THE UNIVERSITY OF HULL

A CASE STUDY OF SURVIVAL AND PRESENTATION OF GASTROESOPHAGEAL CANCER IN LOCAL NEIGHBOURHOODS.

Being a Thesis submitted for the Degree of

Doctor of Philosophy

In the University of Hull

by

Amanda Jayne Lee.

Msc (Curtin, WA) BSc (Hons) (Leeds, UK) PGDip (Northumbria, UK)

(Submitted October 2017)

Abstract

This thesis presents a quantitative case study on incidence, survival and presentation of patients diagnosed with gastroesophageal cancer to evaluate whether where people live affects how they present and survive with a gastroesophageal cancer diagnosis. The focus research evolved from studies on gastroesophageal cancer's 'geographic affiliation' and a desire to review whether patient and population attributes could be harnessed to reveal potential 'hotspots' to inform targeted health intervention strategies. As the most crucial stage for intervention was associated with patients detecting symptoms early enough for intervention, the focus of this case study was narrowed to survival and presentation.

This research analysed data from 2785 patients who presented to a regional referral specialist cancer treatment centre between the years 2000 and 2013. Cohort analysis revealed common attributes and survival, and data were merged with demographic information in a geographic information system to present findings in mapped format.

Descriptive analysis revealed an association between later stage presentation and reduced survival outcome. Emergency presentations tended to have worse outcomes. Survival deteriorated with advancing age. Gastroesophageal cancer diagnoses in the under 54 age group was more common in lower socioeconomic groups and survival outcomes were marginally lower than in those patients from the least deprived areas. Spatial analysis revealed variation in incidence, presentation and survival across the region. Though this case study revealed several new findings on gastroesophageal cancer presentation and survival, there remains no single solution to informing and encouraging earlier diagnosis interventions. Though presenting data at finer scales of resolution is more clinically relevant, it threatens patient confidentiality.

Declaration

THE UNIVERSITY OF HULL

Submission of Research Thesis to the Graduate School

Proof of Turnitin Submission Form

Note: This form must be returned to the Graduate School when you submit your thesis.

Name: Amanda Jayne Lee

Department: Faculty of Health Sciences Degree: Doctor of Philosophy

Full Title: A CASE STUDY OF SURVIVAL AND PRESENTATION OF GASTROESOPHAGEAL CANCER IN LOCAL NEIGHBOURHOODS.

I accept that both printed and electronic copies of the thesis have been submitted in accordance with University regulations as laid down by the Research Degrees Committee.

Signature: _____ Date: _____

Turnitin Declaration:

I confirm that an electronic copy of my thesis was submitted to Turnitin on 171017

and I have viewed the Originality Report.

My Turnitin number is

Signature:

Date;

Name: AMANDA LEE

Supervisor Authorisation

I confirm that I have viewed the Turnitin report, and hereby authorise submission of the thesis.

Signature:_____

Date_____

Name_PROFESSOR ROGER WATSON

Acknowledgements

To my life partner – my love, my rock – Mark Hayter – without your support through the exceptionally difficult times during thesis development, I doubt I would have ever completed!

To my three fabulous children for their never ending words of encouragement and praise. Alexander, Alice and Olivia, I love you more than words can say. You are my inspiration, my 'reason to be'.

To my late dad Norman Lee, and my eternal cheerleader – my Mum - Ann Lee – you both helped me make this happen. This is all because of the encouragement, support, mentorship, praise and fabulous parenting you gave!

To my mentor / coach / academic writing guru and manager of my emotional breakdowns during the development of this thesis - Professor Roger Watson, you encouraged me to write, taught me how to improve and made me realise I was capable. I cannot thank you enough.

To Dr Sam Khulusi – Clinical Gastroenterology consultant, this thesis is here because of your initial interest, your engagement with the research idea and your continual support during data collection and writing up. You are an inspiration.

To Dr Graham Ferrier – Reader – Geography guru and third supervisor. I loved being your student and hope I brought a new perspective to class. Your help with GIS, despite the interdisciplinary nature of our conversations, was invaluable.

Finally, to Leeds Institute for Data Analysis, Drs Paul Norman, Michelle Morris and Pia Wohland – without whom I would have had no data. Your support and encouragement was exceptional.

Contents

Index of figures

Index of Tables

Glossary of Terms

AREA

In this thesis, area is applied as a geographical term to describe the regional referral centre. 'Area' is a multidimensional concept, it describes the width, height and parameter of a given space or place. In this context, it defines the specific parameters of the catchment area which are then broken down into smaller 'output areas' to capture required data in a map (Figure 11 – scales used within this thesis).

BOUNDARIES – GEOGRAPHICAL/SPATIAL AND NEIGHBOURHOOD

A boundary refers to the delineation of spaces, the barrier between areas. Boundaries can separate or encapsulate a set of social, political, administrative areas to assist analysis. They may be artificially formed in a geographical information system to encapsulate homogenous characteristics of populations. In this thesis, the overall geographical boundary is set through the NHS catchment area for the regional referral centre.

BOUNDING

To bound a case for case study analysis means to clearly define 'the case' and to conceptualise and identify its boundaries. These boundaries may be conceptual, geographical, spatiotemporal, and societal or population derived. The case study protocol (Chapter 6) details this process of bounding the case.

DELAY INTERVAL

This is taken from the total interval of time from onset of symptoms, to eventual diagnosis, based on the Aarhus statement (2012). Patient, doctor, system and treatment

level intervals are defined through previous research and integrated into this literature review for analysis of evidence.

DIAGNOSIS (GOC)

Diagnosis refers to the histological or surgical confirmation of gastroesophageal cancer. Diagnosis is further quantified in the results chapter:

EARLY DIAGNOSIS

Early diagnosis in cancer means to detect the disease when at an early stage of progression. It is used here as a descriptive term to identify patients who present early enough in their disease to elicit successful treatment. Early 'stage' diagnosis is generally accepted as a definition which is evidenced through 'TNM' staging (see TNM staging).

LATE DIAGNOSIS

This is where time taken from onset of symptoms, to initiation of treatment is sub optimal and which causes adverse outcome. Similarly with early 'stage' diagnosis, late 'stage' diagnosis is quantified through TNM (where data are available to substantiate this preposition).

ENUMERATION/ENUMERATE

In geographical terms, to enumerate means to extend or expand the area under investigation. To amend the scale of resolution so that data may be reviewed and viewed across (and within) geographical boundaries and landscape.

EUCLIDIAN

Relating to the type of geometry used in this research (after the Greek Mathematician Euclides). The Euclidian distance refers to the straight line distance between two points on a map.

GASTROESOPHAGEAL CANCER

This is defined using the World Health Organisation (WHO) International Classification of Disease (ICD) definitions. The classification is applied to detect all diagnosed gastroesophageal junctional and upper oesophageal cancers within the existing database (WHO classification C15.0 to C16.0). Table 1 in this thesis identifies these in further detail.

INTERVAL

This is the apportioning of timescales between two periods of time in relation to cancer data analysis. The Aarhus statement (2012) is used for the purposes of this thesis to quantify intervals in cancer presentation.

LANDSCAPE

There are two interpretations of the term landscape.

Landscape may be 'characteristically' defined. This approach captures the self-image of inhabitants. It identifies the 'sense of place' that differentiates one region from other regions. It is the dynamic backdrop to people's lives, the demographics, the social structures and the physical environments. (For example, the demographic attributes of inhabitants, socioeconomic groups of individuals, the lifestyles, the cultures, behaviours and aggregated 'types' of inhabitants).

Landscape may also be defined as the visible land features. This more quantitative definition encapsulates the physical landforms, land use, buildings and structures and environmental conditions which have been attributed to the area.

NEIGHBOURHOOD

Defined spatially, neighborhood refers to a specific geographic area. It also refers to a set of social networks. It captures spatial units where social interactions occur, identifying individual settings and situations where residents share common values. Neighborhood differs from landscape in the personal element, as it captures the face to face interactions at local levels.

PICo

The PICo model is a way to develop and define the research question – It separates the question into analysis of **P**opulation, **P**henomenon of interest and **Co**ntext of the enquiry, providing a framework for the development and structuring of a research question. It is widely accepted and applied across a range of peer-reviewed journals and forms a part of the literature review process. PICo is applied to support the systematic literature review presented in chapter 3.

PRESENTATION

Presentation is the constellation of signs and symptoms leading to a specific diagnosis.

PRESENTATION STAGE

This relates to the stage in the cancer journey whereby the physiologic changes are sufficient to produce signs and symptoms.

ADVANCED PRESENTATION

The term advanced presentation generally relates to cancers presenting at higher TNM stage, often this is with metastases. In this thesis, the term advanced relates to any cancer presenting too late in the disease process for curative options.

EMERGENCY PRESENTATION

Where a cancer is diagnosed on presentation to an emergency care setting, due to advanced symptoms.

EARLY STAGE PRESENTATION (see early stage diagnosis)

This relates to optimal time within which presentation to healthcare services occurs, to elicit favourable outcome. Governmental strategies aim to target factors which encourage early diagnosis and early presentation, these terms being the foundations for national drives such as the National Awareness and Early Diagnosis Initiative (NAEDI)

LATE STAGE PRESENTATION (see late stage diagnosis)

A previously adopted UK term which has been hailed as outdated because of its role in potentially stigmatising patients (See late STAGE diagnosis).

PRISMA

The Prisma flow diagram asserts the standardisation of retrieval and search processes undertaken for a systematic literature review. The acronym stands for Preferred Reporting Items for Systematic Reviews and Meta-Analysis – a widely accepted tool applied to standardise information retrieval and assert validity in the literature

searching component of a review. PRISMA supported the systematic literature review presented in chapter 3.

SCALE

Scale refers to the space where ecological processes occur. Scale may be enumerated to a larger or smaller 'space' to allow spatial interpretation of data. For this thesis, scale will be adapted to enable confidentiality, to fully encompass homogeneity of population profiles and to critically engage in analysis of patient level data related to place of residence.

Patterns can result from a variety of processes, and can occur at more than one spatial scale. When mapping data in a Geographical Information System, it is important to apply appropriate scales of enumeration so that data can be weighted against population demographics, lifestyle factors and underlying attributes of the landscape. This research uses a variety of scales set through the Office of National Statistics called 'output areas'. Incidence data are 'pinpointed' on the map via appropriately georeferenced patient postcodes.

SOCIAL ECOLOGY

This captures both the environment, and how it is shaped through structures and relationships with those who reside and have power over the environment. It is a sociological theory which links nature, society and evolution. It is a term used in this thesis to encapsulate the environment, how people create and adapt to its biophysical and structural systems and how culture, society and behaviours can shape that environment. In describing a social ecology, this thesis identifies the institutional organisations, the populations, the power relations, the economy and the moral and social structures linked with the GOC journey.

SPACE

Space is a conceptual term. Physical space is a three-dimensional area; a location; a space which may be defined through the apportioning of representational boundaries. For example, characteristic definitions of space may be drawn through allocation of ecological, social, political, behavioural or environmental boundaries.

Ecological space refers to the society which inhabit the three-dimensional space.

SPATIAL DATA

Spatial data are required to inform a geographical information system. Many forms of data can be layered into a geographical information system through fixed geographical points to further describe the geographical area under study.

In this thesis, the geographical information system applies the World Geodetic System (1984) as the coordinate system. This system has been used to link data files from CASWEB, ONS and Ordinance survey with patient data within the geographical catchment area.

VECTOR DATA comprises points, lines and polygons. For data to be geographically referenced, grids need to identify fixed points.

MULTI SCALE Subject data in this thesis, were apportioned via Eastings and Northings in the GIS. They were layered onto the GIS using the postcode's centroid (or middle) point. Line data represent linear factors, such as roads, rivers and streets. Polygons represent data bounded to an area.

RASTER DATA comprise of a squared grid which represents a phenomenon. These data are useful in identifying groups of continuous spatial data, weighted appropriately to represent a pixelated view (or tessellation) of the phenomenon under

investigation. Using fixed points, raster data can be used to weight data to a tessellated scale. In this thesis, for example, to represent alcohol and smoking data, areas were 'weighted' through geodemographic analysis, to provide representative weights of attributes in given areas. These values were apportioned and layered onto the GIS in grid format.

SPATIAL FACTORS

In this thesis, spatial factors relate to the identification of patterns in the GOC data and how they relate to ecological phenomena. This will take the form of analysing many factors, including;

Detection of potential clustering in certain populations; population profiles and whether they are associated with areas of higher density GOC; evaluating distances from patients' homes, to specialist healthcare providers; evaluating incidence associated with underlying demography – rurality, socioeconomic groups and whether clusters occur in areas of higher industry.

SPSS

This is a software package used for data mining, analytics, management, review and statistical analysis of data. The **S**tatistical **P**ackage for **S**ocial **S**ciences is useful to record all data manipulation which has been required for analysis. Syntax is recorded for this manipulation in the Appendices of this thesis.

TEMPORAL

This refers to time, and specifically, the allocated time between periods during the data collection.

TNM

The internationally accepted diagnostic classification system ‘TNM’ measures levels of **T**umour growth, lymph **N**ode involvement and presence of **M**etastases and provides a ‘stage’ at diagnosis. The TNM classification is updated periodically by international consultation

Chapter 1 Introduction and overview of the thesis

This interdisciplinary case study uses geographical and statistical information systems to explore spatiality, survival and presentation in gastroesophageal cancer.

1.1 Why is the research required?

Gastroesophageal cancer (GOC) ranks as the 5th most common cancer in the UK (Healthcare Quality Improvement Partnership, (HQIP) 2016, Cancer Research UK, (CRUK) 2015). It is the 6th most common cause of cancer mortality (Parkin, 2010, Buas & Vaughan, 2013, Lang & Konda, 2013, ONS, 2014). This cancer has a very high mortality rate and low predicted survival rate following patient diagnosis (Adair et al., 2011, CRUK, 2016, Coupland et al., 2012, NHS, 2008, Nuting et al., 2008, Office for National Statistics, 2005, Orengo et al., 2006, Sloggett et al., 2007). Evidence suggests that this is because most patients present too late for curative options (NCIN, 2012). It is generally accepted that the later the stage a patient presents, then the lower the chances of survival become (Coupland et al., 2012, HQIP 2016). This is reflected in gastroesophageal cancer, where the only curative option is surgical intervention. The pre – diagnostic pathway to this treatment modality is crucial. If patients can present early enough, then surgical intervention is less complex.

There have been many UK calls for further research into presentation, early diagnosis and survival by governmental agencies through the National awareness and early diagnosis initiative (NAEDI, 2016) and Cancer research UK (CRUK, 2016). In response to these calls, this research explores the potential of geodemographic profiling as a way to inform earlier diagnostic cancer interventions at more targeted, and local levels. A case study of gastroesophageal cancer in a local community is presented within this thesis, to explore patient presentation, survival outcomes, and

whether geography may be a relevant tool to predict areas of higher incidences of gastroesophageal cancer. It presents an historical review of data spanning 2000-2013, to reveal areas with increased density in incidence or advanced stage presentations and introduces a GOC specific demographic profile of a regional referral catchment area, to explore whether where patients live has any impact on how they present and subsequently, survive their cancer.

Several UK studies identify an increasing incidence of GOC and indicate that this will have a major impact on health service expenditure in the future (Gatenby et al., 2011, Coupland et al., 2012, Tapp et al., 2013). When compared with Europe, the UK has the highest age-standardized incidence of gastroesophageal cancer and has worse survival outcomes (World Health Organization, 2011). Advanced stage presentation is of major detriment to survival outcomes, as surgical intervention (with or without neoadjuvant therapies) remains the only curative option (CRUK, 2015, Sun et al, 2014, Mariette et al, 2007). The most crucial time for encouraging earlier diagnosis is at patient interval. If health services can identify where disease burdens are greatest and reveal geographical areas of higher incidences in advanced stage presentations, then they can target populations for intervention. Research which studies where patients live and how they present with their cancer, can inform cancer prevention programs and improve service delivery.

For many years, the study of consumer attributes, behaviours and attitudes towards purchasing goods, has provided consumer led services such as retail outlets and food stores with intelligence to indicate most appropriate provision of services. Towards that end, large datasets on consumer information and demographics have been collated and applied to predict the needs of the populations served. Nowadays, a range of computer systems exist which facilitate triangulation of disparate information. These

systems can gather and present data in easily interpretable mapped formats, so service providers are able to make informed decisions relating to ‘who needs what?’ and ‘where do they live?’ This offers an intelligence for many agencies to target individuals specifically.

This thesis presents the case for applying these methods to health-related information. Studying the geography of diseases has, for many years, been the guide of epidemiology and this has led to several breakthroughs in improving health. From John Snow’s cholera epidemic, to today’s regional analysis of cancer profiles across countries and continents. However, to date, studies have been limited to mapping disease clusters at very large scales of enumeration.

There is an opportunity to harness these systems at a more local level. Using pre-existing data from general population surveys and census now has a potential to predict areas with increased needs. This research evaluates common attributes associated with gastroesophageal cancer, against underlying population profiles. It applies geographical information system science to reveal clusters of later or earlier staged presentations to evaluate whether these clusters are explained by underlying population demographics. By merging epidemiology, geodemographics and survival analytics, the case study presents a retrospective analysis of geographical, population and patient data from a regional referral cancer centre, to explore whether this form of analysis can offer intelligence to inform future service provision.

The UK has a rich data source of demographic information, taken from population census and profiling. There are databases on smoking and alcohol propensity and where services are currently situated. Harnessing all these data sources means that an analysis of the relationship between where people live and how they present with

cancer, can inform future delivery of services. The capabilities of Geographical Information Systems enables the integration of a wide variety of data. GIS can analyse spatial relationships between different features and thus produce mapped representation of information from scenario-based modelling.

The GIS allows layering and collation of information and data from several different sources, to be geographically fixed to points on a map. These layers can consist of geographical, patient and sociodemographic data, and of environmental, population and lifestyle profiles. All these layers build to form a mesh of related information for further analysis. The GIS can then extrapolate individual information into a visual (mapped) format to provide easily interpretable displays of data for those responsible for health care planning and provision. Yet this should be considered carefully. Maps offer a visual representation of data which are very easy to read. Though a GIS has the potential to highlight particular spatial aspects, the power of visualisation can introduce erroneous conclusions.

Gastroesophageal cancer incidence demonstrates a marked geographical pattern. Areas of higher prevalence include Asia, the Middle East and Europe. This geographic pattern may be due to a variety of factors, such as genetics, lifestyle choices, diet, culture, or even the environment. Many studies have evaluated how environment impacts GOC, but findings are presented at large scales of enumeration. Results are thus clouded by the large scale ‘average’ maps of incidence against populations. Despite the many calls for information to be detailed at smaller scales (so that local health services can identify areas of increased need), the issue of patient confidentiality has limited this scope of research.

Linking cancers with an underlying geography is fraught with difficulties. Cancers have a long latency period, meaning studies which attempt to prove links between diagnosis and causation are exceptionally complex. However, not all research aims to prove causation. Most geodemographic research seeks merely to uncover potential patterns, or geographic association. The problem comes when finding scales which can both fulfil a meaningful result, but which can also maintain anonymity. The problems with publicly available data sources or 'cancer atlases', is that they must protect the individual patients. In doing so, these atlases merely present area level comparisons of incidence. This means that the publicly available resources can only present data at large scales. These scales are meaningless to inform locally targeted clinical interventions.

1.2 Overview of chapters in this thesis

This thesis seeks to explore presentation and survival in patients with a diagnosis of GOC and expose potential spatial relationships. It provides a retrospective analysis of patients presenting to a regional referral centre with histologically confirmed GOC. Population and lifestyle data from all persons residing in the catchment area are used to classify areas in line with GOC specific attributes and these are compared with incidence. Data are evaluated in a geographical information system which presents findings in a mapped format for easy interpretation.

The aim of this small area health study is to reveal whether population characteristics could be useful in the prediction of GOC incidence, so that interventions may be targeted to facilitate earlier presentation and improved survival. The research is a response to the governmental targets to improve survival in cancer care through

encouraging earlier presentation. The research focus towards evaluating spatiality evolved from gastroesophageal cancer's geographical affiliation. This affiliation has been demonstrated by The World Health Organisation's (WHO) publications which reveal a 'gastroesophageal cancer belt' across certain parts of the world.

The thesis begins with several chapters committed to the development of theory for a study into spatiality in GOC incidence and survival. Chapter 2 explores the government strategies underpinning cancer survival generally. It presents an analysis of data on incidence and survival in gastroesophageal cancer and introduces the underlying pathophysiology and patient attributes.

Chapter 3 identifies the most crucial timeframe for patient survival with GOC. A systematic literature review supports the focus on the 'patient interval' for further research to encourage earlier presentation. It also reveals issues associated with cancer research and identifies patient related factors linked to survival and presentation.

In chapter 4, the thesis presents geographical epidemiology and social marketing theories as ways to explore whether the worldwide spatial patterning, has a potential to be reflected in more localised areas. The research methods appropriate for exploration of spatiality in gastroesophageal cancer are explored. The rationale for this research is to generate information and gather a health intelligence which may underpin targeting of specific populations who can be encouraged to present earlier in their disease processes. Chapter 4 concludes the 'theory generation' component of the thesis, and leads to the methodology used in this research.

The methodology chapter (5) introduces the embedded case study design as a way of harnessing and triangulating data to answer the research question. The case study uses a cohort of GOC diagnosed patients to explore survival and presentation, evaluated

against neighbourhood characteristics, and then applies this to model the regional referral centre's neighbourhood, against GOC relevant characteristics. Case study research allows the development of specialist approaches to answering a query (Yin, 2014). Accordingly, this research began with a question and built an instrument to answer that query. To determine rigour and reliability throughout the case study, Chapter 6 presents the Case Study Research Protocol in a framework as defined by Yin, one of the seminal authors and instigators of rigorous Case study methodology. This chapter details all study methods, procedures, data capture and management as well as the general rules which will be followed in the process of investigation.

Case study methodology allows this protocol to be re-evaluated at any time during preparation, collation and analysis of data, if the data prove further investigation is required (Yin, 2014). Case studies rely on discretionary judgement, rather than formulaic evaluations and this fulfilled the interdisciplinary approach taken in this thesis. If findings become suggestive that further concepts require analysis, then these can be reviewed against the overall case.

Chapter 7 provides the results of this research and is presented alongside the objectives derived to answer the research question. A cohort demographic analysis is followed by quantification of presentation and analysis of survival. Patient neighbourhoods are mapped across a regional referral centre, to reveal areas with higher or lower density incidence. These factors are then triangulated to support the development of a tool which captures the most appropriate attributes which have previously been associated with GOC, into a tool to review whether these attributes can be mapped to populations to reveal areas for targeted interventions.

This case study analyses total cohort of 2785 patients – the number of patients who were recorded with histologically confirmed diagnosis of GOC and who presented to a specialist unit in the North East of the United Kingdom between the years 2000 and 2013. Quantitative data are analysed to reveal general attributes of the cohort, in relation to presentation and survival. Data are analysed to quantify stage at presentation based on survival analysis in this cohort. Stage at presentation is then presented in a mapped format so the reader can identify areas where advanced or extremely advanced stage presentation is more common. Observed versus expected incidence data are presented in mapped format to reveal areas with higher or lower propensity to this cancer. Aspatial scan statistics (data analysis which lies beyond a geographical context) reveal clustering of incidence to offer a crude overview of incidence against subject's residential address.

After this crude estimation, data are analysed further. Areas where there are more men, or older people, or with more dense populations, should display higher incidences that more sparsely populated areas. To account for this, population level data from the Office of National Statistics are applied so that the results are weighted appropriately for population density and attributable factors (such as gastroesophageal cancer's propensity to the older generations).

Neighbourhood level data are then applied to characterise the geographical area, to describe it in relation to factors previously associated with GOC patients. Lifestyle data from Public Health England are presented in a mapped format to reveal geographical areas which display characteristics previously associated with the diagnosis of this cancer. This case study does not attempt in any way, to prove causation, merely, to explore patterns, evaluate links and develop a body of knowledge of how patients presented and survived with GOC in this catchment area.

The discussion in chapter 8 converges results against findings from existing literature. It challenges the original theoretical prepositions to highlight potential limitations of the research and proposes rival theories to examine other plausible factors which may be linked with spatiality in presentation and survival in GOC. The chapter revisits the research question and identifies the contribution this research makes to knowledge.

Finally, chapter 9 concludes the exploratory study. It presents a summary of the research and recommendations for further analysis.

Chapter 2 The demographics of gastroesophageal cancer incidence and survival.

This chapter offers the background to gastroesophageal cancer (GOC) incidence and survival. It is a component of the 'theory generation' element of the thesis, offering information on current drives to reduce delays and improve outcomes in cancers. The chapter reveals how advanced presentation adversely affects outcomes in GOC and identifies the factors commonly associated with these patients. A brief overview of aetiology, pathophysiology and potential causes of GOC are revealed, as well as common attributes of patients. This information is essential to inform area analysis when mapping incidence against underlying community population data.

2.1 What is Gastroesophageal cancer?

Gastroesophageal cancer, otherwise described as oesophagogastric cancer, relates to malignancy of any part of the gastroesophageal junction (NHS, 2008, Dolan et al., 1999, Chadwick et al., 2013, National Audit Office, 2010). The World Health Organisation (WHO) publish the International Classification of Diseases (ICD), an internationally recognised global health information standard to classify diseases. All cancers of the oesophagus and oesophagogastric junction classify in ICD codes C15 and C16 (WHO, 2016). Table 1 identifies the cancer ICD classification, aetiology and applied terminology.

Table 1 ICD (2010) Cancer classifications, descriptions and causative factors

ICD 2010 classification	Description	Causative factors	Terminology
C15.2, C15.5, C16.0	Usual in 40% upper GI tract GOC diagnoses. Male:Female ratio 2.8:1. GOJunction cancers occur in the lower third of the oesophagus and cardia. Usual morphology is adenocarcinoma.	Unclear – some links with Barretts Oesophagus. Linked with Obesity and lack of fruit and vegetables in the diet.	Gastro-oesophageal junction cancers
C15.0, C15.1, C15.3, C15.4	Usual in around 13% of all upper GI tract cancers. Male:Female ratio 0.8:1. Upper Oesophageal cancers occur in the upper two thirds of the oesophagus. Usual morphology squamous cell carcinoma.	Smoking and alcohol consumption. Poor diet Obesity More common in people living in deprived areas.	Upper oesophagus cancers
C16.1-C16.9	Usual in 42% upper GI Tact GOC diagnoses. Male; Female ratio 1.5:1. Occur in the stomach. Usual morphology is adenocarcinoma. .	Previous diagnosis of Helicobacter pylori infection Smoking, Poor diet, Excessive salt consumption. More common in deprived areas	Stomach cancers excluding cancers of the cardia
C15.9	Recorded as unspecified malignant neoplasm	Unlinked	Not possible to determine site
Source International Agency for Research on Cancer (2010) described by author			

Malignancy means any genetically mutated cells which rapidly metastasize and invade and destroy tissues (Moscow & Cowan, 2011). Early treatment of GOC malignancies is necessary because the mutated cells rapidly progress to adjacent structures through the rich lymphatic drainage system in that gastroesophageal tract (Diederich, 2007). The two most common types of GOC include Squamous Cell Carcinoma (SCC) and Adenocarcinoma (ADC).

SCC is the most common form of GOC (Gatenby et al., 2011). It follows a progressive sequence, from mild to severe dysplasia, and then through to carcinoma *in situ*, which subsequently develops to an invasive carcinoma. During these phases, fungating, ulcerating or infiltrating lesions may occur, resulting in patients presenting with symptoms (Schlansky et al., 2006b, Kuwano et al., 2005). Early detection and treatment at these initial phases would significantly improve outcome, facilitate surgical intervention and lessen the potential for migration of the carcinoma (Rothwell et al., 1997). ADC is less common, however several studies identify the incidence is rising significantly (Orengo et al., 2006, Gatenby et al., 2011, Wolf et al., 2012, Dubecz et al., 2012, Defoe et al., 2011, González Ortiz & Toro, 2009).

2.2 Evidence on incidence of gastroesophageal cancer and survival from the disease.

Cancer Research UK (CRUK) identify the 2014 age standardised incidence rates for GOC in the UK far exceed European rates. They identify 22.3/100000 males and 9.2/100000 females were diagnosed in 2014 (CRUK, 2014). Incidence of this cancer is predicted to rise significantly over the next few years and treatments will have significant financial impact on health services (Gatenby et al., 2011, Coupland et al., 2012, Tapp et al., 2013).

The 5-year survival rate in GOC ranges between 27-33%, depending on literature sources and country of residence. Median survival rates are generally cited around 21.4 months (Thrift et al., 2012, Wolf et al., 2012, Kayani et al., 2011). Survival rates significantly decrease when patients present with advanced tumours, where there is lymph node involvement, metastases or where the tumour has infiltrated to surrounding sites (Kayani et al., 2011, Holscher et al., 2011, Shinagare et al., 2012,

Homs et al., 2012). Thus, earlier detection and presentation can improve outcomes and increase survival (Wolf et al., 2012, NCIN, 2015).

To encourage earlier detection and presentation, referral and treatments, patients and clinicians need to be aware of early warning signs of this cancer. Nowadays, health services must judge whether activities and interventions to improve outcomes are both clinically effective and financially viable (NCIN, 2014). To support this, a more targeted approach to deliver services where they are needed most is both appropriate and timely (CRUK, 2017).

There is a significant amount of evidence revealing the geographic propensity of gastroesophageal cancer. The International Agency for Research on Cancer Statistics (IARC) work with the World Health Organisation (WHO) to produce 'Globocan' – global cancer statistics at worldwide levels (IARC, 2015). These 'Globocan statistics' illustrate a geographical belt across Asia, the Middle East and certain parts of Europe and support the hypothesis of geographical propensity in the disease (Figure 1).

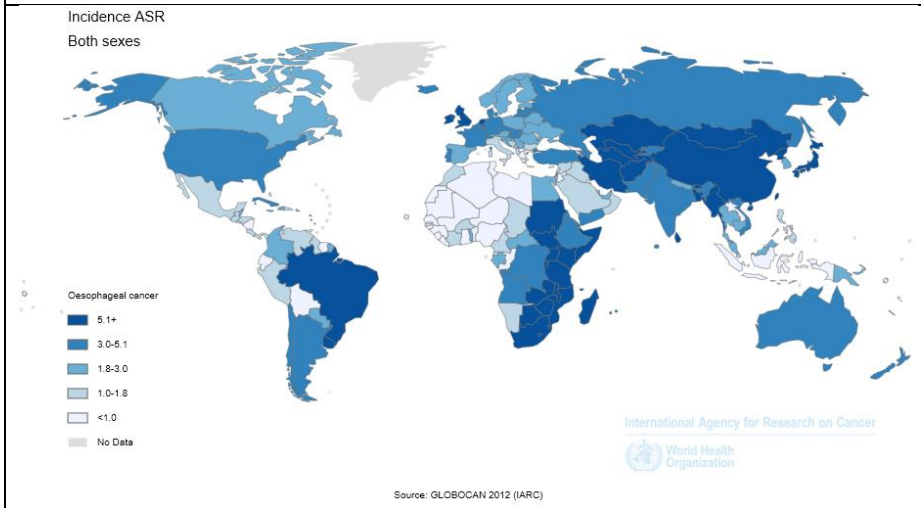
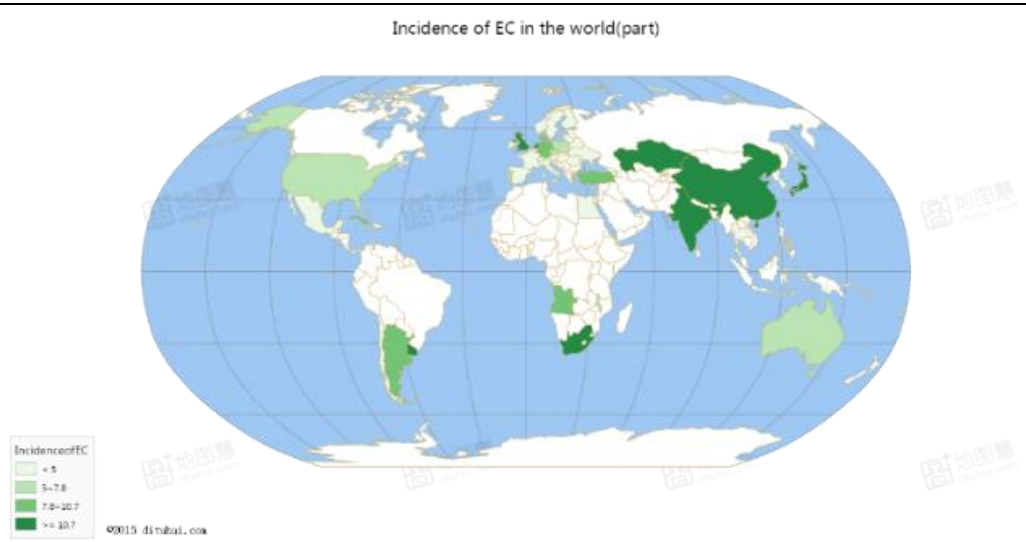
2010 WHO/IARC ‘Globocan’ statistics –incidence of Gastroesophageal Cancer	2007 WHO/IARC ‘Globocan’ statistics –incidence of Gastroesophageal Cancer
	
International Agency for Research on Cancer (IARC) (2015), CancerBase No. 10 [Internet]. Lyon, France: International Agency for Research on Cancer; 2012. Available from: http://globocan.iarc.fr, [accessed November 2015]	

Figure 1 Worldwide cancer statistics

Although rates of any cancer may have strong regional and geographical variations, their causation may be attributed to a variety of factors. Genetics, environmental, cultural or behavioural attributes have all been linked with causation (Sharp et al., 2014b, Tannenbaum et al., 2014, Goli et al., 2013a, Yang et al., 2013, Elebead et al., 2012). As a result, there have been many calls for further research into the geographical nature of GOC (Coupland et al., 2012, National Cancer Intelligence Network (NCIN), 2010).

In the United Kingdom (UK) alone, there are significant regional variations in incidence and survival (Coupland et al., 2012). The NCIN's two publications identify these gross regional variations (Figure 2) (NCIN, 2014, NCIN, 2010). This important information reveals the disparate burdens of GOC diagnosis across UK's cancer networks. However, the extremely large scales offered by the presented data mean that targeted strategies are ill informed. 'Problem areas' need a more clinically relevant scale. From these maps, it is impossible to consider where primary healthcare practitioners should begin to implement strategies to encourage earlier interventions. The smaller the scale of analysis, the more targeted the approach may be. GP surgeries, for example, possess the ability to review their own cancer statistics, but again, they are limited by the large scales of resolution.

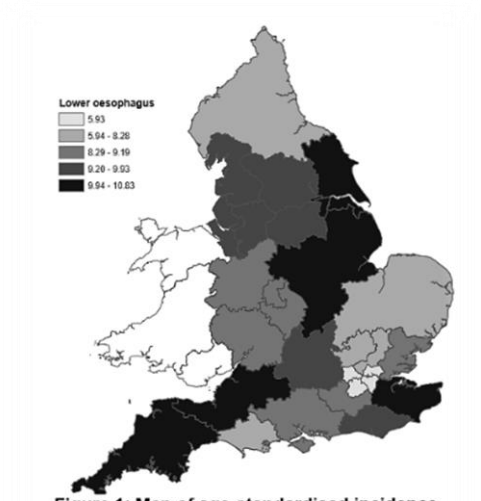
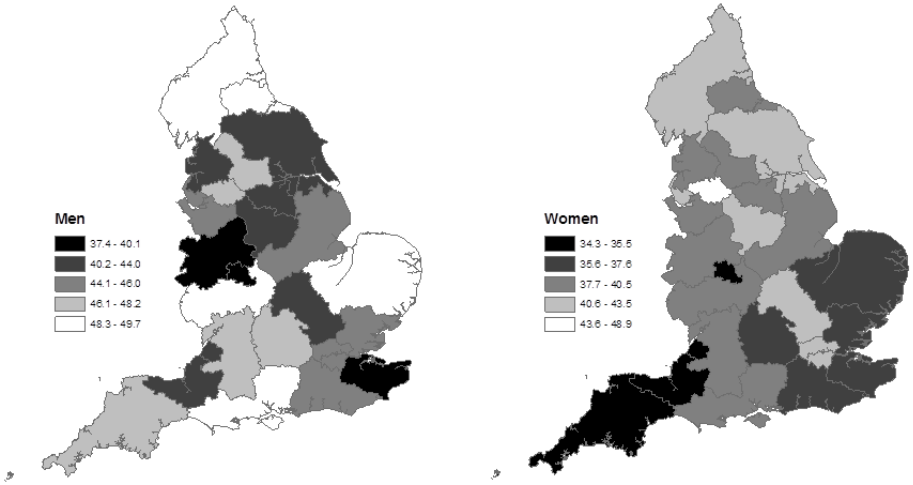
NCIN (2010) Geographic variation in cancer of the lower oesophagus NCIN Data Briefing available at; **http://www.ncin.org.uk** [accessed 062016].	Office for National Statistics (2014) Cancer survival by NHS England Area Team: Adults diagnosed 1997-2012, followed up to 2013 | 16 December 2014 available at **https://www.ons.gov.uk** [accessed 0616].
 **Figure 1: Map of age-standardised incidence rates of lower oesophageal cancer (per 100,000 European Standard population, ASR(E)) by cancer network, males, England, 1998-2006.**	 Contains Ordnance Survey data © Crown copyright and database right 2014 Contains National Statistics data © Crown copyright and database right 2014

Figure 2 UK geographical variation in gastroesophageal cancer (2010 and 2014)

The maps presented in Figure 2 support the requirement to explore the geographical nature of GOC. A smaller scale, more localised analysis may be relevant to inform more targeted interventions.

There are many calls for further research into presentation and survival to encourage earlier diagnosis for cancers across the world (NCIN, 2015, ONS, 2014, Lewis, 2017, Richards, 2009). As there is a geographical propensity of GOC, further investigation of factors relating to its geography at more local levels is relevant.

2.3 National strategies to address early diagnosis in cancer research.

It is essential to any successful treatment outcome for patients diagnosed with cancer that prompt diagnosis, referral to a specialist centre and appropriate interventions are initiated as soon as possible (Elliss-Brookes et al., 2012, Richards, 2009b, Richards, 2009a). Since the 2007 DoH cancer reform strategy was announced, the National Awareness and Early Diagnosis Initiative in England (NAEDI) has developed schemes to synchronise research, gather evidence and instigate any interventions which are associated with earlier diagnosis and interventions (Elliss-Brookes et al., 2012, Richards, 2009b, Richards, 2009a). The purpose of NAEDI was to address the issue of advanced presentation, to encourage earlier diagnosis to improve cancer survival outcomes. However, the recent publication '*Improving Outcomes: A Strategy for Cancer*' (DoH, 2011) shows that England remains below the European averages for cancer survival. This disparity supports a need to further evaluate the UK systems and processes which are associated with patients' journeys to cancer diagnosis. The 2007 cancer reform strategy strongly advocates a multidisciplinary team (MDT) approach as a quality measure for cancer treatment. Regular cancer multidisciplinary

team (MDT) meetings allow a range of professionals to work collaboratively assess clinical decision making, expedite effective treatments and engage in research activities to promote earlier diagnosis and interventions (DoH, 2014).

To study 'early diagnosis' in cancer research, it is important to evaluate community or system mediated factors which may have an impact on the diagnosis (Elliss-Brookes et al., 2012, Abdel-Rahman et al., 2009, Baughan et al., 2009, Macdonald et al., 2006, Kotz et al., 2006, Macleod et al., 2009, Richards, 2009c, Weller et al., 2012). A body of knowledge can be generated which supports and informs targeted intervention strategies. Although there is a wealth of evidence identifying patient, system and community level factors affecting early diagnosis for many cancers, there is a paucity of literature specifically relating to GOC. Furthermore, the diverse timescales appropriated in cancer research do not support meta-analyses of findings. It is difficult to ascertain the extent of a delay in presentation or referral when there is a lack of homogeneity on definitions and timeframes. Many studies rely on patient recall for identifying time to first symptom, others rely on date of histological confirmation, others on date of presentation. This leads to confusion and irregularity.

There is also a diverse nomenclature relating to 'presentation' in cancer studies. Terms in general use, such as 'advanced, delayed, late or later staged' are ill defined and frequently rely on the researcher's perspective.

2.4 The many definitions of presentation, diagnosis and delays in the patient's cancer journey

Literature relating to cancer and diagnosis requires objective interpretation of timelines and the establishment of mutually accepted nomenclature. This section identifies some common cancer terminologies, and underpins how terms such as

‘presentation, diagnosis and survival’ are used in this thesis. This is important, because the research makes several assumptions when defining and quantifying survival in the cohort study. Terminology such as late and advanced stage presentation, frequently found across cancer research literature, are subject to misinterpretation, so clarity is essential.

The term *‘diagnosis’* usually relates to the time at which confirmation of the suspected cancer is undertaken. However, diagnosis can be determined in several ways. In empirical studies, diagnosis may refer to the period when the Primary Care Physician, / General Practitioner, or the system gatekeeper, initially suspected cancer, resulting in discussion with the patient and referral to specialist services. Diagnosis is quantified through the date of histological confirmation. The term ‘diagnosis’ is used interchangeably and is not limited to medical diagnosis alone. Many sources of information construct diagnosis to demarcate the time at which the patient noticed first symptoms, (referring to ‘patient diagnosis’). The term is also confounded by the many diagnoses of cancer which are made post-mortem. Many studies on survival in cancers chose to omit post-mortem diagnoses as they significantly affect survival outcome data.

Presentation relates to a constellation of clinical signs and symptoms which may have been identified by the patient or a clinician, resulting in said patient ‘presenting with symptoms suggestive of cancer’. This leads to subsequent diagnosis and treatment. Presentation s*tage* relates to the stage in the cancer journey whereby physiologic changes are sufficient to produce signs and symptoms. Defining when patients presented can be considered as a variety of ways;

- when the patient first noticed symptoms and sought help.

- at the time of histological confirmation
- when the healthcare provider recognised symptomology and referred to specialist services
- as a presentation in an emergency care situation.

Late diagnosis in the literature is a term generally quantified through 'staging' or assessment of tumour and lymph node involvement. Thus, 'late' – or more specifically – 'late staged' tumours underpin poorer survival outcomes. Yet when considering the array of evidence on cancer diagnosis and survival outcomes, this 'staging' information is often incomplete (Kotz, 2006). In the absence of a complete dataset (where tumour and lymph node involvement is absent), the term 'stage' is often removed, and replaced with an accepted terminology such as 'advanced presentation' which remains an essentially contested concept, but captures the essence of a diagnosis which occurred too late for curative treatment.

The term *advanced presentation* generally relates to locally advanced or metastasised cancers, or those presenting too late for curative intent. Often, the term encompasses patients at higher 'TNM' stages – The TNM stage is a diagnostic classification system specifying and grading the tumour, lymph node involvement and extent of metastasis. 'TNM' definitions classify cancers to a globally recognised standard and consensus opinion. The Union for International Cancer Control (UICC) hold responsibility for maintenance and updating of the TNM definitions. However, many cancer diagnoses do not have available TNM staged data, so this has limitations (McGhan et al., 2012, March et et al., 2011, Nagtegaal et al., 2011, Warneke et al., 2011). In this research, the term 'advanced' will relate to any cancer presentation made too late in the disease process for curative options. *Early presentation* relates to the optimal timescale in which presentation to healthcare services occurs, so that a favourable outcome can

occur. In this thesis, early presentation is quantified through retrospective survival analysis from the cohort of GOC patients, and further depicted as 'treatment facilitated, or survivor' status.

The terminology 'delayed presentation' or 'delayed diagnosis' is widely debated in the literature and in cancer collaborative centres (NCIN, 2015). These terms lead to a potential patient or system mediated stigma. Patients may blame themselves for not 'presenting earlier', feeling blamed for their own survival prospects (Day, 2012). Linking services and health systems with 'delay' is contentious, as the label imparts a culture of blame. Subsequently, there has been a drive to cease using delayed presentation and delayed diagnoses in cancer terminology. Health professionals should, therefore, minimise any potentially stigmatising nomenclature and consider the terms they use in documentation, data collection and publications.

Emergency presentation refers to any patients who presented to an emergency care setting, generally with advanced symptoms of the disease. These diagnoses are useful in that they provide a consensus opinion that emergency presenters are at an advanced stage of the disease. *Emergency diagnosis* is a newer term gaining increasing popularity in the literature. Where patients are diagnosed on emergency presentation, this is usually indicative of an advanced stage cancer presentation. Survival outcomes for emergency presentation are drastically worse that those patients who are diagnosed through referral pathways.

For the purposes of clarity, this research will quantify presentation groups through survival analysis of the cohort. By calculating the number of days between histological confirmation and eventual death – specific groups will be identified. These will underpin a quantitative quartile grouping of subjects based on their survival times.

2.4.1 The routes taken from presentation to survival in gastroesophageal cancer diagnosis.

It is widely accepted that the route from presentation to diagnosis of any cancer is exceptionally important to survival outcome (Weller et al., 2012). There is a significant amount of research supporting this claim DeAngelis et al, 2014, IARC, 2015, Lewis, 2017). Many studies and methodological approaches aim to describe the multifaceted routes patients and healthcare services undergo to assist timely diagnoses. Yet the underlying datasets applied across cancer research studies are confusing. Published research uses many differing measurements of time points, meaning intervals between detection and presenting with a symptom and eventual diagnosis are not standardised in the research literature (Weller et al., 2012). This means that direct comparison of evidence relating to presentation and survival is hampered by disparate definitions and nomenclature.

2.4.2 Elaborating on the ambiguity in current data relating to date of diagnosis

There have been attempts to standardise definitions in cancer research; however, quantifying the actual time between symptom recognition and eventual treatment in the correct specialist service, has methodological limitations. Key points in time used to pinpoint date of first symptom, diagnosis, treatment and referral are lacking in consistency (Elliss-Brookes et al., 2012, Weller et al., 2012, Forbes et al., 2014, Dwivedi et al., 2012). The wide range of confounding variables associated with each point in time has perhaps led to confusion in the evidence on the route to cancer diagnosis (Elliss-Brookes et al., 2012).

Many studies rely on empirical observations, relying on cancer patients' ability to recall the time from the initial onset of their symptoms. This is fraught with the methodological issue of recall bias. Patients may also be unaware of the sociocultural

component of ‘symptom allocation’. The question is raised as to when a normal physiological condition, such as a sore throat, becomes assigned as a symptom which encourages a person to seek medical assistance. Contextual analysis of the events which lead to a correct diagnosis need to account for the many confounding variables which may be associated with delays in treatment.

There is currently no widely accepted theoretical framework in the literature underpinning time and date definitions to study early awareness and diagnoses in cancer research. Walter et al., (2012) provide a theoretical model to describe processes and pathways previously associated with detection, presentation, diagnosis and treatment for cancer. Their diagram identifies the complex nature of this problem. This is simplified in Olessen et al.’s (2009) model, which identifies specific intervals in this process (Figure 3). Both models were developed into cancer specific frameworks through analysis of Anderson et al.’s (1995) ‘patient delay’ model.

Although the frameworks of Walter et al. (2012) and Olessen et al. (2009) describe the journeys, there is still a requirement to evaluate research and how each study uses data to define specific time intervals. Beginning with patient, doctor then system related issues, Olessen’s (2009) model reduces these intervals to primary and secondary care, or to ‘diagnostic’ or ‘treatment’ related components. This is important when considering methodological processes which attempt to highlight underlying causative factors in delays during the initial onset of cancer. To assess, for example, a patient journey to presentation demands a qualitative perspective, contextualising events leading to eventual presentation. System delays on the other hand, lend to more quantitative evaluation, as data are available from referrals to diagnosis.

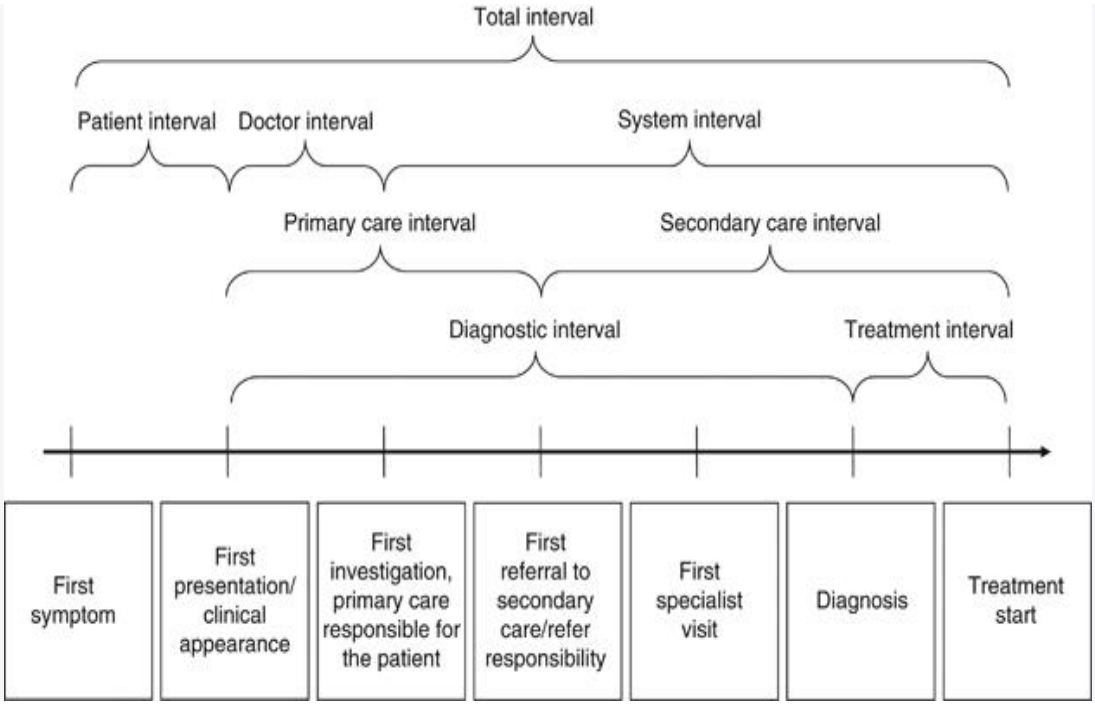


Figure 3 Time from first symptom detection to treatment (Olessen et al. 2009)

Olessen's (2009) model is limited through a very basic description of events. Walter's et al. (2011) model provides detail (Figure 4). It provides a more patient focused and behavioral analysis of events leading to diagnosis. Patient, provider, system and disease factors are listed as important contributing factors. Walter's model uses terms 'appraisal, help seeking, diagnostic and pretreatment' intervals to describe the patient's journey from their first symptom to the start of their treatment.

