calculated age standardised incidence rates of urological malignancies registered. They found a connection between drinking water nitrate load and urothelial cancer occurrences in both genders. They found an inverse association to testicular tumours and no correlation with renal, penile and prostatic tumours. Gulis, Czompolyova et al. (2002) carried out an ecologic analysis to determine if nitrate levels in drinking water were associated with non-Hodgkin lymphoma and digestive and urinary tracts cancers in an agricultural area (Trnava District; population 237,000) in Slovakia. They collected nitrate drinking water data and classified villages as low (0-10 mg/L), medium (10.1-20 mg/L), or high (20.1-50 mg/L). Standardised incidence ratios (SIRs) for all cancer and selected cancer occurrences for each village were also calculated. They found that SIRs increased for all cancers in women from villages with a low rate to those with a high rate. The same pattern was true for all cancer in men for low and medium rates but not high. This trend in the SIRs was also observed for stomach cancer in women, colorectal cancer in women and men , and non-Hodgkin lymphoma in women and men. They found no correlation for bladder or kidney cancer. Gulis, Czompolyova et al. conclude that this ecologic study supports the hypothesis that a positive correlation between non-Hodgkin lymphoma and colorectal cancer and nitrate in drinking water exists. A similar ecological study was undertaken by Sandor, Kiss et al. (2001) to explore gastric cancer mortality in a set of small villages supplied by drinking water with high nitrate content. They found that groups with greater than 88 mg/l average nitrate concentration demonstrated considerable risk elevation and although their analysis was an ecological study and therefore has particular limitations, it also supports the hypothesis that high levels of nitrate in drinking water is associated with the development of gastric cancer. In a case-control study in Taiwan Yang, Cheng et al (1998) found a significant positive association between gastric cancer mortality and high nitrate levels in drinking water. Valencia in Spain has the highest concentration of nitrate in drinking water in all of Europe. Morales-Suarez-Varela, Llopis-Gonzalez et al. (1995) examined the relationship between nitrate levels and stomach, bladder, prostate and colon cancer. They found that the rate of gastric cancers in both men and women and the rate of prostrate cancer rose with increased exposure to nitrates.

The following set of studies did not find a link with drinking water and certain types of cancer. Zeegers, Selen et al (2006) conducted a cohort study in the Netherlands to establish if there was a link between nitrates and bladder cancer. They concluded that although a correlation between nitrate exposure and bladder cancer risk was conceivable from a biological point of view the result of their study did not support this. Leeuwen, Waltner-Toews et al. (1999) undertook an ecological study in Ontario examining the effect of both atrazine and nitrate. They found nitrate levels, were negatively associated with occurrences of stomach cancer. Similarly Loon, Botterweck et al. (1998) completed a cohort study in the Netherlands. The study included 120,852 men and women aged between 55 and 69. They also found a negative association between nitrate ingestion and the risk of gastric cancer. Barrett, Parslow et al. (1998) performed an ecological study in the north of England. They did not find a relationship between oesophageal and stomach cancers and nitrate levels. They did find higher levels of brain and central nervous system cancers link with higher nitrate levels and advocated an individual level study to confirm the results. A number of studies were undertaken in Iowa. For the most part the research undertaken in Iowa did not find positive correlations with cancer however they dealt with low levels of nitrate (Roos, Ward et al. 2003; Ward, Cantor et al. 2003; Coss, Cantor et al. 2004; Ward, Cerhan et al. 2006).

Conflicting results exist in the literature reviewed for this thesis. This conflict seems to exist regardless of the type of study conducted e.g. ecological Vs case-controlled. Some of the studies listed were undertaken in populations where there was a low level of nitrate while others had high rates. The different results may be due to the different exposure levels

(Sandor, Kiss et al. 2001). One pattern that appears to exist is that if the levels of nitrate are significantly elevated correlations are more prevalent.

2.6 Literature Review Conclusion

The literature review has highlighted, with examples, the rich variety of spatial methods and models available for health research. It has reviewed data considerations and looked at GIS for health in Ireland. It has also discussed nitrates and their possible etiological connection to cancer. One conclusion of the literature review is that valuable epidemiological research does not necessarily need sophisticated statistics. Spatial health research should focus on the combination of access to quality data at appropriate scales with sound epidemiological considerations and methodologies in order to answer research questions (Bailey 2001). There is a need for an awareness and understanding of the data and thoughtful analysis when dealing with spatial health data (Pickle, Waller et al. 2005). In order to reap the benefits of GIS and fully exploit it's potential collaboration is needed between medical geographers, GIS specialists, epidemiologists, statisticians and environmental consultants (Vine, Degnan et al. 1997; Nuckols, Ward et al. 2004).

This author hopes to illustrate the benefits of GIS in an Irish context by conducting a case study of nitrates in drinking water and their association with cancer risk. Relevant data will be sourced from disparate organisations and GIS techniques will be assessed to determine the most appropriate for analysis. Advice will be sought from a range of disciplines namely GIS specialists, environmental engineers, statisticians and cancer data experts. The next chapter describes the methodology for the case study in detail. It outlines the data sourced, the technologies used and the methodological approach implemented.

3. Materials and Methods

The previous chapter outlined the findings of the literature review relating to the techniques, concepts and use of GIS and health in general with a specific evaluation of the Irish situation. It also examined the research relating to nitrates and their connection with cancer. In order to illustrate the advantages of using a GIS for an epidemiological study in Ireland a case study was undertaken. This case study reviewed nitrates in drinking water and possible links with cancer. To facilitate the case study data was sourced from different relevant organisations. GIS techniques were evaluated to determine the most useful and applicable. Advice was sought from environmental engineers, GIS consultants and statisticians. This chapter outlines the methodology for the case study. It outlines the data sourced, the technologies utilised and the methodological approach adopted.

3.1 Datasets

Four datasets were required for this research. These datasets are summarised in this section.

3.1.1 Electoral Divisions

Ireland is divided into administrative areas called Electoral Divisions (EDs). Townlands make up EDs and EDs in turn make up local authority areas that then roll up into counties. There are just over 3,400 EDs in Ireland and this areal unit is generally the smallest area for which data is available from the Central Statistics Office (CSO). Figure 4 illustrates EDs and local authorities in the Dublin area.

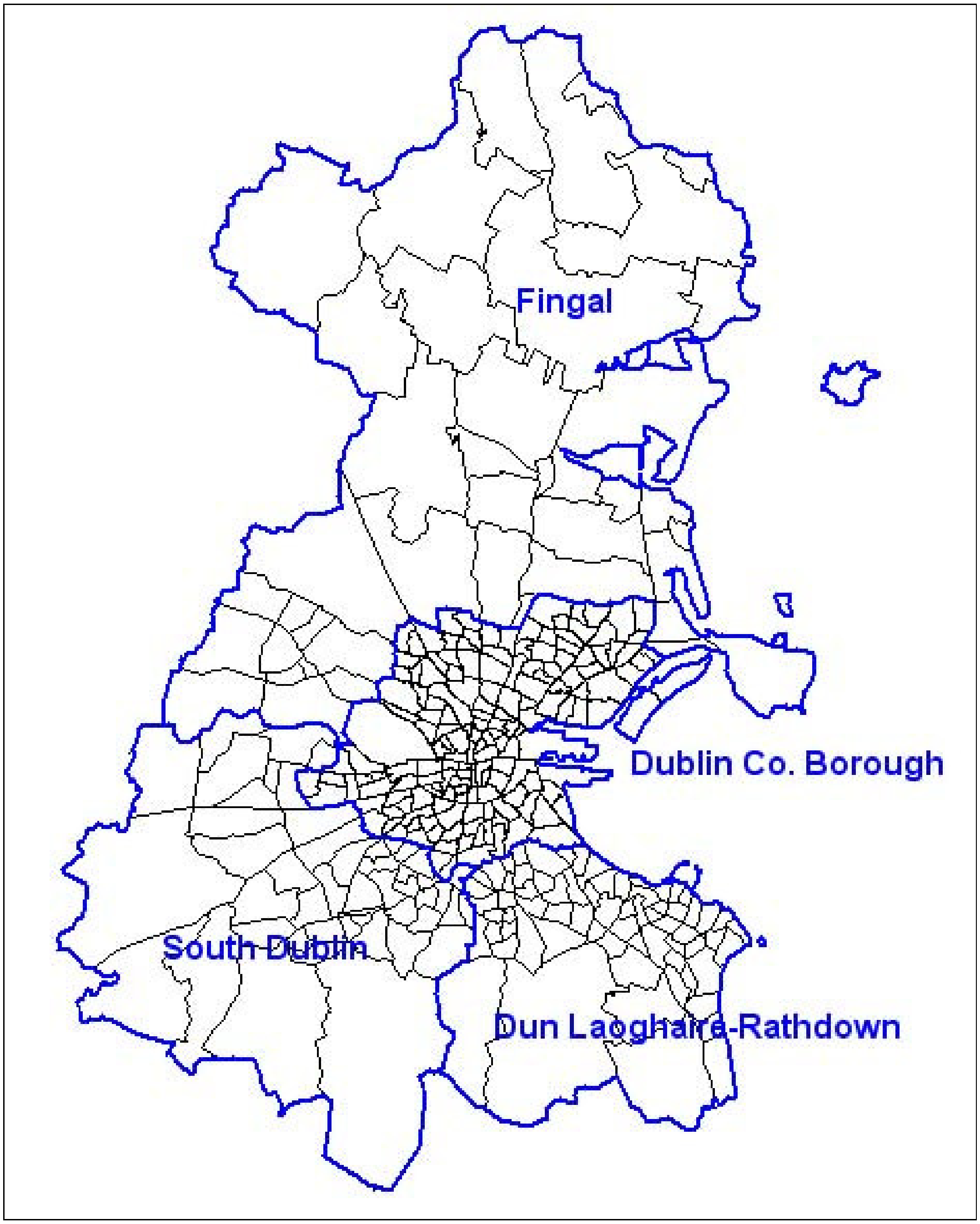

Figure 4 Dublin region showing the four Local Authority Areas and EDs

The CSO publishes small area population statistics (SAPS) at ED level. SAPS provide detailed analyses of ED populations. The 2002 SAPS contain 1,161 variables Electoral Divisions (CSO 2002). ED boundary maps were obtained from OSI. EDs were used as the unit of analysis for the case study for two reasons. Firstly the most detailed cancer data was available at this level. Secondly Kelly and Rybaczuk (2002) under took a study to review the influence of collection units and recommended to use ED level in Ireland.