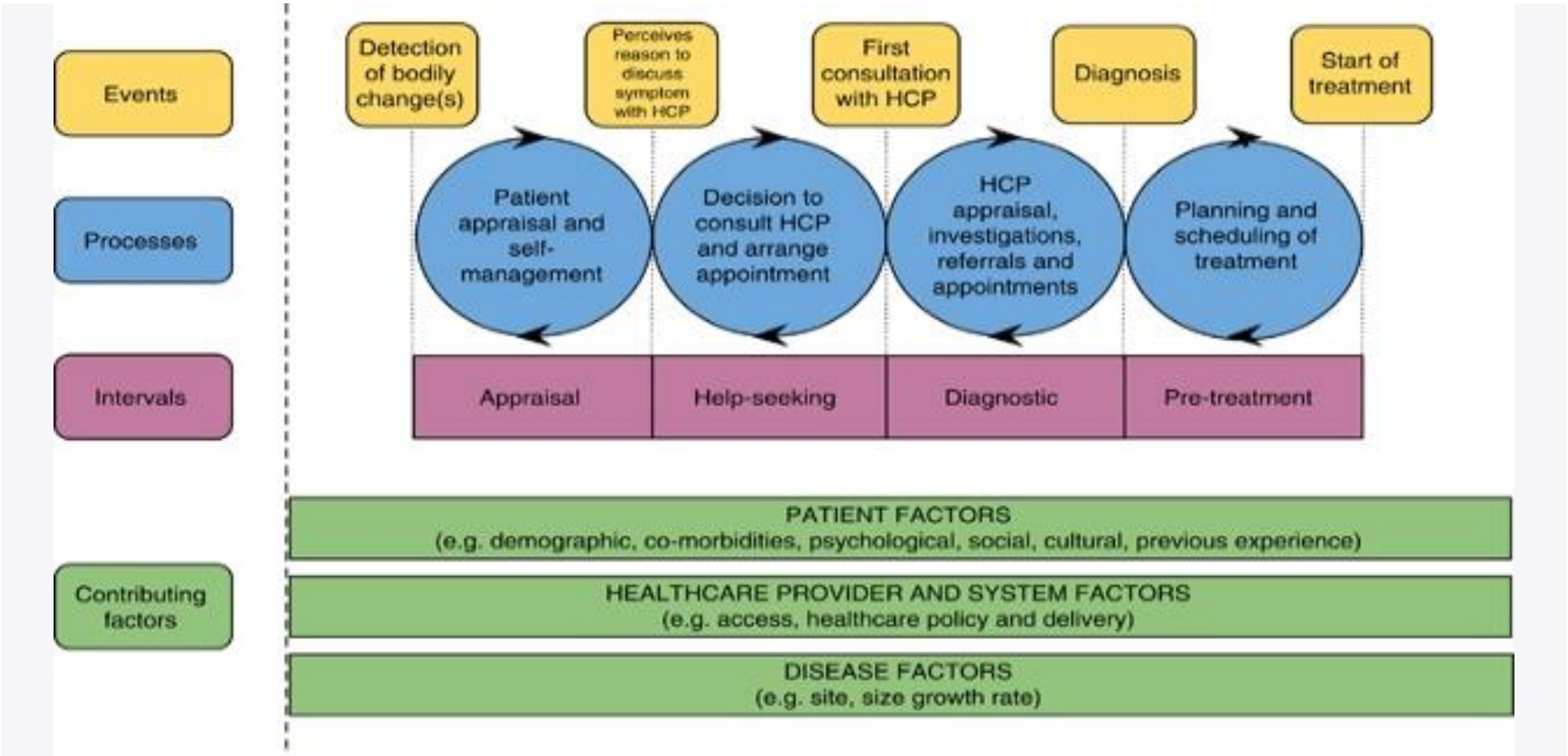


Figure 4 Model of pathways to treatment (Walter et al. 2012).

Both models are useful, however, in providing a theoretical framework to underpin the journey of cancer detection and diagnosis, to eventual treatment. However, they do not standardise the measurement tools used to delineate these intervals. The 'Aarhus' checklist was developed and published in 2011 and this details a standardised tool for definition and recommendations to support research associated with early cancer studies (Table 2).

Table 2 Aarhus key time intervals for early cancer research (Aansen et al, 2011)

Label	**Description**
Date of first symptom	Time when first bodily changes occur or symptoms are noticed (date of first 'alarm' signal identified which caused help seeking – new nomenclature identified and the term 'appraisal interval' has replaced 'patient delay')
Date of first presentation	The time at which signs and symptoms history and risk factors made it possible for clinician to have started referral or investigation (healthcare worker or patient's perspective? Patterns of symptoms leading to event)
Date of referral	The time of transfer to gatekeeper or specialist (cross referrals/complexity of route to diagnosis)
Date of diagnosis	1st histological confirmation, date of hospital admission due to malignancy. Date of initial OPD consultation, a differently established date of diagnosis, date of death (if this occurs before diagnosis)

This checklist also provides guidance into early cancer research. It is a tool to promote transparency and consistency in methodology and measurement. It was developed through a multidisciplinary conference where consensus opinion was formed through systematic review of literature. It clearly sets criteria and measurement parameters to underpin early diagnosis cancer research, presented in a checklist format. For this reason, Aansen's (2011) 'Aarhus' checklist for date of diagnosis at first histological confirmation is applied in this thesis.

To study early diagnosis in cancer, an overview of the full patient journey is necessary. There are many studies relating to the GOC patient journey to diagnosis, but these are hampered by the very high mortality rate. Most cancers commonly reviewed in the literature use follow-up studies which rely on recall of patients. However, the high mortality rates found in GOC patients results in high levels of missing data, as many subjects are lost to follow up through events such as death. The subsequent 'survivor bias' will impact prospective studies of GOC patients. For this reason, many prospective cancer studies present data relating to other cancers. GOC becomes 'lost' as it tends to be grouped with gastrointestinal or head and neck cancers to reach adequate power and sampling to support quantitative analysis.

Yet advanced stage GOC remains a major concern and further study is essential to explore factors associated with encouraging earlier presentation. The literature review presented in chapter 3 will be the first to assess all relevant studies with a focus on GOC. It will uncover factors associated with GOC presentation and reveal the most appropriate interval to encourage earlier diagnosis. The next section presents general information on signs and symptoms, incidence, mortality and survival with GOC.

2.5 Signs and symptoms of Gastroesophageal Cancer

Patients diagnosed with GOC often present to services with dyspepsia, dysphagia, nausea and vomiting and these symptoms are attributable to the fact that the tumour is occluding the gastroesophageal tract. Many studies identify up to 50-60% occlusion is required to elicit dysphagic symptoms and this suggests that the tumour may be at a more advanced stage (Rothwell et al., 1997, Broker et al., 2009, Parsons, 2010, Fransen et al., 2004, Ojala et al., 1982). Alarmingly, dysphagia is exposed as a symptom which is only experienced at 70% occlusion of the oesophagus and infiltration beyond the oesophageal wall is required to cause the physical symptoms of dysphagia. (Di Pietro, 2013, Jayasekera, 2012, Lambert, 2012, Yang, 2012). Many patients present with drastic weight loss, or features suggestive of gastrointestinal bleeding (Wolf et al., 2012). This drastic weight loss (cachexia) is a multifactorial syndrome, usually signifying an advanced stage tumour (Dhanapal et al., 2011). Tumour infiltration to surrounding structures causes a propensity to gastrointestinal bleeding and complications such as anaemia will impair recovery. All these factors suggest that early detection and rapid referrals are necessary so that interventions can be expedited to improve outcomes.

2.6 Survival, treatment, screening and surveillance in gastroesophageal cancer.

The median survival time for GOC diagnosed patients remains at less than 2 years (Thrift et al., 2012, Wolf et al., 2012, Kayani et al., 2011). Survival significantly decreases with larger tumours at presentation, with lymph node involvement, metastases or when the tumour has infiltrated surrounding sites (Kayani et al., 2011, Holscher et al., 2011, Shinagare et al., 2012, Homs et al., 2012). In the UK, one-year survival rates range between 14.8-40.8%, with 5-year survival a mere 3.7-15.6 % (Coupland et al., 2012, Dubecz et al., 2012). One-year survival rates in China are 20%

(Zhang et al., 2008), and 15.4% in the United States (Wang & Wongkeesong, 2005). A Eurocare database analysis (66 registries covering 24 EU countries) of 51149 GOC subjects identified one-year survival was only 33.4%, dropping to 9.8% by 5-years (Gavin et al., 2012).

The only curative option for GOC is surgical removal of the tumour, with or without neoadjuvant therapy. Surgery (with curative intent) increases the 5-year survival rate up to 10-20% (Coleman et al., 2012, Walters et al., 2011, Coleman et al., 1999). In asymptomatic GOC diagnosis, the 5-year survival rates can reach up to 98% (Wang et al., 2002, Tachibana et al., 2006). However, asymptomatic GOC detection is extremely rare. GOC is presently only diagnosed through endoscopy (CRUK, 2015). One is left to question the validity of 'asymptomatic' when a highly invasive procedure is carried out without underlying symptomology, or significant family history.

These data on survival times for GOC are skewed by inconsistencies in presentation and staging data. Patient related factors linked with GOC include advancing age and presence of a range of age related co-morbidities. Many patients may not necessarily die from their GOC, but from other factors which affect their health. For these reasons, survival and prognosis data should be fully scrutinised. Studies into GOC need to acknowledge the range of confounding variables which may be associated with this cancer.

A Cochrane review revealed there are no effective strategies to screen populations fully for presence of GOC (Yang et al., 2012). Balloon or sponge yield cytology through endoscopic examination; endoscopy with iodine or methylene blue staining; radiology and faecal occult blood sampling are currently under evaluation. However, none to date have proven efficacy in predicting GOC. Thus, surveillance of 'at risk'

patient groups (such as those with Barrett's Oesophagus) remains the only current preventative screening measure. (Yang et al., 2012, di Pietro & Fitzgerald, 2013b, Hammoud et al., 2014, Roshandel et al., 2013, Lao-Sirieix & Fitzgerald, 2012, Zhao et al., 2012, Salaspuro, 2011, Tomizawa & Wang, 2009b).

These 'at risk' groups include patients with reflux and dysplasic diseases of the gastrointestinal tract. They cause pre-cancerous cellular changes to the oesophagogastric tract, so surveillance and screening is required (di Pietro & Fitzgerald, 2013a, Choi & Hur, 2012). However, the evidence supporting these remains contentious, as positive yield cancer diagnoses remain relatively low (Choi & Hur, 2012, Booth & Thompson, 2012, Varghese et al., 2012, Caygill et al., 2011, Griffiths, 2011, Muthusamy & Sharma, 2011, Shammas, 2011, Tomizawa & Wang, 2009a, Badreddine & Wang, 2008, Corley, 2008, Brocklehurst, 2013, Di Pietro & Massimiliano, 2013, Lambert, 2012, Yang, 2012, Chang, 2011).

Bhat et al. (2011) used the Northern Ireland Barrett's oesophagus (BE) register alongside the Northern Ireland cancer registry, to analyse incidence of adenocarcinoma of the oesophagus and gastric cardia in 8522 patients with confirmed BE in a retrospective trawl of seven years of data. From the 8522 BE patients, only 79 were diagnosed with oesophageal cancer. They concluded that the risk of malignant progression among patients with BE was not significant. These findings are backed in several studies (Choi & Hur, 2012, Gopal et al., 2004, Mahon, 2009, Allum et al., 2011).

2.7 Attributes of patients with gastroesophageal cancer

There is a male to female ratio of 2:1 in GOC (Buas & Vaughan, 2013, Dubecz et al., 2012, Chen et al., 2013, Otterstatter et al., 2012, Wei et al., 2011, Qiu & Kaneko, 2005,

Carr et al., 2013, Levi et al., 2013). The mean age at diagnosis for both genders is 70 years (SD + 20) (Coupland et al., 2012, Dubecz et al., 2012). Several studies have linked lifestyle factors with GOC diagnosis. Smoking, alcohol and obesity offer a biologically plausible explanation to causation in many cancers. Dietary habits and a lower socioeconomic status have been linked in retrospective studies of GOC. Analysis of PAF – (the percentage of attributable factors) suggests that up to 89% of GOC may be linked with lifestyle factors such as smoking, obesity and alcohol (CRUK, 2015).

Tobacco smoking and alcohol are significant factors in development of GOC and many large scale studies identify the strong association of smoking and alcohol with GOC (Kamangar et al., 2006a, Brown et al., 2008, Wang et al., 2013, Chen et al., 2011, Qiao et al., 2009, Tran et al., 2005). There is a cellular level explanation for the issues related to carcinogenic nature of alcohol and cigarette smoking. Noxious substances are in contact with the lining of the epithelium during consumption, a process significant to GOC. Alcohol and cigarette smoking causes mutations of P53 tumour suppressor genes (Kuwano et al., 2005, Chen et al., 2012, Kato et al., 2001, Hardwick et al., 1997, Brennan et al., 1995, Abedi-Ardekani et al., 2011). This P53 gene is the most frequently mutated gene sequence in all human cancers (Kato et al., 2001, Brennan et al., 1995, Abedi-Ardekani et al., 2011, Leeuwenburgh et al., 2010, Leeuwenburgh et al., 2010, Li et al., 2012, Bektas et al., 2001). Cigarette smoking is strongly associated with any cancers and inhalation of substances during smoking causes up to 25% male and 4% female cancers (Nyren &Adami, 2002). Parkin et al. (2010) estimate up to 66% of male GOC and 63% female GOC cases are attributable to smoking.

Although evidence identifies the link, there is no quantification on the amount of tobacco or alcohol which is required to cause these cellular changes and malignancy (Adair et al., 2011, Coleman et al., 2012, Lubin et al., 2012, Zheng et al., 2010a). Measuring the extent of exposure to alcohol or smoking is also exceptionally difficult. This is confounded through a potential environmental exposure through passive smoking for example. Reliance on 'self-reporting' of intake is also limited through patients need to conform and therefore, to underestimate intake (Courtney et al., 2013, Blakely et al., 2013b, Foster et al., 2013).

There is a wealth of literature supporting GOC's association with obesity (Levi et al., 2013, Tarleton et al., 2014, Hong et al., 2013, Blom et al., 2012, Hoyo et al., 2012, O'Doherty et al., 2011, Smith et al., 2008). There is a biological plausibility to this association, because adipose tissue has a higher propensity to malignant changes (Buas & Vaughan, 2013, Levi et al., 2013, Hong et al., 2013, Robertson et al., 2013). GOC risk is 2.5 times higher with people of excessive abdominal adiposity (Singh et al., 2013). Meta analyses revealed GOC risk as 13% higher with each 5% body mass index score (Smith et al. 2008, Hoyo et al., 2012). A high body mass index is also strongly linked with gastroesophageal reflux disease, which remains closely linked with GOC (Tran et al., 2005, Smith et al., 2008, Kubo et al., 2013, Lagergren et al., 2013).

In addition to obesity, other diet and lifestyle factors have close links with GOC diagnoses. The International agency for research in cancer (IARC) identify the following as 'may pose a risk' to development of GOC:

- Ingestion of hot foods,
- carbonated drinks,
- pickled vegetables

- an Asian drink (Mate)
- Diets rich in red meats, salt or smoked meats, burned foods and high salt intake.

The biological plausibility to the red meats, salts and burned food relates to the damage incurred through oxidative stress and the production of N-nitroso compounds which occurs at the buccal membrane. N-nitroso compounds found in these substances are converted to the carcinogenic 'nitrosamines' which are very damaging at cellular level. (Bartsch et al., 1992, Kamangar et al., 2009, Ward et al., 2012, Wu et al., 2013). Again, the mechanism of exposure is very difficult to prove and this limits the possibility of further study. Nitrosating agents are of interest in GOC studies as those people with poor oral hygiene (and therefore, with higher amounts of endogenous nitrosating agents in the buccal cavity) may also be at higher risk of GOC (Dar et al., 2013, Abnet et al., 2008, Islami et al., 2004, Tran et al., 2005).

Drought and nutritional deficiencies are strongly linked with GOC in several studies (Wang et al., 2013, Jeurnink et al., 2012a, Suh & Pezzuto, 2012, Jeurnink et al., 2012b, Mulholland et al., 2011, Thomson et al., 2003, Cheng & Day, 1996, Blot et al., 1993, Blot & Li, 1985, Van Rensburg, 1981). Selenium and zinc deficiency have been widely studied in prospective Chinese trials, yet research was inconclusive as to protective effects of supplementation (Qiao et al., 2009, Tran et al., 2005, Wang et al., 2013, Van Rensburg, 1981, Khomichuk et al., 2011).

Epidemiologic studies have linked the gastric changes associated with Helicobacter Pylori, as a protective factor in oesophageal cancers (Whiteman et al., 2010, Wu et al., 2003, Hunt et al., 2001, Abrams et al., 2013, O'Connor & O'Moráin, 2013, Xie et al., 2013). Human Papillomavirus (HPV) has been linked with GOC in several studies

(Kamangar et al., 2006b, Zheng et al., 2010b, Sitas et al., 2012, Guo et al., 2012). The evidence suggests further study is required to ascertain any causative elements.

Socioeconomic groups are determined through income, or status in a community. There are many variations in how this is defined. GOC has been associated with lower socioeconomic status in several studies (Levi et al., 2013, Ljung et al., 2013a, Drewnowski et al., 2012, Ellis et al., 2012, Launay et al., 2012). Yet refuted in others (Coupland et al. 2012). Higher educational levels reduce risks of GOC, but living in densely populated areas increases GOC risk (Ljung et al., 2013a).

The problem with apportioning socioeconomic status to causation of cancer lies in the significant amounts of confounding variables. Socioeconomic status effects how people live, how they shop for food, and their lifestyle choices. For example, people who live in areas of socioeconomic deprivation are more likely to smoke (Singhal et al., 2013). They are less likely to present for treatment or investigation (Wan et al., 2013, Uphoff et al., 2013). Their diets are less likely to include a variety of fresh fruit and vegetables (Levi et al., 2013, Drewnowski et al., 2012, Ellis et al., 2012, Risser & Miller, 2012, Ljung et al., 2013b, Gentil et al., 2012). These variables all link to GOC aetiology, and are difficult to separate from socioeconomic labels.

2.8 The effects of advanced presentation on the survival of patients with gastroesophageal cancer

Survival of GOC is significantly reduced with advanced presentation (Kmietowicz, 2014). This is because treatment and curative options require invasive medical or surgical procedures and therefore specialist hospital care. As time progresses, any delays leaves the tumour time to grow and to infiltrate to surrounding tissues

(Whitehead et al, 2018). This has a potential to cause worsening of the underlying condition which makes treatment more complex.

Despite the need for early intervention, many GOC patients delay seeking treatment and investigation (Grotenhuis et al, 2010). A retrospective analysis of 92 GOC patients revealed patients waited 13-16 weeks after initial 'alarm signals', before they received histological confirmation of diagnosis (Maconi, 2003). Many others have multiple GP consultations before diagnosis is made (Lyratzopoulos et al, 2012). The National Oesophagogastric cancer audit (RCS 2014) found just one out of every 20 cases of GOC was diagnosed early enough to facilitate complete surgical resection. These factors merely support a need to examine ways to encourage earlier diagnosis in patients.

A recent statistical bulletin published through the Office for National Statistics (ONS) and Public Health England (PHE) identified that earlier diagnosis (made at TNM stage 1 or 2) improves survival outcomes at 1 year in many cancers (ONS & PHE, 2016). A systematic review of a range of cancers reviewed 209 studies into survival outcomes and diagnostic times (Neal et al, 2015). Despite acknowledging the lack of consensus in timescales found across the cancer literature, authors concluded that earlier diagnosis of symptomatic cancers are likely to benefit improved survival and quality of life. This is more evidenced in cancers such as colorectal, head and neck breast, testicular and melanoma. Neal et al (2015) suggest cancer studies should concentrate on survival or mortality and evidence outcomes based on delay intervals with specific timelines.

As discussed previously in this thesis, the varied nomenclature of cancer diagnostics and delay intervals and the lack of clarity in timescales over the diagnostic journey

can be extremely confusing. Neal's et al (2015) study reflected this stance, identifying the mixed methodologies and varied timescales presented across all the studies. Data linking adverse outcomes with later staged presentation needs to be taken in context. Since the Aarhus statement on key time intervals (Weller et al, 2012), there have been many calls for transparency and consistency in early diagnostic cancer studies, to facilitate meta-analyses and assimilation of information (Tørring et al, 2012).

Historically, there has been a wealth of evidence linking earlier diagnosis with higher mortality rates (Richards et al, 1999, Guzman-Laura et al, 2011, Holmang & Johansson, 2006, Rupassara et al, 2006, Lurie et al, 2010, Foulc et al, 2003, Tørring et al, 2012, Elit et al, 2013, Crawford et al, 2002). However, these studies require consideration in a clinical context. The evidence may be somewhat misleading. Data are impaired by a range of confounding clinical issues. Patients who present with more obvious symptoms will be given priority referrals. Yet their less favourable outcomes create datasets linking higher mortality rates to early diagnostic intervals (Tørring et al, 2013, Zafar et al., 2012, Thornton et al., 2016, Sharpe et al., 2010). Patients presenting symptomatically, die earlier. This may be because symptoms of more advanced cancer are far more likely to be investigated quickly. The 'red flag' indicators reported by patients engage medical staff to expedite referrals (Grotenhuis et al, 2010). Consequently, data reflect earlier demise with expedited referral. This 'waiting time paradox' (Tørring et al, 2013) may be explained clinically, but it does skew non-contextualised data reported through survival analysis alone.

Many patients seek multiple consultations in primary care before their cancer is diagnosed (Lyratzopoulos et al, 2012). Clinical triage may be hampered by patients presenting with vague symptoms and this results in multiple consultations with primary care before diagnosis is made and treatment is offered (RCGP, 2011). This is

a factor in GOC, where signs and symptoms are vague. Tumour growth and infiltration must be extensive to produce signs of dysphagia so that patients are able to detect symptoms (NCIN, 2015).

Linking improved survival rates with earlier cancer diagnosis is also hampered by a range of confounding factors. The variety in cancer phenotypes, the aggressiveness of tumours, whether there is any involvement of surrounding tissues, and the extent of metastases in each patient all have a bearing on survival outcomes (Lyratzopoulos et al, 2013). Additionally, patient status at diagnosis, their pre-existing conditions, the older age profiles and lifestyle choices all have an effect on survival potential in cancers (RCOGp, 2011, Ramos, 2007, CRUK, 2017, Neal et al, 2015).

Despite the complexities and paradoxes in studies of presentation stage and mortality outcomes across the cancer literature, it may be concluded that advanced stage presentation in gastroesophageal cancer causes impaired outcomes. The argument falls from a biological plausibility that removal of advanced tumours require more complex treatment. Most GOC patients present with locally advanced disease requiring surgical resection (Altorki & Harrison, 2017). Symptoms are not evident until there is major tumour infiltration and survival with GOC remains poor (NCIN, 2018, Thrift et al, 2012). During the first year after surgery, tumours commonly re-present and metastases are found in over 80% of these patients (Altorki & Harrison, 2017, Whitehead et al, 2018). There are improvements in treatment options, and evidence supports addition of neoadjuvant therapies in patients who have had complete resection have the best survival (Whitehead et al, 2018). However, this means patients should present early enough to receive curative treatment.

Despite all the evidence linking later staged diagnosis with earlier demise, there remains a consensus opinion that earlier diagnosis improves outcomes in many cancers. New, clinically relevant evidence is challenging data linking earlier diagnosis with increased mortality. This is driving many campaigns for earlier diagnosis in cancer care (Neal et al, 2015). There are many calls at Governmental level to encourage earlier diagnosis. The National Cancer Intelligence Network (NCIN) and Cancer Research United Kingdom (CRUK) identify that the relatively recent '2-week wait referrals systems' and improvements in referral processes have resulted in slowly increasing survival rates (CRUK, 2014, NCIN, 2015). This 2-week wait system has now become a measurement of clinical efficacy for health systems (Thornton et al., 2016).

As previously discussed, GOC is more common in males. There is a wealth of evidence which identifies males are less likely to seek advice on health-related issues (Zheng et al. 2010a, Uphoff et al. 2013, Ellis-Brookes et al. 2012, Adams, 2008, Galdas, 2005, Gerritsen et al, 2009, Wang et al, 2013). Though studies on gender and advanced diagnosis are not evident, this may be a factor in GOC requiring further investigation. Advanced stage presentation is also linked with lower socioeconomic status in many studies (Jansen et al., 2014, Bus, 2012, Mayor, 2014, Wang et al., 2015, Islami et al., 2009, Brown, 2001, Dar, 2013b Macdonald et al., 2006). This is discussed further later in the thesis.

Advanced presentation in GOC impairs treatment options. Early surgery is required to resect the tumours fully and therefore, improve survival, yet people with GOC delay seeking consultations with their GPs (Mayor, 2014). Yet, determining how to measure the extent of patient wait time, is complex. Confounding biological factors, tumour subtypes, delays in seeking treatment and diagnosis all impact on survival. TNM

staging – as described earlier in this thesis, offers a solution to identify the extent of tumour growth and proliferation. However, many GOC patients opt to expedite surgical resection, seeking immediate treatment. As a result, staging data are often missing in datasets (Tentzeris et al, 2011, Neal, 2009, Siau et al, 2015). Staging requires diagnostic procedures and these procedures take time, thus potentially delaying actual surgical removal of the tumour.

With GOC, expedited treatment is necessary. The clinical specialists and multidisciplinary team must make the clinical decision either to perform staging, or surgery, and the latter is often chosen. One evaluation of the UK PHE's 'be clear on cancer' campaigns to encourage earlier diagnosis and referrals in GOC concluded that, although two-week wait referrals increased by up to 48%, stage at presentation was not significantly affected (Siau et al, 2016). However, it should be noted that in this study, 29 patients were diagnosed pre-campaign, whereas 41 patients were diagnosed post campaign.

Kabir & Khoo's (2016) larger retrospective analysis revealed 1143 two-week wait referrals were made before the campaign, and a further 1448 after the public campaign. However, despite the significant impact on endoscopy services, their study concluded that most tumours were still detected at later stages. Both studies concluded that there is a need to detect and diagnose cancers earlier, but that successful clinical outcome is dependent upon the stage at presentation. They raised a contentious issue that increased public awareness significantly impacts service delivery. These campaigns may offer higher rates of diagnoses, but they do not necessarily impact survival outcomes. However, as 'snapshot' studies limited to 1-year follow up, both studies failed to assess the longer term impact of diagnosing the cancers earlier.

In summary, this chapter presented the underlying background of GOC, revealing the problems impacting on survival and attributes associated with incidence. The chapter identified worldwide strategies and aims to improve outcomes in GOC, by reducing delays in diagnosis. However, there remains an element of uncertainty about which part of the patient journey is the most crucial to reducing the delay in diagnosis. In building the theory to support this research, it becomes evident that the patient journey, from detection of disease to diagnosis, must be considered. Identification of the most crucial stage can then define the focus of the research.

The literature review presented in chapter 3 reveals the most crucial interval for exploration in this thesis. It identifies the specific factors which have been linked with presentation and survival in GOC.

Chapter 3 Which is the most crucial interval to improve diagnosis and impact survival with gastroesophageal cancer?

Chapters 2 and 3 begin the 'theory generation' component of this case study, to explore evidence on gastroesophageal cancer (GOC) and patient presentation and survival. The previous chapter identified several common attributes in GOC diagnosed patients. It revealed the cancer's extremely high mortality rate and common social and demographic factors associated with diagnosis. These included male gender, obesity, smoking, alcohol misuse and lower socioeconomic groups. Survival in GOC is dependent on early surgical intervention.

This systematic literature review of empirical evidence presents a narrative synthesis of evidence to identify the most crucial stage to encouraging earlier diagnosis. It also discusses the methodological limitations in cancer research. The findings are relevant to any cancer because earlier diagnosis is essential to survival. Knowing which interval is key in improved survival can guide further research and a more targeted clinical intervention. This thesis focusses on diagnosis and presentation in GOC and this review details that diagnostic journey. It uncovers details on how patients present, what symptoms they have and how they then navigate their journey through to treatment or palliation. The findings are transferable to other cancers because earlier diagnosis is essential to survival. Knowing which interval is key to encouraging earlier diagnosis and increasing potential to curative surgical treatment, can guide further research and a more targeted clinical intervention. This chapter is important to the theory generation component of this thesis, as it substantiates the focus of the research towards understanding patient environments and how they impact on presentation and survival in gastroesophageal cancer.

GOC has an extremely high mortality and very low predicted survival rate following diagnosis (Adair et al., 2011; Allum et al., 2002; NHS 2008; Nuting et al., 2008; Office for National Statistics, 2005; Orengo et al., 2006; Sloggett et al., 2007; Zheng et al.,

2010; Medical Research Council, 2002; National Institute for Health and Clinical Excellence, 2005). GOC patients often present at an advanced stage, meaning tumours are infiltrated to such an extent that surgical cure is not possible (Tørring, 2012). A recent meta-analysis of all cancer literature linked increased time to diagnosis with a poor outcome, citing advanced stage presentation as a major concern in many cancers (Neal et al., 2015).

Given the recently published definitions to support consensus measurements of timescales in early cancer research, a structured analysis of literature related to the mechanisms of presentation in GOC is presented. This is the first systematic literature review specifically to focus on advanced stage GOC since the publication of Weller et al.'s (2012) 'Aarhus checklist'. The Aarhus checklist offers a tool to address the lack of consistency in definitions applied in early cancer diagnosis research. It is applied in this narrative synthesis of evidence as an adjunct to CASP for quality appraisal.

3.1 A literature review of empirical evidence to reveal the most crucial interval to an earlier diagnosis

The aim of this review was to identify the most crucial interval for encouraging earlier diagnosis and increasing potential for curative surgical treatment in GOC and to reveal any factors linked to GOC presentation. Objectives of this study were to:

- Undertake critical evaluation of all evidence evaluating presentation stage and survival in GOC.
- Present a narrative synthesis of evidence through Olenssen's 'delay interval framework'.

- Use the main characteristics and attributes linked with the delay periods to make recommendations for further study into early diagnosis.

The York Centre for Reviews and Dissemination guidelines handbook (2009) was used to underpin a mixed method systematic review. Results of the initial search was structured through the Preferred Reporting Items for Systematic reviews and Meta Analysis (PRISMA) statement. The PRISMA report (first published in 2009), was updated in 2015 to include protocols (Liberati et al., 2009; Moher, Liberati, Tetzlaff, & Altman, 2009; Moher et al., 2015; Shamseer et al., 2015). PRISMA is an effective and widely accepted tool in the standardisation of the information retrieval processes (Moher et al., 2015). For quality appraisal, relevant tools were taken from the Critical Appraisal Skills Programme (CASP), adapted to incorporate the Aarhus checklist for early diagnosis research (Weller et al, 2011) To address the distinct lack of consensus in definitions of diagnosis and staging in the literature, results are presented as a narrative synthesis, through Olessen's delay interval framework (Olessen et al, 2009).

A systematic literature search was undertaken for any papers published from 2000 - 2014 available through CINAHL, Medline, Psychinfo, EBSCOHOST and Academic Search Primer. This date range built on evidence available following Macdonald's seminal systematic review on upper gastrointestinal cancers in 2006 (Macdonald *et al.*, 2006). The UK clinical research network had two relevant studies in progress, but no published results to date. The National Research Register and National Institute for Health Research (NIHR) clinical trials gateway revealed no current trials on advanced stage presentations in cancers. Discussions with Cancer Research Charities UK and the Cancer Research (gastroesophageal group) network identified only one study protocol at an initial phase, so no results were available. Referenced citations were

reviewed and followed up to identify any further studies of significance to this methodology.

Several frameworks exist to assist the process of structuring a research question (PICOs and PICo (JBI, 2014), MIP (Stretch et al., 2008) and SPIDER (Cooke, 2012). As the aim of this review was to identify the most crucial stage for interventions to encourage earlier presentation with GOC, PICo was used Figure 5

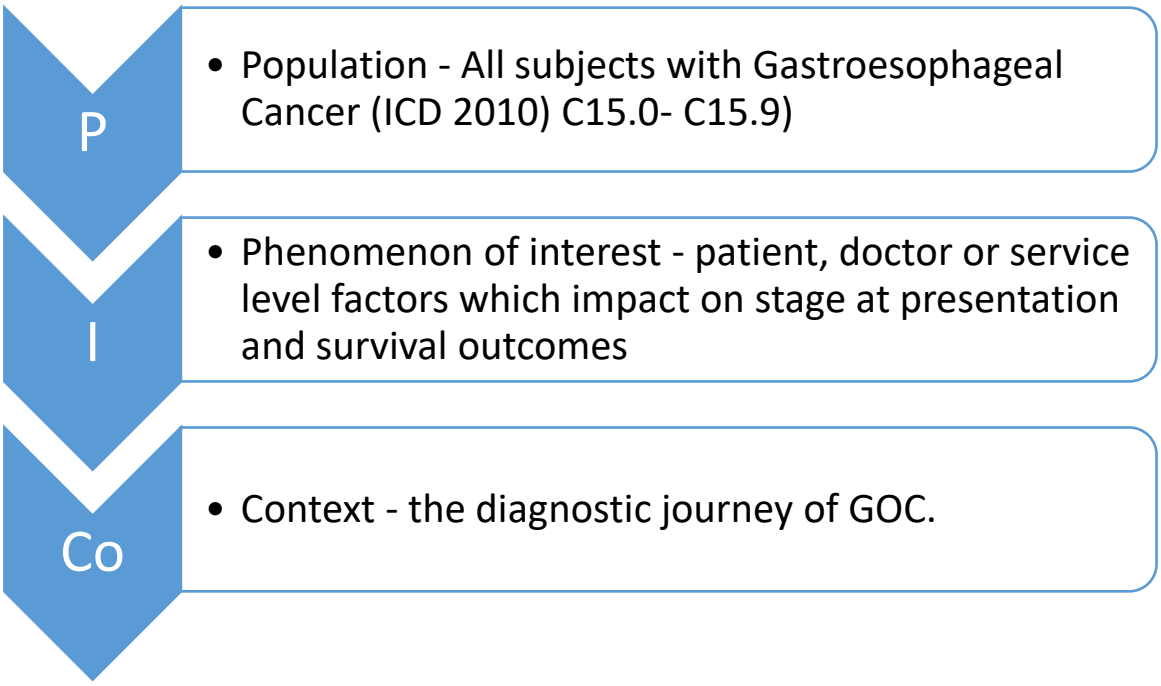


Figure 5 Developing the research question through a PICo framework

The PICo framework (Figure 5) assisted the development and clarification of the research question. The aim of this review was to identify the most crucial interval for encouraging earlier diagnosis and increasing potential for curative surgical treatment in GOC and to reveal any factors linked to GOC presentation.

Explicit inclusion and exclusion criteria supported retrieval of relevant studies. The population was set as all patients with a histological confirmation of GOC. The

phenomena of interest were any factors which impacted on patient survival outcomes. Therefore, any studies which included patient doctor or service level interventions, survival analysis, and/or revealing potential delays in the diagnostic journey were included. Studies without specific GOC focus were omitted from this review.

Studies conducted in any country were included, which were published in the English language. Dates were limited to studies published after 2000 only to maintain clinical and temporal relevance commensurate with improvements in care and treatments. Papers which did not fulfil the appropriate CASP quality appraisal criteria were discussed with the academic supervision team and then rejected where consensus opinion was reached.

A Boolean search strategy of CINAHL, MEDLINE and Academic search primer yielded 12 papers. Synonyms and specific expressions for this were drawn from an initial scoping database review. MeSH terms are identified in Table 3 MeSH terms. Only empirical studies from scholarly journals were selected for review.

Table 3 MeSH terms

#1 *esophagus	#12 advanced presentation
#2 *esophageal	#13 delay* presentation
#3 *esophagogastric	#14 #10 through #13 [and]
#4 GOC	#15 delay factors
#5 gastroesophageal	#16 interval delay
#6 upper gastrointestinal tract	#17 patient delay
#7 cancer	#18 doctor delay
#7 neoplasm	#19 hospital delay
#8 malignancy	#20 service delay
#9 #1 through#8 [and]	#21 #15 through #20 [and]
#10 late presenting	#22 early diagnosis
#11 late presentation	#23 late diagnosis
	#24 #22 and #23
All searches were limited to English language only; all adult 19+ years; humans; published post 2000 with the subject subset limited to 'cancer'	

The search strategy is presented in the PRISMA flow diagram Figure 6. This followed the Equator guidelines for transparency in reporting of systematic reviews. PRISMA offers a framework which includes a minimum set of items which require reporting in any systematic review or meta-analysis and is identified as the 'gold standard' (Moher et al., 2010). All papers focussing on treatment modalities or that were not GOC specific were rejected. After duplicates were removed, abstracts were reviewed. Those meeting inclusion criteria and addressing the research question were chosen for further analysis. Of these 24, a total of 12 empirical papers were selected for critique based on methodological rigour, appropriateness to the research question and definition of timescales. One audit detailing GOC incidence in the UK was appropriate to the review.

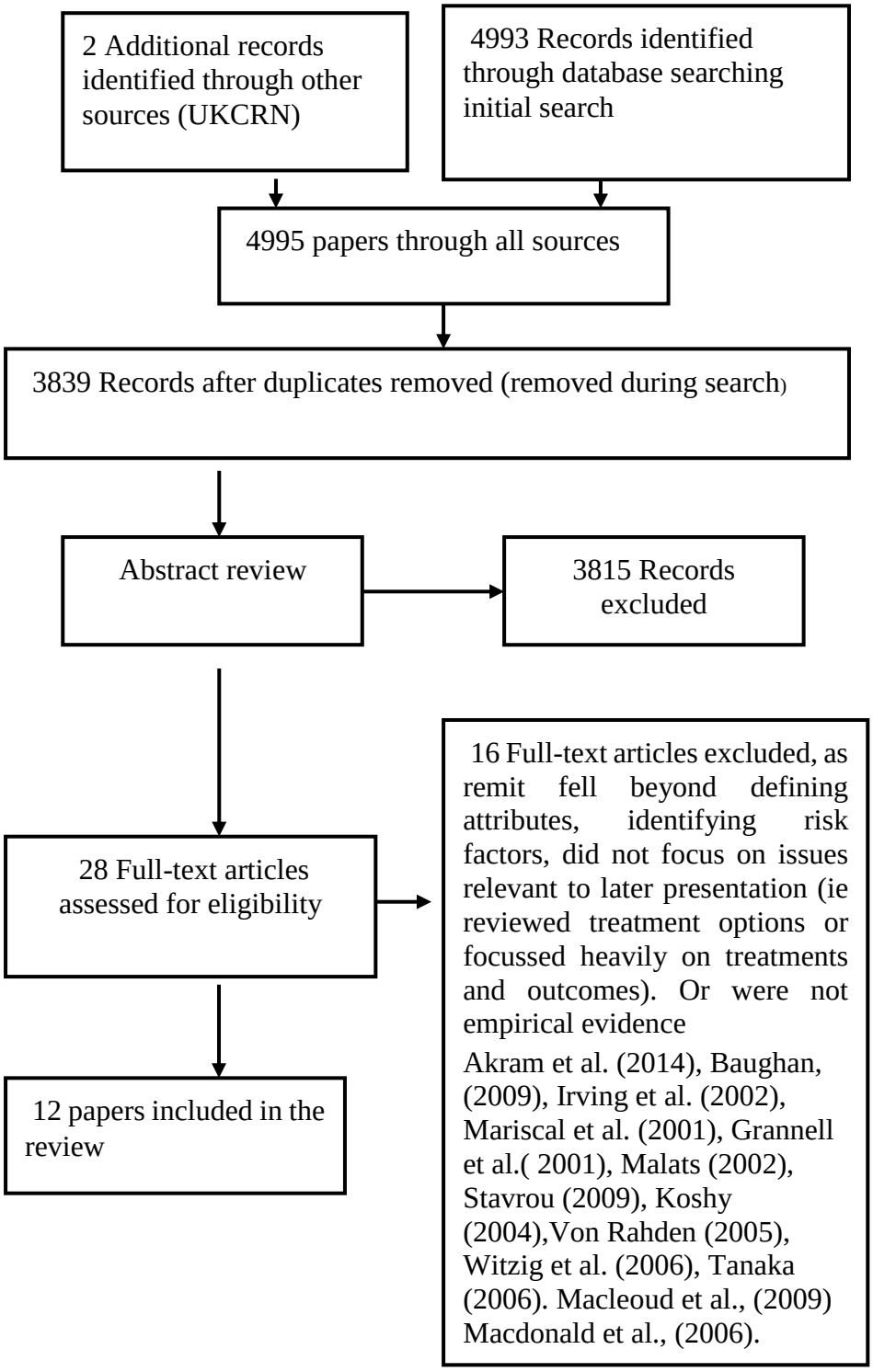


Figure 6 PRISMA flow diagram

In general, most studies were retrospective reviews of cancer registry data (survival studies), or single centre cohort studies. The national registry data studies focussed on events, survival and times and had large datasets (overall n = 26445; mean = 6611 median = 1507). The single centre cohort studies sought to explore factors which may be associated with any delays but relied on smaller numbers (n = 48-491). There were two previous systematic reviews and a GP audit which either merged GOC with other cancers, or focussed on factors associated with delay which were relevant to the discussion.

3.2 Appraising the quality of evidence on patient delay intervals

This study included mainly cohort studies and papers on survival outcomes, so the appropriate CASP tool was used to appraise results. Crowe & Sheppard (2011) have highlighted problems with having to access a range of different tools to appraise evidence on a single research question. They highlighted a need to apply a rigorous approach to appraising quality of evidence. Historically, subjective 'author assessments' of the quality of evidence based on methodology and reporting have been applied to evaluate the strength of the evidence. These are generally undertaken by applying several critical appraisal tools relevant to the methodology. Though the CASP tool for cohort studies encourages analysis of validity, results and local application, it is a generalised tool. Therefore, components of the Aarhus checklist were applied to the quality appraisal. Table 4 identifies the quality appraisal tool.

Table 4 Adapted CASP quality appraisal tool

Are results valid (recruitment and exposure)		**What are the results**	**Are they locally applicable**	**Score**
TIMESCALE DEFINITIONS Beginning and endpoints. Complexity of timepoints Date of first symptom Complexity of recognising first symptom Nature of referrals Date of diagnosis	MEASUREMENT Healthcare context Questions derived from stated definitions Theoretical framework derived from measured time points? Valid instruments for patient data? Data completeness Data extraction methods Rigour of analysis	 Precise Valid Bias Adjusted for confounders Consistent Biologically plausible	Healthcare context	 Insufficient Moderate Strong

All papers were initially assessed by the primary researcher who discussed findings with an academic supervisory team consisting of a professor of nursing and a gastroenterology medical specialist. Where two researchers disagreed in relation to quality appraisal, the third was consulted and the presented evidence discussed until consensus was reached that the appraisal of evidence was appropriate.

Results were conceptually mapped to Olessen's delay intervals (Olessen et al, 2009). This was used to identify the most crucial period to encourage early diagnosis. This theoretical approach allows data to be presented through pre-determined themes which are drawn from the studies which have been selected (Braun and Clarke, 2006; Dixon-Woods *et al.*, 2006 a & b). Results of the multiple studies were assimilated into a data extraction form which incorporated both CASP and Aarhus checklist components.

3.2.1 Outcomes of the literature search and quality appraisal

A total of 3839 records were extracted to RefWorks citation manager which assisted removal of duplicates. These were narrowed following initial and abstract review, to 28 full length papers which were assessed for eligibility. The PRISMA diagram (Figure 6) identifies these texts and rationalises why they were not chosen. The search strategy revealed a significant lack of available evidence with sole focus on gastroesophageal cancer and delays in the diagnostic journey. Much of the research evaluated efficacy of treatment, or survival outcomes but did not apportion the specific time of delay. Table 5 presents a summary of studies included, alongside their quality appraisal results.

Table 5 Studies for appraisal

Author	Study type	Participants	Validity (timescale definitions and measurement)	Results	Locally applicable	Delay interval	Appraisal in relation to research question
Subasinghe (2010)	Single centre prospective cohort – patient. questionnaire & retrospective analysis of records	N = 48	No specific timescales and end point histological confirmation delays from patient first noticing symptoms – histological diagnosis. Relied on patient recall, no survival analysis	3/12 average delay between patients noting symptoms to seeking treatment. 2/12 delay between presentation and treatment Presented with progressive dysphagia	Sri Lanka	Patient	Moderate

Schlansky et al. (2006)	Single centre retrospective cohort descriptive analysis of disease characteristics & diagnostic patterns.	N = 131 (data ranged 1999-2004)	TNM criteria applied. Separated histological types of carcinoma,	most patients presented with symptoms >TNM3 (through analysis of records & GP notes & TNM staging)	USA	Patient delay	Moderate
Gibbs et al. (2007)	Single centre cohort analysis Tumour registry data analysis	N = 307 (data range 1991-1996)	TNM staging confirmed histological diagnosis Subjective patient recall of symptoms to medical alert	Most patients presented at TNM stage 3 or above Average 3/12 symptomatic before presentation to GP ADC presents dysphagia & weight loss/SCC with smoking/alcohol dysphagia & weight loss	USA	Patient doctor & system (through to diagnosis)	Moderate

Smithers et al. (2010)	Retrospective cancer registry analysis	N = 1100 (2002-05)	Follow up secondary dataset, patient self-completed questionnaire – reliance on patient recall, TNM staging only available for 27% cases, AJCC staging in only 7% records	Most patients presented at a later stage in disease with self-reported dysphagia. No timescales identified in the study.	Australia	Patient derived from case notes but no timelines explicit	Moderate
Wang et al. (2008)	Single centre retrospective comparison Purpose to measure delay and stage at treatment	N = 80 (Jan-July 2007)	Recollection of symptoms. All patients given TNM staging at surgery or via CT. unclear as to pre-diagnostic assessments as staging done at surgery	Most patients presented stage 2 or above – patient delay median 2.1 months with 11% waiting up to 6/12 (elicited from interview and patient recall) Average 1.2 months between doctor and diagnosis and 0.25 months taken from histological confirmation – to treatment	China	Patient doctor & system	Strong

Bus et al. (2014)	Retrospective cancer registry analysis	N = 1914 (data range 1990-2008)	Retrospective survival analysis to determine characteristics of survival. TNM staging by pre-diagnostic criteria.	The higher the TNM stage, at presentation the more likely death occurs within 1 year	Netherlands	Patient	Moderate
Hashemi, DiMarino & Cohen (2009)	Single centre retrospective cohort	N = 242 (1994-2004)	Comparative analysis of presentation aged under 50 and over 50. Unsure how TNM stages identified, no data on timescales	Advanced TNM stage at presentation worsens outcome. Younger patients (<50) experience dysphagia for longer before seeking treatment (4.5 months versus 2.5 months)	USA – single centre	Patient	Moderate

Grotenhuis et al. (2010)	Retrospective case note review	N = 491 (1991-2007) Explicit I/E criteria.	Analysis of delay intervals in diagnosis and treatment – ASA classification TNM – no identification of how 'onset of symptom data elicited from patients	ASA classification iii and iv have more hospital delays than ASA i and ii Average patients wait 3/12 with symptoms Higher stage at presentation – less delays in treatment Mean wait to treat from diagnosis to 28 days	Netherlands	Patient Retro analysis of notes to elicit information. and system	Strong
Abdullah, Karim & Goh, (2010)	Single centre retrospective case note review	N = 143 (1998-2003)	Histological confirmed diagnosis, TNM staging but no description of how this took place.	Late staged presentations have a very high mortality rate Most patients here presented too late for curation SCC GOC linked with smoking and alcohol	Malaysia	Patient	Poor

Alimoghaddam et al. (2014)	Single centre retrospective analysis patient records (stomach included in this but only results for GOC extracted)	N = 368 (1995-2011)	TNM classification – time of diagnosis (histological confirmation – to death. Patient files used to collate data.	Over 60% patients present at stages iii and iv.	Iran	Patient	Poor
Coupland et al., 2012	Retrospective cancer registry analysis Focus on survival analysis and histological subtypes.	N = 133804 (1998-2007)	National cancer registry data analysis Survival analysis by morphology, site, gender and IMD	Only 8.3% 5-year survival recommends patients need diagnosis early stage to improve outcomes Obesity, smoking and alcohol linked with poor survival	UK	Total	Moderate

Kotz et al. (2006)	Retrospective cancer registry analysis	N = 632 (1995-2000)	Diagnosis by date of biopsy (verified through histological confirmation) TNM classification through retrospective note trawl. KM survival – censored appropriately	Noted larger delays to surgical resection, but without detrimental effects on survival. Most patients presented at TNM iii and ii. Those presenting at later TNM stage received longer delays to surgery. Median diagnostic delay 47 days (0-287)	UK	System delay	Moderate

3.2.2 Studies covering the total delay interval in gastroesophageal cancer

Six studies presented in this review presented findings across all the delay intervals. Bus et al. (2013) presented a study on 1, 3 and 5-year survival in GOC, from which factors relating to presentation may be drawn. Their data analysis reflected Coupland et al.'s (2012) methodology of retrospective analysis of cancer registry data. Bus et al. (2013) revealed patients presenting with limited lymph node involvement had improved survival outcomes. They also revealed that gender and absence of comorbidities were linked with increased survival. Numbers of subjects in the 1, 3 and 5-year cohorts were 703, 551 and 436 respectively, so the study was large. As with the UK, cancer notification is mandatory in the Netherlands, so there were no omissions from the dataset. The study relied on older data for survival analysis (some of these data were 24 years old at the time of publication) and this can only reflect the treatments available at that time. However, the study showed the ability to treat with curative intent increased survival significantly. Presentation at later stages and with comorbidities significantly reduced survival outcomes. Advancing age was not linked with impaired survival for those in the 3-year and the 5-year groups, but it was significant in those people dying before 1 year. They did not undertake further analysis with histological subtypes, but found a wide disparity in survival between different treatments (surgery, neoadjuvant therapy) and the extent of tumour infiltration and lymph node involvement at presentation, concluding that the earlier patients present with their GOC, the more likely they are to survive.

Coupland et al.'s (2012) research focussed on GOC diagnosis across the UK. A sample of 133804 GOC subjects presenting between the years 1998 - 2007 were analysed in relation to presentation and survival. This study revealed over half of middle and upper GOC were in females. This contradicts the usual 1:2 ratio of male to female GOC

which is prevalent in the mainstay of GOC literature. The study also highlighted differences in survival relating to age, socioeconomic deprivation and stage of disease and called for further research into these factors. The authors conclude that early diagnosis is crucial to survival and encourage further studies to encourage earlier presentation, referrals and treatment.

In the US, cancer registry data are handled differently. Hence more reliance on single centre studies, or state led studies which focus on private or public held datasets. A single centre study in the United States of America by Hashemi et al. (2009) compared age groups and survival in 242 patients presenting between 1994 and 2004. They found younger patients tended to have higher lymphatic spread, but exhibited similar survival outcomes. Advanced stage presenters had worse outcomes, and the authors identified patients detecting symptoms and seeking consultation is crucial to improved outcomes.

Another Netherlands study by Grotenhuis et al. (2010) revealed the impact of pre-hospital and hospital delays on survival. They identified that late onset of GOC symptoms often predisposes a delay in presentation. Grotenhuis's (2010) study also revealed that a shorter timescale between diagnosis and surgical treatment significantly improved survival outcomes. This single centre cohort study of 491 patients presenting between the years 1991 and 2007 found improved survival outcomes occurred with expedited treatment. Their recommendation was to avoid delays, expedite referrals and encourage earlier presentations.

In Malaysia, Abdullah et al. (2010) studied 143 patients presenting between 1998 and 2003 to a single centre. This study identified most patients in Malaysia present at very advanced stages and had very low survival outcomes. They also identified diagnoses

of GOC was more prevalent in lower socioeconomic groups. This study was limited as a single referral centre, but found that over 71% of subjects presented at TNM stage IV. These statistics differ significantly from Western studies, meaning results may not be applicable to UK healthcare. However, the study did identify that patients had not interpreted their symptoms quickly enough to permit surgical intervention. The link between improved GOC survival and later staged diagnosis focusses on the patient interval. With a similar demographic profile, Alimoghaddam et al.'s (2014) retrospective analysis of 368 medical notes in a single centre, linked later staged presentation with impaired survival as most subjects presented with very late staged GOC.

3.2.3 Studies covering the patient delay interval in gastroesophageal cancer

Five studies presented data relating to advanced stage presentation which could be linked to the patient interval. Subasinghe et al. (2010) undertook a 24-month study of patients presenting to a regional centre in Sri Lanka to review where delays may occur in the patient journey. They identified three periods. These included the time from patient first detecting symptoms, to first contact with the health service, then the patients' first contact to endoscopy, then from endoscopy to histological confirmation. The study identified the patient interval as the most crucial delay period. They identified that the time taken between patients recognising their first symptom, to them seeking medical consultation, accounted for up to 82.2% of the delay when compared to the remainder of the cancer diagnosis and treatment intervals. However, as with the Malaysian study by Abdullah et al. (2010), the ability to generalise studies from Malaysia and Sri-Lanka to the Western context is potentially hampered as there are stark differences in healthcare provision. The study by Subasinghe et al. (2010) did acknowledge that there was a lack of available resources to support histology and

endoscopy investigations in Sri Lanka. Subasinghe et al (2010) have not specified how they determined the patient's first symptom detection and there is no information on when patient interviews were timed. There is also no evidence of triangulation of self-reported data. Authors did not elaborate on methodological approaches and data analysis, so this remains unclear. Their cohort also presents a gender ratio of 1:1, which differs significantly from other studies.

A single centre retrospective evaluation of patient notes in America by Schlansky et al. (2006) presented data on tumour stage and presentation symptoms in patients with gastroesophageal cancer. They reported on any subjects who had undergone oesophagogastroduodenoscopies who had not been previously diagnosed or treated for GOC. Medical notes were reviewed to reveal disease characteristics and patterns of diagnosis. They found most presentations were staged at T3 or above, revealing a large 'patient interval delay'. The study is somewhat limited through reliance on patient notes to elicit data on clinical symptoms at presentation.