3.1.2 Ground and Surface Water Chemical Data

Two chemical drinking water datasets were sourced from the Environmental Protection Agency (EPA). The EPA is responsible for the compilation and verification of monitoring results from all drinking water supplies. The EPA collates results from all local authorities and prepares an annual report at national level. This report outlines the overall quality of drinking water in Ireland. Amongst other things the EPA advises and assists local authorities in preparing data and it checks the analytical quality control systems used in laboratories undertaking analysis of drinking water (EPA 2005).

The datasets represent discrete samples from monitoring points around the country. One dataset contained surface water data and the other ground water data. The surface water data came from rivers rather than lakes. This dataset also included other chemical parameters including pH, temperature, ammonia, chloride, nitrates and phosphates.

3.1.3 Cancer Data

The Minister for Health established the NCRI in 1991. The NCRI collates and manages information relating to cancer incidences in Ireland. It collects data about each newly diagnosed patient and each tumour that occurs. The NCRI encourages and supports the use of this data for research and to assist with service provision and management (NCRI 2007).

A cancer dataset was sourced from the NCRI. This dataset contained the number of cases, the expected number of cases and the SIR for a range of cancers. The data included figures for male and female. It was provided at ED level. The cancers included in the dataset were oesophagus, stomach, colon, rectum, head and neck and colorectal.

The literature review highlighted the fact that analysis of disease occurrence should be based on rates. A crude rate is the relationship between the number of individuals displaying a particular characteristic and the total population. An SIR is often used for small areas. To gauge the position of an area in relation to disease occurrence, it is fitting to try and determine what disease occurrence should be 'expected' locally in the area and then to draw comparisons with the actual occurrence by calculating an SIR for the area (Lawson 2001). The SIR adjusts for the age distribution of the population and also helps to relate rates in individual areas to an overall rate for the entire area being studied. An SIR is calculated by dividing the observed cancer occurrence by the expected occurrence. If expected and observed occurrences were the same the SIR would be 100. Ratios higher than 100 indicate more cancers than expected (also called elevated), and those lower than 100 could be less than expected (also called depressed) (Goldstone 1983; Elliott and Wartenberg 2004;

Comber 2005). The cancer dataset received includes SIRs for each ED for six types of cancer. This SIR tells us if the rate in the ED is greater than or less than the rate for Ireland as a whole.

3.2 Technologies

Three commercially available products were used for the data analysis. MapInfo Professional was used to undertake the spatial analysis. MapInfo Professional is a Microsoft Windows-based desktop application that enables GIS professionals to easily visualise and analyse relationships between geography and data. Using MapInfo Professional, sophisticated and detailed data analysis can be performed by leveraging location (MapInfo 2007). MapInfo Professional is used by various government agencies in Ireland. An alternative to MapInfo Professional is ESRI's ArcGIS.

Microsoft Access was used for querying, grouping and amalgamating the chemical datasets. Microsoft Access is a relational database desktop application from Microsoft. It allows a user import data from external sources, analyse and organise the data by running queries and providing the user with the ability to sort, filter and group information. (Microsoft 2007).

Microsoft Excel was used for statistical analysis. Microsoft Excel is a spreadsheet desktop application from Microsoft. It is a effective tool that can be used to create and format spreadsheets and analyse available data (Microsoft 2007).

3.3 Methodology

The previous two sections described the data and technologies used for the case study. This section describes the methodology used.

3.3.1 Data Preparation

The surface water chemical file received from the EPA contained 27,664 records representing various chemical parameters for surface monitoring points from 1995 to 2003. The relevant data was identified and non-nitrate parameters were deleted resulting in 3106 records. These records were then grouped and the average for each monitoring point was calculated. The original file had a minimum, median and maximum level for each record. The median was used to calculate the average. This produced an average nitrate surface water level for approximately 2000 locations. The ground water chemical file received from the EPA contained 319 records representing the average nitrate level for ground water monitoring points from 2001 to 2003.

Once the averages for surface and ground water points were available the two datasets were combined into one table which contained the time period, location, x and y coordinate, average nitrate level and source i.e. surface or ground. Each record in the table represented an average nitrate level for a point in Ireland. Each point was assigned a nitrate risk level of low, medium or high. Table 2 illustrates the nitrate level associated with each risk level. These levels are based on the methodology used by Gulis, Czompolyova et al (2002). This part of the process resulted in a nitrate risk level being assigned to each monitoring point for both the original ground and surface water datasets. This analysis was completed using Microsoft Access.

Nitrate Level (mg/L)	Risk Level
Less than 10	Low
Between 10.1 and 20	Medium
Greater that 20.1	High

Table 2 Nitrate Risk Level Values

3.3.2 Spatial Analysis

In the previous step a nitrate risk level was associated with various monitoring points around Ireland for both surface and ground water. The next step was to determine which EDs contained these points and what risk level should be assigned to each ED. One of the key benefits of a GIS is that disparate data can be opened as layers of data in a GIS and these layers can be linked using the geography of each feature or row value as the key.

MapInfo professional was used perform the spatial analysis required to assign a nitrate risk score to each ED based on the point data prepared. The OSI ED file was opened and overlaid on this was the point file containing the chemical monitoring point data. An initial high-level assessment of the data was undertaken to ensure all points were within the bounds of the Republic of Ireland. The average risk rate for every ED was then calculated. All points within an ED were identified and the average risk rate of these points was associated with the ED the points were within. The ED layer was updated with this value. A choropleth map was then generated to review the risk rate at ED level. The image in Figure 5 illustrates this choropleth map. It can be seen from the legend that 62% of EDs do not contain a rate, as there were no monitoring points within these EDs.

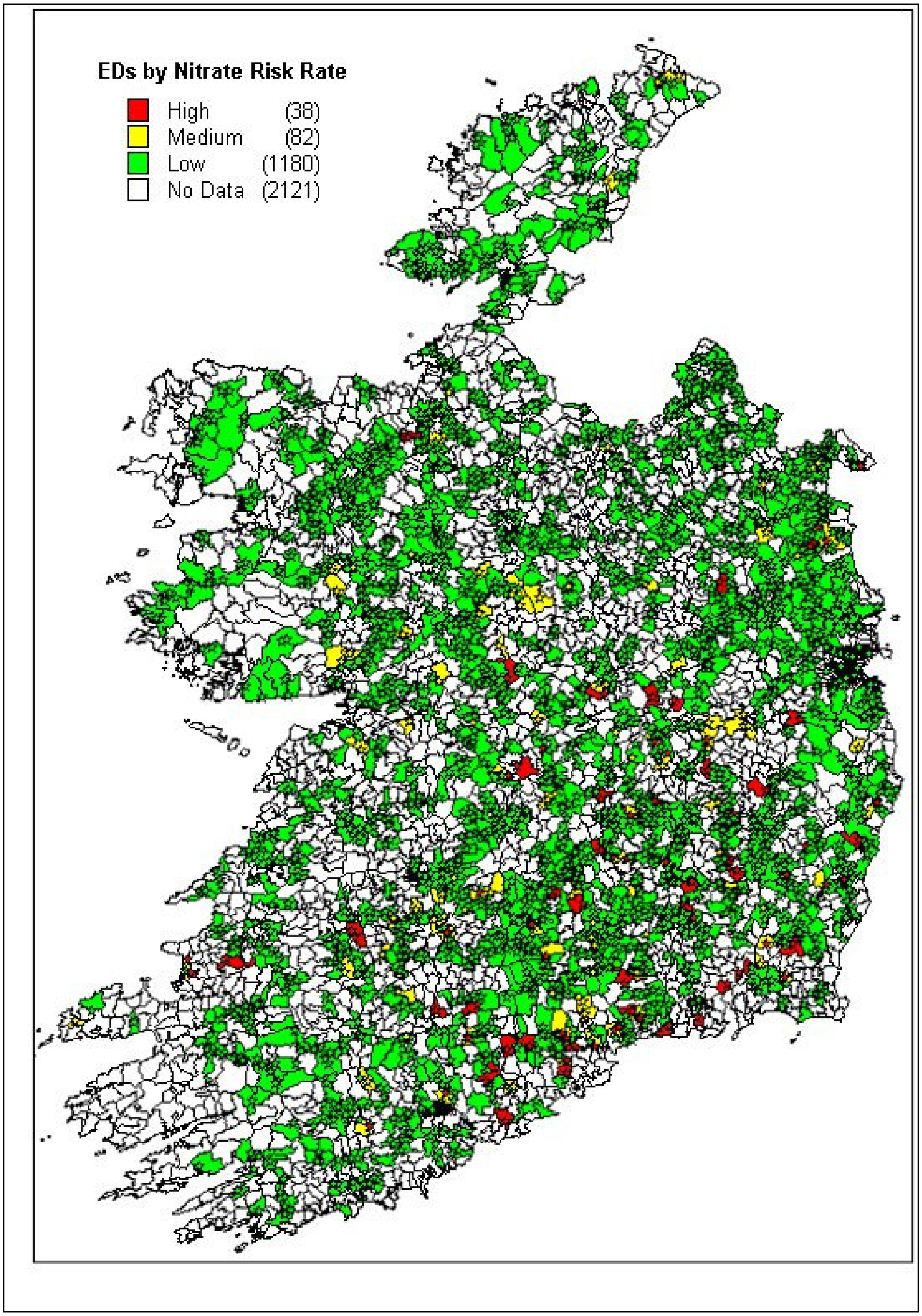

Figure 5 Choropleth Map of Nitrate Risk levels for EDs in Ireland.