Another American single centre study by Gibbs et al. (2007) provided evidence from a retrospective analysis of 307 GOC patients who presented between the years 1991 - 1996. They acknowledge their data is based on subjective patient response on when symptoms were initially detected, but conclude that survival worsens when the time between patient detection and seeking treatment is increased. The study is limited in that authors provide no rationale for choosing the date range of 1991 – 1996. It must be noted that data were old when the research was published in 2007. There is also no elaboration made on sample coverage and completeness of information available in the cancer registry data which was used for extraction.

Smithers et al. (2010) presented a secondary analysis of data collected for the larger Australian Cancer Study. This study began with a sample of 3273, but many of these subjects were either non-contactable, or had died before the study took place. The strength of this study lies in the pre-validated questionnaires used to elicit information on patient symptoms and presentation. However, the total sample was skewed towards survivors and earlier stage presenters. Of 1100 subjects, only 831 had a recorded presentation and symptom history. The authors identify staging information was only available for 7% of patient records, but it is unclear whether this is 7% of the full cohort of 1100, or the 831 with recorded full data. The authors apply a range of cross data analyses to generate TNM staging, generating a figure of 50% of the cohort with staged data. With this, they draw the conclusion that most patients with GOC will present with late stage disease. All the American studies (Hashemi et al. 2009, Schlansky, 2006 and Gibbs, 2007) are extremely useful, but their generalisability is hampered by their single centre status and therefore localised sampling.

The study by Wang et al. (2008), offered a comparative analysis of TNM stage I and II diagnosed patients, with TNM stage III and IV to compare delay intervals and their impact on survival. Using similar timeframes as Subasinghe et al. (2010) they conclude that symptomatic patients generally wait on average 2-3 months before seeking assistance and that those who present earlier, present with smaller and more localised tumours. They acknowledge the limitations of patient recall and, rarely, are they able to offer a dataset with complete TNM records.

3.2.4 Studies covering the treatment delay interval in gastroesophageal cancer

Treatment interval relates to the interval between diagnosis and treatment and involves service level delays. Kotz et al. (2006) presented a study to identify whether delays between diagnosis and surgery may have an impact on survival outcome. Data on 800

patients presenting to a single centre from 1995 - 2000 were analysed to assess whether delays between diagnosis and surgery effected survival outcomes. This study found the time taken between histological diagnosis, clinical decision-making and rationalising interventions through multidisciplinary team discussions, improves survival outcomes. Kötz *et al.*, (2006) reiterate the importance of undertaking multidisciplinary deliberation on treatment strategies and found that clinically considered interventions receive more favourable outcomes. This is supported in other studies which addressed the total delay interval (Grotenhuis *et al.*, 2010).

3.3 The most relevant interval to encourage earlier diagnosis.

This literature review of 12 empirical studies links a need for earlier diagnosis with survival and five studies specifically focussed on the patient interval, identifying it as a crucial time to engage in preventive measures and encourage earlier diagnosis (Subasinghe et al, 2010, Abdullah et al, 2010, Schlansky et al, 2006, Gibbs et al, 2007 Smithers et al, 2010). A total of 6 papers specifically identified the link between early diagnosis and improved survival outcomes (Coupland et al, 2012, Bus et al, 2014, Schlansky et al, 2006, Gibbs et al, 2007, Smithers et al, 2010, Wang et al, 2008).

Most patients present at too late a stage for surgical curation of gastroesophageal cancer, and survival times were reduced through delays in presentation, thus intervention. This supports the UK government and cancer research UK (CRUK's) drive to reduce the delays between presentation and diagnosis with cancer. Six of the studies linked less favourable outcomes with advanced stage tumours, and identified the patient interval as essential to the process of initiating the required medical and surgical interventions (Coupland et al, 2012, Bus et al, 2014, Schlansky et al, 2006, Gibbs et al, 2007, Smithers et al, 2010, Wang et al, 2008).

3.3.1 Reflections from the wider literature

This is supported in the wider cancer literature. Despite the 'diagnostic paradox' studies which link increasing mortality statistics with earlier diagnosis, there remains a consensus opinion that encouraging earlier diagnosis is essential to improving cancer survival (Hiom 2015, Neal et al., 2015). A survey of GPs in Scotland published in 2009 identified referral disparities which were based on cancer type (Baughan et al, 2009). They found General Practitioners (GPs) referred potential breast and skin cancers far more rapidly than other cancers, including those of the upper gastroesophageal, prostate and lung. Kötz et al. (2006) identified that GPs were more likely to refer males than females for investigations of suspected GOC, suggesting females had a higher propensity to system delay. Whereas Macleod et al. (2009) and McDonald et al. (2006) identified males and females exhibited similar wait times to visit a GP with suspected symptoms.

Mcleoud et al. (2009) undertook a literature review of a variety of symptomatic cancers which investigated the factors which caused delays in both presentation and referrals. The study highlighted clinical factors such as symptom severity, patient's appropriate awareness (and interpretation of) their symptoms. Though this literature review did not focus specifically on GOC, the authors noted that patients needed to be aware of their own clinical symptoms and act accordingly. McLeod et al (2009) also noted the effects of a patient's emotional status and availability of support networks on how patients present. Delays in doctor and system levels were attributed to personal and demographic attributes of patients, how they presented and provided histories, and how practitioners responded to cues. McDonald et al. (2006) undertook a systematic review of evidence published between the years 1970 - 2003, to reveal factors which may be attributable to delays in diagnosis of upper gastrointestinal cancers. The

authors identified the main patient related factors associated with advanced presenting GOC included low socioeconomic status and non-white ethnicity. They also identified the clinical manifestations such as presence of pain or bleeding, or any symptoms effecting general functional status. The main system related factors to delays in diagnosis, related to the initial misdiagnosis of common symptoms. Authors surmised that older males from lower socioeconomic groups were less likely to receive faster referrals. Many studies report socioeconomic deprivation as an attribute to GOC diagnosis, but these studies are based on limited analysis of site of diagnosis (Abdullah et al., 2010, Bus, 2014, Baughan et al., 2009, Mao, 2011). Akram et al., (2014) identified living in a rural area has been linked to GOC incidence, but this may reflect limited access to services. Many studies link rurality with environmental factors and studies into these are ongoing (Mao, 2011, Zhang, 2013, Mohebbi et al., 2011, Aragones et al. , 2007).

All studies in this review identified that advanced presentation and delays in diagnosis caused worse outcomes. The Royal College of Surgeons' (RCOS) National Oesophagogastric cancer audit (2014) noted significant links between 'emergency hospital admission diagnosis and extremely high mortality statistics. Again, this highlights the need for earlier presentation and supports the case that patients need to seek medical consultation before symptoms are severe enough to generate emergency admission.

The problem with the patient detecting a potential cancer early enough for treatment, is that GOC has a very insidious onset and clinical signs may not be clinically detectable until the tumour has reached extensive infiltration (Di Pietro, 2013, Jayasekera, 2012, Lambert, 2012, Yang, 2012). The usual clinical signs are dyspepsia, dysphagia, nausea and vomiting, weight loss, or gastrointestinal bleeding (Wolf et al.,

2012). Many papers identify the presence of 'alarm' signals caused patients to seek health advice more rapidly (Hashemi et al., 2009; Gibbs et al., 2007; Akram et al., 2014). Symptoms of weight loss and pain were linked with reduced survival outcomes, but these symptoms were key to patients seeking help (Macdonald et al., 2006; Schlansky et al., 2006; Macleod et al., 2009; Wang et al., 2008). However, cachexia – the dramatic loss of weight with cancers – tends to signify the tumour is at an advanced stage (Dhanapal et al., 2011).

3.3.2 The complexity of integrating global evidence on GOC diagnosis.

Much of the evidence on presentation of GOC relies on retrospective cohort studies. Of the 12 papers, in this literature review, four offered a retrospective analysis of existing datasets. These datasets rely on country specific definitions of timescales. This literature review evaluated geographically diverse studies, and these should be considered in context. In China, for example, patients can present directly to the specialist hospital, whereas in the UK, GPs are used as the gatekeepers of care.

Public funding and access to healthcare services also has an impact on patient presentation. The UK offers a free national health system to the total population, whereas the United States, for example, uses health insurance and a private health system to enable basic or advanced level access to health-related services. Statistics from each country depend on how patients navigate their journey towards diagnosis. Data will subsequently be skewed by patient's socioeconomic status, access to services and by healthcare funding systems.

Different health systems may not be accurate to capture and record all patient data which is relevant to the underlying condition. This was evident in one study relating to GOC which was excluded from this review due to methodological limitations.

Baughan et al., (2009) study, noted significant disparities in data collection across health boards and that this factor resulted in impaired analysis of the first year of results. Baughan's study was omitted from this review, due to these limitations in data collection. The UK National Audits had similar issues. They retrieved data from 99% of individual trusts across England and Scotland, but Wales did not have sufficient data to complete the national 2013 GOC audit (AUGS, 2013). Even with hospital episode statistics, the audit could still only claim an overall case ascertainment rate for newly diagnosed cancers, to an 85% accuracy (AUGS, 2013).

This review revealed that most studies identified patient interval as most crucial to earlier diagnosis. However, the evidence presented is based on disparate measurements of timescales. For example, Kotz et al. (2006) identifies the date of endoscopic biopsy as the date of diagnosis and date of death as the end-point, whereas Gibbs used the date of histological confirmation. Others did not identify how this was defined (Coupland et al., 2012, Alimoghaddam, 2014). There was also a range of different diagnostic criteria applied to evaluate survival. Even this was inconsistent – as many studies identified a significant lack of staging data, or presented results for only subsets of cohorts under evaluation. This lack of data consistency directly affects underlying quality of the evidence (Weller et al., 2012, Liberati et al., 2009).

There was also a lack of clarity on how and when staging took place. Both Alimoghaddam et al. (2014) and Subasinghe's (2010) studies identified missing TNM staging data. Abdullah et al. (2010) reported a 100% data yield, but did not identify how or when staging was undertaken. This was also the case in studies by (Schlansky et al. (2006) and Grotenhuis et al. (2010). The UK Oesophagogastric Audit reported data were missing in 2819 patients (Royal College of Surgeons, 2014). These

variations in the processes of measurement severely threaten internal validity and prevents meta-analysis.

Reliance on the degree of tumour proliferation as a measure of how long the cancer has been *in situ* depends on how metabolically active the tumour is. Subasinghe's study (2010) highlights that some tumours may have been present for up to ten years before the patient presents and with little or no suggestive symptoms. Yet Dutta et al. (2012) describe a 'doubling time' where extremely metabolically active gastroesophageal tumours can grow and spread extensively. In these cases, early diagnosis and rapid removal is essential to a more favourable outcome. These factors will affect the findings of this literature review, which relied heavily on evidence generated from survival studies.

3.3.3 The complexity of interpreting patient histories and defining the referral process.

Throughout the patient journey there are many factors affecting presentation. The patient must suspect there is a problem, the GP must pick up on diagnostic cues, referring appropriately and healthcare systems must run effectively to instigate rapid diagnosis and treatment. The 6 studies which identified 'patient interval' as a critical period for earlier diagnosis linked later staged diagnosis, with impaired outcomes (Coupland et al, 2012, Bus et al, 2014, Schlansky et al, 2006, Gibbs et al, 2007, Smithers et al, 2010, Wang et al, 2008).

Generally, the patient interval is measured from the date the patient first noticed a symptom, to the date they presented to their GP (Weller 2012). However, there is a significant disparity relating to 'date of first symptom'. It cannot merely be assumed that patients will detect a set of symptoms and will then present to health services to

be diagnosed with GOC. There are specific symptoms of GOC presentation, (alarm signals) which are present in almost 50% of cases (Thrift et al., 2013; Wallace et al., 2001; Fransen et al., 2004). However, by the time the patient has symptomatic presentation, the tumour may be far too advanced for curative surgery (Fransen et al., 2004; Jones et al., 2007)

Patients may present with just one sign of GOC (for example, with 'pain') and subsequently receive analgesia or health related advice. They may then re-present at a different time, with the alarm signals resulting in referral and subsequent diagnosis. The time taken for either the patient or the referral centre to 'notice' GOC related symptoms and seek further assistance, is subject to recall bias from all parties involved. Patients may be poor historians, they may not identify the full extent of the complaint. Any study investigating delays in the journey attributed to when a patient or a referrer 'noticed' original symptoms needs to be thoroughly scrutinised. The operational definitions underpinning this type of research require objective consideration. This review used evidence from several retrospective case note analyses and findings are dependent on how 'well' the original assessment was documented. Any reports on when patients 'noticed' signs can only be considered subjective, unless symptoms occurred within days, minutes or hours before presentation. The evidence recording 'time from noticing a symptom, to time to diagnosis or cure' has a significant potential bias.

3.3.4 The evidence on GOC in comparison to other cancers

This literature review also revealed the issue that gastroesophageal cancers tend to be overlooked in the mainstay of literature. Baughan's (2009) study identified 600 cases of oesophageal cancer, but only reported outcomes for the more common breast, colorectal lung and prostate cancers, rather than gastroesophageal cancer and this is

common in many cancer research articles. The significant disparity in focussing on the less common cancers is evident. Gastroesophageal cancer patients die early, therefore, follow up is difficult. As a result, this cancer is subject to survivor bias, because patient data are skewed by early subject demise.

Many papers on cancer survival report a 1, 3 or 5-year outcome. However, most GOC diagnoses result in death before two years (Kötz 2006). This affects the data and findings significantly and misses the large number of GOC subjects who die within the first 6 or 8 months. A reclassification of diagnostic criteria would allow for this skew in GOC survival.

Gastroesophageal cancer diagnoses occur across the globe, but is more prevalent in Middle Eastern and Far Eastern countries, where English is not the primary language. Subsequently, there may be evidence published in other languages which was not identified in this review. As identified in the discussion, the range of different timescales and nomenclature applied in cancer research have an impact on results. To identify when a patient actually detected their first symptoms is hampered by several confounding factors. The patient journey through many different healthcare systems is also complex. For example, how patients may present and navigate treatment with symptomatic GOC in China may differ significantly to that of US, or UK populations.

This review of GOC specific evidence identified 6 papers which supported the patient interval as the most crucial time to encourage earlier presentation (Coupland et al, 2012, Bus et al, 2014, Schlansky et al, 2006, Gibbs et al, 2007, Smithers et al, 2010, Wang et al, 2008). Treatment of GOC includes medical intervention and surgical removal of the tumour – removal is far easier when the tumour is smaller, or where infiltration to surrounding structures has not occurred (RCOS, 2014).

This review also highlighted the disparate processes of measurement and sampling in cancer research articles. There are a wide range of methodologies and measurements in timescales and outcomes. The nomenclature is diverse and confusing. The study noted a significant lack of specific focus towards gastroesophageal cancers. The incidence of GOC is increasing and survival remains poor, yet studies tend to group the condition with head and neck, or gastric cancers.

Consequently, clinicians and researchers must begin to focus on this increasingly prevalent cancer. The key finding of this literature review is that the patient interval is a significant part of a patient's cancer journey towards diagnosis. It is crucial for patients to seek a diagnosis and access surgery as soon as possible, so that their chances of survival are improved. With this in mind, clinicians must strive to explore ways to encourage this early diagnosis.

3.4 Are there any crucial factors which influence diagnosis, survival and premature mortality in cancer diagnosis?

The literature is scant regarding system level interventions and referral processes, but with the access to hospital episode statistics, the evidence may build over time. The UK NAEDI programme have recently published an updated hypothesis (Figure 7), based on evidence gained from several different cancers. The new hypothesis identifies factors influencing cancer diagnosis, survival and premature mortality.

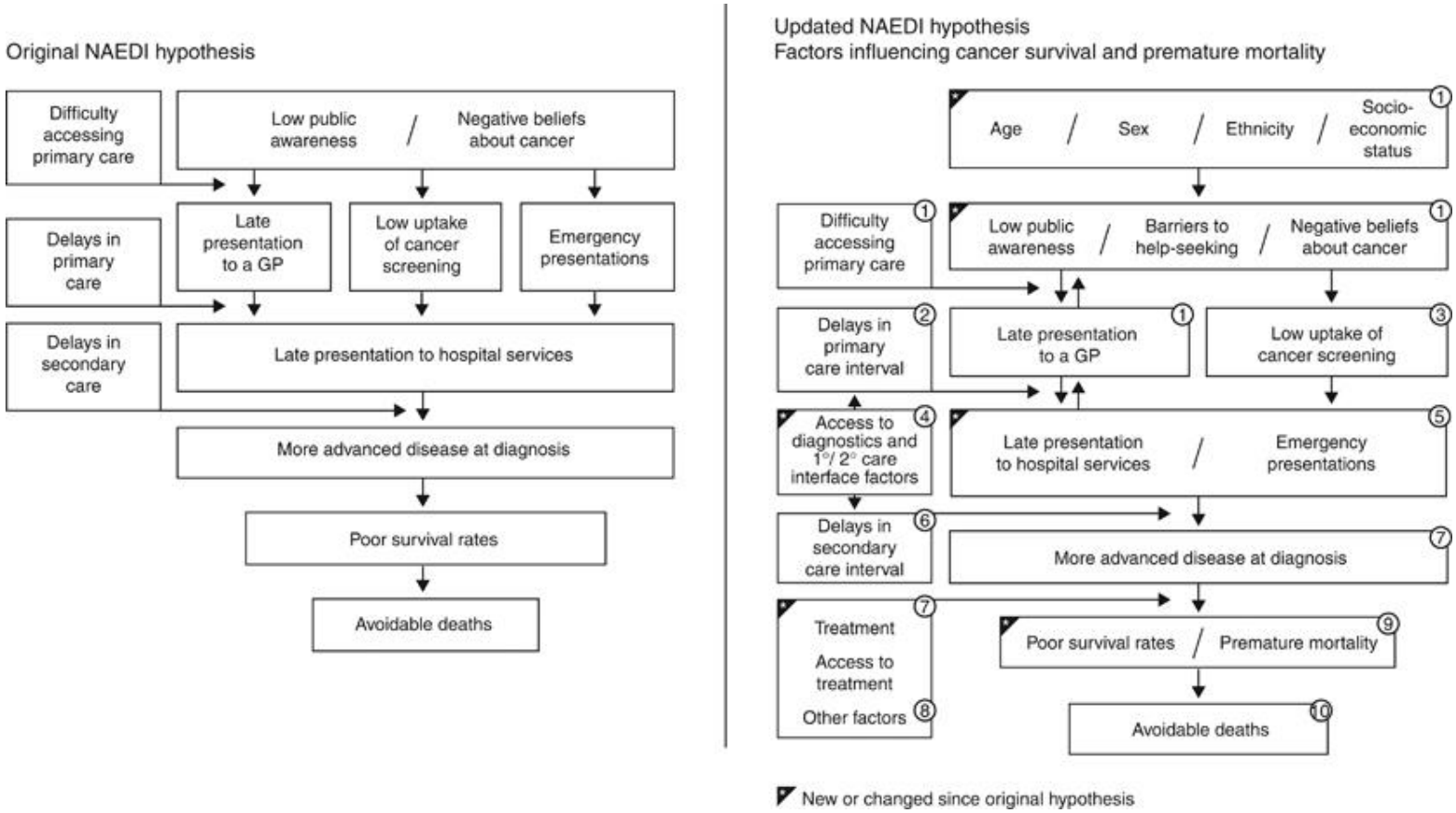


Figure 7 The original and the updated NAEDI hypothesis (Cited in Hiom 2015)

NAEDI calls for further studies into these factors influencing cancer survival (Hiom 2015). Given that this literature review identified the patient interval as the most crucial for encouraging earlier diagnosis, then further study into patient related factors is required. One factor is highlighted across the UK, that there is a major disparity in incidence, presentation and survival across the country. The National Cancer Intelligence Network (2010) have identified a need for further studies to evaluate this geographical disparity.

The updated NAEDI hypothesis identifies that age, sex, ethnicity and socioeconomic status are intrinsic factors which have the potential to affect the patient interval. There are several behaviourally modifiable risk factors which may be associated with GOC presentation. The cancer has been previously linked with smoking, obesity and alcohol misuse (Abdullah et al., 2010, Macdonald et al., 2006, Parkin, 2011). Some studies link a lower socioeconomic status with the diagnosis (Macleod et al., 2009; Abdullah et al., 2010). Though this is contested in Coupland et al.'s (2012) study. These authors undertook further evaluation on the cancer site and revealed lower socioeconomic status was more prevalent only in those with upper and mid oesophageal tumours. Many studies report socioeconomic deprivation as an attribute to GOC diagnosis, but these studies are based on limited analysis of site of diagnosis (Abdullah et al., 2010, Bus, 2014, Baughan et al., 2009, Mao et al., 2011). Akram et al. (2014) identified living in a rural area has been linked to GOC incidence, but this may reflect limited access to services. Many studies link rurality with environmental factors and studies into these are ongoing (Mao, 2011, Zhang, 2013, Mohebbi et al., 2011, Aragones et al., 2007).

Advancing age, male gender and lower socioeconomic status are strongly associated with GOC, and harnessing these for further research may be useful. In identifying the patient interval as most crucial to survival, this review identifies a need for further research into how patients present. By exploring these patient's neighbourhoods, it may be possible to elicit information on where these patients may reside, whether there are clinically relevant geographical pockets at smaller scales to inform more localised and targeted strategies and encourage earlier presentation. The following chapter evaluates current evidence on where people live and whether patients' environments affect how and when they present and concludes with the research question which underpins this thesis.

Chapter 4 - Neighbourhoods and presentation in gastroesophageal cancer developing the research question.

This chapter is the final part of theory generation in the case study. It introduces geodemographics and population analysis techniques as tools to investigate spatial elements which may affect presentation and survival in gastroesophageal cancer. It identifies the government drives to engage research and encourage survival in cancers and proposes social marketing techniques to explore patient behaviours. The chapter introduces the geographical information system as a tool for the storage and analysis of wide ranges of disparate data and explores the geographical scales which are used in this research. The chapter then presents the refined research question and proposes a study methodology.

The theory that a later stage of presentation restricts surgical intervention which adversely affects survival in gastroesophageal cancer (GOC) is upheld in the literature. The previous chapter concluded the patient interval as a crucial time period for further study. GOC patients must be able to access treatment early enough for surgical intervention – but this means they must first be able to recognise symptoms as serious enough to require medical review, then they must act on that recognition. They must present to health services, have these symptoms properly diagnosed and be referred accordingly. Then clinicians must implement the most beneficial investigations and treatments to improve survival outcomes and patients' health status must be secure enough to manage the full journey. These factors present many foci for research and further evaluation.

As part of the theory generation element of the case study, this thesis introduced the physiology, attributes and factors which have been associated with GOC. A literature review of 12 studies focussing on GOC revealed patient interval as a crucial time to act and encourage patients to seek earlier diagnosis. Previous chapters have introduced the geographic nature of GOC and identified the requirement for further investigation.

Since the writings of Hippocrates, the concept of studying population characteristics and disease spread in populations has underpinned a wealth of research into a range of diseases. Through systematic analysis, researchers can generate understanding of the issues and problems in communities and use this intelligence to improve health (Kuhn 1996, Elliott & Savitz, 2008, Wang et al., 2010, Hahn, 2014, Tanaka, 2014, Jacquez, 2004, Elliott et al., 2009). The incidence of GOC is shown to cluster across certain geographies. Incidence is often described using terminology such as 'the geographical belt' and World Health Organisation figures illustrate this cancer's geographical tendency (Figure 1). However, there have been no published 'local level' studies which evaluate this phenomenon.

For many years, private sector organisations have used population characteristics to provide insight into consumer behaviours and target specific social markets. They have harnessed geodemographic research (neighbourhood analysis) with geographical epidemiology (exploring the how and where of incidence) to predict where the 'most likely' customers are. Their social marketing techniques can offer a predictive intelligence to inform service delivery (McCandless, 2014, Le Sage et al., 2011, Laszkiewicz et al., 2014). This approach may be useful to healthcare and supports a drive to analyse neighbourhoods in relation to patient presentation and survival outcomes in gastroesophageal cancer.

4.1 The government drive to consider neighbourhoods in cancer survival outcomes

The National Awareness for Early Diagnosis and Intervention (NAEDI) strategy have called for further research into patient environments and how these may affect access to care (CRUK, 2010, Hiom 2015). Yet to date, there have been no studies to investigate whether patient neighbourhoods may be useful in predicting presentation

and routes to survival in gastroesophageal cancers. This identifies a gap in the evidence. An analysis of the rich cancer datasets available on cancer diagnoses in the UK, has a potential to predict areas with higher GOC incidence. It may even be possible to identify neighbourhoods where more people present later, or earlier in their cancer journey.

There are a wealth of studies linking neighbourhood factors with presentation and survival yet their focus has related to more commonly presenting cancers (Gentil et al., 2012, Lian et al., 2000, Lyseen et al., 2014, Zhang, et al., 2014, Aguilar et al., 2013, Ahari et al., 2013, Blakely et al., 2013a, Brocklehurst et al., 2013, Goli et al., 2013b, Ka et al., 2013, Wang et al., 2013). None of these studies have specifically focussed on GOC, despite this cancer's well established 'geographical' propensity.

Accordingly, further study to evaluate spatiality, presentation and survival with GOC in patient neighbourhoods is appropriate. A geodemographic analysis of presentation and survival in gastroesophageal cancer has the potential to reveal information on whether neighbourhood profiles can be useful to predict incidence and mechanisms of presentation. There is a potential that this study can reveal 'hotspots' where cancers may cluster, or where patients present at earlier or later stages of their disease processes. This valuable information offers healthcare providers with an intelligence to predict areas which may be targeted for intervention.

Early presentation is essential for improved outcomes in GOC, so further study which focusses on whether there are any spatial elements to incidence and presentation in this cancer is relevant. Social marketing techniques use 'neighbourhood studies' and population profiling to predict areas of demand. Using these approaches to evaluate gastroesophageal cancer may be useful to identify where GOC specialist services are

required. Thus providing intelligence to inform health promotion strategies targeted to the areas where they are needed most.

4.2 Geodemographic analysis to reveal neighbourhoods and patient attributes

Geodemographic research harnesses population demographics to capture elements of given populations, characterising them into areas of homogeneity (Norman, 2010, 2010). It segments these groups into geographical areas representing a range of attributes. Geodemographic analysis is a way of gathering data on how people live in their surroundings, how they behave and what choices they make (Norman, 2010). The Office for National Statistics in the UK undertake a census, every 10 years. This information offers a wealth of intelligence to characterise neighbourhoods, their population, lifestyle or socioeconomics, and their health and social status. The underlying presupposition of a geodemographic analysis is that people who live in the same area tend to display similar cultures, attributes and lifestyle choices.

Marketing and sales departments in businesses have been harnessing geodemographic research to provide intelligence on customer demand. This form of research has been used for many years – applying population profiling as a tool to predict consumer behaviour (Abbas et al., 2009).

This research proposes a geodemographic approach to evaluating patterns in presentation and survival in gastroesophageal cancer. There is a potential that population demographics, specific to a disease, can inform local intervention strategies. Thus targeting populations deemed most ‘at need’.

There are a wealth of studies which apply geodemographics to cancer statistics, yet to date, GOC has tended to be merged with the more common cancers such as respiratory, head and neck cancers (Sharpe et al., 2014, Richardson et al., 2006, Casseti, 2008).

Many GOC focussed studies have already revealed a spatial patterning in incidence, yet results are published at very large scales (Wu, et al., 2007, Hexi et al., 2010, Kamangar et al., 2009). The larger scales are useful to inform a geographical propensity, but they do not capture the local level incidence. Data presented at finer scales of resolution can be more clinically relevant, and far more useful to inform targeted strategies.

Given these assumptions, this thesis proposes to evaluate neighbourhoods in relation to presentation and survival of GOC patients at a more localised scale. As the first study to evaluate GOC incidence, survival and presentation in a local neighbourhood, it is proposed that this exploratory research has a potential to reveal new information. The process of analysis of these data are complex and assessment and analysis of data from a multitude of sources can be facilitated through a geographic information system.

4.3 Using Geographical Information Systems to explore cancer research data.

The Geographical Information System (GIS) is a software package – a tool for collation, analysis and presentation of data. The GIS presents these data into mapped formats, allowing the researcher to visualise, interpret and analyse data through a 'geographical' lens. There are many forms of geographic information systems, but ArcGIS (ESRI Products) is the platform of choice for this research.

Geographic information systems offer research tools to cross-link and study the complex interaction of cancer data against a range of other health, lifestyle and population information (Wu et al., 2013, Najafabadi & Pourhassan, 2011, Harris, 2016). The development of increasingly powerful computerised systems has facilitated new ways to analyse data. GIS can capture a wide array of geographical,

health and social data (Harris, 2016). The software offers a platform to layer information for systematic evaluation against other disease confounding factors. Many studies identify the potential for GIS to provide a different approach to disease monitoring so that results can inform service developments and generate new knowledge on the spatial components of those diseases (Longley et al. 2011, Lyseen et al. 2014, Cheng et al. 1996, Wu et al. 2013).

Geographical information systems are built, as data are layered into the system and pinpointed to specific components of a map (Harris, 2016). In GIS data are divided into separate thematic layers. Information in these layers is geo-registered to the same coordinate system which allows quantitative analysis of spatial relationships between features across the different layers. A GIS can reveal patterns and incidence of disease in neighborhoods, by geo-locating data onto population profiles (Dulin et al., 2010, Bazemore et al., 2010, Volkman et al., 2010, Fajans, et al., 2006, Nykiforuk & Flaman, 2009, Najafabadi & Pourhassan, 2011). The researcher using a GIS can undertake geospatial statistical analysis of diverse ranges of 'big data' (Sui, 2017). The spatial analysis algorithms implemented in a GIS can result in presentation of findings in map format (Harris, 2016, Langley et al, 2011).

The functionality with a GIS, allows the researcher to reveal potential disease patterns, or spatial or temporal distribution of diseases across differing geographic scales (Harris, 2016). A geographic information system facilitates analysis of diverse datasets, including patient, lifestyle or disease attributed factors. The spatial analysis algorithms allow for appropriate analysis of neighborhood and population attributes, to uncover relationships previously unknown (Najafabadi & Pourhassan, 2011).

The GIS also offers tools to reveal disease trends over time. Population data describing neighbourhoods may be layered onto health data in a virtual 'map' in the GIS. This map can reveal the impact of ecological, socioeconomic or lifestyle behaviours on diseases (Norman, 2010). Thus, the GIS offers a powerful technology to capture measure and refine data for epidemiological study into disease incidence and prevalence (Bazemore et al, 2010, Dulin et al., 2010, Parrott et al., 2010, Norman, 2006). GIS offers a tool for developing general insight into a variety of factors associated with communities, or disease processes (Norman, 2010).

Many studies have applied geographical information systems to inform cancer data analysis (Goli et al., 2013a, Wang et al., 2012, Hendryx et al., 2010a). Linking diseases with exposure has been at the forefront of this form of health related geographical epidemiological research for several years (Jacquez, 2004, Hendryx et al., 2010b, Beyer & Rushton, 2009, Vieira et al., 2009, Huang et al., 2008, Wang et al., 2008a, Schwartz & Hanchette, 2006, Abdalla et al., 2012, Noble et al., 2012). There are many government led drives to integrate GIS into public health and disease profiling and many data are already available in formats which can be uploaded directly to GIS format.

Revealing geographical patterns may be interesting, but establishing whether those patterns are meaningful is another objective. The GIS offers quantitative methodological tools to ensure transparency in analysis. Merely finding clusters does not necessarily uncover underlying issues potentially associated with the disease under study. Many diseases do not show patterns of equal dispersion in geographic terms. Dispersion can be dependent on an array of factors and making sense of these factors can be exceptionally complex (Lycett & Marshan, 2016).

For example, seasonal dispersion is common as many diseases are temporal in nature. Viruses may be weather dependant, as can seasonal diseases, resulting from environmental exposure (pollen for example in asthmatics). Some viruses are more prolific perhaps because of use of heating systems or air conditioners, colds and influenza tend to proliferate in winter, norovirus in summer. Thus, researchers have to capture data across a range of years so that they appropriately represent potential seasonal variations.

Several other factors can be linked to geographical dispersion of disease. Population profiles play a great part in where diseases are more likely to occur (Norman, 2013, Pourihan et al., 2010, Parrott et al., 2010). Certain genetic traits, or cultural norms, lifestyle behaviours, and the varied diets found in different communities will potentially effect incidence where disease is linked to those factors (Wang, 2013). Environmental exposures of substances attributed to the disease will be effected by where and how people live in their environments. For this reason, studies seeking to reveal clusters, ecology, spatiality or causality are cautious in their findings. A GIS offers the platform for exploration of areas against incidence, and offers insights into associated geographical factors.

In fact, studying the geography of disease incidence against populations has revealed some significant associations in recent years. For example, one study revealed a sevenfold increase in childhood leukaemia when proximity to nuclear power plants was identified (Gardner et al., 1990). Another associated coal burning power plants with increased cancers in the geographical areas of Capper Pass and Seascale in England (Baxter et al., 1996).

Though determining causation is unattainable within the boundaries of this research, looking into spatial patterning and neighbourhoods in a regional profile may reveal new insights on gastroesophageal cancer in the communities (Parrott et al., 2010, Dulin et al., 2010, Bazemore et al. 2010). A study specifically relating to gastroesophageal cancer which reveals knowledge on spatiality in incidence, presentation and survival, has the potential to provide intelligence for more targeted health intervention.

The assumption that people remain in their residential postcode for the duration of their existence is incorrect. People move, they are exposed to many environments, they work in areas beyond their postcodes and they may have been raised in areas of very different socioeconomic backgrounds than their residential postcode at the time of the disease diagnosis. Attempting to prove any causation with environment is, therefore, fraught with difficulties when patients are pinpointed to their home addresses. A significant advantage of carrying out this study in Hull/East Riding is that population movement has been very limited until very recently (ONS, 2017).

Some patients may change residency because of health-related needs, or personal circumstances and this must be considered in any spatial research. Especially one which relies on the analysis of data from an ageing population. Many GOC diagnosed patients may be in residential care homes, due to the ageing demographic of sufferers.

These considerations mean that this research proposes to illustrate any spatial patterning in incidence, presentation profiles and survival outcomes of people with GOC. Population profile data from the neighbourhoods in which they reside will be applied to evaluate confounding factors, such as population density, lifestyle factors and levels of deprivation.

4.3.1 Using maps to represent cancer data in relation to patient neighbourhoods.

In this research, a base map will provide the geographical platform on which data are layered. Patients are located to these maps through the longitudinal and latitudinal location of their postcodes within arcGIS.

As the main aim of geodemographic data is to group individuals into homogenous 'sets', it is offered at larger scales of resolution. Artificial boundaries are drawn around these areas of homogeneity and linked to the map through their geographical 'position' as pockets of attributes. Postcode data offer the pinpoints on maps, but even they are at larger scales as they capture several households within an area. Tools to georeferenced postcodes to maps are available, and can then be placed within the GIS via Northings and Eastings – the smaller more accurate destination points on a map. The GIS allows all these disparate layers to be displayed and can manage a variety of scales of resolution.

A GIS assimilates a wide range of complex data and allows the researcher to present findings in a format which is clear and very easy to read. Maps are useful to infer, to generate new hypothesis, and to consider data through a lens previously untapped. Maps from geodemographic analysis reflecting specific disease attributes have the capacity to improve and predict health service planning and interventions (Harris, 2016).

There are several issues when applying these techniques to cancer specific data. Patient confidentiality becomes problematic in diseases of lower incidence (Wang et al., 2006, Pearce, 2007, Bazemore et al., 2010, Najafabi & Pourhassan, 2010). Individuals suffering the rarer diseases may potentially be recognised by postcode representation. All data which currently exists in the public domain to identify areas

with higher GOC burden, are presented at extremely large scales of enumeration to maintain confidentiality.

For example, the Department of Health (UK) routinely publish health profiles comparing service provision across the UK (DoH, 2010), yet large scales used to present these data render them unfavourable to inform small area interventions. Most publications using maps are generalised to such an extent, they are of limited use to the local providers who deliver services to their communities.

Analysis at finer scales has a potential to pinpoint where burden is highest, yet it must be balanced against the principles of quasi identifiability. There have been many calls for finer scale explanatory studies in cancer research (Green et al., 2013, Goodman, 2010, Cassetti et al., 2008, Elliott & Savitz, 2008, Aragonés et al., 2007, Naureckas & Thomas, 2007, Bell et al., 2006). Yet these finer scale studies must account for a potential to reveal individual patients. This research must therefore, consider scales at which data are analysed and presented carefully. Any maps of findings must be clinically informative, but they must retain confidentiality.

4.3.2 Finding the right data for cancer maps.

There is a wealth of evidence on sociogeographical variability in lifestyle behaviours freely available through population census and private providers. Many studies suggest the use of population, lifestyle and behaviour data is useful for targeted health intervention strategies (Bazemore, 2010, Parrott et al., 2010, Dulin, 2010, Nykiforuk & Flaman, 2011, Pourhassan, 2011Wang et al., 2010). The evidence linking lifestyle factors such as smoking, alcohol misuse and obesity is strongly associated with higher incidence GOC (Gossage et al., 2009, Thrift et al., 2012, Coupland et al., 2010, Kollárová et al., 2012). This research proposes to review whether these attributes may

be useful to inform geodemographic classifications to predict geographical sites of higher risk.

The Office of National Statistics (ONS) hold databases with population demographics derived from census surveys held every 10 years in the UK. There is also a comprehensive cancer database registry available in the UK available from Public Health England (PHE). Merging data from these resources may reveal new insights into where these patients live, or whether there are any patterns in incidence and presentation. It could be useful in detecting whether GOC patients' neighbourhoods has any effect on their ultimate survival and presentation. A retrospective analysis of previous incidence and population data may reveal whether neighbourhood profiling could be useful in predicting areas with a higher propensity to GOC incidence.

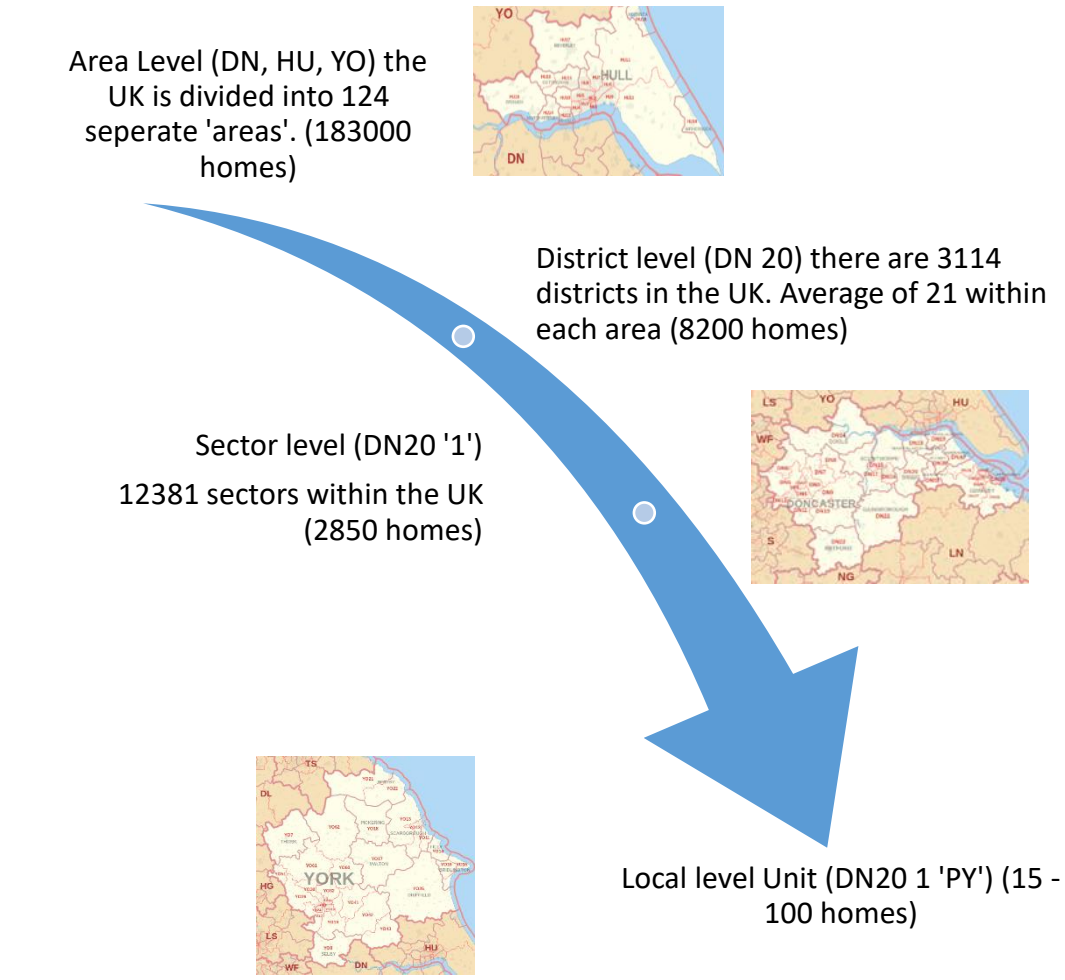
4.3.3 Identifying neighbourhoods with propensity to diseases using patient postcodes.

Presently, the UK has several datasets available on cancer statistics. Local level patient postcodes are applied to locate patients to their neighbourhoods (via the National Cancer Statistics Confidentiality Advisory Group and Public Health England). The UK postcode system attributes groups of households to a fixed point in a map. Figure 8 identifies the differences in local, sector district and area level postcodes. The most central points of a postcode are located to special reference points on maps via the very fine scale 'Eastings and Northing data'. In the GIS, a postcode 'point' on a map represents several households.

Postcodes are the UK standard coding for address location. Each postcode has several layers of geo-referenced information. There are 124 postcode areas in the UK, encompassing 22 million data points. They are stored, maintained and updated by the

Royal Mail under the Postal Services Act (2000). This research captures data from 3 areas (Doncaster - DN, Hull – HU, and York – YO) as all subjects had residential codings of YO; HU; or DN. Each area is separated into districts, then further scaled down into sectors and local units. Eastings and Northings are the most localised points, the most central point of a local level unit postcode.

Figure 8 Postcodes



Population, social and demographic statistics are based on the premise that local people share similar attributes, neighbourhoods have similar characteristics and there is a relative homogeneity of behaviours in these neighbourhoods (Harris, 2016). Labelling, grouping and identifying neighbourhoods with similar attribute profiles can be useful in developing local insights into behaviours and attributes, and may be useful in predicting needs for localised interventions. Thus, postcode based geodemographic profiling will be used in this research.

The approach has been useful in several health-related studies to date (Bazemore et al., 2010, Steevens et al., 2010, Hexi et al., 2010, Mobley, 2010, Dulin et al., 2010, Parrott et al., 2010, Cassetti et al., 2008, Wu et al., 2007, Wang, 2010, Candace, Nykiforuk & Flaman, 2009, Higgs & Gould, 2001, Najafabadi & Pourhassan, 2011).

It is proposed that merging neighbourhood profiles with cancer specific data will reveal new insights into whether population profiles have the capacity to predict incidence and presentation in GOC. By analysing distribution of incidence within a defined area, spatial patterning may be revealed (Noble, 2006, Norman, 2010, ONS, 2009). These patterns have in previous studies revealed significant clustering. One example was Openshaw's discovery of childhood leukaemia clusters in Gateshead and Sellafield (Openshaw et al, 1988).

However, identifying clusters is not necessarily conclusive. Spatial analytics is a research technique, applied to inform, or surmise shapes and patterns in data. When phenomena are located to a point in space (such as postcode points on GOC subjects) then they may be analysed for elements of spatial clustering. Embedding datasets which capture additional attributes, such as survival outcomes, or incidence, has the

potential to reveal links with lifestyle factors (Bazemore et al., 2010, Nykiforuk & Flaman, 2010, Najafabadi & Pourhassan, 2010) and this is the purpose of this study.

4.4 Geographic Information Science; using data to explore and illustrate gastroesophageal cancer.

Geographical information science focusses on how data are represented in spatial terms (Harris, 2016). Digital representation (or symbology) is significantly affected by how it is presented. Symbology describes how the researcher has chosen to present the data for visual representation on the map.

A GIS uses several ways to represent data. Points, lines and polygons can represent objects in a GIS as illustrated in Figure 9 & Figure 11 (Longley et al. 2011).

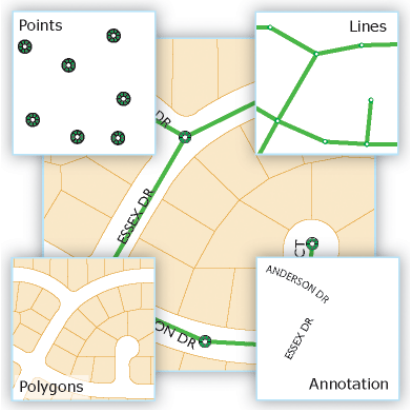


Figure 9 Layers of data in GIS (source Longley et al (2011 p. 42)

Points can be fixed to extremely accurate and small scales. In this thesis, point data are generated by patient postcode, which is linked to underlying geography through Easting and Northing coordinates. This fits with the UK National Grid and ensures a geographic accuracy to link patient postcodes and ‘fix’ them onto a base map.

Polygons are 2 dimensional shapes made of straight lines. They are closed shapes which encapsulate data from a variety of sources. In this thesis, polygons are applied to capture the geodemographic data. For example, a map can be apportioned into a variety of irregular shapes which fit together, but are determined by their contents. Each polygon has a potential to allow for 'weighting' or 'characterisation' of a set of descriptors. For example, the following picture shows a part of a map which has been apportioned into polygons to represent areas with similar population attributes. The polygons sit together within the frame of the map, but 'hold' data in each shape. Once this is done, then each polygon can be 'weighted' dependent on the attribute in question (Figure 10).

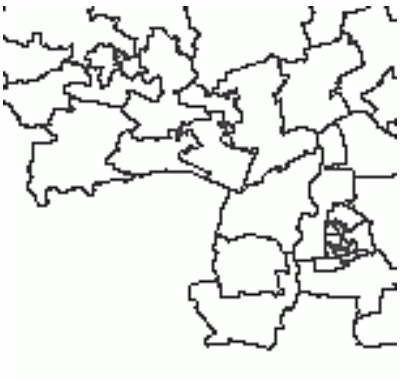


Figure 10 Polygons to frame data

For example, the map can represent degree of rurality by adding colours, or gradients to represent percentage of rurality in each polygon. The same can be undertaken for gender, population density, percentage of persons diagnosed with cancers, age groups of underlying populations and many other factors.

In this research, polygons are used to encapsulate and weight lifestyle data (those areas with higher incidences of smoking and alcohol related problems, those areas which have higher populations of older people, or male gender). In the GIS, data will be layered, 'building' a landscape to geodemographically represent GOC specific data.

Scale is exceptionally important to data analysis at which data are applied in a GIS must be consistent across all layers. The UK is split into a different range of units (scales) for geographical analysis. The scale (or size of the polygon) may be taken from several different classifications. For example, geographies can be separated into different health districts, administrative boundaries, electoral sectors, and census territorial units. This means there are many interpretations or definitions of 'areas' on a map. All research presenting data in mapped format must identify the scale at which data were extracted and applied.

This research links several spatial scales (sizes of polygons) to compare and analyse data. The smallest scales of resolution are Eastings and Northings in this research. They pinpoint data to all layers. As previously mentioned, the UK postal system uses its own polygon system. It applies postcodes to identify areas of habitation, and the longer the postcode, the more accurate they become in relation to where they situate on a map. It is essential to understand that some scales can be merged with others (when deemed 'co-terminus') but others are not.

The Office for National Statistics (ONS) use a different polygon labelling system. These are labelled as 'Output areas'. Beginning with the smallest units of output areas (OA), they merge into the coterminous Lower Super Output areas (LOSAs) then into the larger Mid Super Output Areas (MSOAs).

Health data is available at several different scales. Clinical Commissioning Groups (CCGs) are artificially constructed boundaries which define spaces for analysis, but are coterminous with national statistics offices' OA scales. Figure 11 shows the scales applied in this research, presented in size order, to represent how many people/households are encapsulated in these polygons.

-

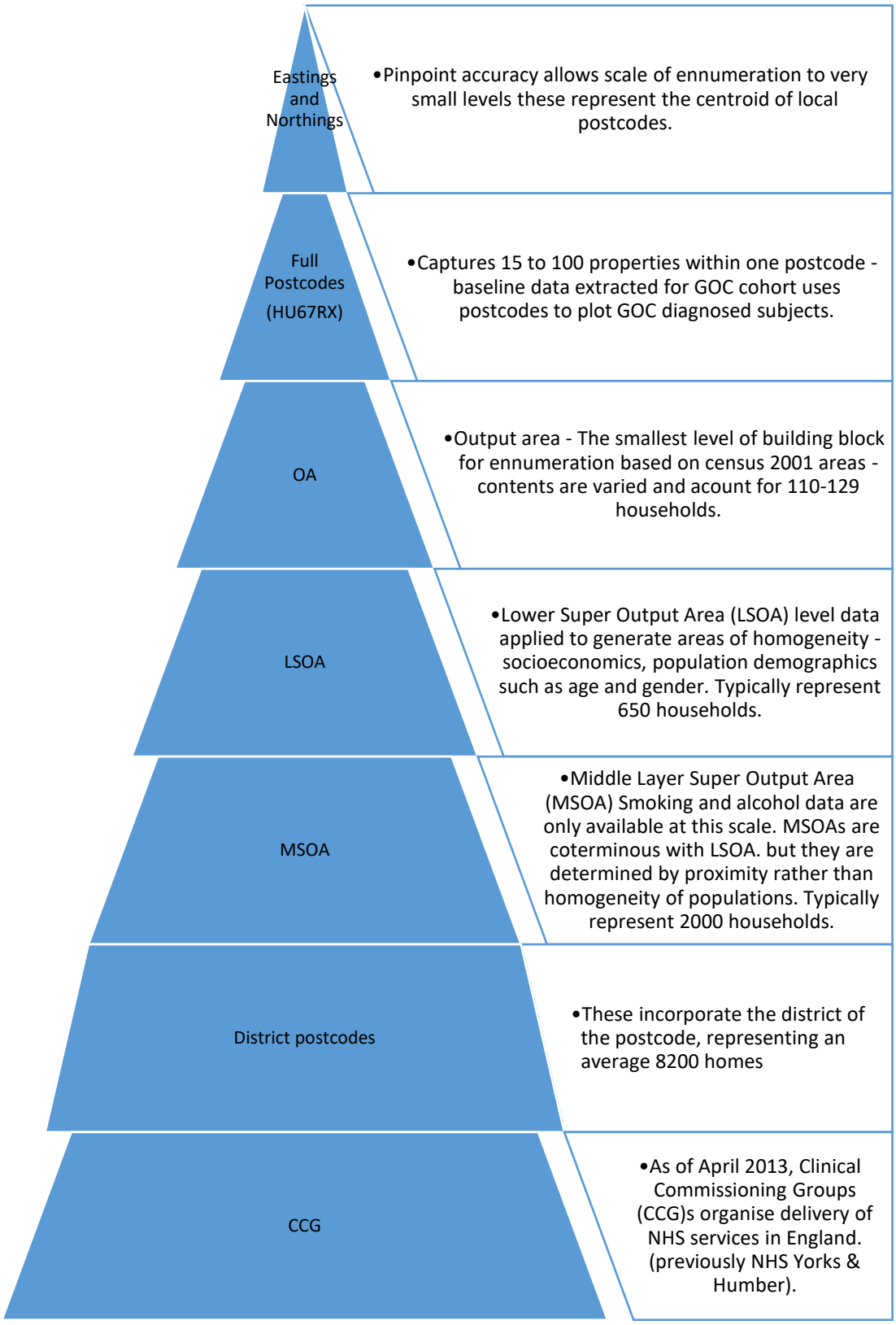


Figure 11 Scales of data applied in this thesis

4.5 Defining the catchment area for data used in this research.

In the UK, Her Majesty's (HM) treasury holds overall responsibility for clinical care and commissioning. HM treasury award the Department of Health (DoH) a budget to apportion to the National Health Service (NHS) England and CCGs. For the UK to plan services, there have been several changes to commissioning and planning of services. The Government white paper – Equality and Excellence – Liberating the NHS' (DoH, 2010) began a major drive to give GPs a commissioning authority so services would streamline to local patient needs. This white paper became law under the The Health and Social Care Act (2012) and 211 CCGs in England were developed to oversee provision and commissioning in clinical services.

Prior to this, Primary Care Organizations or Trusts (PCOs PCTs) were the preferred structures. On 31 March 2013, subsequent to the Health and Social Care Act (2012), the PCT/PCOs were abolished. The Humber and Yorkshire Coast Cancer Network incorporates several of the older PCTs, which would have been referral centers to the HYCCN in the cohort timeframe. These relate to the following areas (Table 6).

Table 6 Yorkshire and Humber NHS centres

North East Lincolnshire
East Riding of Yorkshire
North Yorkshire and York (to incorporate Scarborough and Ryedale**.**
Hull Teaching
North Lincolnshire

Given this, the catchment area defined for this project includes all patients referred to these NHS centres.

4.5.1 Looking for clusters in the geographical information system.

Inputting the ‘mesh’ of data in the GIS develops a descriptive map of underlying structures and populations in this region. This mesh produces a geographical profile of the underlying populations and structures, aligned to GOC patient postcode data. Once all data are inputted to the geographical information system, then further analysis may be undertaken. The first step of the analysis is to identify whether there are any clusters of incidence and whether there are any areas of higher or lower density GOC specific attributes. The merging of data from patients and populations means a geographic information system specifically detailing GOC relevant data is produced.

This retrospective analysis of a cohort of GOC patients will identify any potential clustering of cases against neighbourhood characteristics. It will evaluate whether there is any geographical patterning of incidences unexplained by population demographics. Then the focus will relate to whether any of these clusters may be explained by the characteristics of neighbourhoods, or the built environment.

Subjects will be positioned in the geographical information system ‘map’ through their residential postcode. Neighbourhoods around that postcode will be described in relation to the ‘usual attributes’ which are found in gastroesophageal cancer. Further analysis of potential clustering in incidence or survival patterns will be applied to assess whether there is any geographical patterning in presentation and survival which is not explained though the underlying population demographics. This exploratory study has the potential to uncover previously unknown knowledge on factors associated with GOC and highlight anything particularly relevant to advanced presentation.

The initial evaluation of incidence over the cohort will apply aspatial scan statistics, such as Openshaw's GAM (Openshaw et al, 1988). Then spatial analysis via Gettis Ord (GI*) and Kernel Density mapping. Areas will be defined in relation to their lifestyle and population attributes via K-Means Cluster analysis and these will highlight potential areas where increased incidence should occur. Comparing these analyses with actual incidence will produce a mapped display of gastroesophageal cancer diagnosis between the years 2000 and 2013, within a regional cancer referral centre in Northern UK.

These local studies could offer new knowledge which is specific to GOC, but the process of investigation could be applied to a range of disease processes. Collating data and visualising it in a different way has the potential to yield new information which has previously not been considered. This body of knowledge has a potential to provide a deeper understanding of GOC patients and their environments to further inform the 'early diagnosis' and 'intervention' initiatives.

This information and knowledge has the potential to be useful in health service planning and distribution, in patient intervention strategies and the initiation of specific community engagement with those populations with the highest need (Hardt et al. 2013). There are many studies on spatiality in cancer diagnosis which have highlighted disparities in presentation (Wan et al., 2013, Sharp et al., 2014b, Goli et al., 2013a, Yang et al., 2013, Wagner et al., 2013, Bryere et al., 2014, Goovaerts & Xiao, 2012, Goovaerts & Xiao, 2011, Xiao et al., 2011, Wang et al., 2010, Johnson, 2004, Hanchette & Schwartz, 1992, Krishnatreya et al., 2014, Li et al., 2014, Caprarelli & Fletcher, 2014, Kang et al., 2013, Congdon, 2012, Elferink et al., 2012, Mobley et al., 2012, Boulos et al., 2011). Yet there are limited studies focussing on GOC in particular. As the mainstay of GOC patients are older and male (Gockel 2006),

the disease presents a unique opportunity on which to analyse demographics against an underlying population. The multifaceted journey towards GOC diagnosis requires a range of analytic strategies to unpick, so a spatial analysis is both timely and appropriate.

4.6 The case for using geodemographics to explore population factors in gastroesophageal cancer research

Having earlier identified that later stage presentation can impact upon survival, and that the patient interval is crucial to establishing earlier presentation, this chapter identified the focus to the research. In following the calls by National Cancer Collaboratives and UK Strategies (CRUK, 2012) to reveal new information on patient intervals in cancer diagnosis, it proposes the case for geodemographics to reveal whether incidence and mechanisms of presentation may be predicted by applied and specific population demographics.

Mapping any specific GOC demographics against underlying patient neighbourhood factors in a geographical information system may reveal previously unknown factors on gastroesophageal cancer. The literature review revealed links to advancing age, male gender, lower socioeconomic grouping, rural habitation and lifestyle choices such as smoking and alcohol with GOC patients. By harnessing this intelligence and profiling communities, research can paint a picture, identifying where patients live and where healthcare intervention may be necessary. Presenting these data in mapped form can offer easily interpretable information to those clinicians responsible for developing services.

4.7 The objective of this research – developing the research question from information presented so far.

This thesis has been building and generating theories to underpin the focus of this research. These theories sit on an assumption that survival outcomes are significantly improved with earlier presentation, as GOC patients need to receive timely surgical intervention. The suggestion is to apply geodemographic profiling at small scales, to reveal whether incidence and mechanism of presentation can be predicted – and this intelligence could support more targeted health interventions.

Accordingly, the goal of this study is to reveal whether there are any small scale geographical areas which demonstrate patterns in incidence, presentation and survival of GOC and whether these areas can be predicted through population demographics and neighbourhood attributes.

The specific objectives of the study are:

1- Reveal population factors from a cohort of GOC diagnosed patients by evaluating incidence data, cohort demographics and routes to presentation.
2- Explore survival and presentation in these groups, quantify survival groups and plot where these occur in a GIS, produce specific maps illustrating these findings.
3- Generate a 'neighbourhood description of the area' in the GIS which is GOC focussed.
4- Evaluate whether presentation and survival in gastroesophageal cancer may be linked with patient neighbourhoods, by exploring patterns in data, weighting population profiles and undertaking cluster analysis techniques.

The research hypothesises that the larger scale geographic studies which have linked a geographical affiliation to GOC, may be reflected at smaller, more local scales. In exploring patterns of presentation and survival, the research may identify areas which require health intervention.

Therefore, the following research question is posed:

Are there any patterns in incidence, presentation and survival in gastroesophageal cancer patients' neighbourhoods and is there a potential for these factors to inform targeted interventions?

To answer this question, the research will identify a cohort of patients diagnosed with GOC, it will describe that cohort, then it will review how subjects presented and survived. The research will compare this with underlying population demographics and analyse whether there are any patterns in survival and presentation across the catchment area.

As the focus of the question relates to presentation and survival, then the cohort must be large enough to facilitate statistical analysis of survival and patient demographics. A retrospective cohort is proposed for several reasons. Firstly, it allows survival analytics to appropriately identify surviving patients (to a 5-year span). Secondly, it enables capture of a larger dataset to maintain anonymity. Finally, a longitudinal study of previous incidence and prevalence in an area, will be able to be compared with census data which matches that timespan.

The catchment area must be small enough to be clinically relevant, but large enough to enumerate to scales which maintain patient confidentiality. The longitudinal timeframe must be of sufficient size and span a period which is both clinically relevant and conducive to confidentiality.

This is a UK study, so the neighbourhood must be representative of a 'usual' UK catchment area. Many areas with dense populations (such as London, UK) do not necessarily capture the usual population demographics seen across the rest of the country. The UK cancer data maps identify the Hull and East Yorkshire region as having a higher prevalence of gastroesophageal cancer, but a population profile commensurate with a range of other UK catchment areas.

The following chapter discusses the methodology and processes employed to answer this research question. The chapter introduces the 'embedded case study' as a methodology to address this diverse research question – it explores some of the concepts for the analytic strategy and contextualises the approach. Principles of ethics and validity are discussed, as is the process for generating data. The following chapter underpins the research protocol a tool required in all case studies to assert validity and reliability in the research process (Yin 2014).

Chapter 5 – The Case Study approach as a methodology to explore neighbourhoods, presentation and survival in gastroesophageal cancer.

Previous chapters established the case to undertake an analysis of presentation and survival in gastroesophageal cancer (GOC), specifically targeted to the patient interval and patient neighbourhoods. Investigating spatial patterning and neighbourhoods by geodemographic profiling has been applied in several other studies to target geographical areas of need. Therefore, developing a regional profile specific to gastroesophageal cancers has a potential to reveal new insights on how patients present and how they survive in their communities.

GOC has a stark geographic affiliation which has been unexplored at local levels. The treatment necessitates medical and/or surgical intervention, so encouraging earlier presentation (therefore, diagnosis and life-saving surgical treatment) has the capacity to improve outcomes. Patients are key to their diagnostic and intervention processes, so studying patterns in presentation and survival is appropriate. A study to review whether where people live, affects how they present and survive with GOC, may provide valuable intelligence to support earlier diagnosis interventions.

This chapter provides the underpinning rationale for selection of case study as a methodology. It then introduces the five components of the research design – which are to identify the questions, consider theoretical prepositions, clarify units of analysis – explore the logic on how data links to prepositions and then to develop explicit objectives to answer the following research question:

> Are there any patterns in incidence, presentation and survival in gastroesophageal cancer patients' neighbourhoods and could these factors inform targeted interventions?'

Given the complexity of this question, case study is an appropriate methodology, as it allows multiple approaches to capture and analyse the wealth of data required to explore spatiality, survival and presentation in GOC.

5.1 Why is the case study the ideal tool to answer the research question?

Case studies offer a way to explore and evaluate phenomena through the contextual analysis of events, situations and conditions. A case study allows the systematic study of events to describe the phenomena under investigation (Bromley, 1990). The research methodology offers empirical inquiry to reveal facts about a contemporary phenomenon in its real-life context (Yin, 1989). In exploring relationships between populations, environments and habitation at the point of diagnosis, many different forms of data are required. A comprehensive assessment of patients' neighbourhoods in relation to survival and presentation requires an array of analytical processes. Yin's (2014) embedded case study approach offers an appropriate tool to manage and merge the diverse information.

Case study research can capture many forms of data which are applied to a single 'case' or phenomenon (Yin et al., 2013, Yin, 2014) a case study allows the researcher to tailor design and data collection procedures so they fit with the research questions (Meyer 2001). It is very contextual and appropriate to the exploration of phenomena in a 'real life' context (Meyer 2001). Case studies have previously been useful in exploration of the lesser understood behaviours, systems and processes (Hartley 1994) and this research aims to describe the behaviours of the neighbourhoods, the systems of healthcare service delivery in context of the cohort, and the spatial processes of presentation and survival by merging a variety of data sources.

The first objective of any case study is to establish the theory, to identify whether case study as an approach is relevant and appropriate to answer the phenomenon (Yin, 2014). Case studies may be exploratory, descriptive, interpretive or explanatory (Starke, 1995), so methodological approaches to generation, collection and analysis of data can vary significantly. Case study methodology supports use of new and evolving technology, which enables analysis of large datasets, and application of metadata to offer insight into the 'case of interest'. An exploration of survival and presentation in patient neighbourhoods requires a range of methodological approaches (quantitative analytics on patient data, epidemiology, geodemographic profiling, cluster analysis and geographical information science). These are all key factors for examination of relationships between population, health, neighbourhood and environmental characteristics and methods are drawn from previous geographical and spatial cancer and disease processes (Najafabadi & Pourhassan, 2011, Rushton et al., 2006, Richards et al., 1999).

A case study should be grounded through rigorous design (Yin, 2014). Whereas many other research approaches have explicit designs to identify the steps towards answering the original query, the case study allows the researcher to navigate their own mechanisms. This could be considered a limitation to the methodology, as a lack of set structures of analysis can offer a threat to internal and external validity (Trellis, 1997, Flyubjerg, 2006). However, this lenience in structuring offers the researcher the ability to adapt and design the study so that it most appropriately meets the original brief. So long as the researcher follows set protocols, data management techniques and identifies an explicit research protocol, the methods used to answer the emerging queries can be amended so that they can fully explore the phenomenon under investigation (Yin, 2014). The explicit case study protocol is presented in Chapter 6

for these purposes. The case study is a complex undertaking and requires a thorough identification and evaluation of concepts, phenomena, and the systems and processes which impact on the research. A case study has to be designed around the research question, and underpinned through a strict protocol which details the identification, collection and interpretation of all sources of data (Yin, 2014). The quality of the research design must demonstrate validity and reliability in establishing the research objective, managing data, evaluating findings and exploring rival theories.

5.2 Designing the case study

Yin (2014) asserts the purpose of a case study is theory development (Yin, 2013). In designing the study, the researcher begins with a set of questions and then seeks to answer those questions by designing a study with explicit objectives (Figure 12). These objectives capture the full element of the phenomenon under investigation and they should be drawn from the research design (Yin, 2013).

Any theoretical prepositions must be considered during the conduct of the research. The case can be separated into 'units of analysis' to assist a methodological approach. This allows the researcher to construct the framework of information – and to examine whether there are alternative theories to indicate why the results may manifest.

The logic which links data to the prepositions describes the way information is to be collated and analysed. In this thesis, observed patterns in data can be evaluated against expected outcomes, using the wealth of information on incidence in gastroesophageal caner for example. Finally, the evaluation criteria is made explicit by identifying objectives to answer the question. These evaluation criteria underpin validity and reliability in the research and can be used to determine whether assumptions have been addressed through the research.

Figure 12 The research design

5.3 The case study questions – 'bounding the case' and defining its parameters.

Yin (2013), a seminal author for case study methodology, identifies a requirement to specifically describe the components of a case, otherwise, it can become nebulous and lead to a lack of focus to the study. He cites this as 'bounding the case' (Yin, 2013). Bounding facilitates breadth and depth of study and provides a basis for interpretation of results.

Case study questions form a basis to describe the breadth and depth of the study:

- Are there any GOC patient specific factors which relate to presentation and survival with the disease?
- Are there any geographical patterns in GOC incidence, survival and presentation?
- Could a neighbourhood profile depicting these factors inform targeted interventions?

Determining how to answer these questions is complex as there are many options and foci. An approach is required which fulfils the research objective, and bounds the researcher to present only the most relevant factors. The case relates to the geography of presentation and survival in GOC. Thus, there is a need to establish a geographical 'catchment' area, negotiate appropriate scales of resolution for analysis and consider a relevant temporal profile.

Initial chapters identified the study should focus on GOC only, and this is defined in terms of an ICD (2010) classification. Diagnosis must be confirmed through histological confirmation and patients should be alive at data capture. There is a requirement to clarify presentation groups to be considered during analysis. Socioeconomic deprivation scales will be determined as 'best fit' with patient data,

but they span several temporal iterations. An analysis of confounding variables in survival will be required. In this case, morphology, site of tumour and mechanism of presentation. A 'sub cohort' will be required to analyse survival and allow for censoring to a 5-year outcome. Equally, in cases where data are missing, (for example, those with no TNM staging and those with no data on mechanism of presentation), further sub cohorts will need to be developed.

The timeframe chosen for the purposes of this research needs to be both clinically relevant and allow anonymity of subjects. Geographical boundaries can be determined through the specialist catchment area within the health authority. Previous chapters identified the problems in presenting data to scales which fulfil confidentiality, but remain clinically relevant. With this in mind, it is proposed that an analysis of patient events at two levels of enumeration will occur so the research can present data in mapped format, at a resolution relevant to study, but without violating the requirements to maintain confidentiality. Finally, spatial boundaries are determined by how spaces are defined and managed. This is important in cluster analysis techniques and neighbourhood profiling.

5.4 Establishing the preposition – how does where we live effect how we navigate health?

Landscape has a critical role in how people live and how they make sense of who they are and how they form an identity. Landscape as a concept has social and ideological components. It captures the whole body of lived experiences of the world (Smith & Gazin-Schwartz, 2008). Case study methodology allows the researcher to provide a description of the landscape of GOC patients. Using GIS as a platform, it can capture the essence of 'place' in relation to patient and community level cancer presentations and demographics of the underlying populations. Hood (1996) identifies cultural

landscape as ‘all aspects of culturally defined space’. He identifies landscape as both a physical and a social construct and this highlights the fact that perceptions and definitions of these terms are essentially contestable (Hood, 1996).

Analysing the landscape of a health-related factor is not a new concept. In 400 BC, Hippocrates proposed how the environment has the potential to shape health in his treatise ‘airs, waters and places’. Yet through all his publications, Hippocrates failed to yield a causative mechanism, and so his approach was the subject of major academic debate. This remains true today, it is generally accepted that the health of individuals is shaped through physical, environmental and social conditions, however, how these conditions directly relate to health status across the lifespan is still subject of significant discourse in the literature.

Since early times, health has been linked with habitation and behaviours. Thackrah (1795-1833) led reforms to improve health related conditions and survival outcomes during the industrial revolution, using population descriptors to target areas of need. Edwin Chadwick (1800-1890) reformed poor laws and embarked on several ways to improve health and lifestyle, underpinning an environmental justice system which is still apparent today. The physician John Snow, discovered the source of the cholera epidemic in 19th Century London. He applied cluster analysis techniques to identify and measure outbreaks, pinpointing these outbreaks to find the source of the epidemic. Using geographical methodology in relation to health, he was able to pinpoint areas for intervention and prevent further outbreaks.

This history supports how geography and epidemiological disease patterning can be a key factor in disease control and prevention. Even Florence Nightingale applied geography to health in her study linking unsanitary conditions with deaths in the

Crimea. Geography and health are explicitly interlinked as people are a part of their neighbourhoods. There is a significant body of evidence linking GOC with specific geographical regions across the globe. It is still not established whether this is evident at smaller scales of resolution. Therefore, this exploratory case study has the potential to reveal new insights into how neighbourhood factors could be useful to predict areas where GOC presentations tend to be later, or where incidences are higher or lower.

An embedded case study approach (Yin, 2004) allows for the analysis of multiple units of analysis, so information is collated with a view to answering the research question. It enables a range of methodological approaches to explore the research question. In this thesis, they arise from the disciplines of geography, geodemographics and epidemiology. Answering the research question also requires descriptive and survival analysis to reveal attributes associated with GOC. All these methodologies are underpinned by a philosophy built on facts, empirical observation and logic, hence quantitative research processes. This thesis presents only data which are relevant to a spatial analysis of neighbourhoods, and for that reason, it will not require subject interviews. The rationale for avoiding patient interviews, is to maintain focus on geodemographics and offer analysis at sufficient breadth and depth within this quantitative framework. A proposal for this work was presented to a UK (Northern region) multidisciplinary expert forum of gastroenterology experts (Clinicians, surgeons, consultants, nurses and GPs). They agreed that a study to evaluate geographical patterns in diagnosis and presentation was both timely and appropriate. Capturing qualitative data would be exceptionally time consuming. For the purposes of this research, it was felt that patients' neighbourhoods were adequately captured through place of habitation (post code address at time of diagnosis). As an exploratory

study, it was felt that qualitative information would have potentially removed the focus from a geographical and spatial approach.

Therefore, to articulate and fully investigate elements of spatiality in presentation and survival in gastroesophageal cancer, quantitative methodology was felt to be the most appropriate research to underpin this study.

5.5 Identifying Units of Analysis

The units of analysis in this research relate to the study population and neighbourhood characterisation. There are several data sources to inform the study and develop a 'picture' to describe survival and presentation across the geographical catchment area. The study population, merged with population and lifestyle information across the geographical boundary will be applied to explore areas displaying higher or lower than anticipated incidence and/or presentation. All data will be triangulated in a geographical information system to present a profile of the neighbourhood for comparison against the cohort incidence data.

5.5.1 The study population

Adequately defining the population is a significant factor. Bhopal (2014) suggests that the quality of any epidemiological research is judged through its contributions to its goals, with the researchers having to prove a systematic analysis of data has taken place to draw assumptions on the population under study. This study will draw data from all GOC diagnosed patients presenting to a single regional referral centre in the UK.

To capture all 'survivors' fully for the purpose of this research, a retrospective cohort is necessary. For the purposes of this research, 'survivors' are determined as those people who remained alive 5 years after diagnosis. This means that data are

retrospective to a five-year period, but that data are obtained at the time of study. For this reason, the data requested include all cases presenting 2000-2013, (accessed in 2016). All subjects will have received histologically confirmed diagnosis of GOC conforming with International Classification of Diseases (ICD) (IARC 2010). Their criteria for diagnosis will fall under categories C15.0 to C16.0 (as described in table 1, chapter 2). All subjects will have received care, or treatment within the regional referral centre and postcoded addresses will fall within the geographical catchment area.

Subject's postcodes will provide Easting and Northing point data for input to the GIS. These data will include age, date of birth, sex, date and mechanism of presentation [where this exists]. The survival (in days) will be calculated using SPSS V20 'date and time wizard', survival will be calculated as the number of days between the date of histological confirmation to the date of death (where this has occurred). This research results in a descriptive surveillance GIS on incidence, and presentation of GOC within a regional referral centre spanning 2000-2013.

Alongside postcode point data, enumerated population profiling data will also be used to provide underpinning regional profiling. Estimated populations will be taken from census 2011 at lower super output (LSOA) level. The reason for choosing these data is that they draw from both census 2001 and 2011 and this most closely represents the time period on which the cohort is drawn. Another advantage is that ONS publish data relating to socio demographics on the 2011 census, at LSOA levels. These data contain LSOA level profiles of age, sex and population weighting. They also include behavioural characteristics such as smoking and drinking and dietary habits in local communities (although at different scales of enumeration).

5.5.2 Characterising neighbourhoods

There is a wide and varied discourse related to ‘place’ in research. ‘Place’ may be defined through geographical fixed points, or may be socially determined and described though its characteristics. The ONS use output areas, to determine places of socioeconomic homogeneity, they aggregate ‘places’ using polygons based on postcodes, to determine geographical ‘areas’ considered similar. Subsequently, place can be assigned a definition which is directly proportional to its surrounding areas. The GIS allows a quantitative assessment of place; it cannot fulfil the complexities of defining ‘place’ as a qualitative perspective would. However, the GIS enables the layering of data which captures many facets of ‘place’. It allows data to be layered so that it can form a ‘lattice’ of information regarding ‘place’, so that analysis may be undertaken on a reasonably comprehensive quantitative definition.

The geographical boundaries for the GIS are determined through subject’s postcodes. Base map data will be uploaded from the Environmental Systems Research Institute (ESRI) and this will include postcode point data, as well as LSOA enumerated geodemographic information on the region.

There are several issues relating to enumeration and small area health statistics, namely, that populations are altered by small numbers and that assumptions may be drawn where there are none to be made (Elliott & Savitz, 2008, Tannenbaum et al., 2014, Wang et al., 2012, Gregorio et al., 2005). However, recent studies using methodological techniques of ‘smoothing’ and point scale prediction can assist in reducing this risk of bias in small area enumeration.

Space is a multidimensional concept and one requiring significant consideration. For space to be interpreted appropriately, the researcher must clearly articulate how

‘space’ is defined in the research. For this research, point level data are taken at patient full postcode level, apportioned to Easterlings and Northings via the geoconvert application, which links postcodes to census data. Point data may then be linked in with higher geographic scales, so that comparisons may be drawn from demographics versus incidence.

One would expect to see larger numbers of GOC in densely populated areas (the larger the population, the more people there are to diagnose). Spatial structuring in GIS, allows the researcher to account for different densities, however, this is a complex issue. For the purposes of this research, values of populations are aggregated to postcode levels, and direct standardised incidence rates drawn from epidemiological methodology. This is so that they can truly represent actual cases of GOC on a per 100000 patient year estimation. This is important, because scales used to define the ‘space’ where later stage presentation is occurring, need to be considered carefully.

For example, to use Euclidian distance in measuring access to healthcare, is unacceptable as it may not capture the complexities of how a patient actually gets to the healthcare centres (Wang et al., 2010). Patients will use road and transport networks, they will use public or private transport, and travel times will vary dependent on vehicles and modes used. Merely measuring how far away a person is from the regional centre will not capture all these complexities, but it does offer at least an insight. There is a wealth of geographical literature which identifies ways access can be measured, in terms of service provision and the case study protocol presented in the following chapter clarifies all the methods which will be applied to reveal whether there are any spatial disparities in later stage presenting GOC.

5.5.3 – Applying tools to profile the GOC cohort, and characterise the region.

This research proposes several approaches to triangulate patient and population profiling data within the GIS. There are several steps required to build a dataset within the GIS and to present questions. Figure 13illustrates how the GIS will be built.

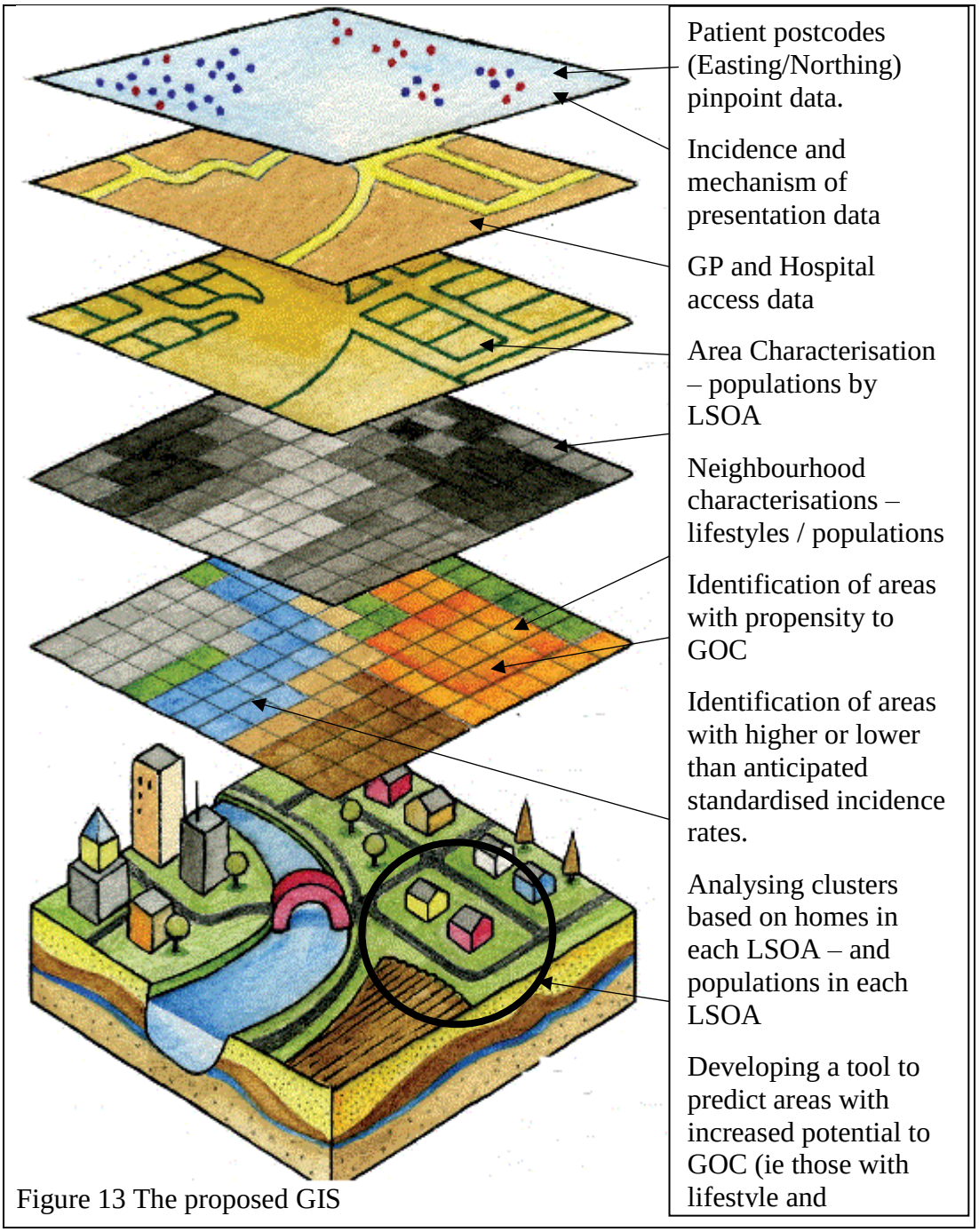


Figure 13 The proposed GIS

Data on lifestyles, populations and GOC patients will be layered into the GIS and several tools will be applied to facilitate further analysis. All descriptive analysis of the GOC cohort will be undertaken in SPSS. GOC specific profiling will identify the overall cohort characteristics. This will include subject demographics (gender, socioeconomics, age ranges, GOC specific treatments and mechanisms of presentation). SPSS will then be used to assess survival in the cohort, detailing how certain characteristics or treatments impact survival outcomes in subjects. These statistics will be applied to derive 'presentation groups' which will determine how many subjects survived and for how long. Subject demographics will reveal whether confounding factors (such as increasing age, tumour morphology, socioeconomic status and surgical interventions) had an impact on survival.

These will all be saved as layers within the GIS (pinpointed into the GIS via patient postcodes). The GIS will then enable enumeration of patient data to a range of scales within the resulting maps.

Information on neighbourhoods and lifestyles will be derived through the Office of National Statistics and a range of health tables. These will be applied to the GIS at LSOA levels, to characterise areas displaying a higher or lower propensity to GOC.

EU Standardised Incidence rates will be compared at LSOA levels on a 'per 100000' capita population rate, with actual incidence data, calculated to LSOA populations. An historical map of actual incidence rates will be presented to identify areas where incidence rates were higher between the years 2000-2013. This provides a crude estimation of previous incidence, normalised to the underlying population.

To explore clustering in incidence and presentation, several tools will be used. Assessing how many people were diagnosed within the catchment area and identifying

potential clusters will be undertaken through Openshaw's Geographical analysis machine, a tool to predict and reveal any areas of clustering based on the number of homes per LSOA. A more refined, Kernel density map will then evaluate this further, based on population statistics at LSOA levels.

The final unit of analysis relates to the generation of a GOC specific prediction tool, which can then be compared against these final three cluster analysis approaches (The Standardised Incidence map, the GAM and the Kernel density map). The GOC specific prediction tool is based on key criteria, weighted to population attributes in GOC and applied to the underlying population characteristics across the area.

These weightings provide a quintile classification of neighbourhoods with a propensity to higher, or lower incidence in GOC. By comparing this tool with the three maps depicting an historical incidence, inferences may be drawn as to whether the tool could be useful to inform future strategies in encouraging earlier diagnosis, by targeting the areas with the highest potential for incidence.

5.5.4 Capturing data and establishing the retrospective cohort timeframe for this research

Census data

This relates to the years in which subjects presented with GOC to the referral centre. Ideally, the population demographics should truly represent each year subjects were diagnosed. However, this is not possible. OCNS statistics are limited to decile census survey data, so specific data techniques will be used to account for the years where population data were unavailable (for example, in the years between census surveys). It is usual in epidemiological studies, to consider the actual (direct) population as the denominator when analysing cases in a region (Bhopal, 2008). Where these are

unavailable, inter censual population counts will be derived from the census surveys of 2001 and 2011 – using aggregate statistics from ONS via casweb and infuse (ONS, 2015, ONS 2011). Using these population estimates has been shown as effective in previous studies (Elliott, 2009, Norman, 2010). Additionally, as this research is a surveillance study, aimed towards highlighting spatial elements of incidence and presentation, it is felt that using census derived population estimates will not have a significant impact on this case study. However, it should be noted as a potential limitation to the research.

Patient data

Capturing all live patients who presented to a regional referral centre with a histological diagnosis of GOC is possible through the national cancer strategies and databases held. Data are captured from 2000-2013. This 13-year duration enables a sub cohort for survival analysis spanning the years 2000 – 2011 to allow a 5-year 'survival' cut point. It also allows incidence data to be enumerated to finer scales. The decision to seek such long temporal data capture was based on the requirement to manage the quasi-identifiable nature of postcodes and patient data. Confidentiality is a significant factor in geographical research where there are limited events. Any patients who were diagnosed at autopsy were excluded, as these data would affect the survival statistics for quantification of advanced presentation.

This case study encapsulates a regional cancer referral centre covering a geographic region determined through trust hospital referrals. The region covers a population with a range of community attributes which offer a similar profile to many other UK regions (in terms of socioeconomic status, health outcomes and demographic profiling). Capturing all cases through NCIN data provides a sound cohort for analysis and

histological confirmation of diagnosis means all cases are valid for purposes of analysis.

5.6 The logic of evaluating neighbourhoods in relation to incidence, presentation and survival in gastroesophageal cancer patients

There is a socio-geographical variability in health behaviours already established in much of the literature in health geography. The defined case can be assessed against population characteristics, to assess the potential predictive mechanism of a population profile. In comparing the 'spaces' or geography of risk to the population at risk, factors can be identified which have potential to have an impact on accessibility of services for all. 'Spaces' can be defined through neighbourhood boundaries, artificially apportioned through administrative definitions. Alternatively, spaces can be considered against demographic and social profiles, with boundaries determined through homogeneity and proximity to health centres. The case study allows a range of approaches to data manipulation, so that geographical research methodology may be used to rigorously evaluate whether differences in incidence exist which cannot be explained by known risks.

In GOC for example, older males, who are smokers with a history of high alcohol intake, and who come from lower socioeconomic backgrounds have a higher propensity to this cancer. A spatial analysis of GOC at local level has the potential to uncover any factors unexplained through patient demographics. Specific features mapped to populations, can be analysed against disease events, to uncover any spatially relevant factors associated with advanced presenting GOC. This enables an exploration of the interplay between resources and how they are experienced, used and perceived at local levels.

The context also relates to time, and health services and care provision improves with time, as services expand, knowledge of effectiveness of interventions increase and treatments improve. Temporal analysis of incidence will be undertaken to review any health service related factors which may be evident in the case.

5.7 The criteria for interpretation of findings

The initial chapters in this thesis presented the case for further evaluation of patient location data in relation to incidence, presentation and survival in GOC. The substantive preposition that GOC patients' environment may play a part in how they present and how they survive. Further study is therefore, necessary to reveal any potential patterning which is unexplained by population demographics.

The research hypothesises that exploring patterns of presentation and survival may identify areas which require health intervention. The most appropriate way to explore presentation and survival is by looking at historical data and reviewing whether patterns could have been predicted through population demographics. A retrospective cohort analysis allows the researcher to merge census and attribute data with patient survival and presentation alongside neighbourhood profiling so that it captures factors intrinsic to the cohort during the defined timescale.

Given the relatively low incidence of GOC, a longer-term retrospective study is necessary. Adding years to longitudinal studies will increase incidence numbers and provide sufficient scales to maintain anonymity. A larger cohort will allow for analysis at finer scales of resolution, so that results are more clinically relevant. However, the dates chosen to delineate the longitudinal study must also reflect current treatments and outcomes. Any subjects presenting before the year 2000 for example, would not adequately reflect the many advances in cancer care and presentation strategies. The

other consideration is that Public Health England (PHE) can only release complete cancer incidence datasets three years following collection, so the most current incidence data would be three years old at capture.

Previous chapters identified a lack of parity in how presentation is defined in gastroesophageal cancer. Given this, there is a need to define the essentially contested concepts such as 'later or advanced' staged presentation. It is proposed that retrospective incidence data which includes demographic information on when the patient presented, and when they died, can be investigated through quantitative survival analysis techniques. The results can then inform the remaining research and delineate specific presentation groups to quantify survival.

To reveal any potential patterning, patient postcodes may be used to 'pinpoint' patients onto a map (Figure 13) so that they can be evaluated against underlying demographic data on areas in that map. Geodemographic profiling of areas may then be compared with actual incidence, to reveal any potential patterns unexplained through population structures (Lewandowsky et al., 1995). Other forms of cluster analysis can provide a crude observation of whether there are any patterns or clusters in incidence (Norman, 2010, Lewandowsky et al., 1995). Merely plotting incidence data (with patient postcodes) to a geographical area, can reveal spatial patterning, clustering and events. To undertake all these forms of analyses, use of statistical analysis packages (SPSS), geographical information systems (GIS), and Geographical analysis machines (GAM) are required. These packages provide platforms to analyse the large datasets and an additional lens from which to review and compare findings.

5.7.1 Analytical strategy – how findings will be presented in this research

Geographical information systems offer a way to present the data, as well as provide a framework for analysis (Longley et al. 2011). They have been hailed as policy decision support systems (Norman, 2010, Longley, 2012) as they allow analysis, hypothesis generation and representation of phenomena which may be spatially distributed. Presentation plays a significant factor in interpretation and analysis, so clear principles taken from both disciplines need to be considered for the purposes of this study. The GIS allow layering of metadata, to produce a visual display, which can then be further analysed alongside the other contexts of the case study.

There is a potential for bias in these visual displays, so cartography and data management techniques will be explicitly stated in the thesis. Visually displaying any data must account for the issues associated with confidentiality and ethics is discussed later in this chapter. The process of visual representation must also be carefully considered (Beyer & Rushton, 2009, Beyer et al., 2010, Brewer, 2006, Parrott et al., 2007, Aronson et al., 2007, Frye, 2001). Generating maps which effectively portray the required information is a methodological process in itself (Frye, 2001, Collins et al., 1998, Lewandowsky et al., 1995), as use of particular colours, fonts and 'fills' in data presentation can present issues with interpretation during the publication process (Frye, 2001).

5.7.2 Survival analysis

To analyse survival, a model depicting duration of time to a given event is required. In this cohort study, the event for analysis is death, modelled against the survival duration. However, there are many confounding variables associated with survival outcome in GOC patients. Age, comorbidities, gender, race and certain lifestyle factors have been linked to survival outcomes (Hiom, 2015, Alimoghaddam, 2014, Bus, 2014;

Wikman, 2014, Hong, 2013, Liu,Shu-Zheng, 2013, Maringe, 2013, Rametta, 2013, McLoughlin, 2013, Wu, 2013, Coupland, 2012). The longitudinal frame of this dataset means that new and emerging technology, enhanced clinical informatics, improvements in services and referral processes have a potential for those subjects with more recent diagnoses, potentially had improved survival prospects. These factors informed the choice of survival analysis methodology.

The assumption that all subjects had the same chances of survival throughout the study is violated by new national cancer strategies, by new and improved clinical treatments and perhaps even by an increase in public awareness, because populations react to drives in public cancer informatics (DoH, 2014, Lyratzopoulos et al, 2012). Many subjects receive clinical interventions which have a potential to improve survival, so survival analysis through Cox regression will reveal whether demographics or treatment have impacted survival outcomes. Overall cohort survival, stratified by age, gender and socioeconomic grouping will be undertaken using Cox proportional hazards indices.

5.7.3 Geodemographics

Geodemographics is an approach to map patterns of social and behavioural data to generate homogenous landscapes, or 'places'. It is 'the art and science of analysing socioeconomic and behavioural data about people in the context of space (Grubesic, Miller & Murray, 2014). The different approaches in geodemographics provide a profile of different communities in relation to single or multiple attributes, characterising them in accordance with underlying population descriptors (Abbas, 2009). The methods applied to characterise and profile space and place in context of this GOC cohort are used in this thesis and explained further in the study protocol chapter.

These methods offer a tool to answer the theories generated in the initial chapters. The information in these chapters revealed the multifactorial pathogenesis of GOC and the common attributes associated with the disease. These attributes guided the collation of population data. Therefore, population data relating to smoking and alcohol use, socioeconomic groups, age and gender stratified density could inform the population profiling and potentially identify areas which may have higher incidence GOC.

An artificially clustered dataset, through a K-Means characterisation technique will be presented and this is based on identification of GOC specific attributes, measured against population attributes. This characterisation can then be assessed against the pre-existing standardised incidence ratio map, and the two clustering techniques (homes and populations), to reveal whether the area characterisation has a potential to reveal areas of more intense incidence.

Identifying whether there are any localised clusters is based on the fact that there is already a geographic propensity to the disease as demonstrated by Globocan's (2008) incidence data and team Tsinguahua International geographical analysis (Tsinguahua 2015).

There is a wealth of research applying geodemographics to cancers (Guajardo, 2009, McEntee, 2008, Mechili, 2014, Sharp, 2014, Tannenbaum, 2014, Blakely, 2013, Buas, 2013, Lian, 2000, Chong, 2013, Goli, 2013, Ka, 2013, Sakai, 2013, Wan, 2013, Elebead, 2012, Lian, 2012, Marzieh , 2012, Mobley, 2012, Roshandel, 2012, Wang, 2012, Bailony, 2011, Joslin, 2011, Christian, 2011, Goovaerts, 2011, Beale, 2010, Hendryx, 2010, Huang, 2010). Many of these apply population demographics, or environmental factors to incidence so that they can produce mapped displays of factors with potential links to the cancers. However, very few are individually cancer specific.

Many of them publish data at large scales of enumeration, so that they maintain patient confidentiality where incidences are low. All this research applies principles of geodemographics through an array of geographical information systems, to map results.

There are issues with applying incidence data at lower scales. As the level of enumeration decreases, so does incidence. To manage smaller scale analysis and conform to requirements to confidentiality, many papers have either expanded their cohort timeframes (spatiotemporal amendments), or enumerated data to larger scales. Others have linked many cancers together, citing common factors as rationale for the choice. For example, (Downing, 2008, Sharp, 2014, Blakely, 2013, Igissinov, 2011, Mahaki, 2011, Congdon, 2012) link a variety of cancers with known causative lifestyle factors. This means they can merge incidence data and produce results at smaller scales of enumeration.

Some studies have specifically targeted GOC (Aguilar, 2013, Ahari, 2013, Silva, 2013, 2013, Levi, 2013, Wang, 2013). In linking socioeconomic, environmental, genetic and dietary data, they have compared incidence rates. However, the relative rarity of this cancer means levels of enumeration in these studies are very high (reported at province levels for example). To overcome the concern of patient identification, this research adopts spatiotemporal aggregation - a 13-year timeframe for the cohort. PHE guidance suggests a minimum Sweeney's K anonymity factor of 5 should be considered prior to any data publishing (Sweeney, 2002). This will be applied to all maps presented within this thesis.

5.8 Units of Analysis and the criteria for interpretation of results. Assessing the quality and validity of this case study.

Any case study must be judged by its ability to maintain quality standards through the research process. Four tests are commonly applied to appraise these standards and these are detailed below with explanations on how they are managed effectively within this research (Figure 14)

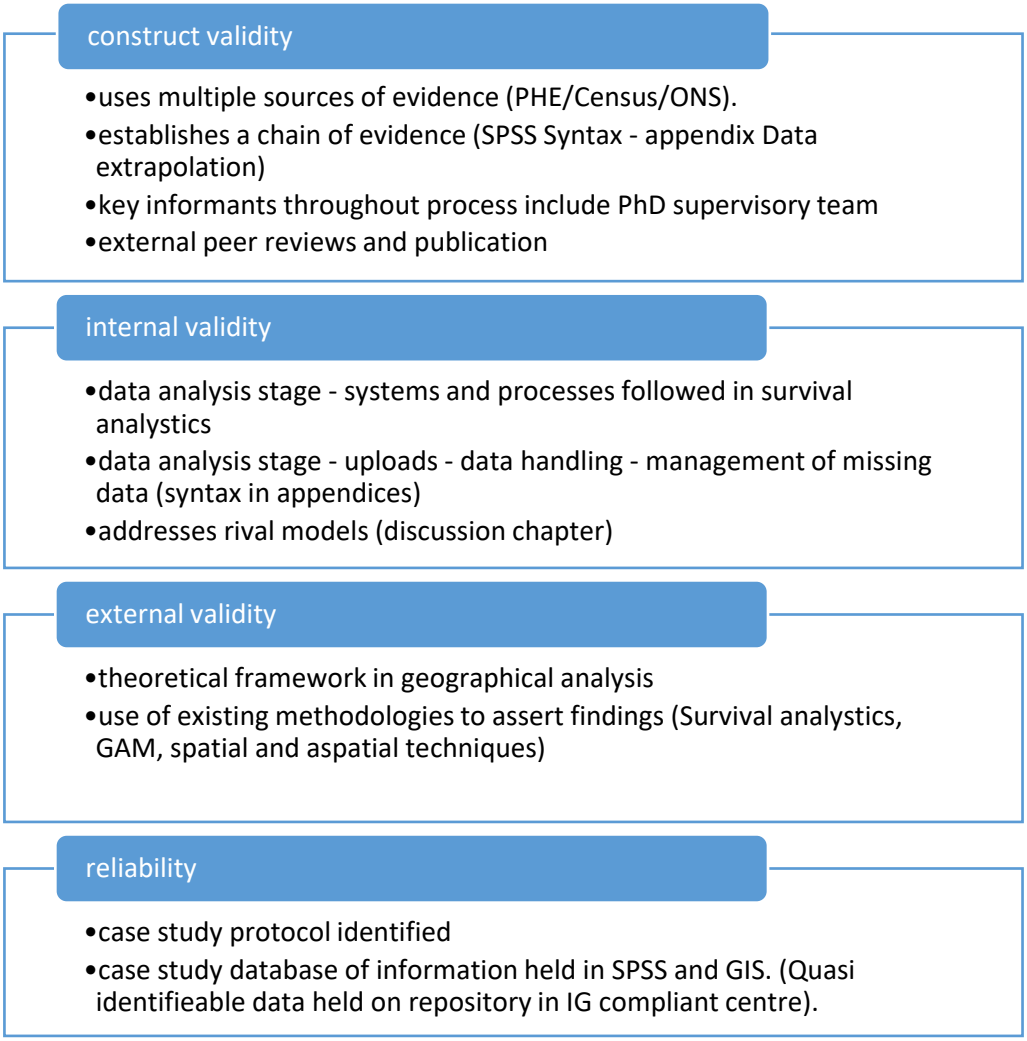


Figure 14 Quality; Validity; Reliability

5.8.1 Presenting the findings

Findings of this research will be presented against 4 explicit objectives as depicted in the study design.

Table 7 developing the objectives for the case study

UOA	Criteria for interpreting findings	Objectives drawn from UOA and criteria.
Data from a longitudinal cohort of GOC patients in a UK regional referral centre will reveal attributes in relation to survival and presentation.	Evaluate attributes of survival and presentation against existing literature	Identify the most relevant population factors Evaluate survival and quantify presentation groups
Population and lifestyle information will be layered into a geographical information system to explore areas displaying higher or lower than anticipated incidence and/or presentation across the geographical catchment area.	Use a range of spatial and aspatial cluster analysis techniques to explore geographical patterns.	Produce a neighbourhood description of the geographical catchment area in relation to lifestyle and demographic factors
Findings triangulated to profile the neighbourhood - so this may be compared to the cohort incidence data.	Present findings using the case study units of analysis.	Triangulate findings from previous objectives, develop a tool displaying areas with higher potential for GOC based on population attributes. Compare this with historical data.

5.8.2 Exploring rival explanations

One of the main threats to this research lies in whether the theoretical paradigm corresponds to the observation. Studying postcode data to identify whether there are any spatially relevant factors in the patient stage of diagnosis does have limitations.

There will be no attempts to establish causation in this research. It is merely an exploratory case study, intended to reveal potential spatial elements in presentation and survival. Cancer has a prolonged disease process, so to attempt to apportion cause and effect related to ecological issues is not the objective in this study. This is the reason for a case study approach, seeking to perform an empirical enquiry as to whether there may be community level factors which can predict areas of need. This research presents a retrospective analysis of a cohort in a regional referral centre, to identify whether population manifests predict areas of higher density diagnosis, and therefore areas of risk. It will not seek to apportion cause and effect related to postcode.

Another threat lies with inferring that nature of individuals may be deduced by apportioning them to groups, based on the area in which they reside. In grouping individuals to postcode areas, or lower output areas, the researcher must form assumptions based on those geographical areas. These assumptions are weighted according to the general 'average' and will not capture all the outliers. This means that however geographical areas are delineated or defined, they can only capture the essence of 'most' likely or most probable descriptors of people captured in the geographical catchment area.

To reduce the potential to ecological fallacy, a longitudinal approach is proposed. Longitudinal approaches provide larger datasets. For example, when considering incidence of GOC within a small area, one could assume that numbers of presentations would increase over time. Capturing 13 years of data provides far larger cohort numbers than capturing just 1 year of data. This research acknowledges the presence of ecological fallacy, as with any other spatial research aiming to make inferences based on geography or habitation.

Larger cohort numbers are very important when analysing incidence at smaller scales of enumeration. The more subjects captured across the scales of enumeration, the less likely it is to be able to identify individuals. Capturing descriptive data on higher numbers of inhabitants who live within smaller scales of resolution means inferential statistics tend to be more rigorous and many researchers adopt the longitudinal approach to facilitate larger datasets (Goovaerts & Xiao, 2011, Luo, 2012, Choi et al., 2010, Hewapathirana & Wijayarathna, 2010, Bell et al., 2012, Lovett et al., 2014, Lawson, 2013). There are techniques to analyse data at many different scales of enumeration and any geodemographic literature confirms that the higher the level of enumeration, the less any assumptions can be made regarding areas under investigation. What this means is that the closer the lens, the more detail can be seen, the higher the resolution, the richer the data can be. Patients are more than the products of their postcodes and this research accepts this as a theoretical preposition. Yet many studies have shown postcodes as reasonable predictors of social status, lifestyle factors and area economics (Aguilar et al., 2013, Bryere et al., 2014, Danesh et al., 1999, Sharp et al., 2014a).

Another threat is the use of secondary and tertiary data. This study relies on datasets from a range of sources, all of which boast high confidence intervals (>95%) for the data they produce. Public Health England has an established rigorous process to record all cancer data. The Office for National Statistics is a recognised centre of excellence in statistical processing of data and census data are considered rigorous. Application of statistical techniques to the dataset to cover periods of time where data are missing, will add further rigour to the case study. However, some of the lifestyle data (smoking and alcohol, air quality and survey data) are captured at larger scales of enumeration,

from tertiary sources. To reduce any potential effects, data have been aggregated from several surveys.

This research is novel in the application of a range of research techniques to uncover any spatial phenomenon associated with GOC. If these techniques yield sufficient output, then the tool may be used to evaluate further cases. Replication logic will not necessarily be in the findings of this research, but the techniques applied to the dataset and generation of the geographical information system will be. The community profile can be used against a range of alternative cancers previously associated with factors such as obesity and increased alcohol intake. The rationale for choosing only one area, is that it still captures a population covering 1.8 million people, yet the processes involved can provide an in-depth analysis.

A data management plan is presented in the appendix and the case study protocol outlines how data are to be analysed in the study. This protocol also identifies how data were captured, and how any inferences are developed and how outliers are dealt with during the data processing phase.

5.9 Ethical implications.

There are several ethical considerations in this research. These relate to managing patient confidentiality, research integrity and data management.

5.9.1 Confidentiality

All research subjects have a right to anonymity. The researcher holds a common law duty of confidence to all subjects in this research as the main risk is through patient identification. Under section 33 of the Data Protection Act (1998), data must be processed so they are not likely to cause substantial damage or distress to subjects or

their families. This Act identifies 8 principles of data management and these are detailed in Table 8.

Table 8 DPA principles

1	Information must be fairly processed
2	Information must be specified for purpose (and limited)
3	Must be adequate and relevant
4	Must be accurate and up to date
5	Not kept for longer than necessary
6	Respect for individual rights
7	Secure and valid
8	Not transferrable to countries outside EU

There is a requirement to manage quasi identifiable data and ensure there are no individuals identified within this research. Published maps displaying presentation stage 'hotspots' will aggregate findings to scales to ensure no individual subjects were identifiable. Survival and cluster analysis processes will also aggregate data to a non-identifiable range prior to any publication or dissemination of this research.

There is no requirement for patient consent to release quasi-identifiable data. This is because the following legislation allows time-limited disclosure of identifiable patient information, without patient consent if it is used for medical research:

- Statistics and Registration Services Act (2007)
- Section 60 of the Health and Social Care Act (2001)

- Article 8 EU convention of Human Rights duty of confidentiality of information
- Section 251 of the NHS Act 2006, and subsequent Regulations

Although obtaining consent was not required for the purposes of this research, there is a professional obligation to consider a potential for identifiability. The possibility that living subjects may be unhappy to share their diagnosis, by being an identifiable pinpoint on a map was considered. With this in mind, Sweeney's K statistic (Sweeney, 1999) will be applied to all data considered quasi identifiable that areas with limited incidence cannot be used to identify individual patients. This data does not refer to any persons by name. However, as postcode level data are being used, these will be aggregated to enable K anonymity factor > 5 (Sweeney 1999). The temporal range of data are 2000 – 2013, so reflects mainly GOC of deceased subjects. However, in areas where there have been very small numbers of cases, data will be aggregated to compensate for potential identification. The researcher will act to maintain honesty and integrity in the manipulation, storage and collection of all data. Any dissemination of results will ensure there are no quasi identifiable variables in published format.