The author sought advice from an environment engineer and a GIS specialist to determine a method of assigning a risk rate to the EDs without a value. One approach is to determine which river catchment and sub-catchment the ED belonged to and to assign a rate based on the average of EDs in the river catchment and sub-catchment. The river catchment, or drainage basin, is all the land from the mountain to the seashore, drained by a single river and its tributaries. However an examination of the data revealed that EDs with risk levels did not have the same level within a catchment so therefore this method would not be accurate. Another approach is to use the value of neighbouring ED and distance to neighbouring points. Again however it was viewed that the rate would at best be an estimate.

Risk Score	Number of EDs	Percentage
High	38	1.11
Medium	82	2.40
Low	1180	34.49
No Data	2121	62.00
Total	3421	100.00

Table 3 Number of EDs with each Nitrate Risk Level

Therefore it was decided to only consider EDs with a risk level in the next set of analysis. Table 3 illustrates the number of EDs with each level. Cancer SIRs were extracted from the cancer datasets for the remaining 1300 EDs. Having assigned a risk level to each ED with data and having extracted the cancer SIRs for these EDs the next step is to determine if there is a correlation between risk levels and cancer rates.

3.3.3 Statistical Analysis

In the previous step a nitrate risk level was assigned to each ED based on chemical data at monitoring points around the country. 1300 EDs emerged for analysis and cancer SIRs were extracted from these EDs for both men and women. The next stage of the analysis is to determine if there is a correlation between nitrate risk levels across EDs and cancer SIRs for each type of cancer. A correlation is an association between two variables which is monotonic, this means it goes in one direction. A positive or direct correlation indicates that an increase in one factor will result in an increase in the other. A negative or inverse correlation indicates that an increase in one factor will result in a decrease in the other. The correlation coefficient is calculated to ascertain the association. This is a number between 1 and –1. The closer the correlation coefficient is to 1 or –1 the stronger the relationship. A correlation coefficient of 0 means that there is no linear relationship between the variables (Goldstone 1983).

In line with similar ecological studies reviewed in the literature Pearson's product moment correlation coefficient (r) was calculated for each cancer type across EDs for both males and females (Chen and Wang 1990; Johnson, Dack et al. 1994; Mills 1998; Mizoue 2004; Jemal, Ward et al. 2005; Ayotte, Baris et al. 2006; Garland and Garland 2006). Once the correlation coefficient is calculated the results are ready for analysis.

This chapter has described the datasets collated for the case study. It outlined the technologies used and described the methodology. The next chapter discusses the results. It outlines drinking water quality in Ireland, it illustrates the possible etiological pathway for nitrates and cancer and it outlines the results of the statistical analysis.

4. Results

A case study was undertaken to demonstrate the benefits of using a GIS for an epidemiological study in Ireland. It reviewed nitrates in drinking water and possible links with cancer. The previous chapter described the data sourced and technologies used for this case study. It also included a description of the methodology used to prepare and analyse the data. This chapter describes the results of the research. It deals with data findings for drinking water quality in Ireland. It outlines why nitrate could be considered a cancer cause. Finally it explains the results of the statistical analysis of an ED's nitrate risk rate and any correlation with cancer incidence in Ireland.

4.1 Drinking Water Quality and Nitrate Levels in Ireland

The EPA produces a drinking water report on an annual basis. This report collates and reviews the monitoring of drinking water supplies in accordance with EU regulations. The production of the 2005 report involved the evaluation and appraisal of over 200,000 individual tests undertaken on over 16,500 samples of drinking water. The EPA uses 48 drinking water standards to determine water quality. On the whole in 2005 the rate of compliance with these 48 standards was 96.7%. However although compliance was high this may not be a true reflection of the actual quality as there was a significant shortfall in the monitoring of water supplies. The report states "9% of public water supplies, 65% of public group water schemes and 41% of private group water schemes were not adequately monitored and in some cases not at all." One of the tests carried out by the EPA is for levels of nitrates. Elevated nitrate levels in Ireland are not common, the 2005 report illustrates that in Ireland there is 99.4% nitrate compliance across all monitored sites. However in 2005 the number of cases exceeding the parametric value was up 2 to 29. The EPA state that more monitoring of nitrate levels should take place in particular in potentially vulnerable areas like private water supplies in rural agricultural areas (EPA 2005).

The figures in the previous chapter are derived from the chemical data collected at monitoring stations. The same data is used in the EPA drinking water reports. The data for this case study illustrated that Ireland does not have widespread nitrate contamination of drinking water sources. Most nitrates found come from organic and inorganic sources. Organic sources include waste discharges and inorganic include artificial fertilisers (EPA 2007). In Ireland the large proportion of land used for grazing and relatively low use of artificial fertiliser means there is a reduced risk of nitrate contamination (European Commission 2000).

Although nitrate levels in Ireland could be considered low, nitrate levels in rivers have risen since the late 1970s. This increase has coincided with the demise of the pearl mussel, which is very sensitive to pollution. Similarly nitrate levels have increased in ground water. Elevated levels appear to occur primarily in the southeast although high levels can also be found in other places like county Louth and Kerry. In some cases elevated nitrate levels can be found at monitoring points that are close to waste discharges. However elevated concentrations seem to be associated more with intensive agricultural practises. This could develop into a extensive problem without adequate planning and management. It could result in drinking water treatment requirements for nitrate or could lessen the utility of ground and surface water as sources of drinking water (EPA 2005).

Two EU directives exist to help protect water sources. These are the Nitrates Directive and the Water Framework Directive. The EU recognises that nitrate levels in drinking water can be linked to agricultural practises like the application of slurry on land resulting in run off into rivers or lakes and the use of nitrogenous fertilisers. The Nitrates Directive is EU legislation aimed at reducing nitrate pollution of water as a result of agriculture (FSAI 2006). In Ireland the Nitrates Directive has been adopted and implemented since 1991. This includes widespread inspection of nitrate levels in waters, the examination of the trophic status of waters, the development, circulation and diffusion of a Code of Good Agricultural Practice to Protect Waters from Pollution by Nitrates and a number of other measures which exist to protect quality of water from pollution by agricultural sources. However the European Court of Justice has found Ireland non-compliant with the Nitrates Directive. The Government plans to take all reasonable efforts to respond to this judgment and to accomplish compliance at the earliest possible date. The government maintains that it is not practical to attain instant compliance with all requirements on all holdings and therefore has developed an action programme to ensure that the suitable measures become operational on a phased basis taking cognisance of the required lead-in times needed by farmers (DOEHLC 2005).The EU Water Framework Directive creates a new framework for community action in the area of water policy. The general aim is to achieve, at a minimum, good status for all waters by 2015. Good status includes chemical and ecological elements of water. The EU Water Framework directive is currently being implemented. It has been transposed into national law (FSAI 2006).

Overall the quality of water in Ireland by European standards is generally good. Nevertheless quality has declined from the condition that prevailed when surveys commenced in the early 1970s. While Ireland has accomplished 99% compliance with the nitrates standard any breaches for water supply is of critical concern. A significant deepening of resolve and effort is needed by all sectors to protect and maintain water quality (DOEHLC 2005; EPA 2005).

4.2 Nitrate Levels and Cancer Aetiology

Nitrates exist naturally in soils, waters, plants, and in meats. They are part of the universal nitrogen cycle and are ubiquitous in the natural world. In healthy persons, nitrates and nitrites are quickly absorbed from the gastro-intestinal tract. Nitrite that is absorbed reacts with haemoglobin to form methaemoglobin. Methaemoglobin is unable to carry oxygen. However, in adults, methaemoglobin is rapidly converted to oxyhaemoglobin. Oxyhaemoglobin transports oxygen to the cells of the body. In babies up to three months old the enzyme reducing system that converts methaemoglobin to oxyhaemoglobin is not entirely developed. Therefore the methaemoglobin formed can rise in the body leading to a characteristic clinical condition called methaemoglobinaemia or blue-baby syndrome (WHO 1978).

The International Agency for Research on Cancer (IARC) examines environmental factors that may increase human cancer risk. Working groups of interdisciplinary scientific experts analyse published studies and review the weight of the evidence that an environmental agent can increase the risk of cancer. Since 1971, in excess or 900 agents have been examined and approximately 400 have been recognised as carcinogenic or potentially carcinogenic to humans (IARC 2007). A working group meeting in June 2006 looked at ingested nitrates and nitrites. It reviewed a number of studies and produced a summary report. The report explains that human contact with nitrate is primarily from exogenous sources through the intake of water and food. Nitrite exposure is principally endogenous i.e. in the body nitrate is reduced to nitrite. Nitrosation reactions can occur when nitrite reacts with amines and amides under acidic gastric conditions in the human stomach to form N-nitroso compounds. Some of the N-nitroso compounds formed under these circumstances are known carcinogens. Vitamin C inhibits nitrosation reactions. The working group concluded that in humans there is limited or inadequate evidence to suggest that ingestion of nitrate in food or drinking water is carcinogenic. However it found that in circumstances where ingested nitrate or nitrite undergoes an endogenous nitrosation process to generate N-nitroso compounds it is probably carcinogenic to humans (IARC 2006).

The suggested path linking nitrate to cancer is complex. Levels of nitrosation will vary from one individual to another, which means various degrees of exposure to N-nitroso compounds. Although nitrosation variances among individuals are not fully understood some factors have been established. Smoking is a possible nitrosation enhancer while vitamin C is a nitrosation inhibitor (Coss, Cantor et al. 2004).

4.3 Nitrate Levels and Cancer Incidence in Ireland

This case study undertaken reviewed a number of cancers and their possible associations with nitrate levels in EDs. These cancers were oesophagus, stomach, colon, rectum, head and neck and colorectal. The cancers were examined against a nitrate risk rate for an ED. EDs with chemical data from monitoring stations were used. There were 1303 of these EDs representing 38% of the total EDs in Ireland. Of the 1303 EDs used in the study only 2.9% had a high nitrate risk level and therefore correlations of statistical significance were not expected. Table 4 illustrates the number of EDs and the associated nitrate risk level.

Risk Score	Number of EDs	Percentage
High	38	2.92
Medium	82	6.29
Low	1183	90.79
Total	1303	100.00

Table 4 Number of EDs (with data) with each nitrate risk level

The cancer dataset used included SIRs for oesophagus, stomach, colon, rectum, head and neck and colorectal cancer. SIRs are frequently used when examining rate in small areas. SIRs are adjusted for the age distribution of the population. They indicate if an area's disease occurrence is more of less than expected (Goldstone 1983; Elliott and Wartenberg 2004; Comber 2005). The next stage of the analysis determined if there was a correlation between nitrate risk levels across EDs and cancer SIRs for each type of cancer. The objective was to determine if an ED had a high nitrate rate did this correspond with an elevated SIR.

A cursory examination of the data did not reveal any obvious patterns. Where a nitrate risk factor was high in an ED in some cases the SIR was elevated but in other cases it was not. Similarly if an ED's nitrate risk factor was low this did not always correspond to a low SIR. Statistical analysis was carried out by calculating Pearson's product moment correlation coefficient (r) for each cancer type across EDs for males, females and both. A number close to 1 indicating that an increase in one factor will result in an increase in the other represents a positive or direct correlation. A number close to -1 indicating that an increase on one factor will result in a decrease in the other represents a negative or inverse correlation. A correlation coefficient of 0 means that there is no linear relationship between the variables (Goldstone 1983).

Cancer	Oesophagus	Stomach	Colon	Rectum	Head and Neck	Colorectal
Female	0.0242	0.0536	-0.0091	-0.0054	-0.0142	-0.0063
Male	-0.0603	-0.0055	0.0149	-0.0136	-0.0353	-0.0006
Both	-0.0445	0.0220	0.0030	-0.0118	-0.0337	-0.0040

Table 5 Correlation coefficient between nitrate risk levels and cancer SIRs in Ireland

Table 5 illustrates the results of this analysis indicating the correlation coefficient between nitrate risk levels and cancer SIRs in Ireland. The majority of correlation coefficients are either close to zero or are negative. This is the case for male, female and combined SIRs. This implies that there was no correlation between nitrate levels and cancer SIRs across EDs in Ireland. As a general rule a correlation coefficient of 0.6 or greater would indicate an acceptable association between variables (Goldstone 1983). In this case the highest correlation coefficient is 0.06. Therefore overall there were no significant findings in relation to cancer incidence and nitrate levels in drinking water in Ireland. However these finding cannot be considered conclusive, as they are a product of a study that had considerable data shortcomings in terms of nitrate risk levels in EDs due to lack of nitrate monitoring. Additionally as drinking water quality in Ireland is high, nitrate levels were low so it is not likely to be a risk factor in an Irish context.

This chapter has detailed the results of the case study. It outlined drinking water quality in Ireland, nitrate and cancer pathways and the results of the statistical analysis of an ED's nitrate risk rate and any correlation with cancer incidence in Ireland. While the case study did not find correlations between nitrates in drinking water and cancer it did illustrate the use of a GIS for an epidemiological study in Ireland. It showed the benefits as a visualisation technique, as an administrative tool for linking disparate data and as a collaborative mechanism for use in cross-functional teams. The next chapter discusses the benefits and limitations of using a GIS for an epidemiological study in Ireland in more detail.

5. Discussion

The previous chapter discussed the results of the case study undertaken to review nitrates in drinking water and possible links with cancer in Ireland. It outlined the quality of drinking water in Ireland, possible pathways linking nitrate and cancer and the results of the statistical analysis of an ED's nitrate risk rate and any correlation with cancer incidence for six types of cancer. The case study undertaken did not find correlations between nitrates in drinking water and cancer incidence for oesophagus, stomach, colon, rectum, head and neck and colorectal cancer. These findings are in line with other studies of this nature that dealt with relatively low levels of nitrates (Barrett, Parslow et al. 1998; Roos, Ward et al. 2003; Ward, Cantor et al. 2003; Coss, Cantor et al. 2004). A path from nitrate levels to cancer aetiology is biologically plausible. The results represented by the case study do not mean that there is no link. They simply mean that in Ireland, for the 38% of the country evaluated, nitrate levels were not high and as a consequence correlations with cancer incidence were not found.

What the case study did illustrate was that the use of a GIS for an epidemiological study in Ireland could add value in a number of areas. These include visualisation techniques to illustrate patterns and gaps in data, linkage of disparate datasets to undertake analysis, and its use as a tool to foster collaboration between multi-disciplinary teams. This case study also highlighted issues in terms of data acquisition and quality, technology and ecological study limitations. This chapter discusses the benefits and limitations in more detail.

5.1 Benefits of GIS

5.1.1 Visualisation and Hypothesis Generation

A GIS is a powerful visualisation tool. Mapping disease rates can highlight previously undiscovered patterns that can help with suggesting reasons and bring to light areas for targeted studies in the field. Various hypotheses can be modelled with different variables and potential exposures to help understand possible aetiologies (Maantay 2002). Aase and Bentham (1998) used a GIS to examine the geographic variations of prostate cancer incidence in Norway. They acknowledge that the unusual international variations in occurrences could indicate that factors illustrating geographical variations may provide etiological clues. However their hypothesis was that a large proportion of local cancers could be due to intensive diagnosing rather than any etiological reason. Having undertaken the study they found that some regional differences could be artefactual due to diagnostic intensity but they also found considerable genuine differences in occurrences across different areas. The most striking finding was the low rates in the northern zone. They concluded that there was genuine basis for searching for etiological clues about the causes of the geographic variations. By visualising the data on a map an entire new dimension was added to the data. In the US tabular mortality data had been published for many years. In 1975 rates were mapped for the first time and it was only then that spatial patterns were seen. For example there was a cluster of high rate of oral cancer in the southeastern states. These were later found to be due to the use of smokeless tobacco (Pickle, Waller et al. 2005).

A GIS goes further than traditional database tables or spreadsheets and facilitates the discovery and conceptualisation of new data patterns and relationships that without a GIS may have remained hidden. The types of visualisation techniques available in a GIS are varied and can be tailored to the problem being addressed or the data available. These techniques are helpful when reviewing both the geography of healthcare systems and the geography of disease. The geography of healthcare systems deals with the development, running and delivery of health services. Part of this includes determining the needs of a catchment or community and ensuring adequate patient access. Simple maps can display locations of health facilities and associated road, rail and public transport networks to help the reader visualise gaps or areas that are underserved. Geographical visualisation of health services can help identifying inequities in health care delivery between socio-economic groups and different regions. The visualisation can facilitate the allocation of staff based on actual needs, and help with determining locations and specifications for new healthcare facilities or in planning extensions to existing ones (Boulos, Roudsari et al. 2001).

For epidemiologists GIS visualisation techniques offer a range of facilities in the study of the geography of disease. Visualisation is the primary step in exploratory data analysis. A GIS can, for example, quickly and easily create maps of morbidity and mortality patterns linked to population size and class, possible negative exposures and geographic areas (Jacquez 2000). Disease maps that can be generated in choropleth and isopleth form assist in understanding disease occurrence and clusters (Berke 2004). The nature and transmission of a disease can be mapped over time and space to understand patterns and flows (Clarke, McLafferty et al. 1996).

In the case study undertaken the generation of a choropleth map of nitrate risk levels for EDs in Ireland (see Figure 5) immediately illustrated two facts about the data. Firstly data was only available sporadically around the country. The map is a powerful mechanism to argue for increased monitoring of surface and ground water. Secondly it showed instantly that nitrate levels in Ireland in general were good and compliant with international standards. A map displays how “a picture paints a thousand words.” Externally the reader focuses on a map’s facts and function while internally uses cognition to perceive a maps meaning. Information is conceptualised and processed easily by the reader. This allows for faster and more cohesive decisions to be made about a set of information presented by a GIS (Bell, Hoskins et al. 2006).

5.1.2 Collation and Analysis of Disparate Data

The information input into a GIS database can come from a variety of sources. The data received can have different spatial resolutions representing anything from points or addresses to townlands, EDs or counties right up to national level. The GIS arranges each data source as a layer. Once the layers are loaded and placed on top of each other the software clarifies the geographical relationships between the different sets of data (Betts 1997). The multi-layer GIS database is an important feature of GIS in general. If a map contains a layer of administrative areas like EDs then other information layers can be superimposed and shown simultaneously.