5.9.2 Data integrity and ethics approvals

For this study, research, proposals have been submitted to the National Health Service Integrated Research Information System (NHS IRAS), and the University of Hull as sponsor, in addition to the University of Leeds as data repository centre.

Appropriate agreements for data use have been sought and granted by the following agencies: The University of Edinburgh online data service (EDiNA); ArcGIS mapping data from the UK ESRI; The UK Office for National Statistics (ONS) and the Public Health England Confidentiality Advisory Group (PHE CAG). All sources have been

identified in line with documented licence terms of use, export compliance, data attributions and privacy policy.

Subject to National Health Service (NHS) data requirements, the project proposal was submitted via the Integrated Research Information System (Appendix 2). The data repository centre is overseen by the Confidentiality Advisory Group. This group assesses the information governance (IG) status of the host institution and releases data subject to PHE, IRAS and NHS approval. CAG only release data to Information Governance compliant institutions and individuals. This means the researcher had to complete the Medical Research Council (MRC) research data compliance program.

During the processes of data release, the research centre had to be amended. Data were held at the Leeds Institute for Data Analysis, (an IG compliant centre). This meant that a further NHS Ethics application and notice of substantial amendment was required. Another Confidentiality Advisory Group application was sought, and data release was granted following substantial discussions. Ethical approval was granted for this project by the sponsoring university (Appendix 3). The IRAS project ID was granted in Sept 2015 (ID 161434 – submission number 15/YH/0318). This project requires management and processing of quasi identifiable patient data (postcodes, dates of diagnosis, death and birth), so consultations with the NHS confidentiality advisory group (CAG) and Public Health England (ODR1516-063), were undertaken. These ensured the researcher could source the required patient information and that the study was compliant with NHS information governance and data management principles. This project has been available for consultation on line at the Health Research Authority and can be accessed via the Health Research Authority website. (Appendix 5).

A research contract was also obtained by the researcher, to facilitate research at the LIDA site. The researcher successfully completed modules on the Medical Research Council data management program. The processes of data cleaning and manipulation was overseen by a team of researchers in the university. CAG released data in July 2016 having been assured of governance procedures.

Data must be processed and managed in line with mandates through the Data Protection Act (1998). To reduce the possibility of identification, a data management plan is presented in Appendix 1. Any publications resulting from this research have to follow strict de identification procedures. Data were stored in Information Governance compliant password encoded repositories. These all conform to information governance standards as decreed by the Confidentiality Advisory Group (CAG) and Public Health England (PHE). These standards are informed by the following legislation, governmental papers and codes of conduct:

- The Data Protection Act 1998
- British (International) Standard ISO 27001
- The Caldicott Report 1997
- Information: To share or not to share? The Information Governance Review March 2013
- The Freedom of Information Act 2000
- Section 251 of the NHS Act 2006 (originally enacted under Section 60 of the Health and Social Care Act 2001)
- Confidentiality: NHS Code of Practice 2003
- NHS Records Management Code of Practice (Part 1, 2006 & Part 2, 2009)
- Health and Social Care Act 2012
- The NHS Information Security Management Code of Practice 2007
- The Computer Misuse Act 1990
- The Electronic Communications Act 2000

- The Regulation of Investigatory Powers Act 2000
- The Copyright, Designs and Patents Act 1988
- The Human Rights Act 1998
- The NHS Care Record Guarantee 2011 (Version 5)
- The Social Care Record Guarantee 2009
- Anonymization Standard for Publishing Health and Social Care Data
- Anonymization: managing data protection risk code of practice (ICO 2012)
- A guide to confidentiality in health and social care (HSCIC 2013)
- Code of Practice on confidential information

5.9.3 Other considerations in relation to data and results of this study

There is a possibility that this cancer may uncover areas of higher than anticipated advanced presentation and this may reflect on services in the health authorities. All researchers will maintain strict levels of confidentiality in these instances, ensuring services are made aware of potential environmental, social or healthcare priorities in GOC management.

To maintain independence and impartiality, all researchers will work within the boundaries of their professional bodies. The supervisory team will act to maintain data integrity and uphold ethical principles. The researcher will use the supervisory team to engage in academic debate and maintain and update the data management protocol. The university, in their capacity as research sponsor, will provide the required research training and supervision, as well as necessary data storage and insurances in accordance with their role.

Summary of the chapter

This chapter has explored case study as a methodology to capture all elements of the research. As case study can be extremely complex, the chapter introduced the study

design, presenting the need to refine the study to a single centre, retrospective cohort analysis of patients in a fixed geographical area.

Ethics of consent, confidentiality and data management were discussed, but it must be noted that this section is supplemented by a wealth of information presented through ethics approval committees and NHS centres. The next chapter presents the case study protocol, which is used to operationalise the research and offers prescriptive detail how the research methods will answer the question.

Chapter 6 The Case Study Protocol – how this research explored neighbourhood factors, presentation and survival in gastroesophageal cancer

This chapter presents the case study protocol to detail how the research question is answered throughout this thesis. Case study protocols must detail procedures and general rules to be followed through the process of the research (Yin 1993, Yin 1989, Yin 2014). The chapter summarises previous chapters to provide an 'overview – or background' to the case study which details how the theories have been generated to develop the research question. It then presents the research objectives and describes rationales for data sources chosen so that the reader can clearly follow systems and processes involved in the research. Many case study authors suggest that the case study protocol forms the 'backbone' of case study research. The protocol validates the systematic processes taken to answer the original research question.

Following the suggestions by Yin (2014), the main elements of a research protocol for case study are to provide an overview of the study, to identify data collection procedures, propose data collection questions against research objectives, and introduce how results are to be reported. This chapter acts as a fundamental component of the total thesis, as it embeds the research methodologies employed in the case study, to reveal whether there are any geographical patterns in gastroesophageal cancer incidence, presentation or survival unexplained by population demographics.

6.1 Background – How the research approached patterns in incidence presentation and survival in patient neighbourhoods.

Patients with GOC have improved outcomes when they present early enough in their disease process to facilitate surgical removal of the primary tumour (Smithers, 2010, Tachibana et al, 2006). The patient interval was deemed crucial to encouraging earlier presentation and diagnosis, and previous chapters called for further study into patient factors. Gastroesophageal cancer's geographic affiliation was previously un-

investigated at local levels. GOC patient groups displayed some very similar attributes, offering demographic profiles to support a more localised geodemographic analysis.

Geodemographics offered useful tools to predict ‘high demand’ areas (they are widely applied to a range of social marketing strategies). Therefore, this research harnessed GOC’s geographic propensity and proceeded to investigate whether local areas of need could be identified through these techniques.

The goal for this research was to reveal whether there were any small scale geographical areas demonstrating patterns in incidence, presentation and survival of patients with GOC. A tool was developed to characterise areas by population attributes affiliated with GOC, and this was compared with actual incidence from a retrospective cohort of patients presenting to a regional referral centre.

Despite the wealth of information on the use and applicability of geodemographic profiling in predicting areas of need, there were no GOC specific studies published to small (therefore, clinically relevant) geographical scales.

This research sat on the theoretical preposition that the spatial distribution of incidence should be reflected in population profiles which reflect the characteristics of GOC. By exploring factors on incidence, survival and presentation in gastroesophageal cancer, the research identified intrinsic population and lifestyle characteristics associated with the cancer.

These were then mapped into a geographical Information system to characterise the regional referral area – specifically to GOC. Evidence linked certain attributes with an increased propensity for diagnosis. These included advancing age, male gender, lower socioeconomic grouping and lifestyle choices such as smoking and alcohol. These attributes were mapped at community levels using the GIS and then were compared

with actual incidence. The aim was to offer a novel intelligence to predict the most appropriate geographical areas for interventional healthcare. To do this, the research objective was to explore any patterns in incidence, presentation and survival in gastroesophageal cancer patients' neighbourhoods and to identify any potential for these factors to inform targeted interventions.

The research question called for a number of processes.

1) Identifying a cohort which offered incidence rates to support the need for geographic confidentiality (spatiotemporal parameters).
2) Description of the cohort attributes to support demographic profiling and compare with the literature
3) Identification of mechanisms of patient presentation
4) Analysis of survival and revelation of any factors impacting survival
5) Definition and grouping of presentation and the stages at which patients presented for treatment.
6) Geographic description of the catchment area for the regional referral centre
7) Identification of the most appropriate scales of resolution to present data
8) Presentation of observed versus expected incidence (using EU standardised incidence rates)
9) Cluster analysis to reveal aspatial and spatial clustering across the geography.
10) Characterising the area using population demographics related to GOC.
11) Finally, comparison of the characterisation tool, with actual incidence across the area, to reveal whether small scale areas may be targeted for healthcare interventions which encourage an earlier patient presentation, therefore, diagnosis.

The question also required several methodological approaches, drawing on survival analysis, demographic profiling, geodemographic analysis and population characterisation. To capture all these elements, the case study approach offered an overarching methodology to capture all these elements in one research study.

A retrospective analysis of incidence, presentation and survival in cohort of cancer patients presenting to one regional referral centre offered data spanning clinically relevant timescales, and allowed a smaller scale geographic analysis with the consideration of confidentiality.

6.2 Developing the objectives of the study

Case study methodology required the development of specific objectives to answer the original query. The approach was taken to collate and integrate a range of methodological approaches and these are explained later in this protocol. The objectives underpinning this case study on presentation and survival of gastroesophageal cancer in local neighbourhoods were:

1) Identify the most relevant population factors previously linked with GOC.
    - Demographic analysis of a cohort of GOC patients to reveal factors associated with incidence (age, gender, cancer morphology, site, and socioeconomic grouping).
    - Evaluation of more recently collated data on routes to presentation, to clarify whether advanced stage presentation is contributable to reduced survival in this cohort.
2) Evaluate survival in GOC and quantify ‘advanced presentation’.
    - Determine a ‘survival cohort’ with days censored to manage groups and account for a 5-year cut off point.

- Apply Aarhus et al (2012) standardised timescale framework for early cancer research as ‘date of first histological confirmation’ as a marker for diagnostic date, then define days survival. (days survival between date of diagnosis and date of death). This can be used to determine ‘presentation groups’ within the cohort, based on survival time.
- Analyse all subjects with complete TNM data to determine whether this calculation is effective to denote the developed presentation groupings.
- Quantitative evaluation of the developed ‘presentation groups’ against factors previously linked with GOC incidence (age, gender, cancer morphology and site, routes to diagnosis and socioeconomic groups).
- Undertake a temporal analysis to determine homogeneity of incidence across the longitudinal timeframe.

3) Produce a neighbourhood description of the geographical catchment area, in relation to lifestyle and demographic factors.
    - Define the catchment area in the NHS cancer network boundaries to represent the cohort.
    - Identify overall population figures by PCT, gather population figures at LSOA and MSOA levels to evaluate which scale of enumeration fulfils data confidentiality and supports a clinically relevant scale.
    - Apply age standardised EU incidence data against cohort data to reveal ‘observed versus expected’ incidence across the region (MSOA level).
4) Triangulate findings from objectives 1-3 to reveal any geographical sites displaying clusters in incidence or survival outcomes.

- An initial crude cluster analysis to reveal areas with higher or lower incidence through Openshaw's Geographical analysis machine (nearest neighbour analysis of incidence per LSOA, against the number of households in the region). This normalises incidence data by the number of houses per LSOA, but is not smoothed to actual populations.
- Reveal population smoothed incidence with kernel density mapping, so that incidence can be assessed against underlying populations. Aggregate data to incidence per LSOA over the longitudinal study to reveal potential clusters in geographical areas.
- Characterise the catchment area to reflect incidence – Use findings from objective (1) to support a tool to evaluate the 'population potential' of areas with higher or lower GOC incidence density.
- Compare the findings from this tool, to actual retrospective cohort data, to reveal its effectiveness and relevance to future healthcare planning strategies.

6.3 Methods and data collection processes for this case study on presentation and survival in gastroesophageal cancer

An embedded case study design (Figure 15) was applied to identify the case, its boundaries, its context and the data sources. GOC cancer data was triangulated and compared with existing neighbourhood population and lifestyle demographics in a geographical information system.

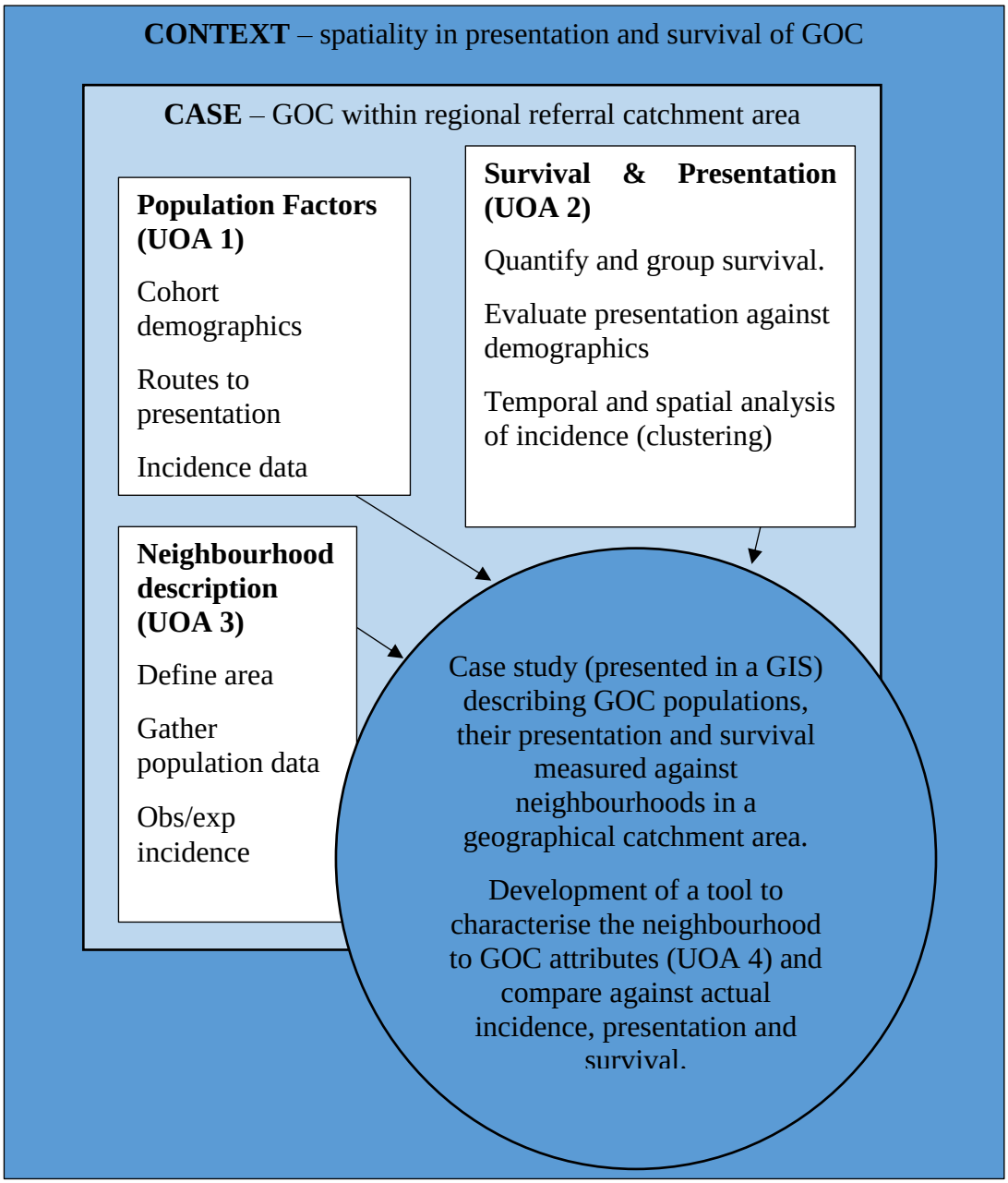


Figure 15 The embedded case study

This case offered a study of spatiality, presentation and survival in gastroesophageal cancer and developed a tool to characterise the neighbourhood to investigate a potential for population demographics as a way to predict areas in need of health intervention.

The case was bound by both geography (the catchment area) and time (longitudinal cohort). The context was to explore patient neighbourhoods and whether they could

inform clinical intervention. Each patient in the cohort was 'mapped' to a neighbourhood through the residential postcode which was identified at the time of histological confirmation of GOC diagnosis. Analysis of clusters in incidence offered maps for comparison against the characterisation tool.

A 13-year cohort of GOC diagnosed patients from a single UK regional referral centre was chosen as it offered a clinically relevant spatiotemporal snapshot for analysis against GOC relevant population profiles. Data from both Public Health England (PHE) and the Office of National Statistics (ONS) census was used to develop a geographical information system (GIS) for further geographical analysis.

This GIS was used to produce mapped analysis of data relating to GOC and population profiles across the referral area's geography. However, a range of data were required to support this. In case study methodology, these are labelled as 'Units of Analysis'.

6.4 Data retrieval, management and analysis for each of the Units of analysis

The four units of analysis relate to GOC population factors, survival and presentation, neighbourhood profiles and descriptions, and the development and analysis of the characterisation tool. Data collection, analysis and methods applied to each unit of analysis is described within this section.

6.4.1 UOA (1) – GOC population factors.

These data were retrieved from Public Health England (PHE) to capture all cases of histologically confirmed GOC presenting to a regional referral centre between the dates 01/01/2000 and 31/12/2013. Table 9 identifies all data obtained from PHE and lists the rationale for the request.

Table 9 Data retrieval and rationale to support data for objectives (1 &2)

Data	Field description	Rationale and OBJECTIVE
Full patient postcode	Postcode of patient residence at tumour diagnosis	Full postcode provided to pinpoint patients into a map on GIS. This will be used in geographic information system to identify incidence against O/A demographics as denominators Postcode lookup tables applied to pinpoint to Eastings and Northings within GIS.
5-year age band	5-year age band of patient at histological confirmation	To enable demographic analysis and age banded survival analysis of the cohort. 5-year banding reduced quasi-identifiability (1)
Gender	Gender of patient at diagnosis (M/F)	Gastroesophageal cancers are male dominant - this skews data, so gender is necessary. Gender may change over a lifetime, so gender at time of diagnosis is most relevant to this study. (1)
Tumour Count	Count of every tumour assigned to this Patient ID in range C00-97 excl C44	A specific request - all patients presenting between Jan 2000- DEC 2013 with clinically confirmed diagnosis of ICD10 coded C15.2 - C16.0 to the regional referral centre (Queens Medical Centre – Hull).
Vital Status Death Date (the best)	Vital Status of patient (Dead / Alive) Date of death of the patient	Death date used to calculate days between date of histological confirmation and date of death (where this has occurred). Underpins survival analysis. (1 & 2)
Death date flag	A flag set to inform if any part of the death date has been imputed	Increase validity of death date (2)
Post Mortem	Indicates whether a post-mortem took place	All patients diagnosed at PM removed from survival analysis cohort. (2)

Tumour ID for every patient	Unique tumour ID in ENCORE	To identify all C15 and C16.0 presenting in the timeframe and whether there were other tumours present at diagnosis. (1)
Unique identifier per patient	De identified patient number	Non NHS – Unique identifier for subjects (data management).
Date of Diagnosis	Histological confirmation of diagnosis (date)	Date of presentation to date of diagnosis will provide information to underpin the survival analysis and stage at presentation analysis. (2)
Year of Diagnosis	Year of diagnosis	This provides an overview of dates in years, for spatiotemporal profiling of the community. (assists in geodemographic profiling analysis and temporal analysis) (1)
Basis for diagnosis	Basis of diagnosis of the tumour, according to all data received by the registry. The definition provided conforms with the international requirements specified by the European Network of Cancer Registries (ENCR)	To confirm whether diagnoses were made on histological confirmation, at PM or on clinical examination. (2)
Site of cancer	Site of the cancer, in the coding system that the tumour was originally coded in Description of the code in SITE_CODED	To provide further detail in survival analysis (1 & 2).
Tumour morphology	Morphology of the cancer	To identify tumour type (code and full text versions) (1)
TNM stage (where available)	Pathological stage at diagnosis	TNM stage at diagnosis is essential for descriptive analysis of the cohort and to underpin survival analysis. (1)

LSOA/ MSOA geographical area	2001 Lower Super Output Area of patient residence at time of diagnosis	For linkage of data between ONS demographic and patient postcodes
Geography – strategic clinical network name and code Strategic cancer network code	Name of the Strategic Clinical Network the patient was resident in when the tumour was diagnosed Cancer network code the patient was resident in when the tumour was diagnosed	This is a cohort analysis - so data relating to area of health authority will underpin the subjects' demographic analysis The cohort is limited to the Yorkshire and Humber Strategic Clinical Network
IMD data	All IMDs offer an overall measure of multiple deprivation experienced by people living in an area. LSOAs in England have been allocated to deprivation categories using the Index of Multiple Deprivation. 2004, 2007 and 2010 iterations applied.	The cohort timeline is 2000-2013 – therefore, the correct IMD classification over this timeframe requested. (1)
Events during treatment	Description of treatments and events	Allows for censoring of survival analysis for events such as chemo/surgery/radiotherapy (1, 2)
Date of diagnosis	Diagnosis date of the patient, as defined by the UK ACR	Date of histological confirmed diagnosis Survival analysis depends on date of diagnosis to date of death. This forms the basic component of this study, whereby survival grouping is calculated (see above) (1, 2)

Data on 2785 subjects were retrieved from Public Health England in June 1986. They were released in Excel® format. Appendix 6 lists the SPSS syntax applied to group variables for further analysis. There are several sub-cohorts within the subject groups to facilitate analysis. These are:

(1) The total N – all subjects presenting to the regional centre with histologically diagnosed GOC between the dates 1st January 2000 and 31st December 2013. (N 2785)
(2) Routes to diagnosis cohort – all of the aforementioned subjects who presented after 2006 with recorded 'route to diagnosis' data. (N 1097)
(3) Survival cohort –all subjects who presented between the dates 01;01;2000 – 08;06;2011 (N 2215).
(4) TNM cohort - those subjects from the survival cohort who had recorded and complete TNM staging data (N 121).

Initial SPSS analysis revealed some missing data, but these related to TNM staging and presentation routes. All demographic information was available for the full dataset. The survival cohort was developed from Around 45% of subjects within the total cohort had very sparse and/or incomplete staging data, the rest had no staging data recorded at all. Complete datasets depicting the full 'TNM stage' (graded at 1-4), was only present in 121/2215 subjects in the survival cohort.

Descriptive analysis was undertaken to reveal cohort attributes, such as age, gender, surgical intervention, cancer morphology and site of tumour. Iterations of IMD were compared so that the most relevant was chosen for further analysis. IMD status was compared to age and gender profiles within the total cohort. This revealed any

associations between socioeconomic status and GOC diagnoses, by gender. Mechanisms of presentation were analysed by IMD status, and tumour types.

The following table collates all information gleaned from major literature sources on population attributable fractions (PAF) in GOC causation. PAFs provide a quantitative appraisal of how different exposures and lifestyle behaviours impact cancer diagnosis. They are calculated by multiplying the proportion of the population exposed to the risk factors, by the relative risks which have been previously associated with that risk factor. GOC has multiple risk factors, and adding all factors together would potentially overestimate the attributable proportions, resulting in very high estimations of risk. Instead, risk was apportioned based on the studies highlighted in Table 10.

Table 10 Risk factors associated with GOC

Study	Linked attribute	Results Presented as odds ratios (LR studies) PAFs	Weighting Positive cause (+ low++ mid +++ high) Negative cause (0 low/-00Mid/000 High)
Kollarova et al. (2012) All histological subtypes explored via Logistic regression modelling (n88 GOC vs n200 control group).	Age over 65 Male Gender Smoking Overweight Alcohol Increased veg intake	(O/R via LR) Ratio 6:1 2-3.5 fold increase 2-3 fold increase 3-5 fold increase 000	++ ++ +++ ++ ++ 000
Parkin, Boyd & Walker (2011) All cancers – but results for GOC separately	Smoking Overweight Alcohol X-ray exposure Veg intake	65.5 21.7 20.6 2.7 46.1	+++ ++ ++ + 000
IARC list of classifications by cancer sites AGENTS CLASSIFIED BY THE *IARC MONOGRAPHS*, VOLUMES 1–118 [accessed 2017]	Smoking (tobacco/non tobacco and betel quid) Alcohol X-ray exposure	World Health Organisation monographs on cancer (responsible for listing agents specific to cancer sites)	+++ ++ +
WCRF American Institute for Cancer Research Oesophageal cancer [accessed 2017]	Smoking Overweight Alcohol Mate drink HPV Processed meat Vegetables Physical activity	Review 46 studies across the world on attributable risks for GOC – last updated Feb 2015	+++ +++ ++ ++ ++ + 00 00

As these studies contain a range of recent meta-analyses on PAFs associated with GOC, they were felt to be relevant to geodemographic 'weighting' of areas to support Objective 4.

6.4.2 UOA (2) – evaluating survival and presentation across the catchment area.

Data to underpin this unit of analysis was taken from the GOC survival cohort and EU age standardised incidence data (CRUK, 2014). The smaller cohort (post 2016) depicting mechanisms of presentation was also analysed against the survival data.

The 'survival' cohort (N = 2215) was developed so the study could account for the timescale required for a 5-year cut point to demarcate 'survival'. Survival status is apportioned to those patients remaining alive at 5 years post histological conformation. This 'survival' cohort contains only those subjects who were diagnosed before 08.06.2011 (n = 2215). The rationale for this is that the full dataset was released on 08.06.2016 and those diagnosed after this date could not be censored to the 5-year interval. The 5-year timeframe was calculated as 1826 days or 365 days X 5 + 1 day for leap year. The variable for days between histological confirmation and date of death was developed using the date and time wizard in SPSS. Drawing from previous calls for specific time frames to inform cancer research studies (Aarhus, et al, 2012), survival time was recorded as the number of days between histological confirmation of GOC and the date of death.

Analysis of survival revealed that over 60% of subjects had died within 6 months of diagnosis, and this degree of mortality is not adequately described through a 1-year survival statistic. It was felt that a quintile group was required to adequately capture and describe 'survival' across the cohort. Percentile analysis on censored variables was undertaken to support this grouping.

The presentation groups offered a crude measure of 'presentation' in the absence of actual staged data. Names apportioned to these groups were based on the physiological presupposition that an earlier presentation allows surgical intervention and removal of the tumour to occur thus, potentially, improving survival outcomes. Later staged presentation results in poorer outcomes, because surgical removal of the tumour becomes increasingly complex with progressive tumour growth and infiltration of surrounding structures. Thus, those subjects who died earlier (within 6 months) were labelled as extremely advanced presenters, those dying within a year, as advanced presenters. The subjects dying between 1 and 5 years were considered as 'treatment facilitated' and this captured the clinical supposition that subjects had received clinical input with either palliative or curative intent.

A smaller, 'routes to diagnosis cohort' (N = 1097) was developed for this research as PHE have only been collating data on routes to diagnosis since that date. The 'routes to diagnosis sub cohort is therefore taken on all subjects presenting to the regional referral centre between the years 2006-2013 as statistics pre-2006 were unavailable.

Once these survival groups had been developed, parametric and nonparametric methods were applied to reveal between group differences. Cohen's (1988) criteria was applied post-hoc to reveal effect sizes – (0.01 = small, 0.06 = moderate and 0.14 = large effect).

Methods for testing independent samples included the one way repeated measures ANNOVA – applied to compare scores relating to age and gender and their effects on survival. Kruskal-Wallis was applied to nonparametric data to reveal any associations between groups (IMD and survival, mechanisms of presentation against survival). Cox regression curves were developed to identify effects of several variables on survival

outcomes, censored for date of death. These curves illustrated effects relating to IMD status, cancer morphology and site of cancer.

A final analysis of data relating to subjects who did have complete TNM status recorded (N121) was presented and analysed against mean survival in days.

6.4.3 UOA (3) – Describing and clarifying the regional referral centre – capturing data to inform and describe the neighbourhood.

This required relevant data to describe the geographic catchment area, its population and to measure the observed versus expected incidence rates across populations. Data on lifestyles and populations within the geographical catchment area were apportioned to a base map in the GIS. Based on geographical location of patient postcodes, data on all subjects with GOC residing in the following LSOAs between the years 2000 and 2013 were extracted from PHE's Office of Data Release in June 2016.

- East Riding of Yorkshire
- Hull Teaching hospitals
- Lincolnshire teaching hospitals
- NE Lincolnshire
- North Lincolnshire, and
- North Yorkshire and York.

The catchment area was geographically defined through a variety of iterations on how health service delivery was organised over the cohort timeframe (as boundaries were amended to 'Primary Care Trusts' PCTs and then Clinical Commissioning Groups' CCGs). For the purposes of this research, the area was defined using the Humber and Coastal region profile – the boundary determined through the Cancer network and established by the regional referral centre's catchment.

CASWEB census level data with lifestyle profiles for communities at LSOA levels was used to provide age and gender profiled population data from Census 2001. The UK data service provided data on smoking, alcohol and lifestyles and this was retrieved from census surveys and household datasets. These profiles offered a basis for calculating density of homes and populations and acted as comparator against age standardised EU cancer incidence data. They also provided underpinning information for determining cumulative incidence rates across MSOA and LSOAs.

Base maps were appropriated from the University of Edinburgh (ESRI) 'digimap data services' in order to generate maps and link data through Eastings and Northing pinpoints. All data sources were acknowledged and referenced in relation to this research.

The most relevant geography applied to ascertain boundaries was taken from the primary care trust (PCT) catchment areas. The strategic cancer network was Humber and Yorkshire coast, which is broken down into Primary Care Trusts – these include Hull Teaching hospitals, North Lincolnshire, North Yorkshire and York, East Riding of Yorkshire and North and North East Lincolnshire.

To represent the GOC cohort with the date range 2000-2013, and obtain relevant lifestyle data, the most relevant census output was from the 2001 census. These provide the appropriately scaled data on population statistics by age and gender. CASWEB and InFuse portals were used to access these datasets (UK Data Service Census Support - Casweb, 2016; UK Data Service Census Support - InFuse, 2016). Expected standardised GOC incidence rates will be taken from Cancer Research UK figures and apportioned to the geographical output area to calculate an observed versus an expected rate for each local area (CRUK, 2014).

Smoking and Alcohol related information was collated from several tertiary sources. Local area profiles on tobacco and alcohol provided baseline data. Where more than one research methodology had generated data, results were enumerated so that they most represented the GOC cohort. Public Health England data was used to characterise areas on alcohol and smoking across the region. The LAPE statistical tables identify alcohol related hospital admissions in England based on primary and secondary diagnosis, weighted to partially attributable conditions. These tables are based on data capture spanning the timescale 2008 - 2015. Figure 16 shows how areas fall higher, or lower than UK average within these tables.

Figure 16 District level LAPE data compared to UK average

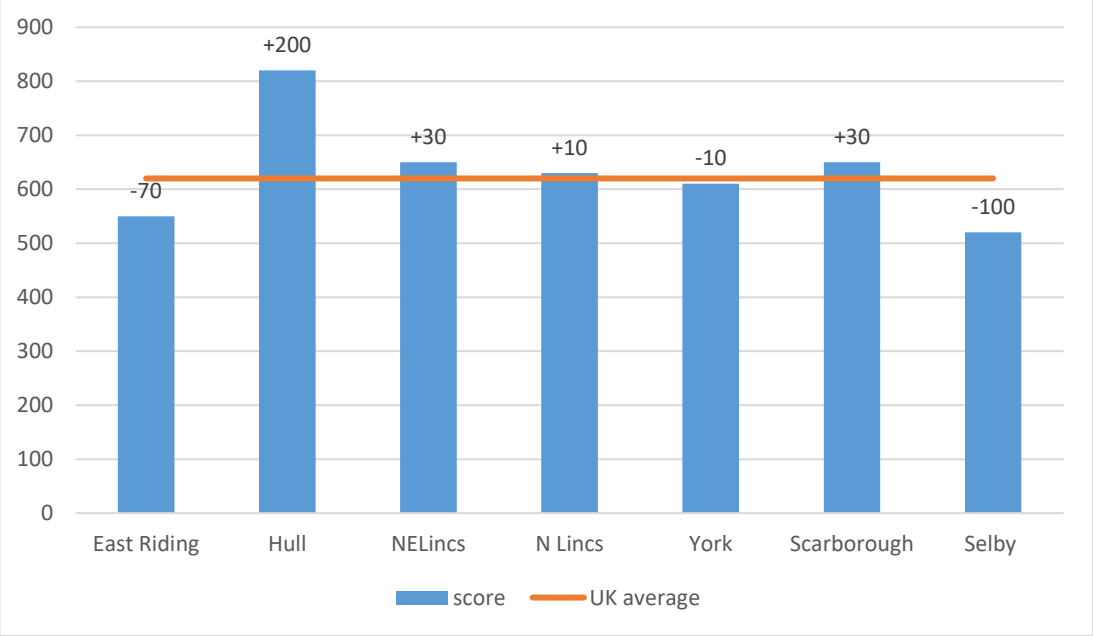


Table 11 illustrates how data published in 2013, 2014 and 2015 were applied to generate a time series data capture. The average numbers of alcohol related admissions per 100000 population were identified to produce a variable for comparison with the

UK national average. Ranking was apportioned by calculating deviations from the UK average.

Table 11 Alcohol related hospital admissions per district

Publication Year	UKAV	East Riding	Hull	NE Lincs	N Lincs	York	Scarborough	Selby
2014/15	640	570	820	660	670	630	680	520
2013/14	620	540	840	640	620	630	620	500
2012/13	610	540	800	660	590	560	650	530
Total	1870	1650	2460	1960	1880	1820	1950	1550
Rounded decile	620	550	820	650	630	610	650	520
Comparison UK average	0	-70	+200	+30	+10	-10	+30	-100
AVERAGE RANKING		3 Less than	1 Very high	2 Above	2 Above	3 Less than	2 Above	3 Less than

Public health England have organised smoking prevalence tables as an average indicator from the integrated household survey, the GP quality outcomes framework (QOF) which details GP surgeries' data on smoking, the GP patient survey data (GPPS) and the Health Survey for England data (HIS) (PHE, 2016). These scores are based on a decile ranking and available at NHS area levels. The following table**Error! Reference source not found.** shows the results of QOF, HIS and GPPS urveys – the mean score of all three surveys was applied to rank areas across the catchment area.

Survey	East Riding	Hull	NE Lincs	N Lincs	York	Scarborough	Selby
QOF	3	10	10	9	3	7	4
HIS	3	10	10	6	6	7	7
GPPS	2	10	10	7	2	7	3
Mean score of deciles	3	10	10	7	4	7	5
Decile Ranking	3 LOW	1 HIGH	1 HIGH	2 MID	3 LOW	2 MID	2 MID

Table 12 smoking survey data

Postcodes have been identified as a marker of social class (Danesh et al., 1999), however, use of postcodes alone to determine social status classification is contentious. There are many measurements of status in society which may or may not truly represent individuals, hence the issue of ecological fallacy. Equally, these types of areal measures can provide information on the context of people's lives. For example, they offer information on the general levels of crime, area level pollution and a variety of factors which have a potential to impinge on a person's social experience of life. Therefore, descriptions of each postcode in relation to social class can offer a contextual description of that area.

Postcode based social status studies frequently rely on self-reported income; access to various consumer goods; social class; home ownership and duration of schooling to gauge 'social class', with results enumerated to postcode levels. There are potential concerns relating to responder bias, reliance on self-reporting and grouping many individuals into similar characteristics (ecologic fallacy), which have the potential to threaten validity for this study. Considering this, individual subject data were evaluated against the area classifications for purposes of this research.

All GOC subjects were classed to IMD status during the process of their care. Between the years 2000-2013, there were three IMD iterations. Sociodemographic 'spacial' area descriptors were presented at LSOA level and identified using data from just one timeframe, (Census 2001 – from ONS, 2013) as this was felt to truly represent subjects IMD through their life course.

The English Indices of Deprivation (2004, 2007 and 2010 iterations) offer a measurement of multiple deprivation released through the Office for National

Statistics (ONS, 2007). They offer area descriptors by combining several distinct domains of deprivation. These domains include income; employment; health and disability; education skills and training; barriers to housing; living environment and crime. The ONS also provide a separate index to measure deprivation affecting older people as an adjunct to the IMD 2007, and reflects all adults over 60 who currently claim pension credits or income support. The numerator for this is based on population of over 60s per LSOA. This dataset was considered, however the score lacked specificity at small scale enumeration, so was discounted for the purposes of this research.

An area classification based on findings from the 2001 Census (Norman, 2010) was applied to group areas into clusters with common characteristics. The LSOA scale was applied to this cohort for subsequent comparative analysis in incidence Table 13.

Table 13 ONS Supergroup descriptors

Supergroup	Consists of
Countryside	Countryside Communities; Rural Economies; Farming and Forestry.
Professional City Life	Educational Centres; Young City Professionals; Mature City Professionals.
Urban Fringe	Urban Commuter; Affluent Urban Commuter.
White Collar Urban	Well off Mature Households; Young Urban Families; Mature Urban Households.
Multicultural City Life	Multicultural Inner City; Multicultural Urban; Multicultural Suburbia.
Disadvantaged Urban Communities	Struggling Urban Families; Blue Collar Urban Families.

Miscellaneous Built Up Areas	Suburbia; Resorts and Retirement; Urban Terracing; Small Town Communities.

These generate one of the many layers in the GIS, to produce a mapped version of neighbourhoods for further analysis.

Access to GP within 15 minutes and a Hospital within 30 minutes by public transport is published by the department of transport (DFT, 2011). These data were used as they offer a measurement of access to clinical services. They are based on a crude measure of access based on the centroid of the LSOA, so were applied to illustrate accessibility by public transport to GP within 15 minutes. To apply these data in mapped format, access was aggregated into three levels, depicting the percentage of homes within each LSOA who could access a GP within a 15-minute timescale (Table 14).

Table 14 Availability of public transport to GP within 15 minutes

LOW	**Less than 50% of households can access GP by public transport within 15-minute journey time.**
MID	Over 50% of households can access GP by public transport within 15-minute journey time.
HIGH	All households are within a 15-minute pubic transport journey to a GP

Similar profiling was undertaken to denote access to hospitals, but this was taken at 30 minutes.

The DFT data is based on Euclidian distance to the referral centres. This is limited in that it is merely a straight line distance between the postcode and the referral centre, so may not truly represent actual travel times. However, as an addition to the DFT (2011) statistics, it did yield a crude measure of access which were felt suitable for the purposes of weighting the number of variables which may be associated with accessing health services.

A crude measurement of observed versus expected incidence was undertaken to illustrate areas with historically high or low incidence rates. By comparing the total GOC diagnoses per MSOA over the 14-year period in the cohort, with population data per MSOA as a denominator against the actual UK age standardised incidence rates, areas of higher or lower density were illustrated. The UK Cancer Research statistics (CRUK, 2014) identify age standardised incidence rates for GOC as 22.4/100000 male population and 8.9/100000 female population (Figure 17).

		UK
Male	**Cases**	6019
	Crude Rate	18.9
	AS Rate	**22.4**
	AS Rate - 95% LCL	21.8
	AS Rate - 95% UCL	23.0
Female	**Cases**	2900
	Crude Rate	8.8
	AS Rate	**8.9**
	AS Rate - 95% LCL	8.6
	AS Rate - 95% UCL	9.2

Source CRUK, Oesophageal Cancer (C15), Number of New Cases, Crude and European Age-Standardised (AS) Incidence Rates per 100,000 Population, UK, 2014 available at http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/incidence [Accessed April 2017].

Figure 17 Oesophageal cancer incidence (UK) by gender, (CRUK 2014)

To generate the same variables in the cohort, incidence rates were calculated by gender to produce rates per 100000 population at MSOA level. Calculations followed Moon

et al.'s (2000) epidemiological formulae. The CRUK figures provided an expected incidence comparator against gender specified incidence across the regional catchment area.

As this was a closed cohort, incidence (density) rates were calculated from the number of cases per MSOA as numerator and total person time of the observation as denominator. The total persons at risk were identified as MSOA populations by gender.

$$\frac{n\ of\ new\ cases\ during\ the\ time\ period\ (per\ MSOA)}{Total\ persons\ at\ risk\ during\ follow\ up\ period\ (per\ MSOA)}.$$

Quintiles were classified through the GIS application. Scales were apportioned manually to capture data representing gender based CRUK standardised incidence rates (CRUK, 2014) and are illustrated in the following table.

Table 15 quintile classifications of incidence across MSOA

Classification of incidence	MALE INCIDENCE	FEMALE INCIDENCE
Extremely low	1.429 - 17	0 – 5.06
Low	17.01 - 22	5.07 – 9.08
Expected	22.01 - 23	9.09 – 13.40
Higher than expected	23.01 - 30	13.41 – 18.55
Extremely High	30.01 – 56.70	18.56 – 24.50

6.4.4 UOA (4) – Analysing clusters and characterising the area for GOC specific attributes. Developing the K-Means area classification.

A distribution of features or attributes within a defined area can make a pattern – and this pattern may be clustered, or evenly dispersed. Identifying patterning can lead to a better understanding of geographic, or demographic phenomenon. Maps can be useful

to illustrate these patterns, but statistics are then applied to determine the extent of patterning, or dispersion. This is relevant to incidence data, whereby one can compare an observed distribution, measuring extents of clustering as a tool to elicit a probability estimate on how data are – or should be, distributed in geographic terms.

To identify whether there were any locally significant incidence clusters, Openshaw's (1998) Geographical Analysis Machine was applied to incidence data to determine the distribution of incidence. This statistical package compared incidence which was pinpointed through Easting and Northing data (Table 16). This means results are not artificially constrained by geographic boundaries. GAM offers a crude measurement to identify where incidence rates were higher across the geographical catchment area. Data are analysed through a series of concentric circles, drawn across the area to 'encapsulate areas of higher incidences. GAM has been applied in a range of studies as a disease profiling mechanism to detect clustering. Radial scales were set to maximum 5000 and minimum 250, incremented to 250m with a circle overlap of 0.5. The rationale for this was to identify any crude clustering of diseases, notifying 'hotspots' where the concentric circles had overlapped.

Table 16 GAM data input

Case ID	Easting	Northing	Incidence (GOC cases per LSOA)	Population per LSOA

To identify the extent of density in incidence, Kernel density tools were then applied to GIS data. This tool produced a weighted map illustrating proportion of incidence per population. This was useful in that it presented information on incidence across the region, but did not violate requirements to patient confidentiality. By producing a smoothed incidence map, it offers a snapshot of incidence, which did not display

identifiable data. It also informed the Gettis-Ord or GI* (polygon) mapping which identifies areas of increased incidence across the geography.

This tool was developed from GOC specific attributable factors and attributes of the cohorts, the characterisations were based on attributes and weighted in order of priority. A quintile grouping which represented the weighting and accorded homogenous descriptors of areas was developed and compared against actual incidence.

K means cluster analysis offers an initial, crude multivariate classification technique to enable comparisons of multiple characteristics. Used widely in many marketing strategies, it seeks to identify the geography of human behaviours, so that goods and services are marketed appropriately. Feasibility studies are used in marketing to identify potential sites, so that products may be placed appropriately to meet consumer demand.

Evidence identified several factors linked with GOC. Previous studies identified advanced age, male gender, environmental factors, socioeconomic deprivation and poor housing and these factors can be applied in the GIS for analysis. These attributes were harnessed to characterise the catchment area, revealing areas which have a higher (or lower) propensity to GOC incidence. The tool was developed from lifestyle data identified previously, for comparison against cluster analysis results.

The characterisation tool was based on findings relating to the most common risk factors in GOC (highlighted in previous chapters). Evidence from a variety of meta-analyses on attributable factors has been summarised in Table 10. Though the cohort data had no evidence relating to subject's lifestyles, demographic analysis supported a GOC predisposition to attributes such as male gender and older populations.

These attributes were apportioned to the geography of the catchment area,and weighted so that they described places in relation to their population potential for GOC. It is common practice in epidemiology and geodemographic studies to label small areas in relation to their populations, features or attributes. Previous data analysis and GIS development were useful to identify potential clustering in incidences. However, the next step was to develop a tool which captured areas of heterogeneous social, economic, population and lifestyle characteristics. In this way, areas classified as the same type, encapsulate those inhabitants sharing similar characteristics.

This methodology draws from geodemographics – and rests on the theoretical preposition that persons living in similar neighbourhoods have similar characteristics, that people with certain behaviours, cultures and lifestyles tend to gravitate towards those with similar behaviours. Geodemographics categorise neighbourhoods through their underlying population characteristics and more than one neighbourhood can be allocated to each category. The first step was to identify and weight the specific attributes of GOC within the populations. Input variables chosen for the purposes of this analysis were alcohol and smoking, IMD 2010, and density of older age populations. Socioeconomic deprivation and access to services were also included in the analysis, drawing from the information presented in previous chapters.

For each of the input variables, data were normalised to Z scores, depicting positive and negative correlations with GOC specific attributes. Five groups were identified to capture the demographic data. Two groups were far smaller than the others (Groups 1 and 2) however, they were relevant because they displayed factors which were either highly associated with GOC, or not at all. For example, the demographic profile of Group ‘2’ would suggest very low risk of GOC.

These Z scores were classified into clusters relating to Lower Super Output Areas with a potential to higher GOC attributes. A maximum seven groups was determined to be the most effective sampling, as this related to a minimum distance between centres of 4.2, with an iteration of 5. The key criteria for the K means analysis were drawn from findings presented in both Chapter 2 and identified from cohort data. Table 17 lists these criteria.

Table 17 Identifying key criteria for the model

Factor	Weighting in order of importance	Links with data from UOA
Alcohol rates higher than average	++	3
Smoking rates higher than average	++	3
Population density over 65 high	+++	3 / 1
Socioeconomic deprivation (poor lifestyle)	++	3 / 1

These key criteria formed the basis for the model as data were layered into the geographical information system.

Finally, an evaluation of this clustering algorithm characterisation tool was undertaken to determine whether it represented actual incidence data (determined through GAM and Kernel density statistics. Maps were compared and evaluated accordingly.

6.5 Links between the 4 units of analysis to the objectives

To reveal whether there are any small scale geographical areas which demonstrate patterns in incidence, presentation and survival of GOC and whether these areas can be predicted through analysis of population demographics and neighbourhood attributes, several objectives were been developed. Table 18 illustrates the links between methods and data sources and identifies how these met the objectives.

Table 18 Linkage between objectives and data

Objectives	Sub-Units	Methods and Data types
Identify the most relevant population factors previously linked with GOC.	PHE GOC COHORT data Literature review findings National Oesophagogastric cancer audits.	Quantitative descriptive & correlational analysis. Demographic data – age, gender, tumour site, cancer type, morphology, routes to diagnosis. Literature review findings – comparison of cohort against literature review
Evaluate survival in GOC and quantify 'presentation stages'.	PHE cohort data	Survival analysis GOC data – date of presentation – date of death. Compare survival against age, site, morphology, routes to diagnosis, presentation stage
Describe the neighbourhood	ONS population level data CASWEB data EU age standardised Incidence rates of GOC	Capture data at MSOA then LSOA level on age, gender, Area classifications on age, gender, smoking, alcohol, deprivation indices.
Triangulate findings and observe clusters for comparison against sociodemographic characterisation tool.	EDINA data (base maps) ONS data Z score characterisation of access/ alcohol / smoking / SEG LAPE tables InFuse & CASWEB data on population demography	Use aspatial and spatial scan statistics to reveal patterns. GAM, Kernel density GI* K Means Cluster analysis

6.6 Establishing the validity of this research cohort data

In the UK, all diagnoses of GOC are recorded by PHE via regional cancer registries. These are subject to 12-18 months delay before all cases are recorded (OCNS 2012). Cancer Information Services identify they can capture a minimum 97 % of data at 18 months, rising to 100% completion at 5-years (OCNS 2013). This level of confidence in data accuracy supported the choice of dates ranging 2000-2013 for this cohort.

These data were analysed in SPSS V23 (IBM corporation, 2016). Descriptive statistics summarised subject data. Where comparisons were made between variables, Chi Square, Mann Whitney U and Kruskal Wallace tests were applied. All SPSS syntax is

listed in Appendix 6 to illustrate how data were cleaned for transfer from Excel to SPSS and to indicate the processes involved in creation of new test variables.

Several data analysis tools were applied to analysis in this research. Excel, SPSS and GIS formatted files were used to facilitate analysis. At each merge, data were evaluated by the researcher's supervisory team to confirm accuracy during translocation. During the data cleaning process, several file formats were required.

Microsoft Excel	**.csv files, XML files.**
SPSS	.sav files
GIS files	Mxd, SHP, dbf, atx, prj, sbn, sbx, shx files.

6.7 Presenting the results of this research.

Results were presented using the objectives stated in the embedded case study design. The initial objective related to population factors, the second to analysis of survival and presentation. A third section presented the neighbourhood description and maps depicting the geographical catchment area, its' lifestyle and demographic factors and incidence of GOC. Finally, all information was triangulated in the GIS and a tool depicting potential areas of higher or lower incidence was developed. These were presented alongside data and maps of actual incidence.

Results chapter 7 –Revealing GOC attributes and identifying patterns in incidence, presentation and survival in gastroesophageal cancer patients' neighbourhoods

This chapter presents the results against the research objectives as shown in Figure 18 (units of analysis UOA are in bold).

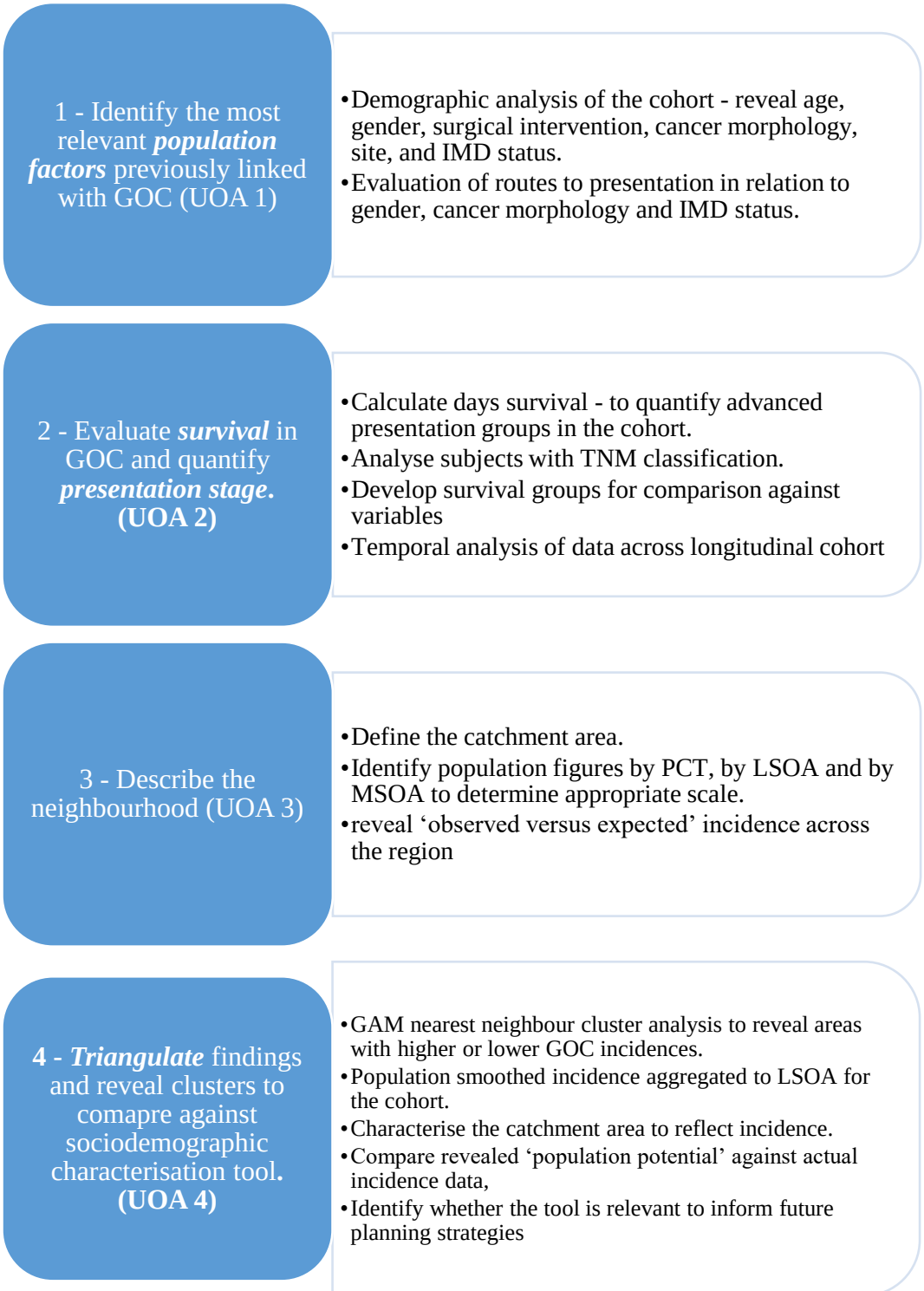
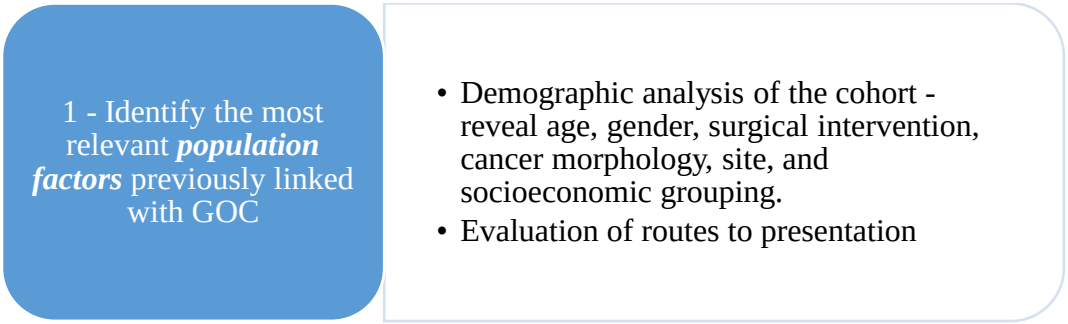


Figure 18 How the chapter is presented

7.1 Revealing population factors previously linked with gastroesophageal cancer (Objective 1).



This section relates only to the GOC cohort, revealing common attributes relating to presentation and demography. By analysing the units of analysis (1) identified in the research protocol, it reveals common factors associated with GOC and identifies potential differences in age, gender, morphology and site of the cancer and the different treatments provide to subjects. To fully answer the objective, this section also collates recently published evidence in support of socio environmental and lifestyle factors previously attributed to development of GOC. This supports weighting of the tool which is generated in objective 4 for subsequent analysis.