Data sources can be rich and varied including administrative data like cancer registries and population census data. Ground survey data collected using a Global Positioning System (GPS), can be utilised if environment measures are collected in the field. These could include chemical data in a river, pesticide levels in soil or air quality data. Also available for use is satellite data from remote sensing or aerial photography showing images of the earth. This form of data is useful for obtaining land use and land characteristics of an area. An example of the use of these different forms of data combined in, for example, a cancer study would be to map residences on areas of crop production and include some random pesticide samples gathered at points on the map using a GPS. One could then overlay cancer data and estimate pesticide exposure in an area and observe correlations with cancer incidence (Boscoe, Ward et al. 2004).

The multi-layer architecture is an important feature of a GIS. In addition to the ability to overlay layers and view them collectively it also enables spatial transformations of data e.g. the selection of all cancer incidences within 200 metres of a river. In this manner a GIS facilitates the step from descriptive to investigative epidemiological studies. This helps the epidemiologist generate hypotheses about different data associations (Boelaert and Stuyft 1998). While simple mapping is valuable a GIS offers much more than this. Overlay analysis empowers a researcher to compute various values of location based on many layer attributes to perhaps identify areas that meet a certain criteria or to test a hypothesis (Clarke, McLafferty et al. 1996).

This case study completed as part of this research collated data layers incorporating administrative areas (EDs), chemical surface and ground water point information and cancer rates for administrative areas. By loading the chemical point information and overlaying it on an ED layer the author was able to assign a nitrate risk level to each ED based on the chemical values for drinking water at points within the ED. Once the risk rate was assigned to each ED it could then be compared with the corresponding cancer SIRs to establish correlations. Put simply this research could not have been completed without the visual, overlay and analytical powers of a GIS.

5.1.3 Multidisciplinary Teams

A GIS can take data from different sources, disciplines and scientific domains and combine it. The combined data can then be analysed and visualised, twisted and turned, modelled and hypothesised. This ability has fostered a setting whereby people from different areas and specialties come together to solve complex and varied problems. Map analysis began in the 1960s and 1970s for assessing resources, land use and planning purposes. The recognition that various components on the earth's surface were not independently functioning entities but rather had effects on complex interrelated systems meant that people wanted to assess them in an integrated way using multidisciplinary teams (Clarke, McLafferty et al. 1996; Kaminska, Oldak et al. 2004). In addition to its powerful visual and analysis capabilities, a GIS provides the circumstance for the combination of multidisciplinary methods and ideas to provide a clear, cohesive and more holistic picture of health (Boulos 2004).

An epidemiological study will be most successful if it incorporates knowledge and expertise from different disciplines. Environmental scientists help verify the accuracy, validity and pathways of exposure data. Epidemiologists select appropriate mathematical models and statistical formulas. They also look for and establish additional data that, if omitted from the correlation test, could confound the results. Clinicians ensure any hypothesis is biologically plausible. The GIS engineer or geographer will geocode or georeference the data so that it can be combined into a GIS (Nuckols, Ward et al. 2004). When a multidisciplinary approach is adopted the most applicable methodology for each study will be used rather than the methodology with which researchers are most familiar (Cockings, Dunnb et al. 2004).

A multidisciplinary methodology will also inadvertently cultivate a data sharing ethos. Agencies or departments usually collect data in silos where each department creates and analyses its own specific data. In many cases research teams can have complementary data and be unaware that the other dataset exists or be unwilling to release the data. When professionals from different disciplines come together awareness of data sets emerge. Relationships develop and trust and respect emerges between teams and disciplines. In this type of environment data sharing occurs naturally aiding existing research projects and creating ideas for new ones. For example, to create a map of road accident injuries and deaths a local hospital could establish a data partnerships with the department of transport, ambulance services and local authorities (Richards, Croner et al. 1999).

In order to complete this research the author liased with cancer data experts, GIS technicians, academics and environmental engineers. By combining the knowledge and experience of these different disciplines it was also possible to determine the appropriate data required for the research. It was possible to establish data availability, sources and contacts. A methodological approach could be established and verified. An understanding of the results could be achieved.

5.2 Limitations of GIS

5.2.1 Data Acquisition and Quality

A key benefit of GIS is its ability to link data from different sources however data acquisition and quality is the single biggest obstacle to the implementation of GIS projects. The main restriction in the uptake of most GISs is the acquisition and implementation of the data layers. This part of the process can account for 70% of the research effort (Vine, Degnan et al. 1997). It generally takes munch longer than the analysis itself. This is often overlooked when planning projects even though data acquisition and quality are critical success factors (Pickle, Waller et al. 2005).

Data acquisition is generally challenging because information on human health needed for research is usually dotted across many different agencies. Records are collected for clinical reasons and usually do not hold demographic information. They may include a home address for billing reasons but would not usually have a list of previous residences or work locations (Pickle, Waller et al. 2005). Problems relating to patient confidentiality, lack of sharing of information between hospitals and doctors and national bodies all conspire against far-reaching national health databases (Maantay 2002). In Ireland a Hospital In-Patient Enquiry Scheme (HIPE) is maintained. The HIPE unit manages the collation, coding, entry, quality and reporting of data from contributing hospitals. A coder translates each patient's chart. The information is standardised for later analysis of the data. Diagnoses and procedures performed are classified using The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes. However only a patients county of residence is recorded (HIPE 2007). County level is too coarse a scale for meaningful epidemiological analysis in Ireland (Kelly and Rybaczuk 2002).

Another issue with data in epidemiological studies can be what is known as the space-time issue. What this means is the study is undertaken at a particular snapshot in time and does not cater for the long latency periods associated with certain diseases in particular cancer (Nuckols, Ward et al. 2004). It's difficult to obtain data with exposure and disease rates over time for a particular place. Health outcomes are as a result of a range of numerous and diverse exposures which often take place over a long time periods and in different places. For example most cancers develop over a period of up to 30 years. They arise due to many exposures intermingling with a person's genetic susceptibility. Latencies can vary based on an individual's vulnerability and also on the type of cancer (Pickle, Waller et al. 2005). The case study undertaken in this research attempted to mitigate against this issue to an extent by using an aggregate of cancer rates from 1994 to 2003 and an aggregate of chemical drinking water data from 1994 to 2003.

Migration can cause a data quality issue as people move around so exposures can vary (Houghton and Kelliher 2003). However arguably the biggest data challenge in the use of GIS in any sector in Ireland is the difficulty associated with geocoding Irish addresses. If data is not geocoded it has no spatial reference point on the earth and cannot be linked in a GIS. In Ireland the geocoding process is particularly challenging for two reasons. Firstly we do not have a postcode system. In the UK a researcher could match all address postcodes to a postcode area file quite simply. In Ireland there is no straightforward or inexpensive way of matching addresses to map co-ordinates. An address database for Ireland called GeoDirectory is commercially available. GeoDirectory was developed by An Post and OSi. It is a complete database of every building in the Republic of Ireland and contains 1.7 million records (An Post and OSi 2006). Some commercial organisations provide a geocoding service. Outsourcing the geocoding process can be expensive and in some cases cost prohibitive. The second challenge in Ireland relates to the fact that non-unique addresses exist. This means that rural inhabitants whose address is a townland have the same address as other residents of that townland because many rural areas do not have road names or house numbers. Mail or post is delivered simply due to the local knowledge of the postman. It is impossible to geocode these types of addresses to an exact point without visiting each address and taking a GPS co-ordinate. The final challenge relates to the fact that townland and ED names can be repeated, sometimes even within a county so verifying where to place an address on a map is difficult (Houghton 2004). Ireland is working on introducing a postcode system and has set a target date for implementation at no later than 1st January 2008 however a number of issues need to be overcome if that deadline it to be met (NPPB 2005). The absence of accurately geocoded datasets results in garbage in garbage out exercises that do not add value (Boelaert and Stuyft 1998).

This author pursued many case study ideas, which failed due to lack of data. Eight months, from October to May, passed before all data was collected for the case study undertaken. The data available only included 38% of the country as chemical data was unavailable for the rest of the country. To successfully complete a GIS research project in Ireland data availability must be established and time must be allocated to the acquisition of this data.

5.2.2 Ecological Studies

The case study undertaken in this research was an ecological study. A GIS is often utilised in this type of study. An ecological study is a piece of research that uses aggregate group data rather than individual level data. Making assumptions that a group observation at an area level holds for a person within that area can lead to what is known as the ecological fallacy (Wakefield and Shaddick 2005). Aggregate data is used to draw conclusions about individuals if individual data is not available. Individual data is often unavailable in health research due to patient confidentiality reasons. The main issue is that different possible values at individual levels can produce the same result at the aggregate level. For example, there are a myriad of ways in which electoral votes for a political candidate can divide among individual voters however regardless of this the same aggregate level of support will still be produced. This issue is one of the most unrelenting amongst researchers in epidemiology, geography, sociology, among others (Schuessler 1999). Results from ecological analysis of groups do not always hold at an individual level. Many researchers contend that ecological studies can be misleading and should be used with caution. Ecological studies should be considered hypothesis generating. Hypotheses can be verified by follow-up work. In fact an initial hypothesis can point the researcher in the right direction making more efficient use of resources in the field. Validation can occur at an individual level through cohort, case-controlled studies or through trial of prevention mechanisms or interventions (Bølviken 2001).

Although issues exist with ecological analysis there is still a case for performing this type of research in particular when known issues are mitigated against. Elliott and Wartenberg (2004) point out that these types of studies have been essential in discovering hypotheses of public concern. They refer to the research by Beasley (1988) which linked malignant hepatoma with hepatitis B infection by studying male Chinese government employees (civil servants) in Taiwan. Elliott and Wartenberg also refer to the work by Keys in the Seven Countries study. This study was the first to methodically evaluate the relationship between diet, lifestyle and the rates of stroke and heart attack in different populations. It illustrated the extent to which diet and saturated fat played a role in the aetiology of heart disease (Keys 1980). Therefore it can be seen that ecological studies can assist with hypothesis generation. The case study conducted in this research was ecological. However it was useful in determining that nitrates were probably not a factor in relation to cancer incidence in Ireland.

This chapter described the benefits of using a GIS for an epidemiological study. It outlined a GISs multidisciplinary nature and its visualisation, collation and analytical capabilities. It also highlighted the challenges faced when conducting these types of studies. These challenges generally emerge due to the nature of the studies, which are ecological and due to the difficulties encountered obtaining quality data.

6. Conclusion

This research set out to assess to what extent a GIS can assist with epidemiology studies in Ireland. A GIS allows an epidemiologist, or researcher, inspect geographical patterns in data and understand the relationship between a disease and other factors. High quality maps provide a facility for an epidemiologist to present a persuasive case for further studies, interventions and service provision (Brewer 2006).

To gauge the usefulness of using a GIS for an epidemiological study in Ireland a case study was completed. This cases study examined cancer incidence and nitrate levels in drinking water in Ireland. Cancer was chosen as it is the second biggest killer in Ireland. It was also used because the NCRI collates information on cancer incidence in Ireland therefore high quality data was available. Relevant data was sourced from OSi, the EPA and the NCRI. The cancers reviewed were oesophagus, stomach, colon, rectum, head and neck and colorectal. Using a GIS a nitrate risk rate for each ED in Ireland was calculated. This risk rate was calculated based on ground and surface water chemical data from national monitoring points. ED nitrate risk rates were then analysed against cancer rates to establish if correlations existed.

The previous chapter discussed the advantages of using a GIS in epidemiological studies. It outlined a GISs visualisation, collation and investigative potential and its multidisciplinary nature. Conversely the chapter also outlined the problems that generally materialise when using a GIS for epidemiological research. These relate to the ecological nature of the studies and the issues encountered obtaining quality data. This chapter contains the major findings of the research. It proposes future studies that could be undertaken. The chapter concludes with the contributions this research has made to the overall body of knowledge in this area in Ireland.

The research showed that there is no significant correlation between incidences of cancer and nitrate levels in drinking water in Ireland. While a path from nitrate levels to cancer aetiology is biologically conceivable no association was found using Irish data. The results do not mean that there is no link between nitrates and cancer. They merely mean that in Ireland nitrate levels are low and therefore associations with cancer incidences were not discovered. Therefore nitrate is unlikely to be a cancer risk factor in an Irish context.

This study illustrated the value a GIS can add in a number of areas. Using a GIS to ascertain or eliminate risk factors is a cost effective method of generating hypotheses. These hypotheses can be validated through field based case controlled or cohort studies (Aamodt, Samuelsen et al. 2006). Powerful visualisation techniques available through GIS can present compelling arguments in terms of data needs and gaps. Disease occurrence can be mapped presenting an extra, and sometimes missed, dimension to data. Patterns can emerge that were previously unseen. A GIS shows how "a picture paints a thousand words." Information presented by a GIS is conceptualised and processed easily by the reader. This facilitates faster and more cohesive decisions to be made about a set of information presented (Bell, Hoskins et al. 2006). The ability to connect unrelated datasets through their geography facilitates linkage and analysis of data that could not occur without a GIS. In linking data from different sources the GIS promotes dialog, partnership and cooperation amongst multi-disciplinary teams (Boulos 2004).

The study also outlined the difficulties encountered when using a GIS in an Irish context and highlighted issues that emerge. Data acquisition is usually difficult because information on human health needed for epidemiological research is usually scattered across different agencies and departments. Records are collected for clinical reasons and usually do not hold demographic information needed (Pickle, Waller et al. 2005). Studies are undertaken at a particular snapshot in time and often do not account for the long latency periods associated with certain diseases in particular cancer (Nuckols, Ward et al. 2004). In Ireland the address system means the process of geocoding Irish data, so that it is spatially referenced and can be used in a GIS, is arduous and expensive. As these studies are ecological they are limited to hypothesis generation they cannot present a definitive argument at an individual level (Elliott and Wartenberg 2004).

Future research naturally emerges from this study. The research evaluated nitrate levels against various cancers. The surface water dataset obtained also contained ammonia, biochemical oxygen demand (BOD), chloride, conductivity, phosphate and pH levels. These chemical levels could be examined against the cancer dataset received and against other cancers. The chemical data could also be analysed alongside other disease occurrences in Ireland subject to data availability. This case study has defined a methodology, which can be reused again and again for different chemical and disease combinations.

This research has made a number of contributions. Firstly and specifically the case study determined that nitrates were probably not a cancer risk factor in Ireland. This fact was previously unknown and this knowledge guides a researcher to look at other areas for cancer etiological clues. One in three Irish people are likely to contract some form of cancer by age 74 (Walsh, Comber et al. 2001). Cancer has touched the lives of almost every individual both in Ireland and internationally. Although some cancer risk factors such as smoking are well known the causes of cancer are for the most part unidentified. Any research that identifies or eliminates a risk factor will assist in the fight to find causes and cures.

The research has highlighted the difficulties in obtaining data in Ireland. It warns and advises a future researcher to allocate time up front for data collection. It illustrated gaps in the data used in relation to drinking water quality. Dr. Harry Comber provided the cancer data from the NCRI. At a meeting in May Dr Comber spoke to environmental health officers about the investigation of cancer clusters. In order to illustrate the point that there is not much environmental exposure data currently he presented the map in Figure 5 which was generated as part of this research. Awareness of data issues amongst the relevant parties will encourage more comprehensive data collection and sharing.

Finally this research has illustrated the benefit and contribution of the use of GIS. GIS is incredibly under utilised in Ireland (Houghton 2004). Many possible reasons exist however the completion of this study may prompt increased usage in health research. Although issues were encountered and persistence was required the results were fruitful in the end. This research could not have been completed without the use if a GIS. It proves that not only is it possible but often it is preferable to use a GIS when completing epidemiological studies in Ireland.

In conclusion cancer is a widespread disease. In Ireland it is the second biggest cause of death. This research used a GIS to conduct a case study that evaluated drinking water nitrate levels and their possible correlations with cancer. The case study determined that in general there were no significant association between cancer incidence and nitrate levels because drinking water nitrate levels in Ireland are low. The study illustrated the benefits of GIS including visualisation abilities, linkage of disparate datasets to undertake analysis, and its usefulness as a tool to cultivate cooperation between multi-disciplinary teams. The study also discovered issues that may be encountered in terms of data acquisition and quality, technology and ecological study limitations. The health of a population will improve the more we understand the geographical nature of disease and possible environmental etiological risk factors. Not only does this aid with disease prevention it also facilitates targeted interventions and appropriate resource allocation to help populations most at risk.

7. References

Aamodt, G., S. O. Samuelsen, et al. (2006). "A simulation study of three methods for detecting disease clusters." International Journal of Health Geographics **5**(15).

Aase, A. and G. Bentham (1998). "The geography of prostate cancer incidence in Norway: Are the patterns real?" Norsk Epidemiologi **8**(1): 65-72.

Aragonés, N., R. Ramis, et al. (2007). "Oespohageal cancer mortality in Spain: a spatial analysis." BMC Cancer **7**(3).

Ayotte, J. D., D. Baris, et al. (2006). "Bladder cancer mortality and private well use in New England- an ecological study." Journal of Epidemiology and Community Health **60**: 168-172.

Bailey, T. C. (2001). "Spatial statistical methods in health." Cadernos de Saúde Pública **17**(5): 1083-1098.

Barrett, J., R. Parslow, et al. (1998). "Nitrate in drinking water and the incidence of gastric, esophageal, and brain cancer in Yorkshire, England." Cancer Causes Control **9**(2): 153-159.

Beasley, R. (1988). "Hepatitis B virus. The major aetiology of hepatocellular carcinoma." Cancer **61**: 1942–1956.

Bell, B. S., R. E. Hoskins, et al. (2006). "Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public." International Journal of Health Geographics **5**(49).

Berke, O. (2002). "Explorative disease mapping: Kriging the spatial risk function from empirical bayesian estimates of regional counts." IBEI - School of Veterinary Medicine Hannover, Research Report July 2002.

Berke, O. (2004). "Exploratory disease mapping: kriging the spatial risk function from regional count data." International Journal of Health Geographics **3**(18).

Betts, K. S. (1997). "Mapping the environment." Environmental Health Perspectives **105**(6).

Bithell, J. F. (2000). "A classification of disease mapping methods." Statistics in Medicine **19**: 2203-2215.

Boelaert, A. and V. d. Stuyft (1998). "Editorial: Geographical Information Systems (GIS), gimmick or tool for health district management? ." Tropical Medicine & International Health **3**(3): 163-165.

Bølviken, B. (2001). Ecological analyses: radioactive elements versus nasopharyngeal carcinoma and multiple sclerosis. Natural Ionizing Radiation and Health Proceedings from a symposium held at the Norwegian Academy of Science and Letters, Oslo, Norway.

Boscoe, F. P., M. H. Ward, et al. (2004). "Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer." International Journal of Health Geographics **3**(28).

Boulos, M. N. K. (2004). "Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom." International Journal of Health Geographics **3**(1).

Boulos, M. N. K., A. V. Roudsari, et al. (2001). "Health Geomatics: An enabling suite of technologies in health and healthcare." Journal of Biomedical Informatics **34**: 195-219.

Brewer, C. A. (2006). "Basic mapping principles for visualizing cancer data using geographic information systems (GIS)." American Journal of Preventative Medicine **30**(2S).