A total cohort of 2785 subjects living in the regional referral centre were diagnosed with gastroesophageal cancers (determined as C15.2-C16.0) between the dates of 1/1/2000 and 31/12/2013. Patient postcodes were linked to Eastings and Northings via geographical lookup tables to facilitate spatial analysis.

7.1.1 Descriptive analysis of cohort

Of 2785 GOC diagnosed subjects, the average age at presentation was 70 (SD 11 years), and there was a 70% male 30% female divide. The age dispersion across both genders displayed similar patterns across all age ranges apart from 65-74 years, where the numbers of males who were diagnosed in that age category increased (Figure 19).

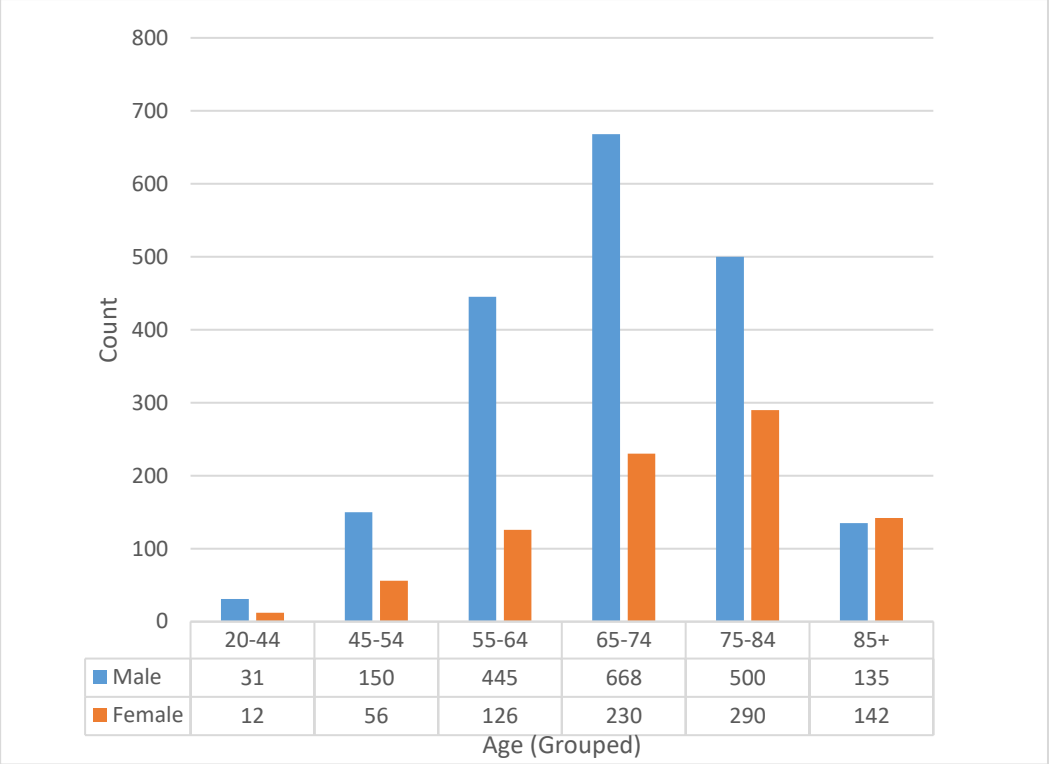


Figure 19 Age and Gender.

There were more males than females in this cohort (N1929 males and N 826 females). Of the total cohort (N2785) - 2509 (90%) had a recorded status of dead at data capture. The deceased subjects included 1740 males and 769 females. Only 276 of the subjects' were 'alive' at data capture (189 males and 87 females).

7.1.2 Surgery

Only 19% (N = 520) of GOC patients received surgery out of the total cohort (n = 2785), however there was no data on whether this surgery was palliative, or treatment oriented (Table 19). A total 56 subjects died within 6 months and an additional 63 within one year. Of the 520 subjects who received surgical intervention, 221 survived beyond one year, and 180 survived beyond 5 years. Surgery was performed in 20% of the males and 16% of the females in the cohort. Though surgery was performed across all age groups, rates of surgical intervention reduced from an average 30%, to less than

21%in the groups over 65 years. Rates of survival after one year were fairly constant across all age groups.

Table 19 Surgical intervention by age

Age groups	Total number of subjects who received surgery	Rate of surgical intervention/age group	% of these subjects who survived over 1 year post surgery	Males receiving surgery	Females receiving surgery
20-44 years	14	32.5%	12 (86%)	10	4
45-54 years	64	31%	54 (84%)	46	18
55-64 years	166	29%	135 (81%)	129	37
65-74years	192	21%	147 (77%)	147	45
75- and above	84	11%	53 (64%)	66	18

7.1.3 Histology/Morphology

The main histological subtypes were recorded as adenocarcinoma (N = 1744 – or 63% of the total cohort), then squamous cell (N = 2785 – 27% of the total cohort) (Table 20). Males were more likely to present with adenocarcinoma (71% of the male cohort). In contrast, females had a tendency towards both adenocarcinoma and squamous cell (43 % and 47% respectively).

Table 20 Morphology of GOC in the cohort

Morphology						
	Total cohort		Males		Females	
Epithelial neoplasms NOS	156	6%	102	5%	54	6%
SCC	739	27%	336	17%	403	47%
ADC	1744	63%	1378	71%	366	42%
Cystic and Mucinous	110	4%	85	4%	25	3%
Mixed neoplasms	22	<1%	19	>1%	3	<1%
Neoplasms NOS	14	<1%	9	<1%	5	1%
TOTAL	N 2785		1929		856	

7.1.4 Site

The most common tumour site in males was GO Junctional (53% males), but there was also a high percentage of unspecified sites in this group (38%). More sites were labelled as unspecified in the female groups (44% of the female cohort). The most common site in females was also GO Junctional (41% of the female cohort). Cancers of the upper oesophagus were more prevalent in females than in males (6% males versus 15% females).

Table 21 Site of cancer by gender in the cohort

Site	Total N	Males	Females
Upper oesophagus	250 (9%)	120 (6%)	130 (15%)
GO Junctional	1476 (53%)	1130 (59%)	346 (41%)
Oesophagus unspecified	1059 (38%)	679 (35%)	380 (44%)
Total	2785 (100%)	1929	856

7.1.5 Socioeconomic groups

All subjects were allocated a quintile IMD status at diagnosis by their diagnosing clinicians, and this was based on postcodes of residence at the time of diagnosis. However, the longitudinal cohort data were captured over several IMD iteration periods (2004, 2007 and 2010). Figure 20 is presented to identify the changes in IMD status of postcodes within the catchment area over time. Given that iterations on 2007 and 2010 offered similar profiles, all spatial data applied within this thesis were based on IMD 2010. These data are available at LSOA level via CASWEB, and present the closest fit with IMD iterations across the timeframe.

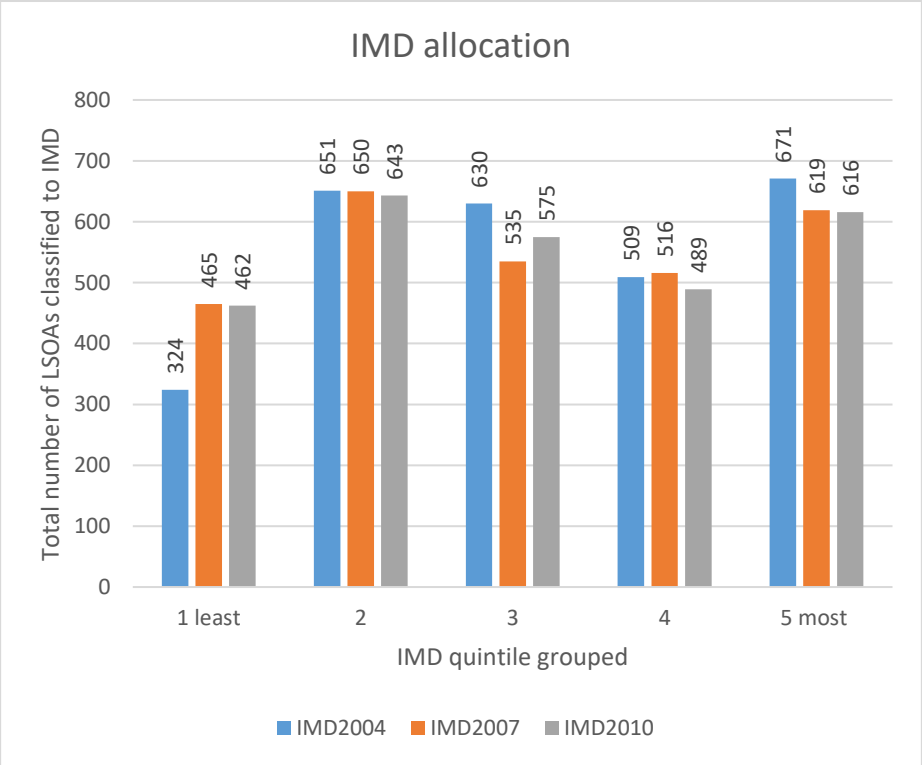


Figure 20 Longitudinal data - differences between IMD status 2004, 2007 & 2010.

To present an analysis of deprivation in the total cohort (N=2785), three separate groups were developed to reveal classifications into deprivation. These were labelled

as ‘most’ (1105 subjects), ‘middle’ (575 subjects) and ‘least’ (1105 subjects). From 55-years onwards, there was little difference in IMD status when reviewed across the age groups (Figure 21).

However, there were far more subjects diagnosed under the age of 54 who came from the most deprived areas. A total 48 % of younger GOC subjects (n = 119) were in the most deprived groups, whereas only 35% (n = 88) were in the least deprived groups. Figure 22 identifies the main differences with age group 45-54.

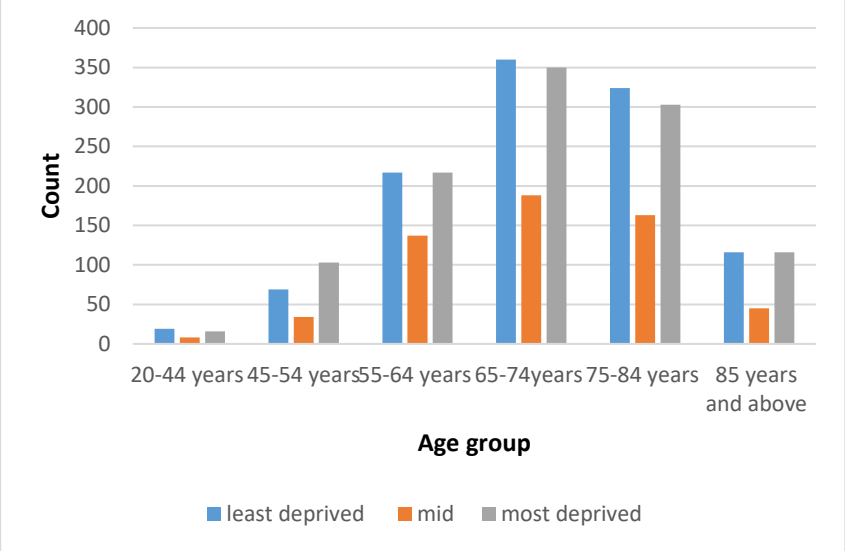


Figure 21 IMD 2010 scores by age within the cohort

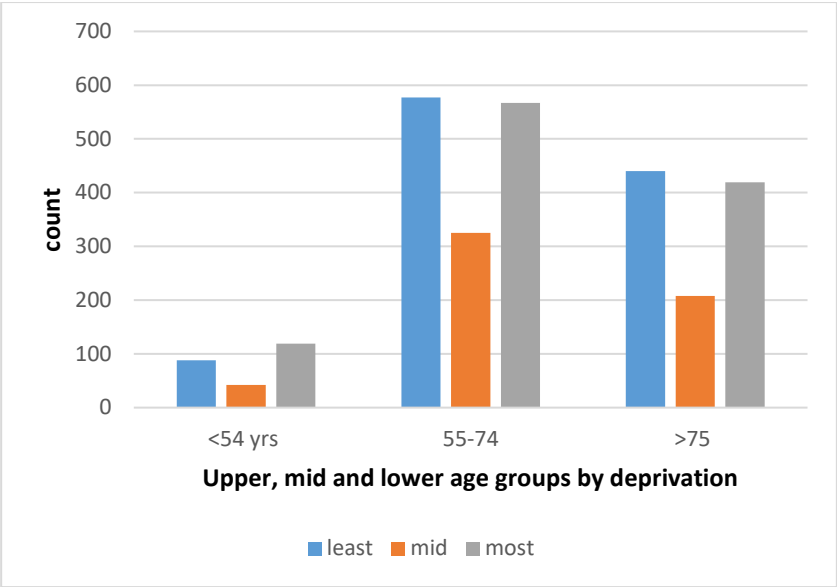


Figure 22 IMD 2010 across age groups in the cohort

The IMD socioeconomic groups in the cohort were similar across genders.

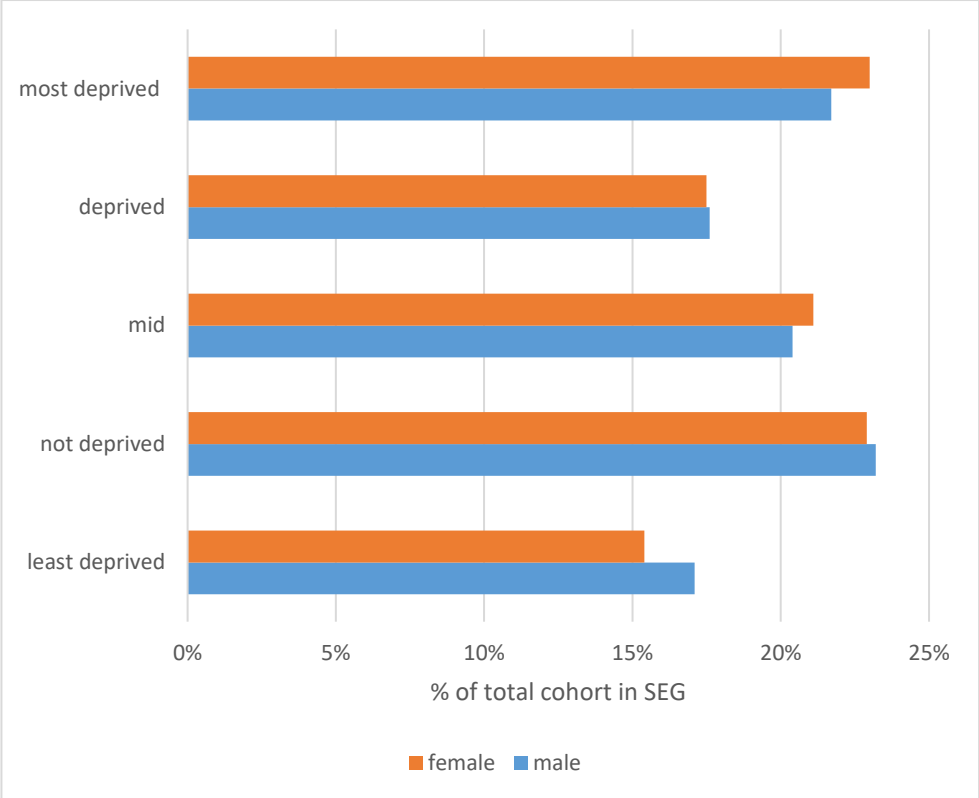


Figure 23 Percentage of cases by IMD 2010, gender specific.

7.1.6 Routes to presentation

Of the 2785 patients presenting with GOC across the timeframe, 1640 had data relating to their mechanism of presentation. These data were only routinely collected and recorded from 2006 in response to the national early diagnosis cancer campaign. Figure 24 shows the main mechanism of presentation was via the 2-week wait, GP rapid referral system. However, a total of 317 patients presented as emergencies and 376 through routine GP referrals, a further 150 through hospital inpatient diagnosis and 57 through other services. Presentation mechanisms were similar across genders.

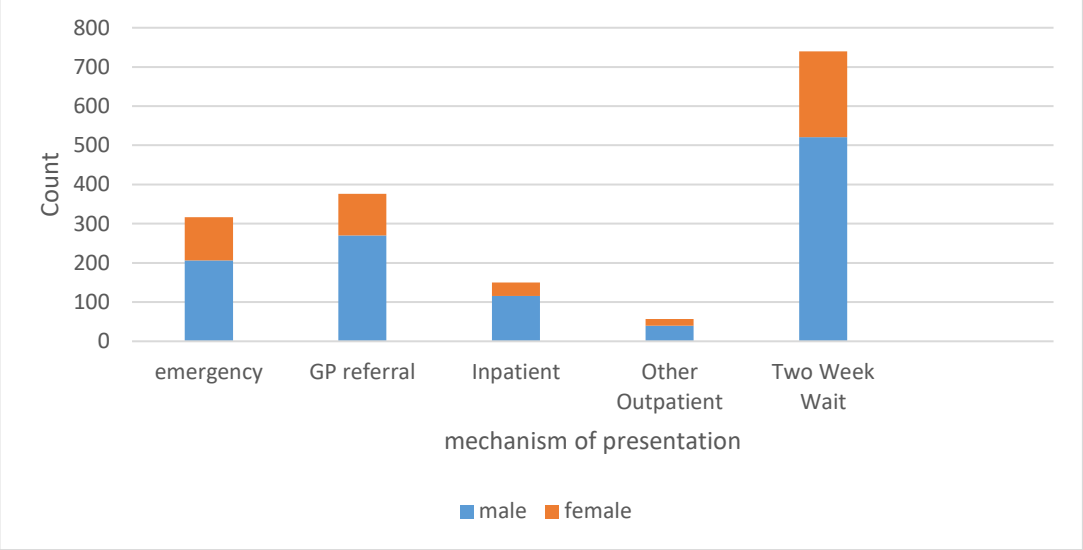


Figure 24 Mechanism of presentation in the cohort (post 2016)

Of the 1640 subjects with records of mechanism of presentation, only 265 received surgery (25 subjects had missing data on surgery).

	Emergency	GP referral	Inpatient	Other Outpatient	Two Week Wait
NO Surgery	300	314	104	48	609
Surgery	17	62	46	9	131

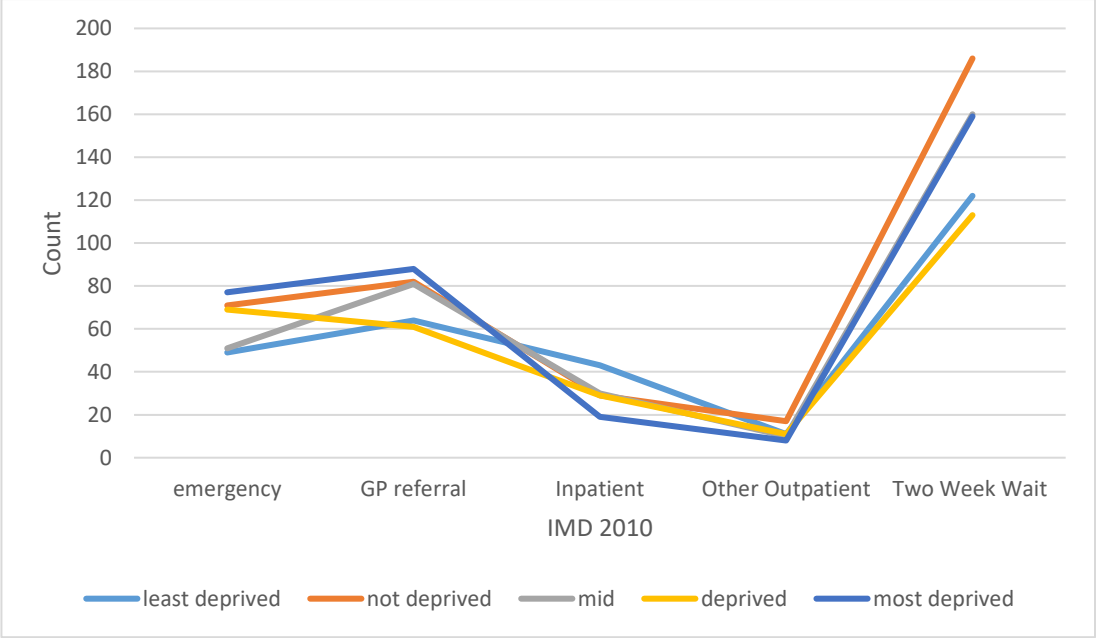


Figure 25 mechanism of presentation by IMD

Figure 25 identifies the mechanism of presentation across the socioeconomic deprivation groups, a chi square test for independence (with Yates Continuity Correction) indicated no significant association between IMD status and referral mechanisms *x*2 (1, N=1640 = .24 p=.50, phi= .171). Figure 26 illustrates data relating to cancer morphology – again, showing no significant between groups associations (*X2* (1, N=1640 = .47 p=.23, phi = .13).

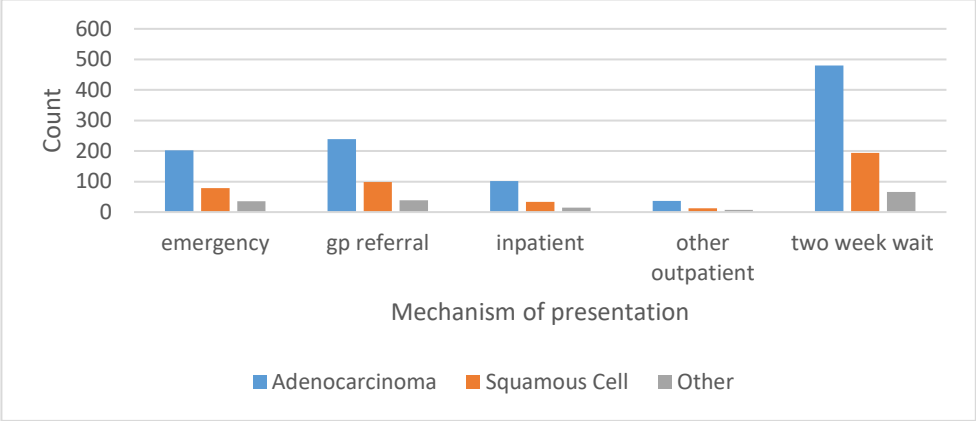
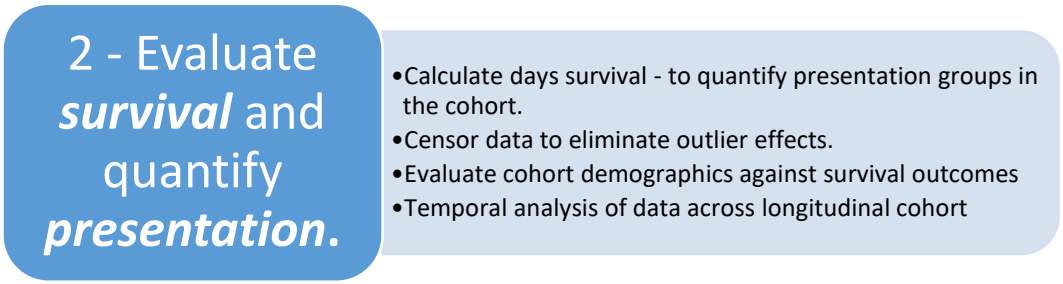


Figure 26 Mechanism of presentation by histology

7.2 Evaluating survival and quantifying presentation stage (Objective 2)



Survival in gastroesophageal cancer is dependent on surgical intervention. Surgical intervention is less complex when the tumour is smaller in size – or has less infiltration to surrounding structures. Therefore, survival outcomes may be linked to stage of presentation. The earlier a patient presents, the more likely they are to have a favourable outcome. There was a significant amount of missing data on stages of presentation (TNM). In the full cohort (N = 2785), only 8% remained alive and 92% were deceased at data capture (08.06.2016).

Of these, 74 subjects were considered as ‘outliers’ (as their ‘days survival’ fell far beyond the 5-year cut point (the longest duration was 5022 days). These outliers are shown in the boxplot in the following figure, which illustrates how censoring was applied to prevent data skew.

Boxplot (Total group - N2785)

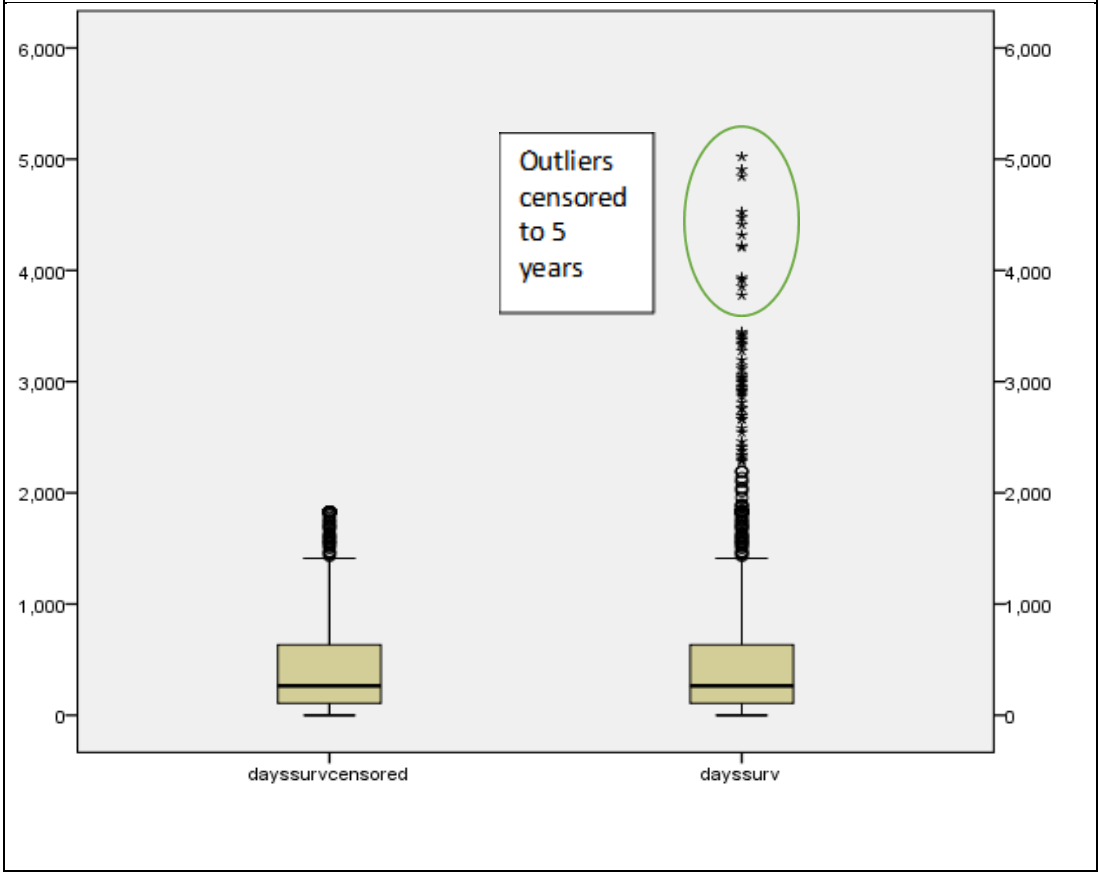


Figure 27 Total cohort illustrating requirement to censor for survival skew

Of the survival sub cohort, (N=2215), 867 (39%) subjects died within the first six months, and a further 471 (21%) after the initial 6 months, but within the first year. Over 60% of the total cohort died on or before 1-year following diagnosis. A further 617 (28%) subjects survived between 1 and 5 years. Only 260 (12%) of the survival cohort were alive at 5 years after the date of histological confirmation. The median survival time in days was 264 days. The following figure illustrates how the survival cohort was generated.

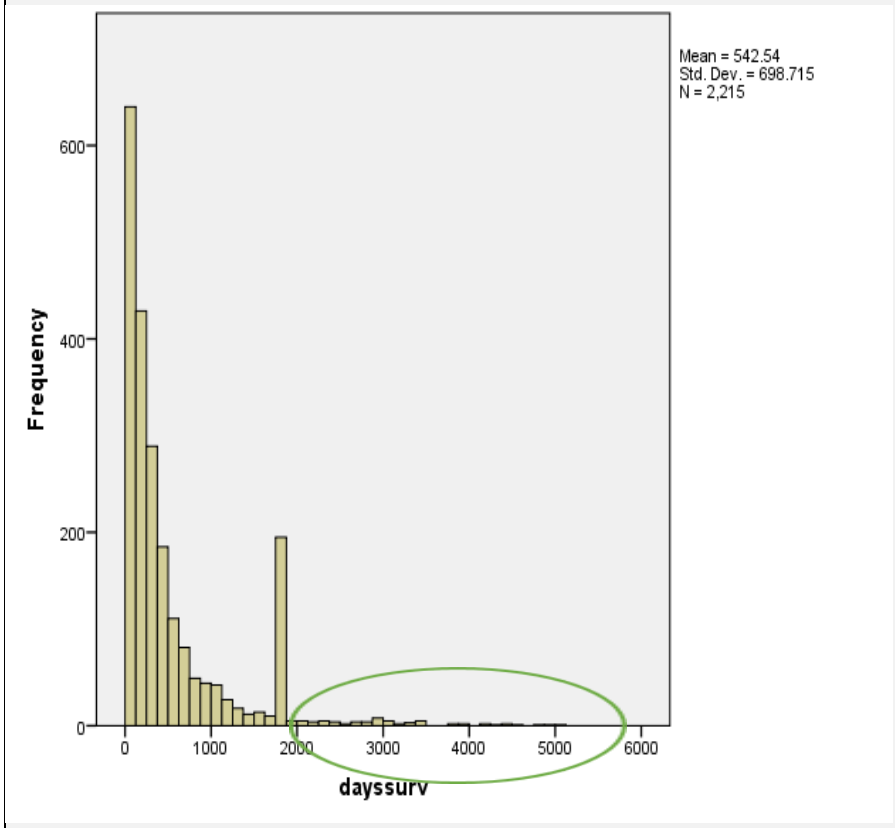
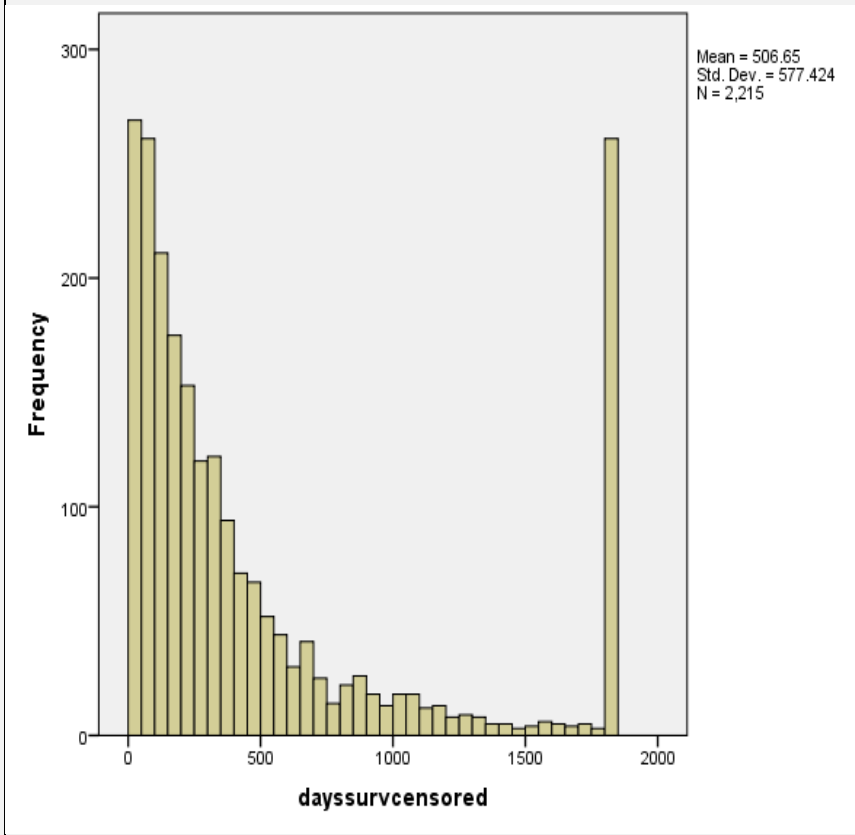
Survival group - Pre-censoring	Survival Group - censored to 1826 days
	

Figure 28 boxplot - Days survival and censored days survival

Given the significant skew where most patients died within the first few months, the median days survival from histological confirmation, to death (where this occurred) underpinned the groups for further analysis. Tukeys Hinges provided the first, second and third quartile which was applied to generate the most relevant groups for this research. Table 22 illustrates this accordingly.

Table 22 Survival percentiles (Days survival censored)

Percentile survival	**5**	**10**	**25**	**50**	**75**	**90**	**95**
Average (weighted)	23	42.6	106	264	634	1826	1826
Tukey's Hinges			107	264	634		

*censored for death and extreme survival status.

Over 60% of the cohort died within the first year. Given this figure, data are unrepresentative of the majority of GOC diagnosed patients who die far earlier than the generally apportioned 1 year cut point prevalent to cancer research statistics. The 1-year survival status which is generally applied in cancer research was felt to lack specificity in capturing a significant proportion of GOC subjects who die earlier.

Figure **29** illustrates this – revealing how a 1-year data capture does not represent the high numbers of patients who die before 6 months.

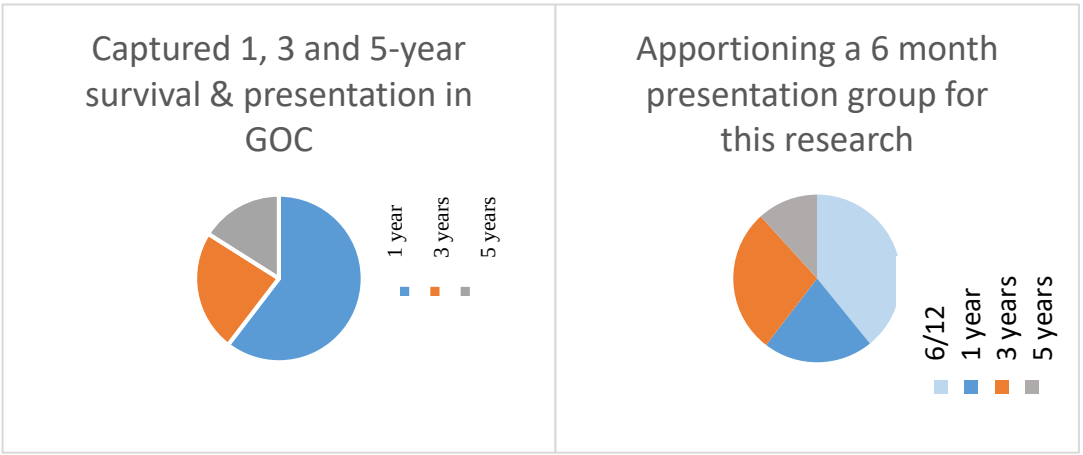


Figure 29 Survival group chart

Given this problem, separate groups were determined as more appropriate to underpin further analysis on presentation and survival in GOC for this cohort. Table 23 identifies how the 'presentation groups' are formed for the purposes of this research.

Table 23 Presentation groups

Date of histological confirmed diagnosis – date of death or censored endpoint	Label	N within group and % of cohort	Clinical rationale for presentation grouping
Deaths *within 6 months* of presentation to cancer specialist services	Extremely Advanced presentation	n = 867 (39%)	Captures those patients presenting with aggressive tumours, or too late for curative treatment
Deaths *up to 1 year*	Advanced presentation	n = 471 (21%)	Captures >6 month presentations), have had surgical interventions, but with limited curative options.
Death *1-5-years*	Treatment facilitated	n = 617 (28%)	Patients who have received surgical intervention of a curative or palliative nature
Survival > 5-years	Survivor	n = 260 (12%)	Considered as patients who have survived, or are in remission of the disease process.

These data are presented in the following map to illustrate how these groups were spread across the region.

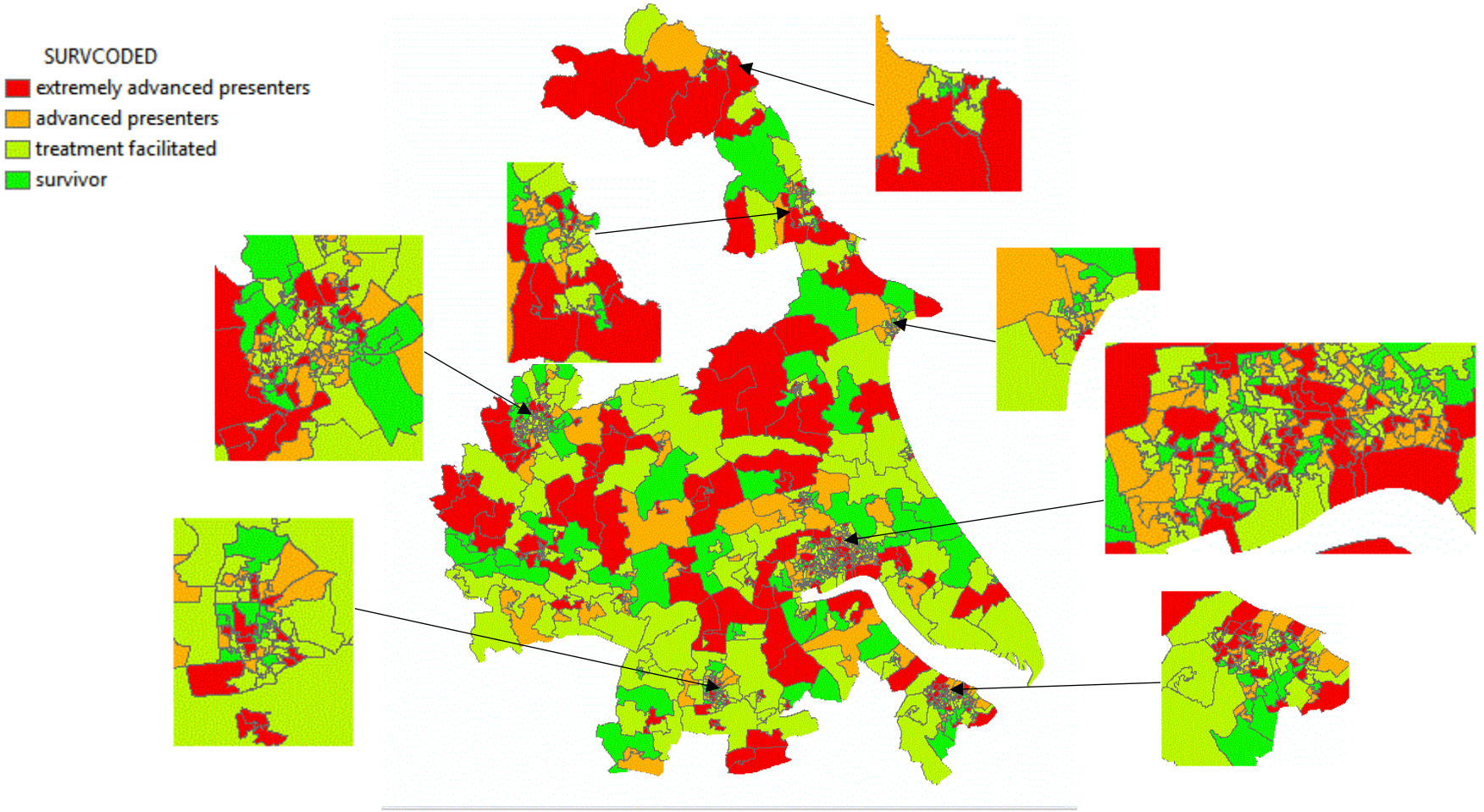


Figure 30 Survival groups across the region

7.2.1 How patients present

Since January 2006, PHE have collated data on routes to diagnosis and this facilitates assessment of survival against route to diagnosis. Table 24 shows routes to diagnosis against the proposed presentation groups. It should be noted that the total number in this analysis relates to the sub cohort (N = 1097) date range 2006-2013 as statistics pre-2006 are unavailable

Table 24 Routes to diagnosis and survival (survival analysis cohort)

	Presentation type						
Presenting Group	Emergency	Standard GP referral	Inpatient	Other outpatient	Two Week Wait (TWW)	Unknown	N
Extremely advanced presenters	134 (36%)	81 (21%)	23 (6%)	11 (3%)	126 (34%)	4 (1%)	379
Advanced presentation	28 (12.5%)	54 (24%)	18 (8%)	8 (4%)	113 (51%)	1 (0.5%)	222
Treatment facilitated	38 (11%)	91 (26.5%)	35 (10%)	11 (3%)	158 (46%)	11 (3%)	344
Survivors	12 (8%)	41 (27%)	23 (15%)	7 (5%)	64 (42%)	5 (3%)	152
Total	212 (19%)	267 (25%)	99 (9%)	37 (3%)	461 (42%)	21 (2%)	1097

This shows that 70% of GOC patients who died within the first 6 months (extremely advanced presenters) were emergency or 2-week wait referrals. Those dying after 6 months, but before 1 year (advanced presentation) tended to present with the 2-week wait referrals. Many presented via their GP, but referrals were not considered relevant for expediting through the 2-week wait system. The survivors and those dying between 1 and 5 years (treatment facilitated) tended to present via 2-week wait, or through a non-urgent GP referral. Overall, most of referrals came through 2-week wait referrals.

As presentation groups were not equally dispersed, a Kruskal Wallace Test revealed a statistically significant difference in survival days across the presentation groups. Those presenting as emergency patients had less median days survival (Media N = 112) when compared to other groups (all in excess of Media N = 359) x^2 (5, n = 1097) = 112, p = < 0.05 (Figure 31). As subjects were not equally distributed across all 6 groups, the nonparametric Kruskal-Wallace test was applied to rank median days survival and compare between groups. This revealed emergency presenters with reduced survival options, when compared with all other groups.

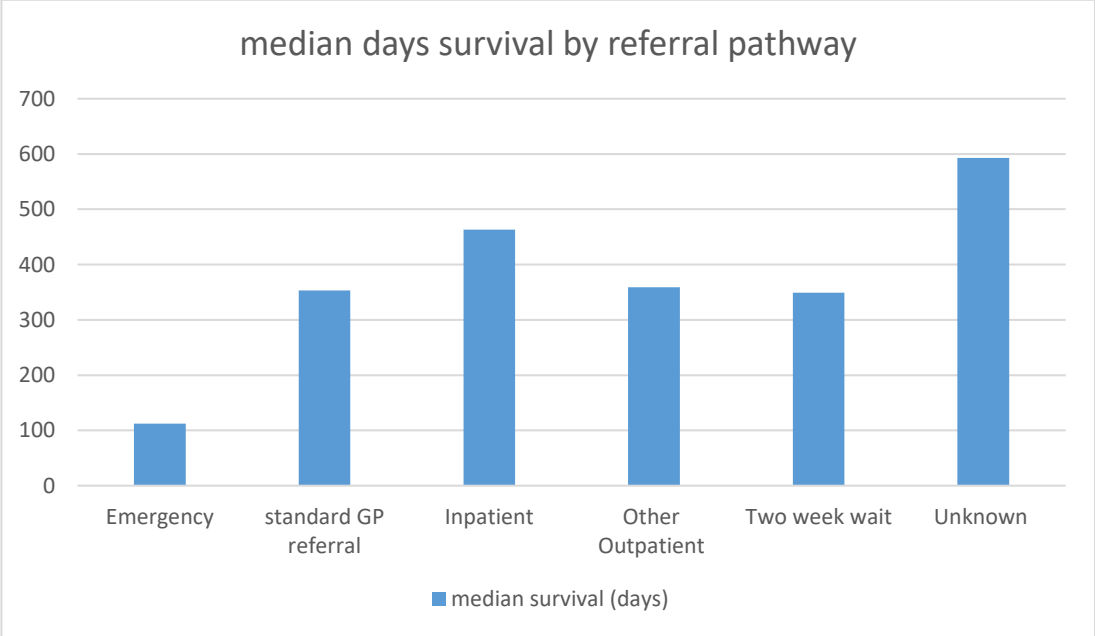


Figure 31 Median survival days by referral pathway.

7.2.2 Exploration of survival against attributes (survival cohort).

This section evaluates whether certain patient attributes impact on survival. It explores demographic variables associated with survival, thus allowing isolation of the effects of treatments, from the effects of other variables. It was used *a priori* as other

variables, such as advancing age, date of presentation and poor initial vital status have been shown to have an impact on long term survival. Cox proportional hazards allowed censoring for subjects who survived beyond 5 years (1826 days – allowing for a leap year day). The survival cohort (N = 2215) is assumed as adequate size for statistical relevance and accuracy, based on Peduzzi et al. (1995) requirement for samples exceeding 250 subjects.

7.2.3 Age, gender and survival

The relationship between advancing age and survival was evaluated through Pearson product moment correlation coefficient. Preliminary exploration of data revealed that there were no violations of the assumptions of normality, linearity and homoscedasticity. Applying Cohen's (1988) guidelines, there was a small negative correlation between age and survival, ($r = -.28$, n2215, $p<0.01$) meaning advancing age decreases survival outcomes, but that only 7% of variance in survival can be explained by advancing age ($r^2 = .729$).

An independent samples t-test compared survival in males and females. There was no significant difference in scores between males (mean 524 days) and females (469 days). t (2215) = 2.07, P 0. 04 (two tailed). The magnitude of the differences in means (95% ci 3.0-106.5) was very small (eta squared = 0.001).

A one way between groups analysis of variance (ANOVA) was conducted to review whether survival in days was different between year of presentation. Here, it must be noted that the year 2011 only represents a half of the presentations in that year, as data were accessed to June of that year. Subjects were defined by year of diagnosis. Levene's statistic revealed the sample violated the assumption of homogeneity of variance (4.066 at P = 0.1), therefore, the alternative nonparametric test was applied.

The Kruskal Wallace Test revealed a statistically significant difference in survival across the years, revealing survival time is increasing over the years when patients presented. X^2 (11, N = 2215) = 31.54, p = 0.01. A Mann Whitney U Test revealed a statistically significant difference in mean survival, meaning those presenting after the year 2006 had improved survival outcomes, when compared with subjects presenting before 2006.

The pre 2006 group displayed worse survival in days (Media N = 234, n 1116), whereas the post 2006 group exhibited improved survival outcomes (Media N = 310, n 1099 U = 535907, Z = -5.1, P = 0.01, r = 0.1). Though statistically significant, this is only a small effect using Cohen's criteria (Cohen, 1988). Figure 32 illustrates median days survival across the longitudinal dataset. The improvements in survival rates may be a result of improved clinical care over the years.

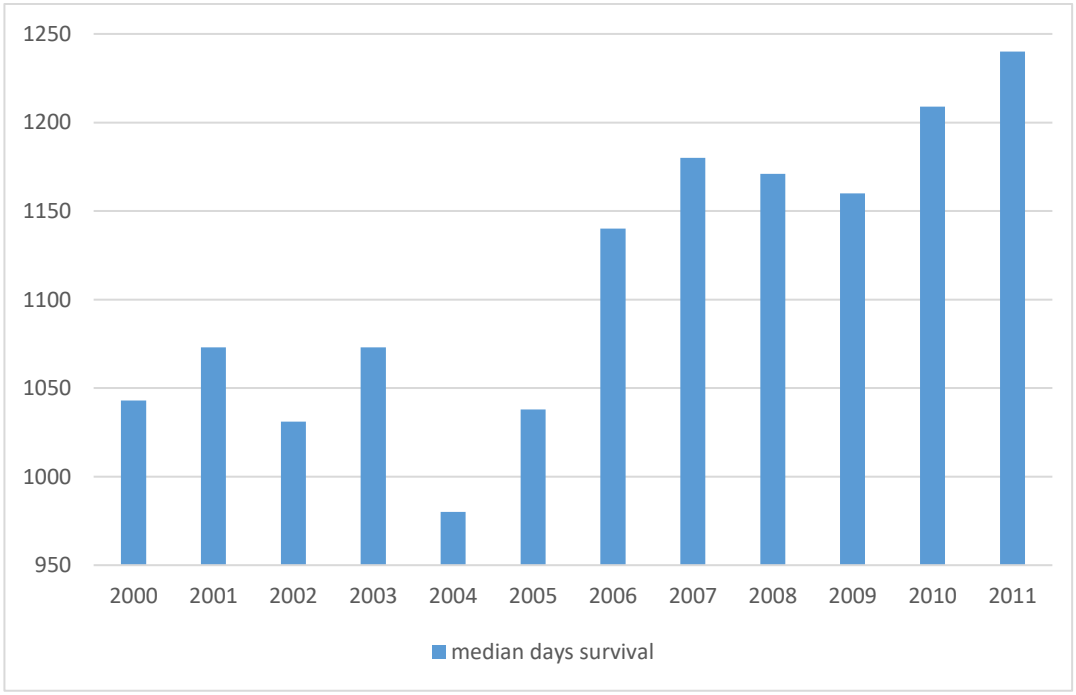


Figure 32 Median days survival across presentation years in the survival cohort

A Kruskal Wallis H test revealed a statistically significant difference in survival days between the 6 different age groups (gp 1 N= 36: 0-44 yrs, gp 2 N=174: 45-54 yrs, gp 3 N=457: 55-64 yrs, gp 4 N=707: 65-74 yrs, gp 5 N=621: 75-84 yrs, gp 6 N=220: 85plus) X2 (5, N=2215) = 20.1 p=0.001 (Figure 33).

A Mann-Whitney U Test revealed a significant difference in days survival in subjects under 65 (Md=350 N=1374) when compared to over 65 (Md171 N=841) U 39661p=0.001. z= -12.4 r= 0.3. Subjects over the age of 65 experienced fewer days alive after their GOC diagnosis.