Brody, J., A. Aschengrau, et al. (2004). "Breast risk and historic exposure to pesticides from wide-area applications assessed with GIS." Environmental Health Perspectives **112**(8): 889-897.

Chen, C.-J. and C.-J. Wang (1990). "Ecological correlation between arsenic levels in well water and age-adjusted mortality from malignant neoplasms." Cancer Research **50**: 5470-5474.

Chung, K., D.-H. Yang, et al. (2004). "Health and GIS: Toward spatial statistical analysis." Journal of Medical Systems **28**(4): 349-360.

Clarke, K. C., S. L. McLafferty, et al. (1996). "On Epidemiology and Geographic Information Systems: A review and discussion of future directions." Emerging Infectious Diseases **2**(2): 85-92.

Cockings, S., C. E. Dunnb, et al. (2004). "Users' perspectives on epidemiological, GIS and point pattern approaches to analysing environment and health data." Health & Place **10**(2): 169-182.

Comber, H. (2005). Cancer in Louth, [Online] Available from: <http://www.ncri.ie/pubs/pubfiles/louthreport.pdf> [Accessed 13 January 2007]

Committee, U. F. G. D. (2003). "Homeland Security and Geographic Information Systems – How GIS and mapping technology can save lives and protect property in post-September 11th America." Public Health GIS News and Information **52**: 21-23.

Coss, A., K. Cantor, et al. (2004). "Pancreatic cancer and drinking water and dietary sources of nitrate and nitrite." American Journal of Epidemiology **159**(7): 693-701.

CSO (2002). Small Area Population Statistics, [Online] Available from: <http://www.cso.ie/census/SAPs.htm> [Accessed 13 May 2007]

DOEHLC (2005). Nitrates action programme under the nitrates directive, <[Online] Available from <http://www.environ.ie/en/Environment/Water/WaterQuality/NitratesDirective/PublicationsDocuments/FileDownLoad,1571,en.doc> [Accessed 01 April 2007].

DOHC (2003). National task force on medical staffing (Hanly Report), [Online] Available from <http://www.dohc.ie/publications/pdf/hanly.pdf?direct=1> [Accessed 26 February 2007].

E-Tenders (2006). Health Atlas Ireland, [Online] Available from: <http://www.e-tenders.gov.ie/search/search_show.aspx?ID=FEB053373> [Accessed 02 March 2007]

Elliott, P. and D. Wartenberg (2004). "Spatial Epidemiology: Current approaches and future challenges." Environmental Health Perspectives **112**(9): 998-1006.

EPA (2005). The quality of drinking water in Ireland: A report for the year 2005, [Online] Available from: <https://www.epa.ie/OfficeofEnvironmentalEnforcement/PublicAuthorityEnforcement/DrinkingWater/DrinkingWaterDownloads/DW%20Report%202005%20Final%20Jan%2007.pdf> [Accessed 13 May 2007]

EPA (2005). Water quality in Ireland 2005: Key indicators of the aquactic environment, [Online] Available from <http://www.epa.ie/OurEnvironment/Water/WaterIndicatorReports/FileUpload,10215,en.pdf> [Accessed 25 May 2007].

EPA (2007). Parameters of water quality., [Online] Available from <http://www.epa.ie/rivermap/docs/Parameters.pdf> [Accessed 25 May 2007].

European Commission, D. E. (2000). Verification of vulnerable zones identified under the nitrates directive and sensitive areas identified under the urban wastewater treatment directive, [Online] Available from <http://www.epa.ie/OurEnvironment/Water/WaterIndicatorReports/FileUpload,10215,en.pdf> [Accessed 25 May 2007].

FSAI (2006). Safety of potable water in Ireland, [Online] Available from: <http://www.fsai.ie/publications/other/water_report.pdf> [Accessed 29 May 2007]

Garland, C. F. and F. C. Garland (2006). "Do sunlight and vitamin D reduce the likelihood of colon cancer?" International Journal of Epidemiology **35**(2): 217-220.

Gatrell, A. and M. Loytonen (1998). GIS and Health Research: An Introduction. GIS and Health. I. Masser and F. Salgé. London, Taylor and Francis Limited**:** 3-16.

Glass, G. E., B. S. Schwartz, et al. (1995). "Environmental risk factors for Lyme disease identified with geographic information systems." American Journal of Public Health **85**(7).

Goldstone, L. A. (1983). Understanding Medical Statistics. London, William Heinemann Medical Books Limited.

Grinsven, H. J. v., M. H. Ward, et al. (2006). "Does the evidence about health risks associated with nitrate ingestion warrant an increase or the nitrate standard for drinking water?" Environmental Health: A Global Access Science Source **5**(26).

Guidry, V. T. and L. H. Margolis (2005). "Unequal respiratory health risk: using GIS to explore hurricane-related flooding of schools in Eastern North Carolina." Environmental Research **98**(3): 383-389.

Gulis, G., M. Czompolyova, et al. (2002). "An ecologic study of nitrate in municipal drinking water and cancer incidence in Trnava District, Slovakia." Environmental Research **88**(3): 182-187.

Hales, S., W. Black, et al. (2003). "Social deprivation and the public health risks of community drinking water supplies in New Zealand." Journal of Epidemiology and Community Health **57**: 581-583.

Health, W. S. D. o. (2007). EpiQMS: Epidemiological Query and Mapping System, [Online] Available from: <https://fortress.wa.gov/doh/epiqms/asp/showmap.asp> [Accessed 02 February 2007]

Higgs, G. (2004). "A literature review of the use of GIS-based measures of access to health care services." Health Services and Outcomes Research Methodology **5**: 119-139.

HIPE (2007). Hospital In-Patient Enquiry Scheme [Online] Available from: <http://www.esri.ie/health_information/hipe/> [Accessed 02 June 2007]

Houghton, F. (2001). "Health GIS in Ireland - Lessons from New Zealand." Irish Geography **34**(1): 96-97.

Houghton, F. (2004). "Looking on from the sidelines: the use of geographical information systems in health boards in Ireland 2003." Area **36**(1): 81-83.

Houghton, F. (2005). "Hiding the evidence: the State and spatial inequalities in health in Ireland." Irish Geography **38**(1): 96-106.

Houghton, F. and K. Kelliher (2003). "The exposure fallacy: migration, mobility and ecological analysis of health status in Ireland." Irish Geography **36**(1): 47-58.

HSE (2006). Health Atlas Ireland, [Online] Available from <http://www.ich.ie/healthatlas.htm> [Accessed 26 February 2007].

IARC (2006). Vol 94. Summaries and evaluations: Ingested nitrates and nitrites, [Online] Available from <http://monographs.iarc.fr/ENG/Meetings/94-nitratenitrite.pdf> [Accessed 05 May 2007].

IARC (2007). Monographs on the evaluation of carcinogenic risks to humans, [Online] Available from <http://monographs.iarc.fr/index.php> [Accessed 05 May 2007].

IPH (2006). INIsPHO: Ireland and Northern Ireland's Population Health Observatory [Online] Available from <http://www.inispho.org/> [Accessed 26 February 2007].

Jacquez, G. M. (1998). GIS as an enabling technology. GIS and Health. I. Masser and F. Salgé. London, Taylor and Francis Limited**:** 17-28.

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

Jemal, A., E. Ward, et al. (2005). "Geographic patterns of prostate cancer mortality and variations in access to medical care in the United States." Cancer Epidemiology, Biomarkers & Prevention **14**(3): 590-595.

Johnson, Z., P. Dack, et al. (1994). "Small area analysis of low birth weight." Irish Medical Journal **87**(6): 176-177.

Kalogirou, S. and R. Foley (2006). "Health, Place and Hanly: modelling accessibility to hospitals in Ireland." Irish Geography **39**(1): 52-68.

Kaminska, I. A., A. Oldak, et al. (2004). "Geographical Information system (GIS) as a tool for monitoring and analysing pesticide pollution and its impact on public health." Annals of Agricultural and Environmental Medicine **11**: 181-184.

Kelly, V. and K. Rybaczuk (2002). The spatial distribution of cancer and material deprivation in the Republic of Ireland: The influence of collection units. 8th EC-GI and GIS Workshop ESDI- A Work in Progress, Dublin, Ireland.

Keys, A. (1980). Seven Countries - A multivariate analysis of death and coronary heart disease. Boston, Harvard University Press.

Kohli, S., H. N. Brage, et al. (2000). "Childhood leukaemia in areas with different radon levels: a spatial and temporal analysis using GIS." Journal of Epidemiology and Community Health **54**(11): 822-826.

Krivoruchko, K. and C. A. Gotway (2002). Expanding the "S" in GIS: Incorporating spatial statistics in GIS. CSISS Specialist Meeting on Spatial Data Analysis Software Tools, Santa Barbara, USA.

Krivoruchko, K., C. A. Gotway, et al. (2003). Statistical tools for regional data analysis using GIS. GIS, New Orleans, Louisiana, USA.

Kulldorff, M. (1998). Statistical methods for spatial epidemiology: Tests for randomness. GIS and Health. I. Masser and F. Salgé. London, Taylor and Francis Limited**:** 49-62.

Lawson, A. B. (2001). "Tutorial in biostatistics: Disease map reconstruction." Statistics in Medicine **20**: 2183-2204.

Leeuwen, J. V., D. Waltner-Toews, et al. (1999). "Associations between stomach cancer incidence and drinking water contamination with atrazine and nitrate in Ontario (Canada) agroecosystems, 1987-1991." International Journal of Epidemiology **28**(5): 836-840.

Levy, Y. and T. J. Ellis (2006). "A systems approach to conduct an effective literature review in support of information systems research." Informing Science Journal **9**: 181-212.