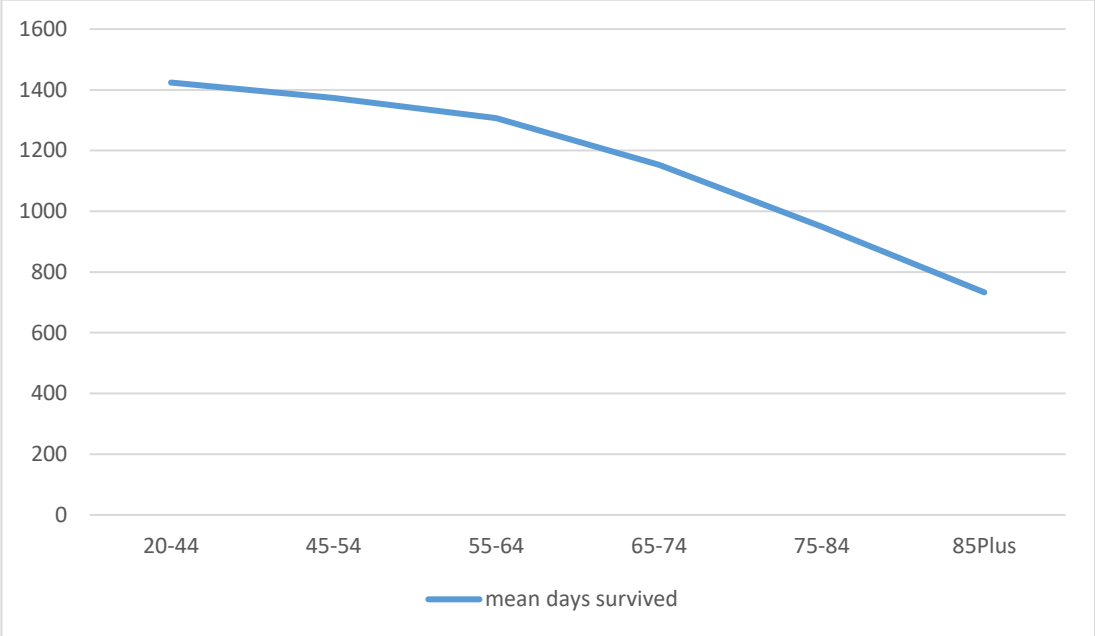


Figure 33 Age and survival across the cohort.

A one way analysis of variance was undertaken to review whether age had an effect on survival for those who had surgical intervention (n520), and those who did not (n 1695). Subjects were divided into age groups for analysis (group 1 under 44 years, group 2 45-54 years, group 3 55-64 years group 4 65-74 years and group 5 65-74 years – data for the over 85 age group was not included due to small numbers

(Figure 34). Mean survival outcomes were greater in patients who were offered surgical intervention (p = 0.005).

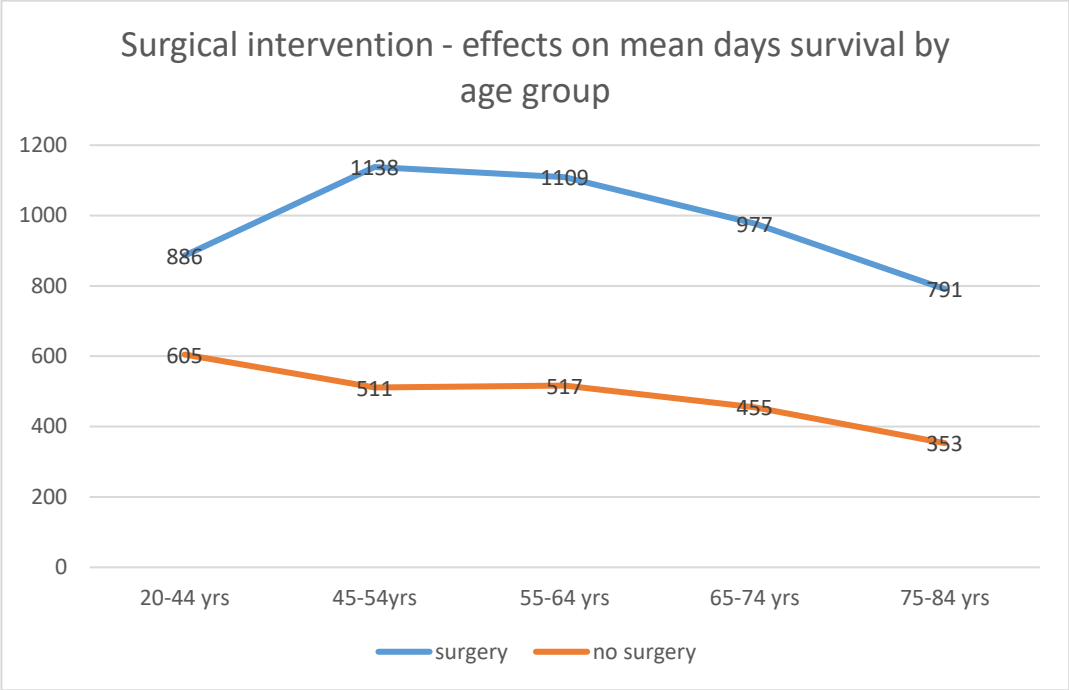


Figure 34 Surgical intervention and survival in days by age groups (over 85 omitted)

7.2.4 Survival and socioeconomics

A one way between groups analysis of variance compared survival outcomes with socioeconomic groups by IMD status. Subjects were apportioned an IMD score from their diagnostic clinicians and these were based on postcode of residence. There was a statistically significant difference at $p<0.05$ level in survival days between IMD 4&5 scores, when compared with IMD 1&2. There was also a statistically significant difference in mean scores between IMD 3 (mid category) and those who were least deprived. Despite reaching statistical significance, the actual difference in mean scores was extremely small (eta squared = 0.02) Cohen (1988) classifies this as a small effect. Figure 35 shows these differences, however, findings must be considered against the lower effect size.

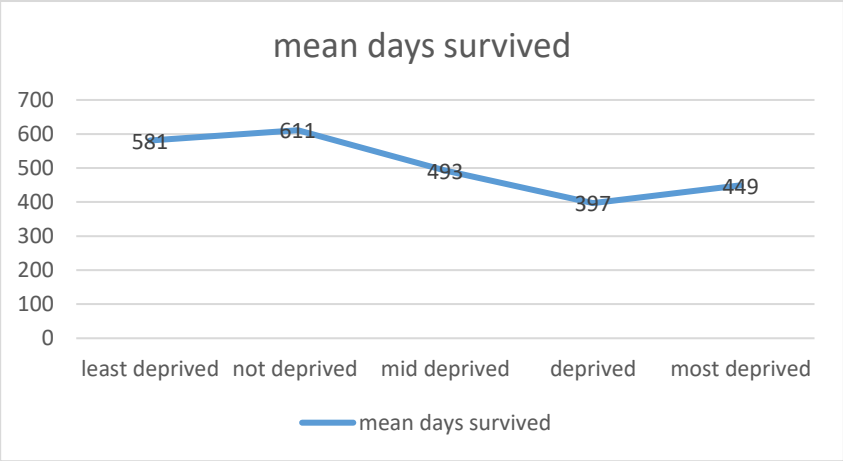


Figure 35 survival in days per IMD classification

These data may be skewed as the sample violated the assumption of homogeneity. The 'mid deprived' group was far smaller than the others. Given this, a nonparametric analysis was undertaken.

The Kruskal-Wallis test revealed statistically significant differences in survival days between groups. x^2 (4, n = 2215) = 38.6, p = 0.05, suggesting an association between lower deprivation scores and poor survival.

Figure **36** identifies cox regression analyses by deprivation groups at both 1 and 5 years.

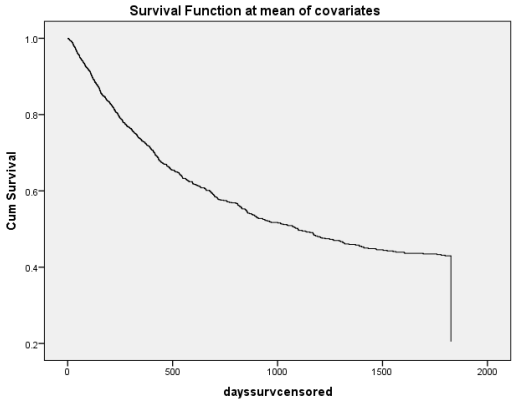
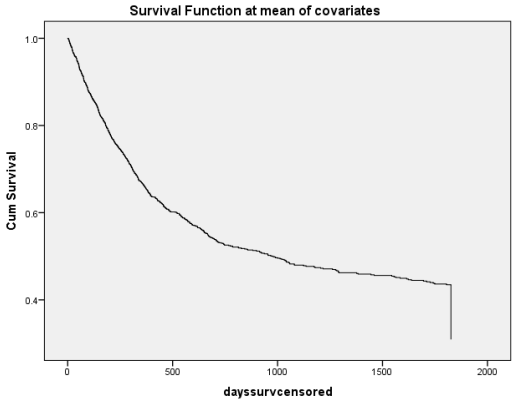
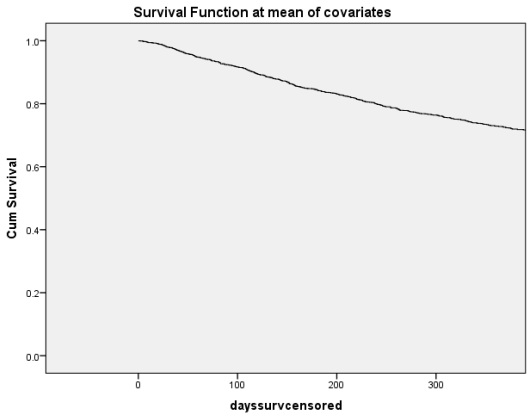
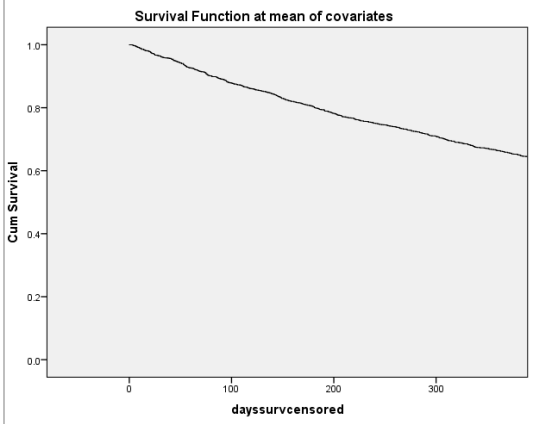
IMD10 groups 1 & 2 (least deprived)	IMD10 groups 4 & 5 (Most deprived)
5-year survival 	5-year survival 
1-year survival 	5-year survival 

Figure 36 Cox regression analysis by deprivation at 1 and 5 years

7.2.5 Survival and presentation by cancer site and morphology

All cancer sites had a significant drop in survival in the first six months of GOC diagnosis in the survival cohort (N2215 - Figure 37).

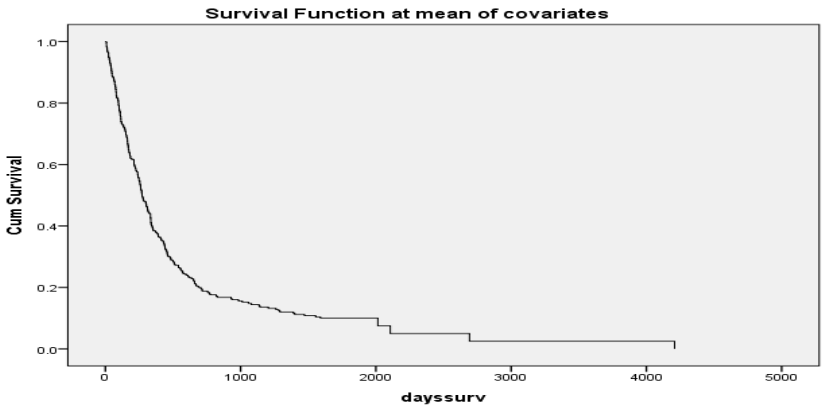
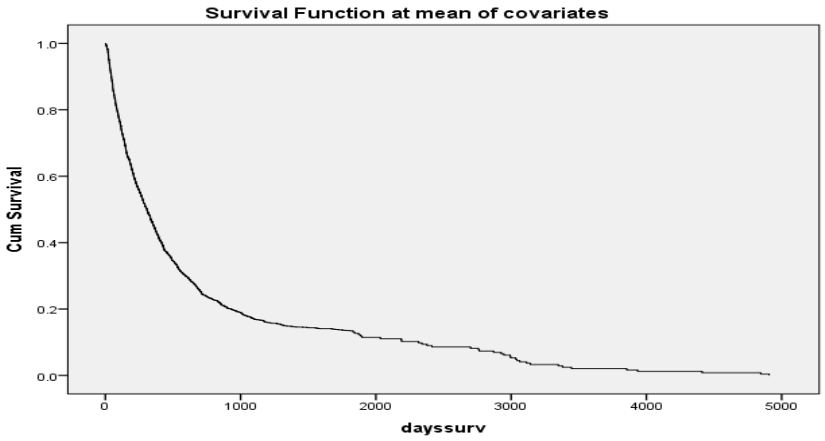
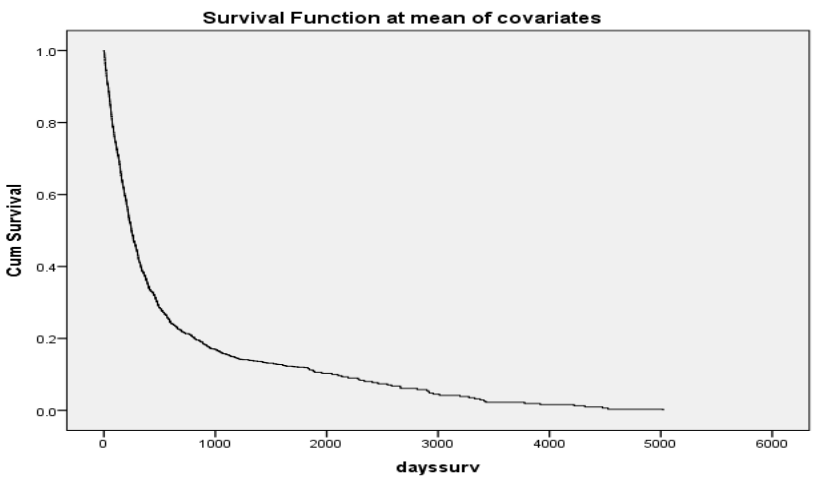
Upper Oesophagus N196	
GoJunctional N1176	
Oesophagus unspecified N843	

Figure 37 survival by site of cancer

Figure 38 identifies the different sites of cancer and the mean days survival (in days) in those groups.

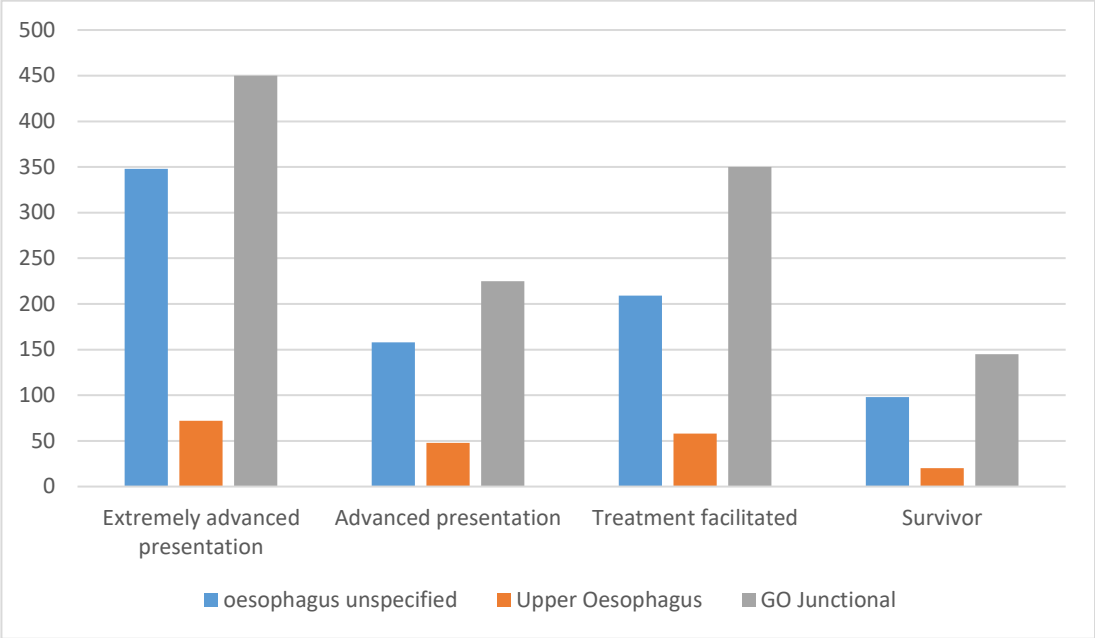


Figure 38 Presentation groups and mean days survival by site of cancer.

The morphology of the cancers in the presentation groups is represented in the following Figure 39. The mechanism of presentation is illustrated in Figure 40.

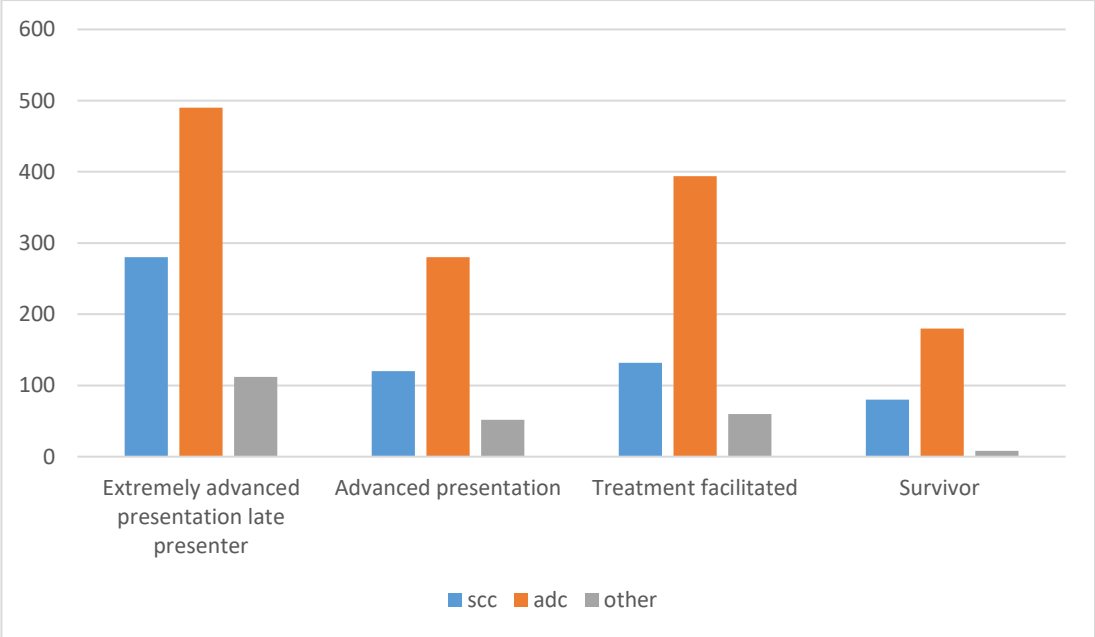


Figure 39 Presentation groups by Morphology

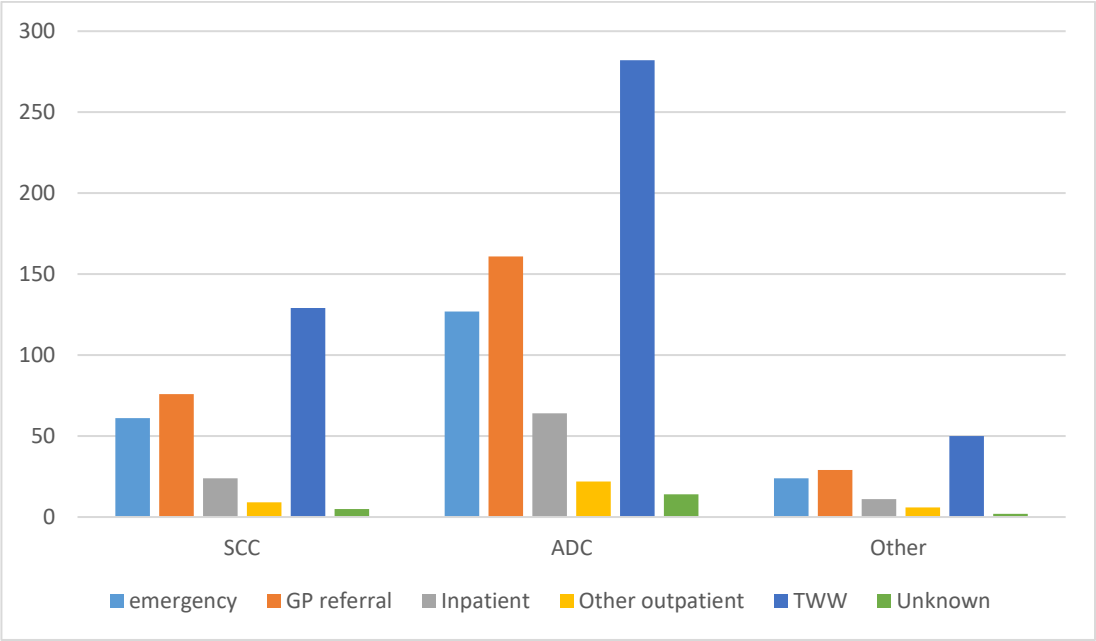


Figure 40 Mechanism of presentation by cancer morphology

TNM staging data were only present for 121 subjects out of the sample 2215. The following figure shows most subjects (89) presented in later stages TNM 3 & 4, and only 32/121 subjects presented in early stages.

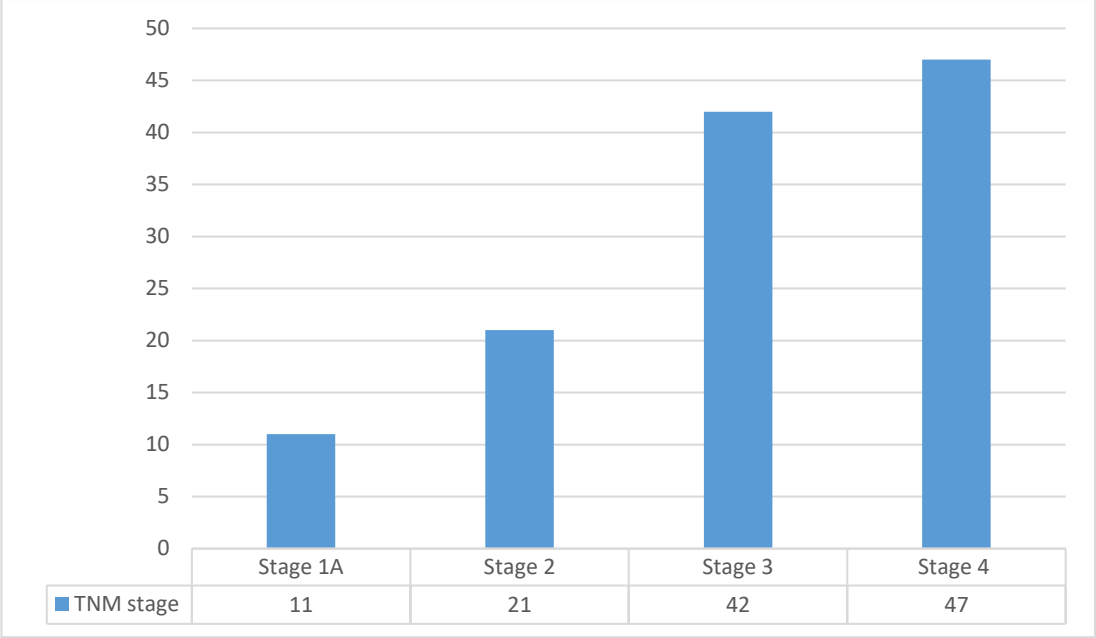


Figure 41 Number of patients presenting across TNM stages.

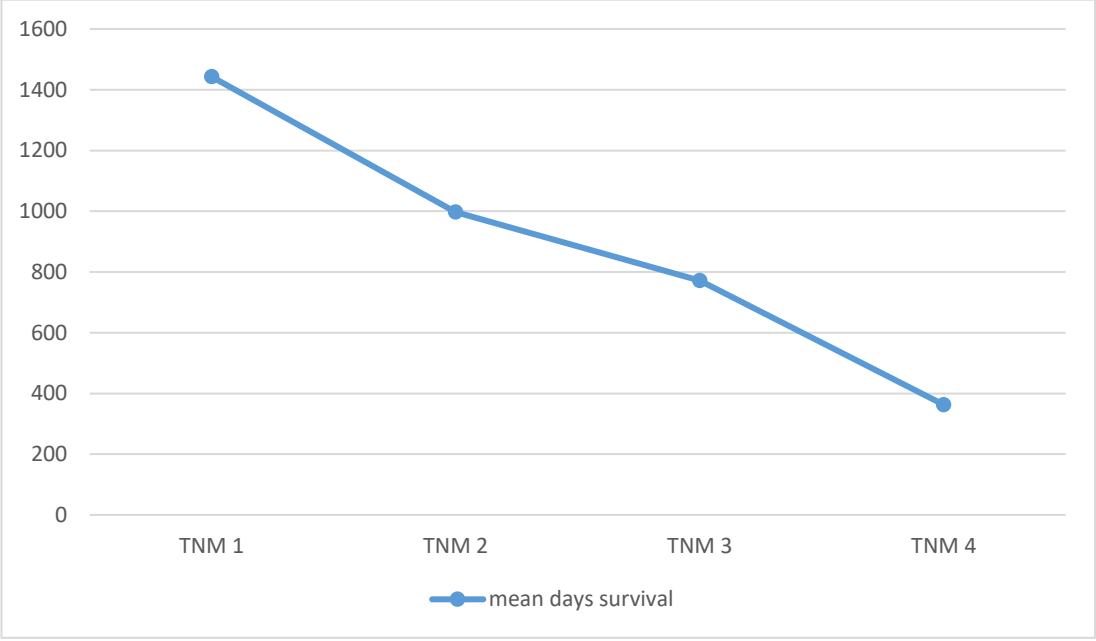


Figure 42 Mean days survival in TNM groups.

Mean days survival across TNM staged subjects (n 121) are shown in Figure 42.

The main GOC morphology in the survival cohort (N2215) was adenocarcinoma (ADC) with papillary and squamous cell (SCC) second. Cancers labelled as 'other' included neoplasms and cystic/mucinous morphology and these were diagnosed in 258 cases.

Mean days survival after diagnosis in squamous cell carcinoma was 485, Adenocarcinoma was 543 and all other cancers was 368 days. For SCC, 61% of the cohort died within a year and 11 % survived to 5 year. 57% of ADC sufferers died within a year of diagnosis and 13% survived over 5 year.

A Kruskal-Wallis test revealed a statistically significant difference in days survival post diagnosis between the three morphology groups (SCC n = 620, ADC N=1337, Other N= 258). X2 (2, N=2215) =16, p=.005. The ADC groups had a higher median score in days survived (MD291) than the other two groups, with values of MD 245 and 212 (Figure 43).

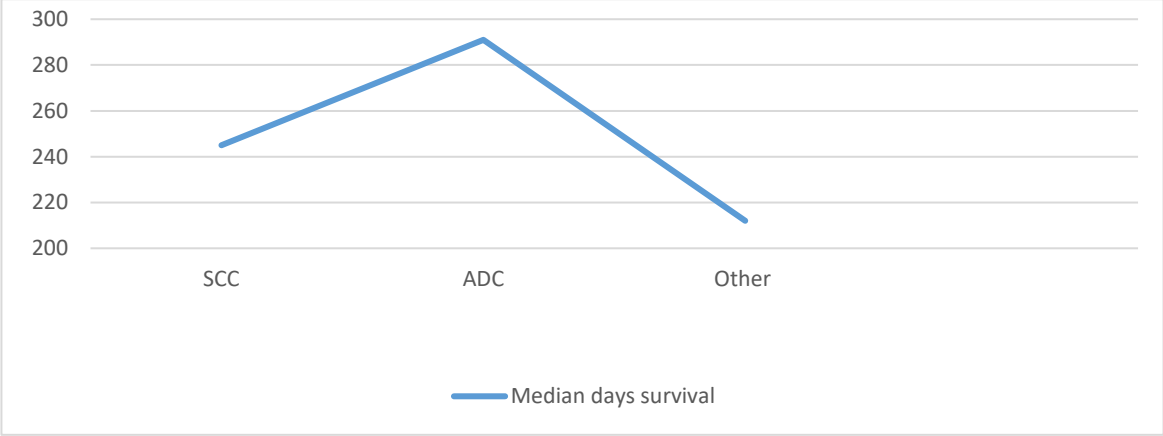
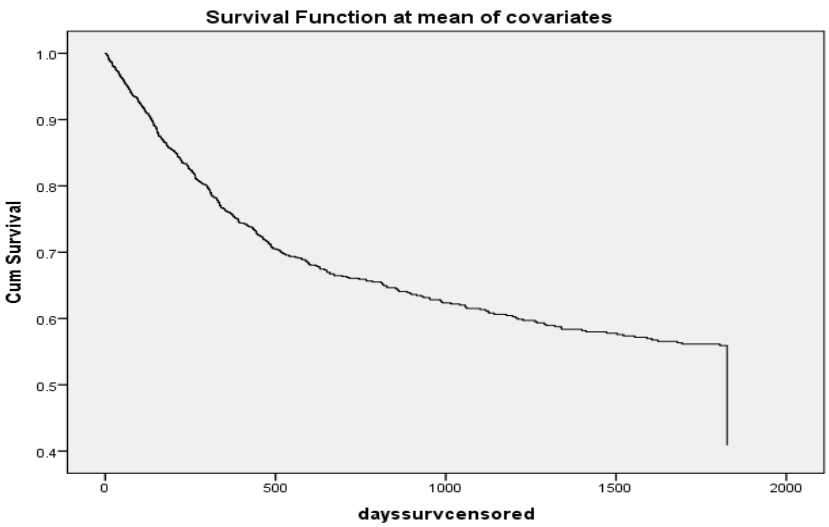


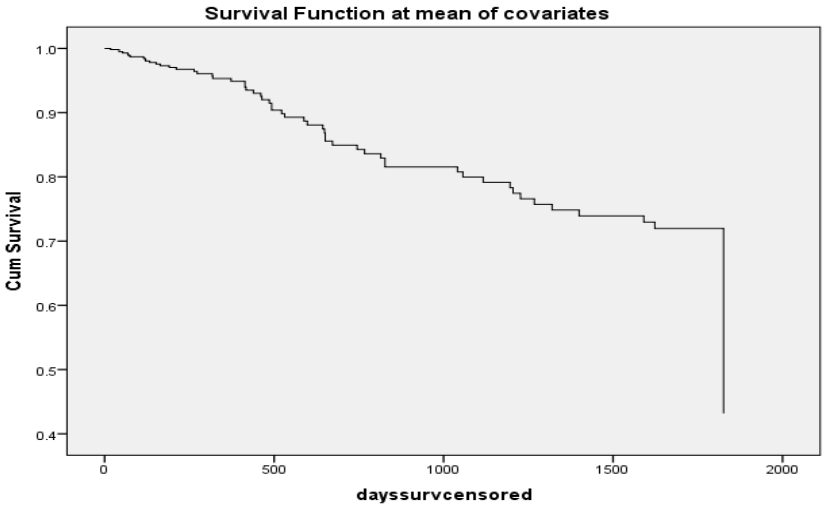
Figure 43 Comparison of survival in morphology groups

The next figure illustrates cox proportional survival times by morphology in GOC presentations, indicating survival functions decrease with time in all 3 groups (CI 95 P=0.05).

SCC (19% had surgery)



ADC (14% had surgery)



Other (17% had surgery)

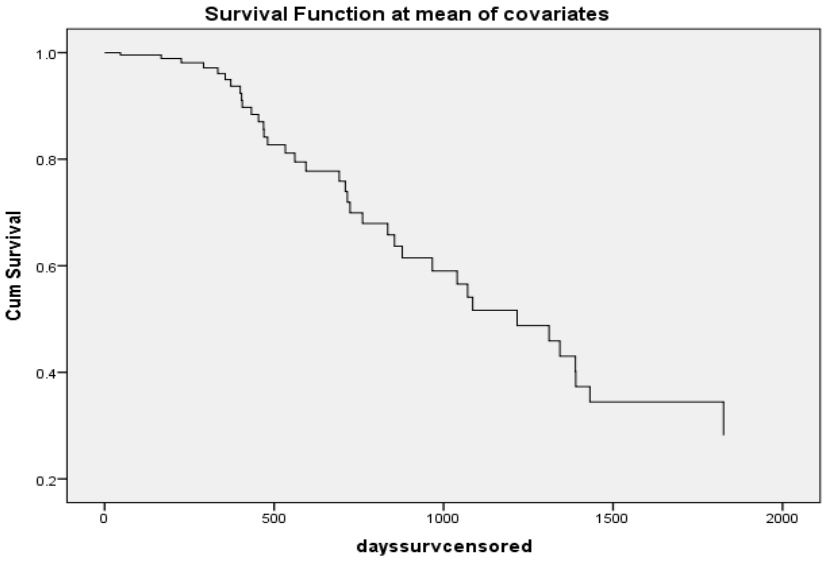
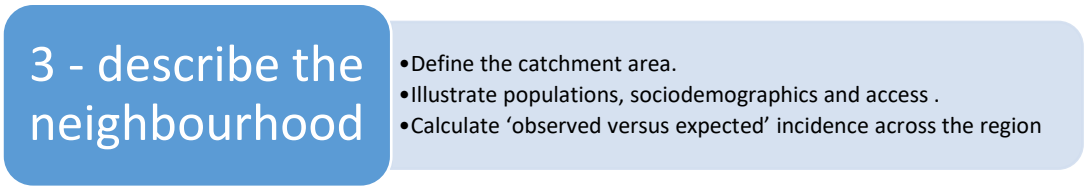


Figure 44 Survival by morphology

This analysis of the GOC cohort, their presentation and survival has identified several significant factors, though effect sizes are relatively small. From these data, it can be assumed that increased age and male gender are significant attributes found in GOC. The literature review highlighted that smokers and those who use alcohol on a regular basis have a higher propensity to GOC. There is also evidence to suggest environmental factors may be associated with a higher incidence (exposure to environmental pollution). The following section of this chapter describes the residential areas of this cohort, the population and lifestyle description of the 'place' in which GOC patients reside.

7.3 Describing the neighbourhood (Objective 3).

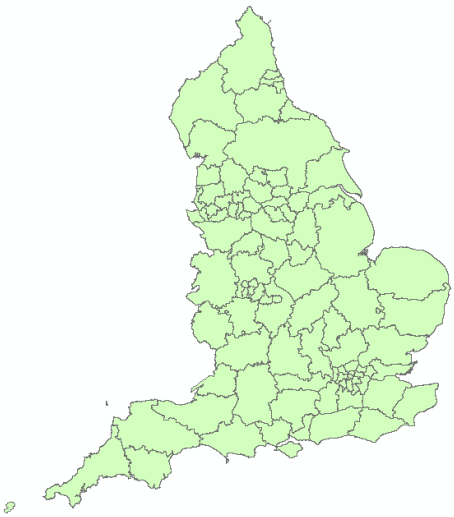


This section of the results chapter presents a neighbourhood description of the geographical catchment area, in relation to lifestyle and demographic factors**.** It applies population data on age, lifestyles, housing and socioeconomics to uncover potential areas with higher propensity to GOC diagnosis.

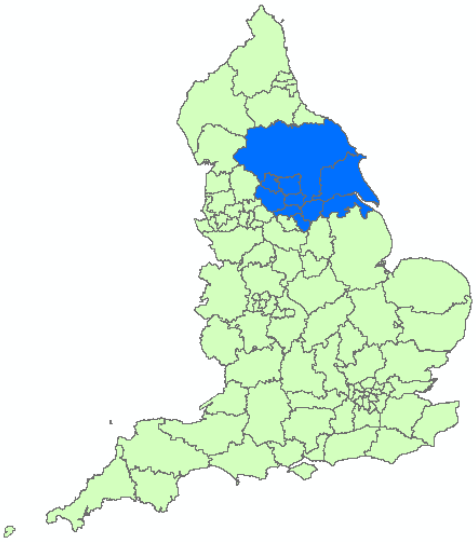
7.3.1 Defining the catchment area

The Humber and Yorkshire coast Cancer network nestles between various regional cancer networks (Figure 45). It sets an artificial boundary for the management of inpatient and community level cancer services provision and illustrates how the cancer network feeds into the varying NHS boundaries. Although some patients may be treated across networks and boundaries, this research focusses on the referral centres in the Humber and Yorkshire Coastal region. This also includes patients from Scarborough.

England health authority regions 2001



Yorkshire and the Humber region 2001



Humber and Yorkshire Coast – catchment area

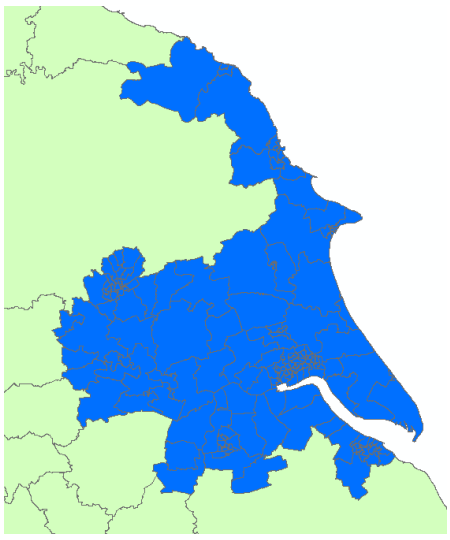


Figure 45 cancer network boundaries

Primary care trusts (PCTs) were the statutory bodies responsible for delineated 'areas' in health districts. They were responsible for budgeting and overseeing all care administered in geographical areas. The Health and Social Care Act (2012) have since replaced PCTs with local Clinical Commissioning Groups, but as this research relates to subjects presenting before 2013, PCT classifications are used in this research. The populations in East Riding of Yorkshire, Hull Teaching, North Lincolnshire, North East Lincolnshire and North Yorkshire and York PCTs are not equal (Figure 46).

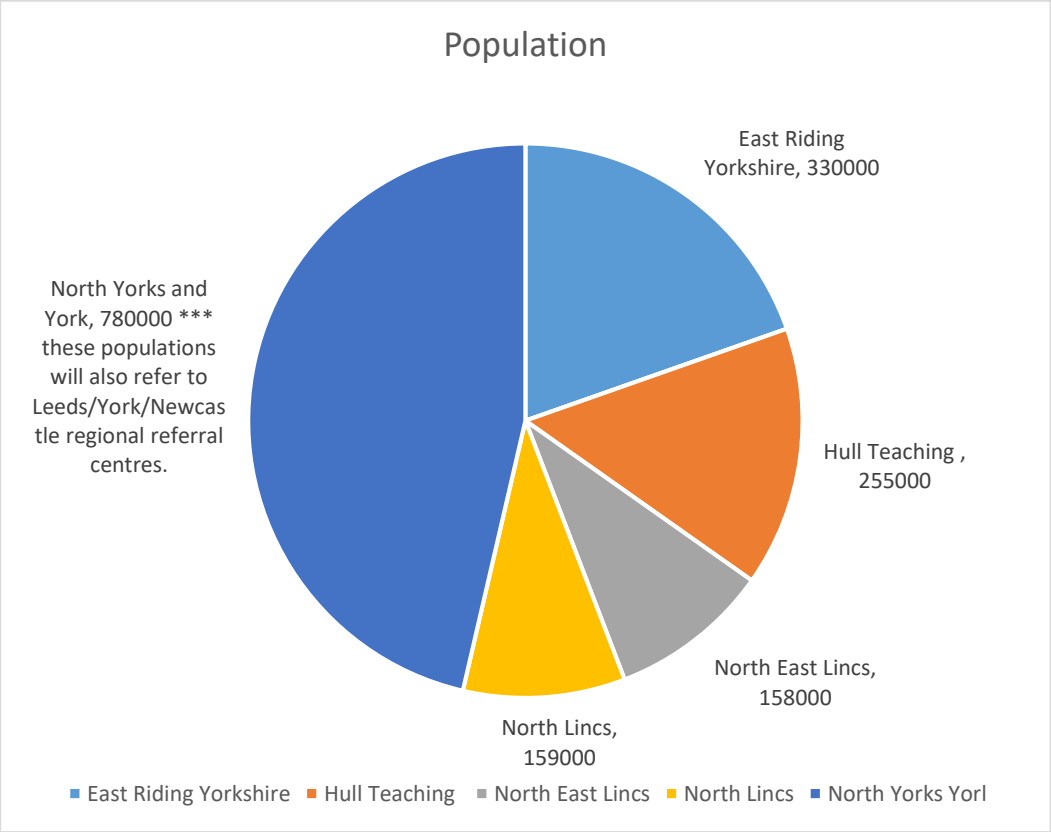


Figure 46 PCT populations (Northern and Yorkshire Cancer Registry and Information Service, 2010)

Most subjects who received histological confirmation of GOC resided in the East Riding of Yorkshire, however incidence rates/100000 population was analysed using the following formula:

N persons diagnosed 2000 – 2013 x 100000 / PCT population 2000-2013

These are identified in Figure 47.

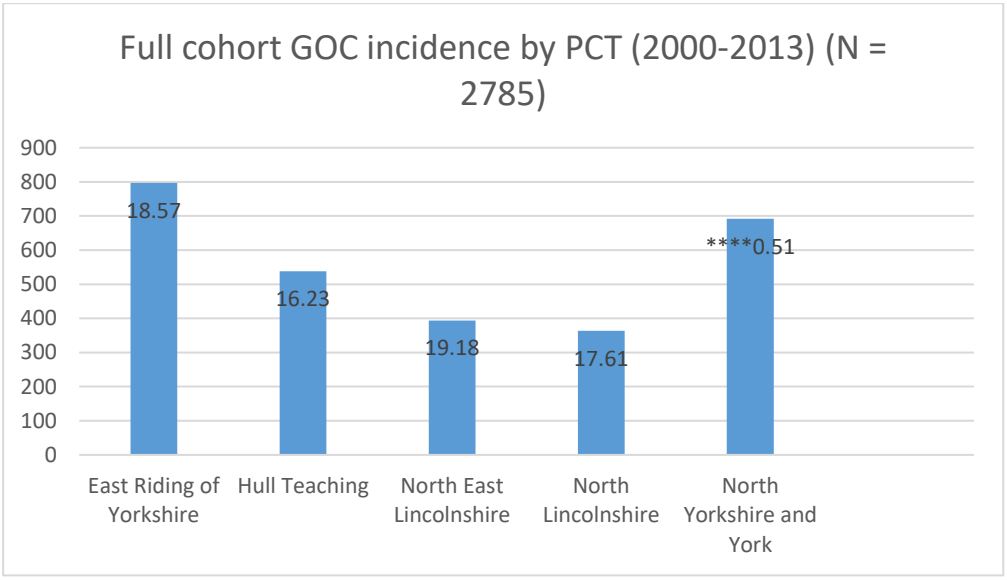


Figure 47 Cumulative GOC incidence and incidence rates/100000 population by PCT

Whereas most diagnoses within the regions of Hull, East Riding, North East and North Lincolnshire would be captured by the regional referral centre, many diagnoses within the North Yorkshire and York would be captured by a neighbouring PCT (through Leeds and York). It must therefore, be noted that the North Yorkshire and York incidence rates identified in this graph are taken only to represent those patients referred to the Hull centre. It therefore, does not represent incidence rates/100000 populations.

7.3.2 The regional referral centre and its population

The regional referral area serves a population of approximately 1,682,000 people. Figure 48 identifies how this is geographically placed.

Figure 48 The regional referral centre catchment area

The UK undertake census surveys every 10 years to gather information on populations. These data are produced to several geographical scales, so the researcher must make choices dependant on how well the data releases can fulfil the research question. For this research, population data must represent the cohort, but the scale for analysis must fulfil both confidentiality and be appropriate to the health service boundaries.

For these reasons, population and geographic data were gathered from the 2001 census. Census 2001 provide details of populations and lifestyles at LSOA and MSOA levels, which allow the researcher to choose which level of enumeration fulfils the brief. Figure 49 identifies these enumeration levels, MSOAs offer a larger boundary than LSOAs. The geographical catchment area for this cohort study contained around 170 MSOAs and over 800 smaller scale LSOAs (LSOAs are coterminous with MSOAs). LSOAs and MSOAs are amended on a regular basis to account for population changes and shifts. This results in some discrepancies between total numbers of output areas across census surveys. MSOA boundaries contained an average 17 GOC subjects (range 0-35), whereas LSOAs contained an average 6 subjects (range 0-21). This meant that incidence data at LSOA level had a high potential for quasi identifiability. Only 10 of the MSOAs contained had less than 10 subjects, as opposed to the 701 LSOAs with small numbers. Consequently, MSOA aggregate levels were identified as the smallest scales which most appropriately maintained anonymity. This provides the rationale for extraction of national statistics' population level data at MSOA level. The server CASWEB provided the 2001 population census data stratified by age and gender for the purposes of this research. The total population of MSOAs varied greatly, with a range between 37115 and 181675 (mean 102058).

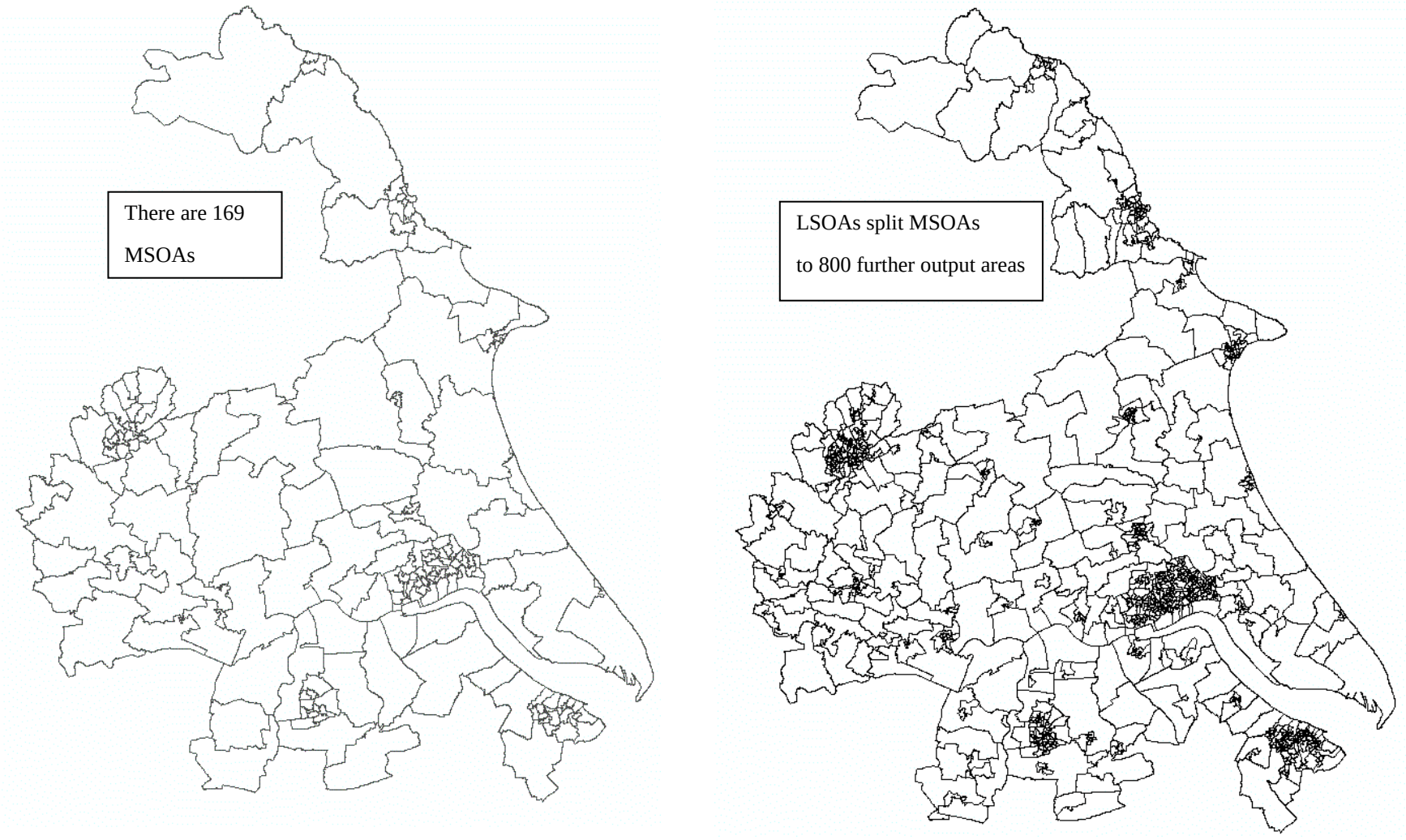


Figure 49 MSOA and LSOA illustration

The following figures identify how many people and how many homes there are per LSOA scale. This is relevant to the analysis of population density and GOC incidence later in the thesis. The maps illustrate areas of increased and decreased density in populations. A Total of 524197 homes are identified in ONS statistics across the area, with a mean 636 homes per LSOA. The following map identifies areas with density in homes per LSOA.

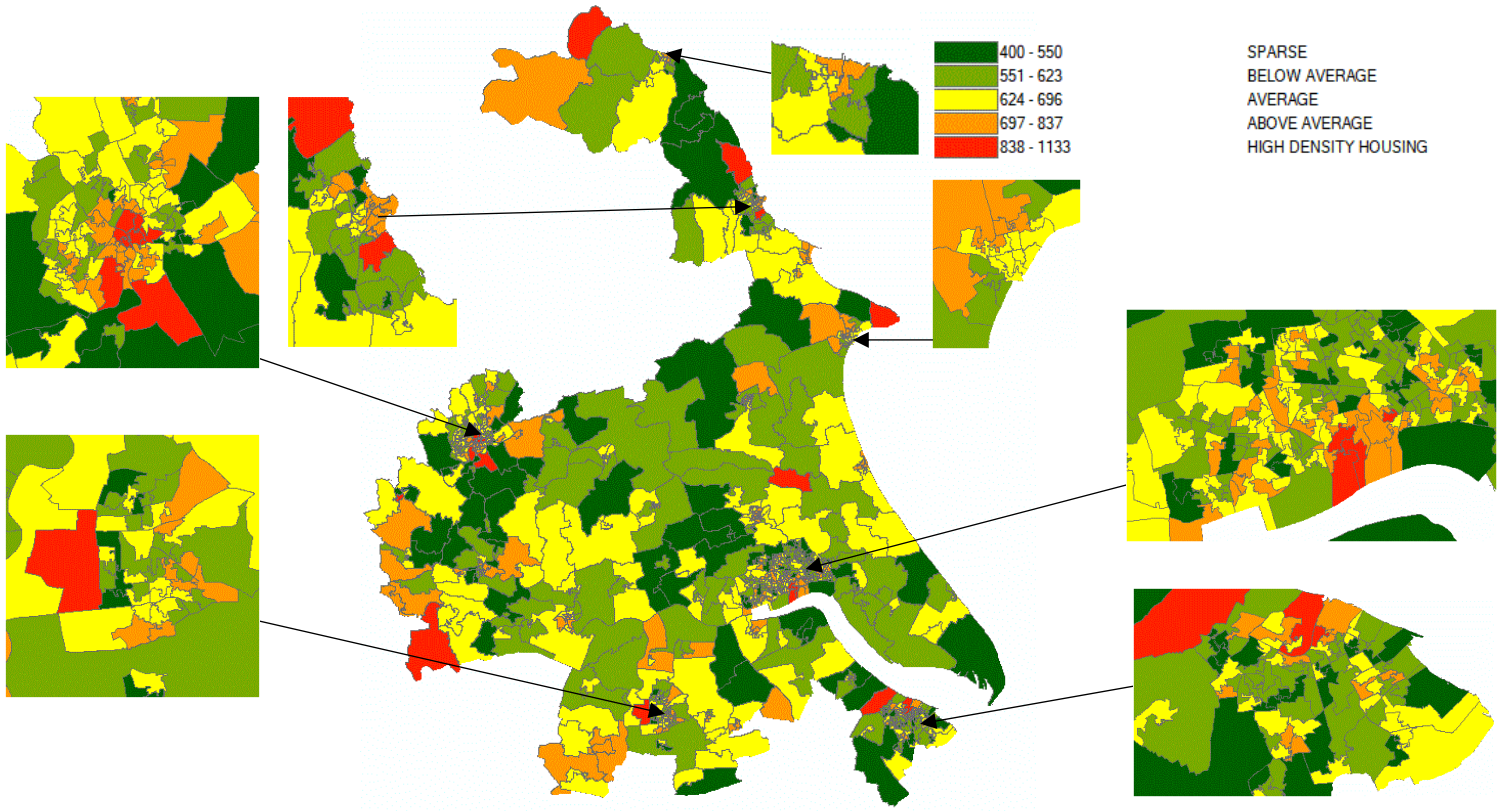


Figure 50 Density of homes per LSOA

The next figure shows mean density of populations per LSOA. The total population across the region was 1241566. The median population of all LSOAs is 1500 (SD 184) persons per area.