Loon, A. v., A. Botterweck, et al. (1998). "Intake of nitrate and nitrite and the risk of gastric cancer: a prospective cohort study." British Journal of Cancer **78**(1): 129-135.

Maantay, J. (2002). "Mapping environmental injustices: Pitfalls and potential of geographic information yystems in assessing environmental health and equity." Environmental Health Perspectives **110**(2): 161-171.

Maheswaran, R., D. P. Strachan, et al. (2002). "A population-based case-control study for examining early life influences on geographical variation in adult mortality in England and Wales using stomach cancer and stroke as examples." International Journal of Epidemiology **31**: 375-382.

MapInfo (2007). MapInfo professional product brochure, [Online] Available from: <http://resource.mapinfo.com/static/files/document/1074660264000/mipro_brochure.pdf> [Accessed 13 May 2007]

Microsoft (2007). Microsoft access product overview, [Online] Available from: <http://office.microsoft.com/en-us/access/HA101656301033.aspx> [Accessed 13 May 2007]

Microsoft (2007). Microsoft excel product overview, [Online] Available from: <http://office.microsoft.com/en-us/excel/HA101656321033.aspx> [Accessed 13 May 2007]

Mills, P. K. (1998). "Correlation analysis of pesticide use data and cancer incidence rates in California counties." Archives of Environmental Health(November).

Miranda, M. L., D. C. Dolinoy, et al. (2002). "Mapping for prevention: GIS models for directing childhood lead poisoning prevention programs." Environmental Health Perspectives **110**(9).

Mizoue, T. (2004). "Ecological study of solar radiation and cancer mortality in Japan." Health Physics **87**(5): 532-538.

Monmonier, M. (1996). How to lie with maps. Chicago, University of Chicago Press.

Moore, D. A. and T. E. Carpenter (1999). "Spatial analytical methods and geographic information systems: Use in health research and epidemiology." Epidemiological Reviews **21**(2).

Morales-Suarez-Varela, M., A. Llopis-Gonzalez, et al. (1995). "Impact of nitrates in drinking water on cancer mortality in Valencia, Spain." European Journal of Epidemiology **11**(1): 15-21.

Murray, C. J., M. Ezzati, et al. (2003). "Comparative quantification of health risks: Conceptual framework and methodological issues." Population Health Metrics **1**.

NCI and CDC (2006). State Cancer Profiles, [Online] Available from <http://statecancerprofiles.cancer.gov/micromaps/> [Accessed 01 February 2007].

NCRI (2005). Cancer atlas of the United Kingdom and Ireland 1991–2000, [Online] Available from <http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14059> [Accessed 26 February 2007].

NCRI (2007). National Cancer Registry, [Online] Available from: <http://www.ncri.ie/ncri/index.shtml> [Accessed 13 June 2007]

Nerriere, E., D. Zmirou-Navier, et al. (2005). "Lung cancer risk assessment in relation with personal exposure to airbourne particles in four French metropolitan areas." Journal of Occupational & Environmental Medicine **47**(12): 1211-1217.

NPPB (2005). An examination of the issues in relation to the introduction of a postcode in Ireland. Report to Mr Noel Dempsey T.D, The Minister for Communications, Marine and Natural Resources., [Online] Available from <http://www.dcmnr.gov.ie/NR/rdonlyres/876D71A5-8115-469A-B2D0-4818F446253F/0/PostcodeReport.doc> [Accessed 03 March 2007].

Nuckols, J. R., M. H. Ward, et al. (2004). "Using geographic information systems for exposure assessment in environmental epidemiology studies." Environmental Health Perspectives **112**(9): 1007-1015.

Ozdenerol, E., B. L. Williams, et al. (2005). "Comparison of spatial scan statistic and spatial filtering in estimating low birth weight clusters." International Journal of Health Geographics **4**(19).

Phillips, R. L., E. L. Kinman, et al. (2000). "Using geographical information systems to understand health care access." Archives of Family Medicine **9**: 971-978.

Pickle, L. W., L. A. Waller, et al. (2005). "Current practices in cancer spatial data analysis: a call for guidance." International Journal of Health Geographics **4**(3).

Post, A. and OSi (2006). GeoDirectory, [Online] Available from: <http://www.geodirectory.ie/> [Accessed 03 March 2007]

Puolstrup, A. and H. Hansen (2004). "Use of GIS and exposure modelling as tools in a study of cancer incidence in a population exposed to airborne dioxin." Environmental Health Perspectives **112**(9): 1032-1036.

Reader, S. (2001). "Detecting and analyzing clusters of low-birth weight incidence using exploratory spatial data analysis." GeoJournal **53**(2): 149-159.

Richards, T. B., C. M. Croner, et al. (1999). "Geographic information systems and public health: Mapping the future." Public Health Reports **114**: 359-373.

Roos, A. D., M. Ward, et al. (2003). "Nitrate in public water supplies and the risk of colon and rectum cancers." Epidemiology **14**(6): 640-649.

Rushton, G. (1998). Improving the geographic basis of health surveillance using GIS. GIS and Health. I. Masser and F. Salgé. London, Taylor and Francis Limited**:** 63-80.

Rushton, G. (2003). "Public health, GIS, and spatial analytic tools." Annual Review of Public Health **24**: 43-56.

SAHRU (1996). The Small Area Health Research Unit, [Online] Available from <http://www.sahru.tcd.ie/aboutus/smallarea.php> [Accessed 26 November 2006].

Sandor, J., I. Kiss, et al. (2001). "Association between gastric cancer mortality and nitrate content of drinking water: ecological study on small area inequalities." European Journal of Epidemiology **17**(5): 443-447.

Sankoh, O. A., O. Berke, et al. (2002). "Bayesian and GIS mapping of childhood mortality in rural Burkina Faso." Control of Tropical Infectious Diseases **Discussion Paper 03**.

Schuessler, A. A. (1999). "Ecological inference." Proceedings of the National Academy of Sciences **96**: 10578-10581.

Snow, J. (1855). On the mode of communication of cholera. Second edition, much enlarged. London, John Churchill.

Staessen, J., H. Roels, et al. (1999). "Environmental exposure to cadmium, forearm bone density, and risk of fractures: prospective population study. Public Health and Environmental Exposure to Cadmium (PheeCad) Study Group." Lancet **353**(9159): 1140-1144.

Sui, D. Z. (2004). "Tobler's first law of geography: A big idea for a small world?" Annals of the Association of American Geographers **94**(2): 269–277.

Talbot, T. O., M. Kulldorff, et al. (2000). "Evaluation of spatial filters to create smoothed maps of health data." Statistics in Medicine **19**: 2399-2408.

Teljeur, C., J. Barry, et al. (2006). "The potential impact on travel times of closure and redistribution of A&E units in Ireland." Irish Medical Journal **99**(7).

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

Vine, M. F., D. Degnan, et al. (1997). "Geographic information systems: Thier use in environmental epidemiologic research." Environmental Health Perspectives **105**(6).

Volkmer, B., B. Ernst, et al. (2006). "Influence of nitrate levels in drinking water on urological malignancies: a community-based cohort study." BJU International **95**(7): 972-976.

Wakefield, J. and G. Shaddick (2005). "Health-exposure modelling and the ecological fallacy." Biostatistics **1**(1): 1-19.

Walford, N. (1995). Geographical data analysis. Chichester, John Wiley & Sons Ltd.

Walsh, P. M., H. Comber, et al. (2001). All-Ireland cancer statistics (1994-1996): A joint report on incidence and mortality for the island of Ireland, [Online] Available from <http://www.ncri.ie/pubs/pubfiles/allirelandjointreport.pdf> [Accessed 20 January 2007].

Wang, F. (2004). "Spatial clusters of cancers in Illinois 1986-2000." Journal of Medical Systems **28**(3): 237-256.

Ward, M., K. Cantor, et al. (2003). "Nitrate in public water supplies and risk of bladder cancer." Epidemiology **14**(2): 183-190.

Ward, M., J. Cerhan, et al. (2006). "Risk of non-Hodgkin lymphoma and nitrate and nitrite from drinking water and diet." Epidemiology **17**(4).

Ward, M., T. deKok, et al. (2006). "Workgroup report: Drinking-water nitrate and health, recent findings and research needs." Environmental Health Perspectives **113**(11): 1607-1614.

Webster, J. and R. T. Watson (2002). "Analyzing the past to prepare for the future: Writing a literature review." MIS Quarterly **26**(2): xiii-xxiii.

Webster, T., V. Vieira, et al. (2006). "Method for mapping population-based case-control studies: an application using generalized additive models." International Journal of Health Geographics **5**(26).

WHO (1978). Nitrates, nitrites and N-nitroso compounds. Environmental health criteria Document No. 5, [Online] Available from <http://www.inchem.org/documents/ehc/ehc/ehc005.htm> [Accessed 25 January 2007].

WHO (2006). Guidelines for drinking-water quality, third edition, incorporating first addendum. Chapter 8: Chemical Aspects, [Online] Available from <http://www.who.int/water_sanitation_health/dwq/gdwq0506_8.pdf> [Accessed 05 February 2007].

Yang, C., M. Cheng, et al. (1998). "Calcium, magnesium, and nitrate in drinking water and gastric cancer mortality." Japanese Journal of Cancer Research **89**(2): 124-130.

Yiannakoulias, N., B. Rowe, et al. (2003). "Zones of prevention: the geography of fall injuries in the elderly." Social Science & Medicine **57**(11): 2065-73.

Zeegers, M., R. Selen, et al. (2006). "Nitrate intake does not influence bladder cancer risk: the Netherlands cohort study." Environmental Health Perspectives **114**(10): 1527-1531.