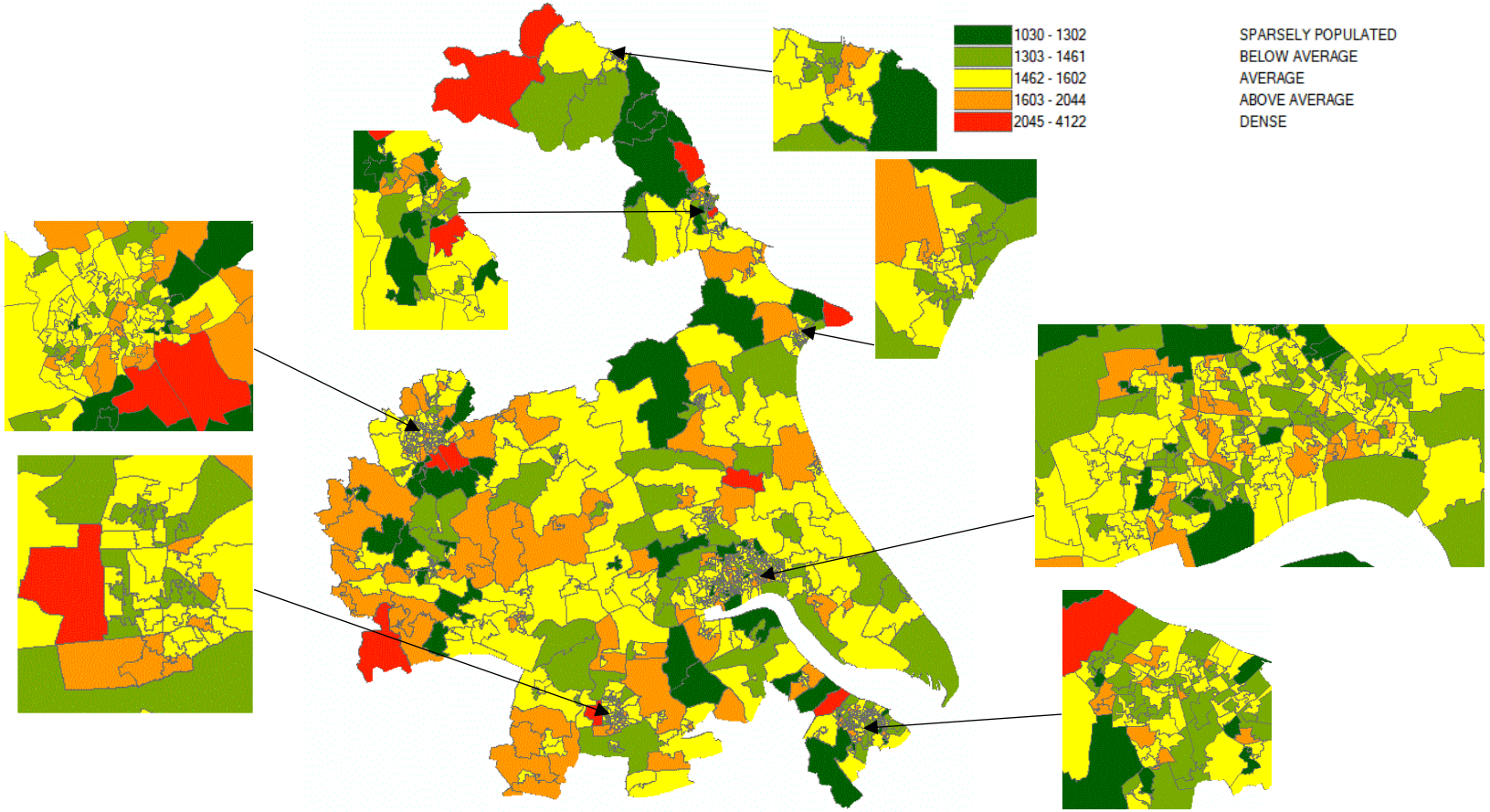


Figure 51 Population counts per LSOA

7.3.3 Illustration of area by incidence (then observed versus expected incidence).

Figure 52 identifies the total incidence rate for GOC over the 14 year period, by morphology (N = 2785). Annual incidence rates varied between 167 and 242 per year (mean 199 SD20). The mean period prevalence rate (2000-2013) was calculated using the following formula;

$$\frac{total\ persons\ with\ GOC}{total\ years} = \frac{2785}{14} = 199.$$

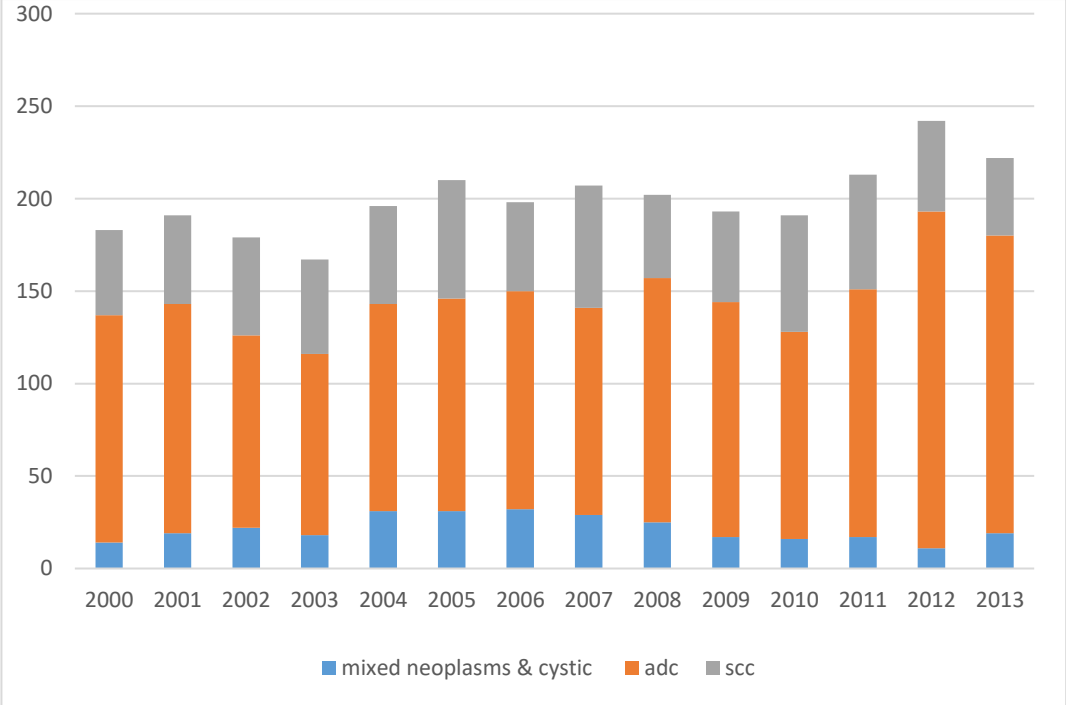


Figure 52 Annual GOC diagnosis (by morphology)

The following two maps illustrate incidence across the regional referral centre by LSOA. There were more males than females in this cohort, which was expected (N1929 males versus N 826 females). The average number of diagnoses per LSOA across the timeframe were 12 males (SD5) and five females (SD3).

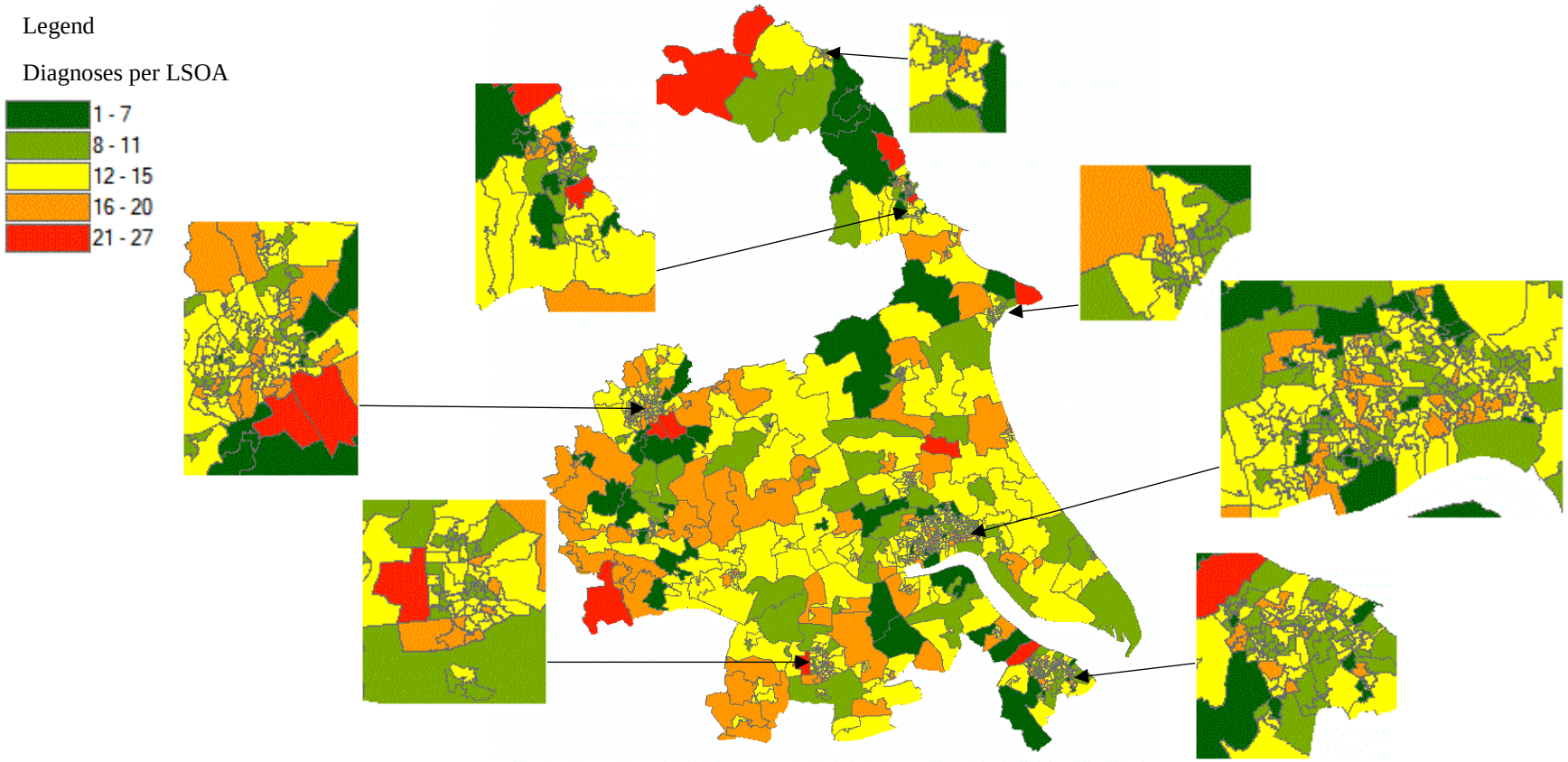


Figure 53 Regional map male GOC (N1929)

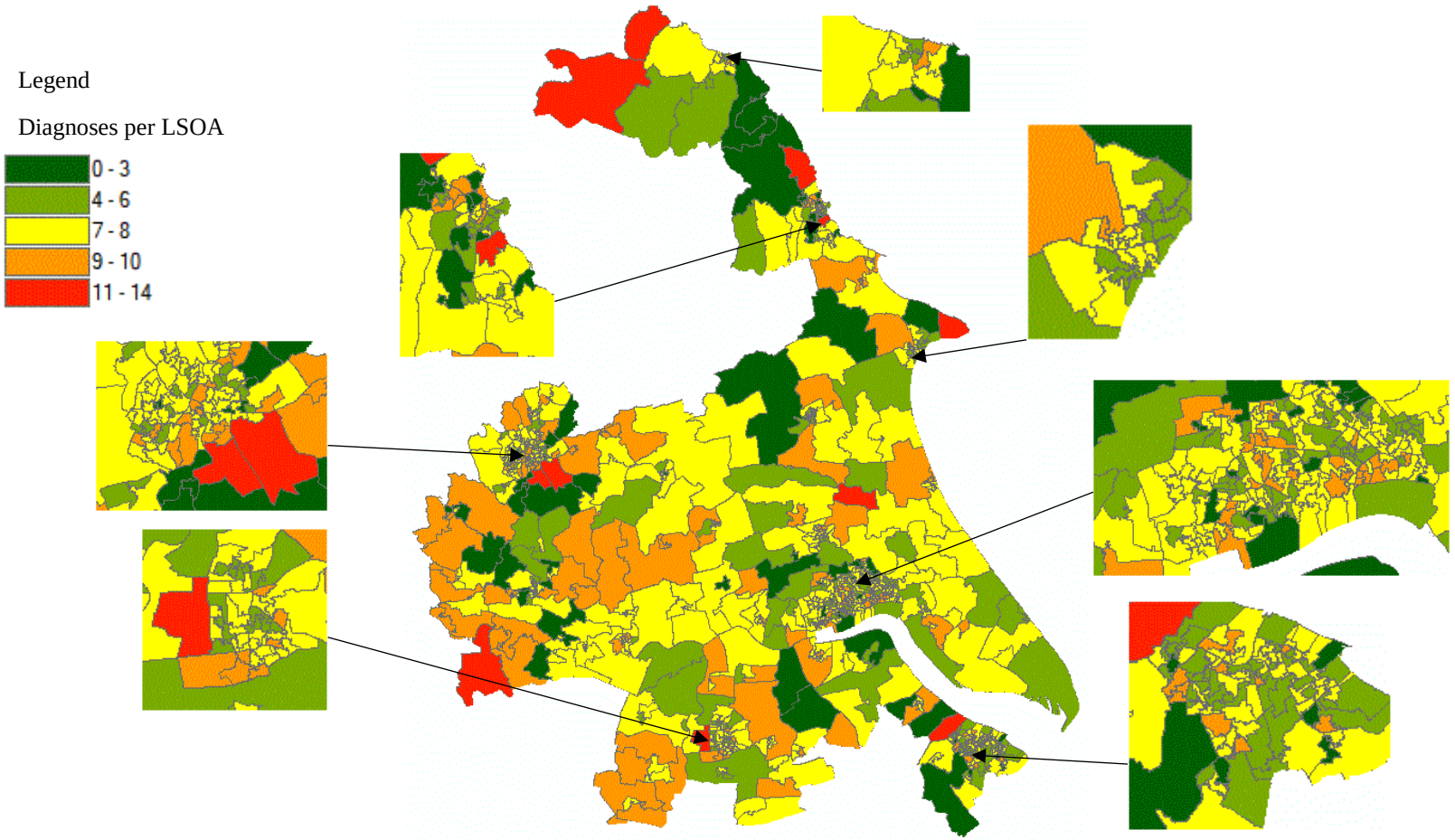


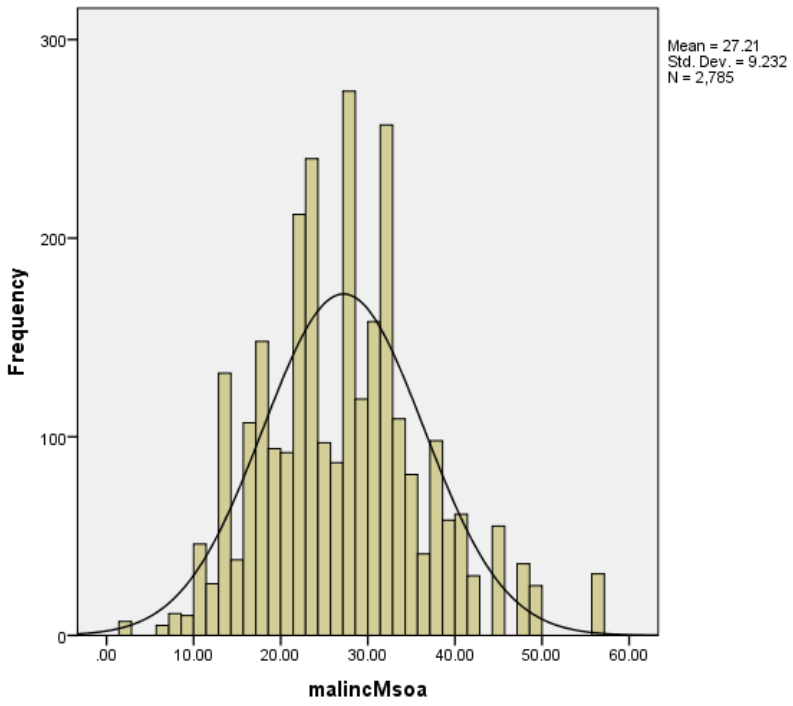
Figure 54 Regional map female GOC (N826)

7.3.4 Observed versus expected incidence

The previous two maps merely identify prevalence of GOC during the years 2000-2013 and indicate where incidences were higher. However, these do not offer a context. Comparison against expected rates is required to identify any areas with higher or lower than expected rates.

To support this, incidence data were presented in the geographical information system and calculated against 2001 populations per MSOA. Quintile classifications were based on SIR data published by cancer research (UK) (CRUK) 2014. This analysis revealed areas with higher than expected incidence (Figure 55). Of 165 MSOAs where diagnoses occurred in the catchment area, 99 MSOAs had higher than expected incidence of male GOC and 56 had a lower than expected incidence of male GOC (Figure 56). Of the 162 MSOAs where diagnoses occurred for females, 93 MSOAs had higher than expected incidence and 69 MSOAs had a lower than expected incidence (Figure 57).

Male Incidence by MSOA (Exp = 22-23)



Female incidence by MSOA (Exp = 8.6-9.2)

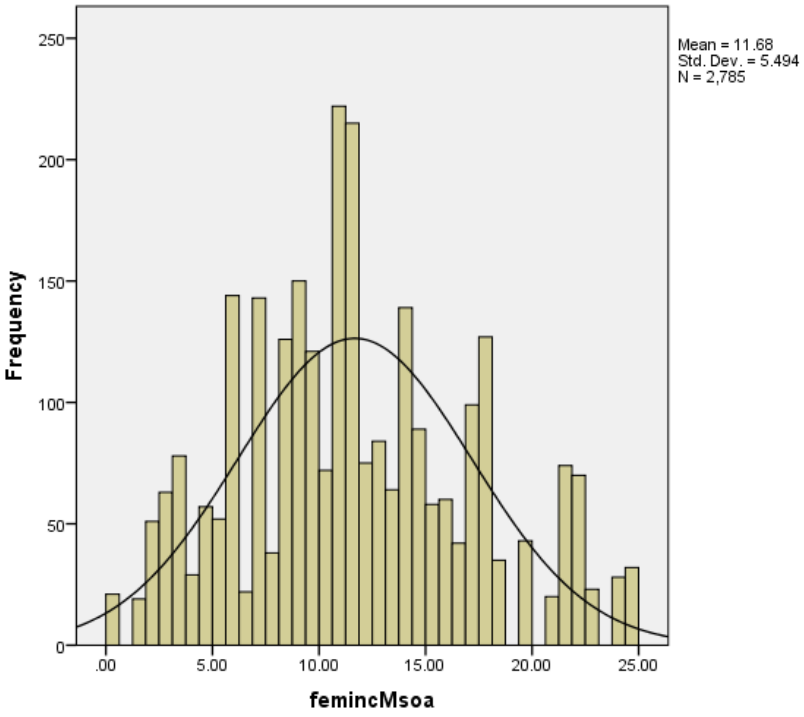


Figure 55 Male and Female incidence of GOC by MSOA

To calculate observed versus expected incidence rates in cancer data, there are several processes to be undertaken. These cohort data span a 14-year period, so temporal nature of these data needed to be considered.

The period prevalence (GOC subjects presenting 2000-2013) was divided by the average size of the populations over the time period. The catchment area population was 1,682,000 people, giving a period prevalence of 604. However, the population structures do not account for the skew to diagnosis of GOC in elderly, male populations, so this had to be considered. An age standardised, profile of MSOAs across the region was applied to gender specific population profiles (giving total persons at risk during the follow up period). This was applied to identify incidence rates per MSOA by gender and then presented in mapped form via the geographical

information system. This must be considered as descriptive data, as inferences cannot be made on findings. The denominator (underlying population at risk) is taken from census survey (2000) and may not be as accurate as required, however, this crude measurement offers some insight into EU compared with local level GOC incidence, with the caveat that population estimation may not be as accurate as required. One factor to consider, however, is that the same claims could be made against EU standardised measurements and their population data.

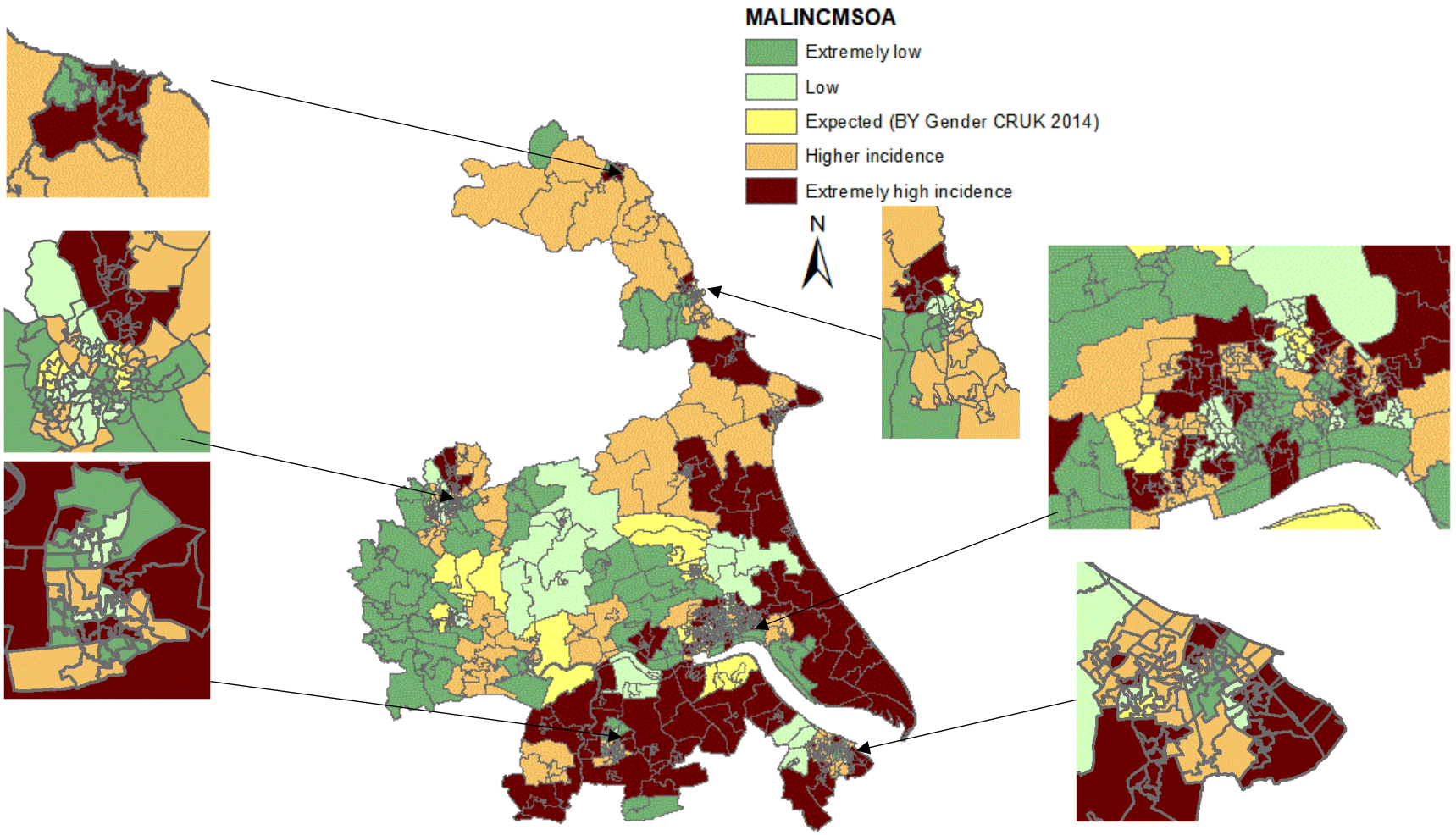


Figure 56 Observed versus expected Male incidence per MSOA (standardised to CRUK 2014)

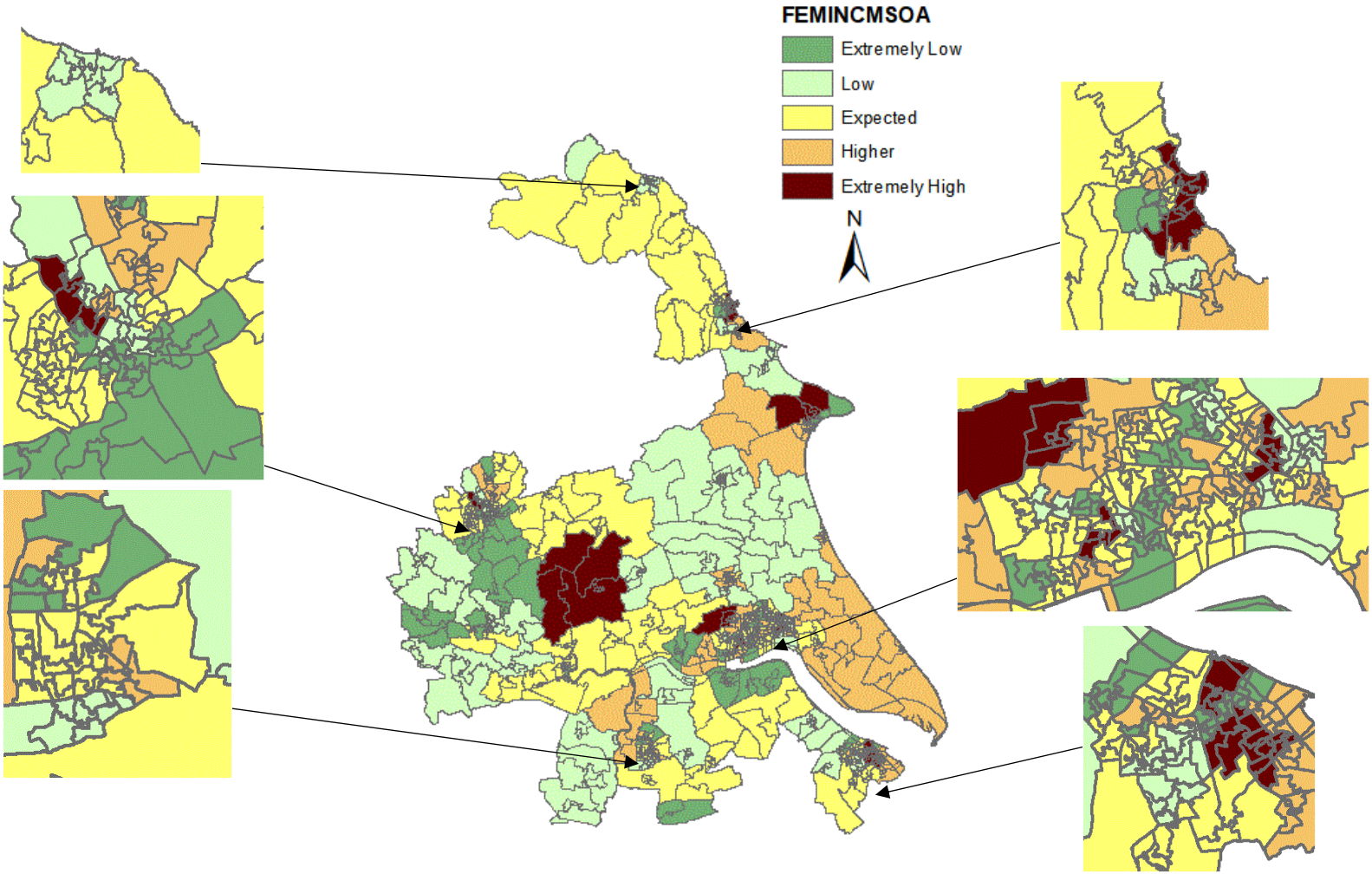


Figure 57 Female observed versus expected incidence per MSOA (standardised to CRUK 2014)

These maps illustrate observed versus expected rates, normalised for populations and gender and compared against age standardised incidence data from Cancer research UK (CRUK, 2014). This methodology has been applied in a range of papers to larger enumeration scales (district and PCT area scaled output) and are commonly available through cancer research websites. However, the larger the scale, the less clinically relevant they become.

By generating maps at MSOA scales, they offer a descriptive analysis of past events. They cannot identify where future events may occur or detect any of higher than expected incidence.

7.3.5 Illustrations of life style and social indicators across the region.

The following maps (Figure 58 and Figure 59) provide social area descriptors across the catchment area. Norman's (2010) Area Classification (ONS-AC) are shown in the first map. These classifications are based on a K-Means clustering algorithm evaluated against national statistics. They provide a geodemographic classification to stratify LSOAs into 'similar' components of the ONS Super-Groups, to highlight areas of deprivation, worsening health and living environments, and to identify the more disadvantaged LSOAs. The map highlights areas of disadvantage based on these descriptors and the map is shown to highlight the significant range of 'countryside dwellers' present across the catchment area.

The second map identifies the IMD 2010 classification across the region. Deprivation is calculated relatively to all other UK output areas. Indices highlight areas which display the population's 'unmet needs', and their lack of available resources. The English Indices of Deprivation offer a quintile classification of deprivation based upon a broad range of issues.

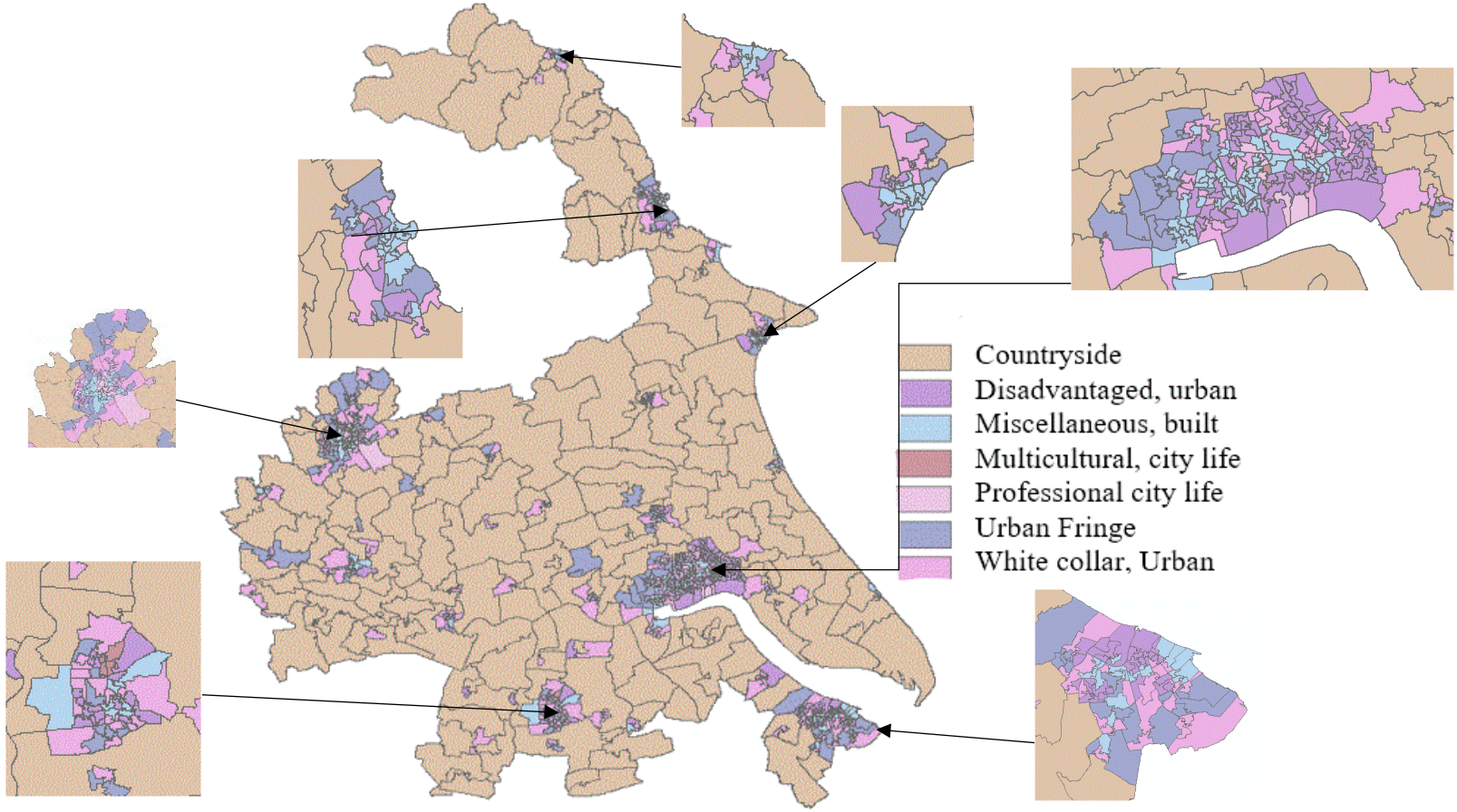


Figure 58 Area descriptions – Norman (2010) Advantaged/ less advantaged areas

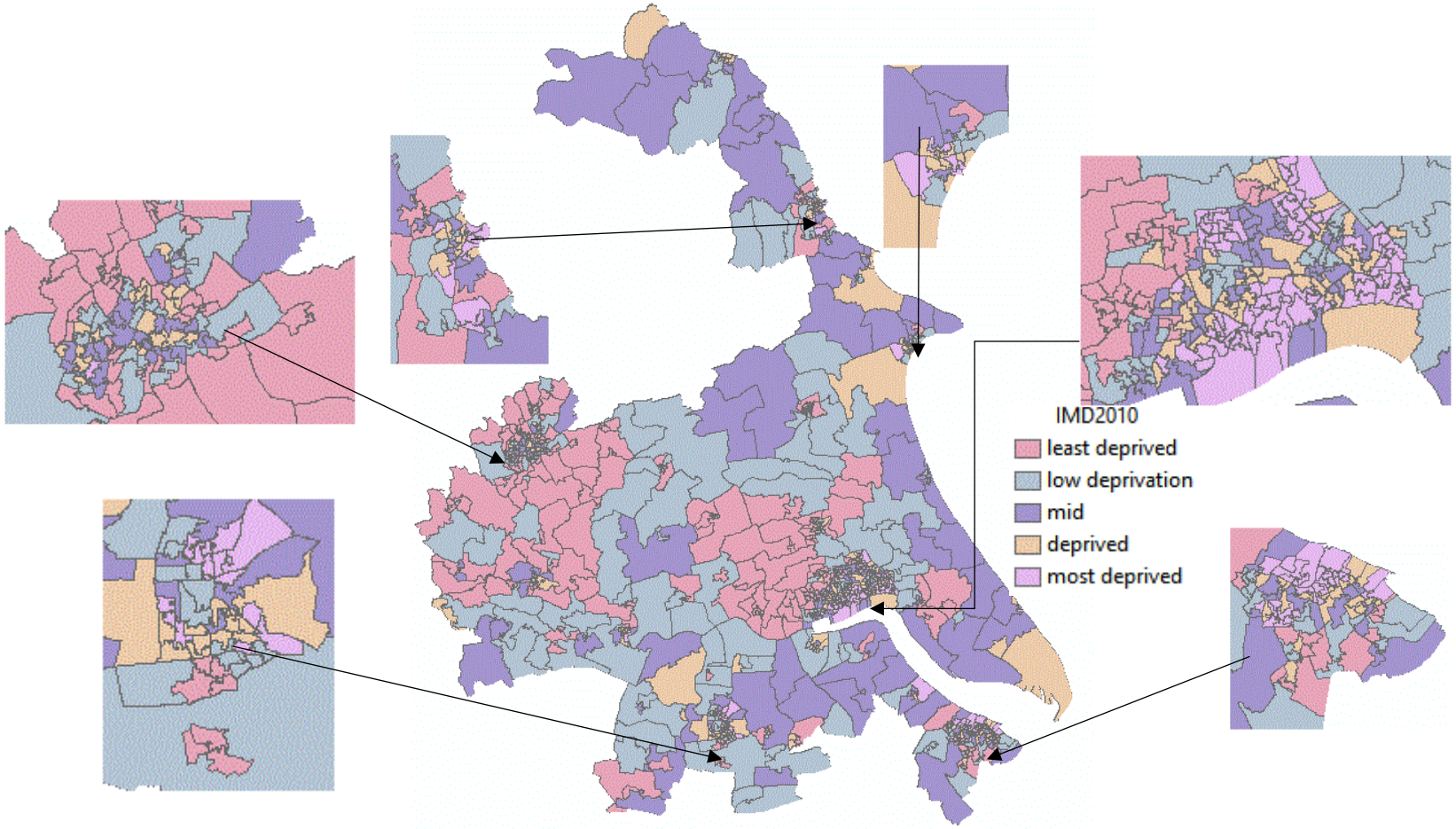


Figure 59 IMD 2010 across the region

The following figures provide a catchment area descriptor for access to GP and hospital services. There are areas across the catchment area where access to a GP is in excess of 15 minutes by public transport (these are in red).

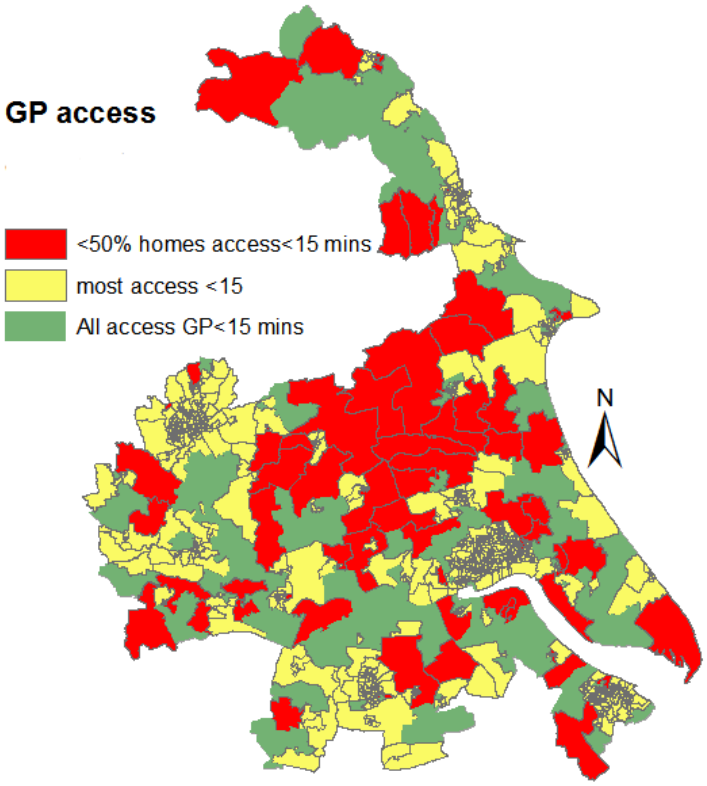


Figure 60 GP access via public transport (within a 15 minute timescale)

Many homes require journeys longer than 30 minutes to get to a hospital when using public transport.

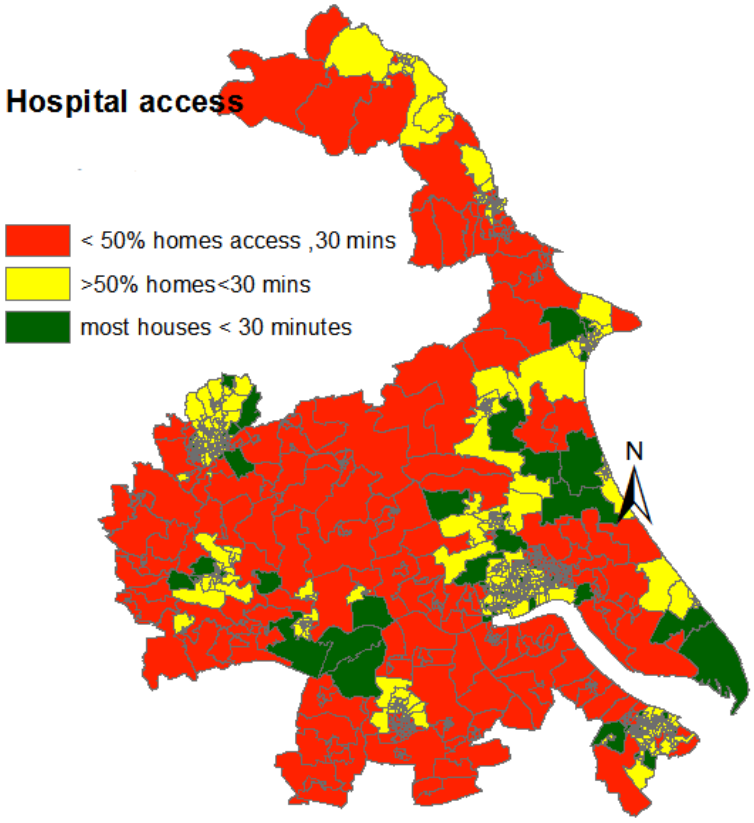


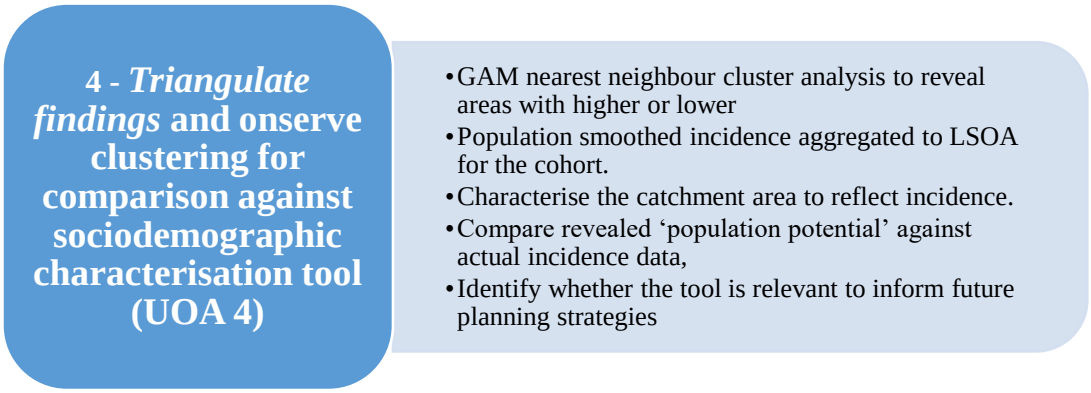
Figure 61 Homes with over 30 minute journeys to hospital via public transport across the region

The following map shows the geography of the catchment area to review against UK averages for smoking and alcohol. These data are only available at very large levels of enumeration so are not necessarily clinically relevant, however, they can be applied to multi criteria modelling in the GIS.

Figure 62 Smoking & alcohol levels (compared to UK average - see tables in protocol chapter)

These attributes will be used to base layers in the multi criteria evaluation model and appropriately weight the model to produce a ‘suitability map’ depicting areas at higher need for intervention.

7.4 – Triangulating findings to reveal any clusters for comparison against sociodemographic characterisation tool (Objective 4).



This section triangulates findings from objectives 1, 2 and 3 to reveal any geographical sites displaying clusters in incidence or survival outcomes.

It is presented to inform patterns and spatial relationships on incidence of GOC across the timeframe of the cohort. The chapter begins with the point distribution of rates, using postcodes of patients to define where they live. It then presents two forms of cluster analysis to identify whether there are any spatial patterns in the data. A local clustering technique through a geographical analysis machine (GAM) is presented to identify the initial question – are there any clusters and if so – where? Once this question is answered, the research can then move on to look more closely at explanations for the patterns exhibited.

7.4.1 Looking for crude clusters in incidence data

Openshaw's Geographical Analysis Machine (GAM) was applied to identify whether there was any geographic patterns to incidence. Openshaw's GAM allows a 'nearest neighbour' statistical analysis of incidence across LSOAs in the region, by applying concentric circles to capture incidences across the region. This nearest neighbour statistic provides a ratio of the observed average distance between nearest neighbours

of a point distribution – against the expected average of nearest neighbour distance. For this GAM, point data were taken as the centroid of patients' postcodes, applied to northing and eastings. The numerator was classified as incidence per LSOA and denominator as number of homes per LSOA. Of all the LSOAs in the region, only 741 had GOC diagnosed patients. To assess 'nearest neighbour' radial scales were set to maximum 5000 and minimum 250, with increments of 250m and circle overlap of 0.5. (Figure 63)

The next approach to reviewing these data was to look for clusters in incidence. Mapping 'hotspots' is a way to visualise the geographic dispersions of data. Any distribution of data across a mapped, or defined area can produce a display which demonstrates a range from complete clustering, or complete dispersion. Identifying these patterns as random, or causative is key to data analysis. There are many approaches to cluster analysis, and many aim to reduce the 'noise' or extraneous variables which have a potential to artificially constrain data (or cause artificial 'clustering').

The geographical information system enables data to be viewed and reviewed outside artificial boundaries. This case study aims to offer an illustration – it presents a descriptive analysis of retrospective incidence data at a very small scale and can therefore make no inferences into clustering (unless this clustering was found to be remarkable). Actual cancer data was plotted into the geographical information system to reveal higher or lower density of cases. The densities cannot be linked with causative factors and there is no suggestion within this research that previous presentations may predict future incidence.

This research applied LSOA level population statistics to reveal areas where observed incidences were higher. However, these were artificially constrained to the artificial MSOA populations. Oppenshaw's Geographical Analysis Machine (GAM) works by plotting concentric circles at each data point of a grid covering the study area. Neighbouring circles are permitted to overlap to capture more cases of the disease and then incidences within each circle are counted. The rationale for this approach was to provide a crude estimation of GOC subjects' residences, against the number of homes within each LSOA. Homes per LSOA was felt to be representative of patient homes within the postcodes, and did not require gender specific plotting.

It must be noted that 'homes' can be defined in terms of households and so their inhabitants change between residencies. This statistic enabled the GAM to be undertaken with the exclusion of places of residence which included areas such as residential care homes, hospitals and prisons. Concentric circles were drawn across the data area and 8 main 'hotspots' were identified.

The initial GAM revealed areas north of the northern part of York, Driffield, Bridlington, North and East Hull. Withernsea and Patrington , South Killinghome and Grimsby districts had potential clusters of incidences in Gastroesophageal cancer between the years 2000 and 2013. However, several factors must be taken into account when analysing these crude results. The age profiles of North York, for example, display higher 'elderly residents' than some other populations across the region. Similarly, gender of residents in the industrial areas noted around Grimsby and districts, associated with transient worker populations, could merely indicate the clusters as artefact. It is essential that any research which displays this clustering must identify potential factors which can allude to the clusters.

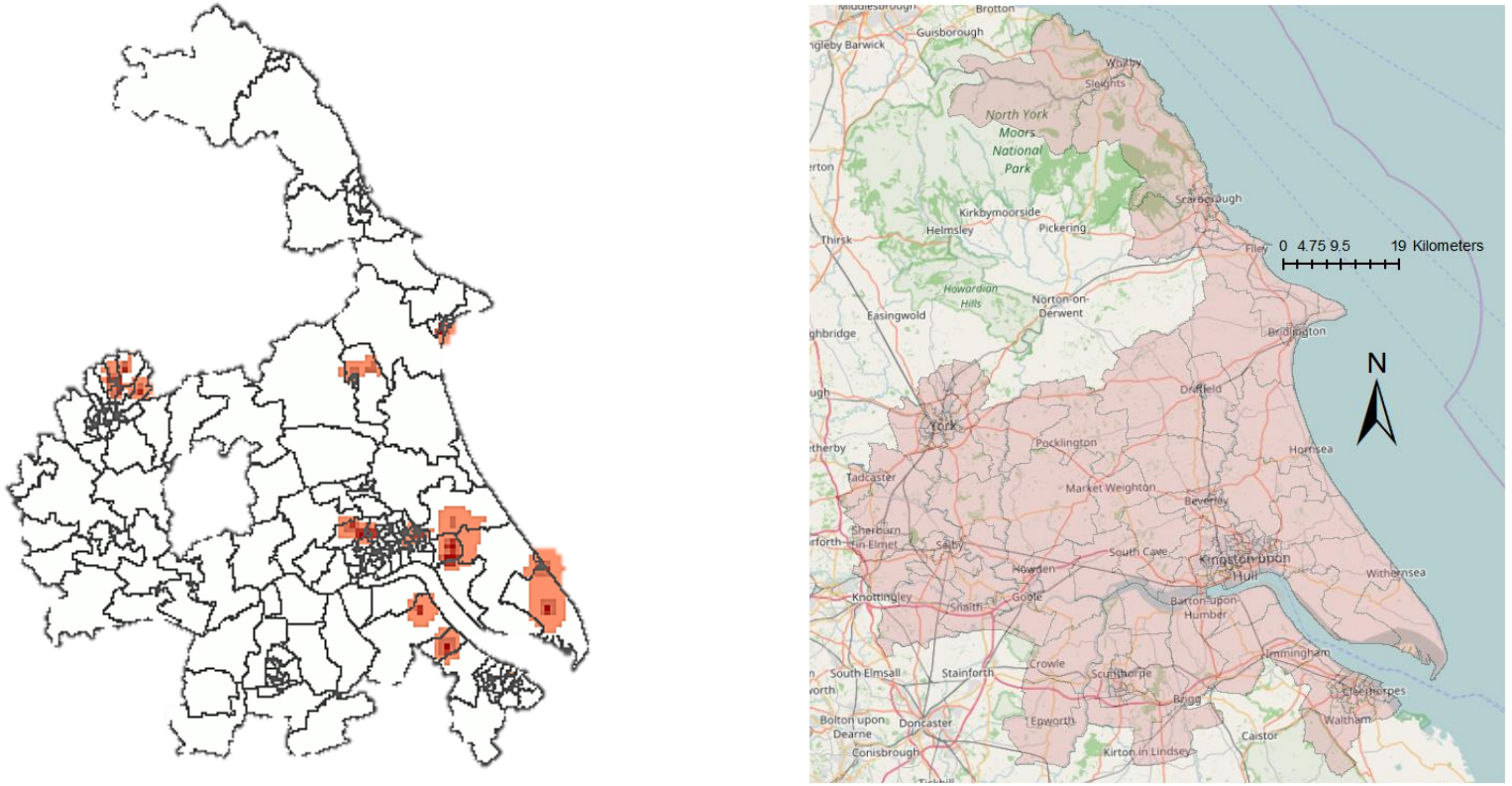


Figure 63 GAM showing potential clustering in North part of York, Driffield, Bridlington, North and East Hull. Withernsea and Patrington , South Killinghome and Grimsby districts.

7.4.2 Weighting these clusters against population density

Though GAM offers a crude measurement of where there may be clusters of higher incidence, Kernel density mapping enables these clusters to be weighted for populations. The following kernel density map was developed by smoothing incidence by populations of LSOA. Clustering becomes evident through these statistics, but these are applying point level data. Further smoothing by LSOA aggregation can go further and identify 'hot spots' for comparison against underlying population demographics. This is why a three-dimensional cluster technique is also applied within this research. Kernel density estimation emphasises the places where groups are most spatially concentrated, using population counts as a field so that results are weighted according to populations. Unlike the point mapping which identified all individual cases, kernel density estimation offered a non-parametric method to estimate density of GOC incidences within the geographical catchment area. Unlike GAM, which offers a flattened, concentric circular estimation of where incidences occurred, this technique builds sets of 'bumps' on the map, where more incidences occur within an area, the 'bumps' become larger and the resulting display more prominent. The highly coloured areas in Figure 64 show the most common areas of GOC diagnosed patients.

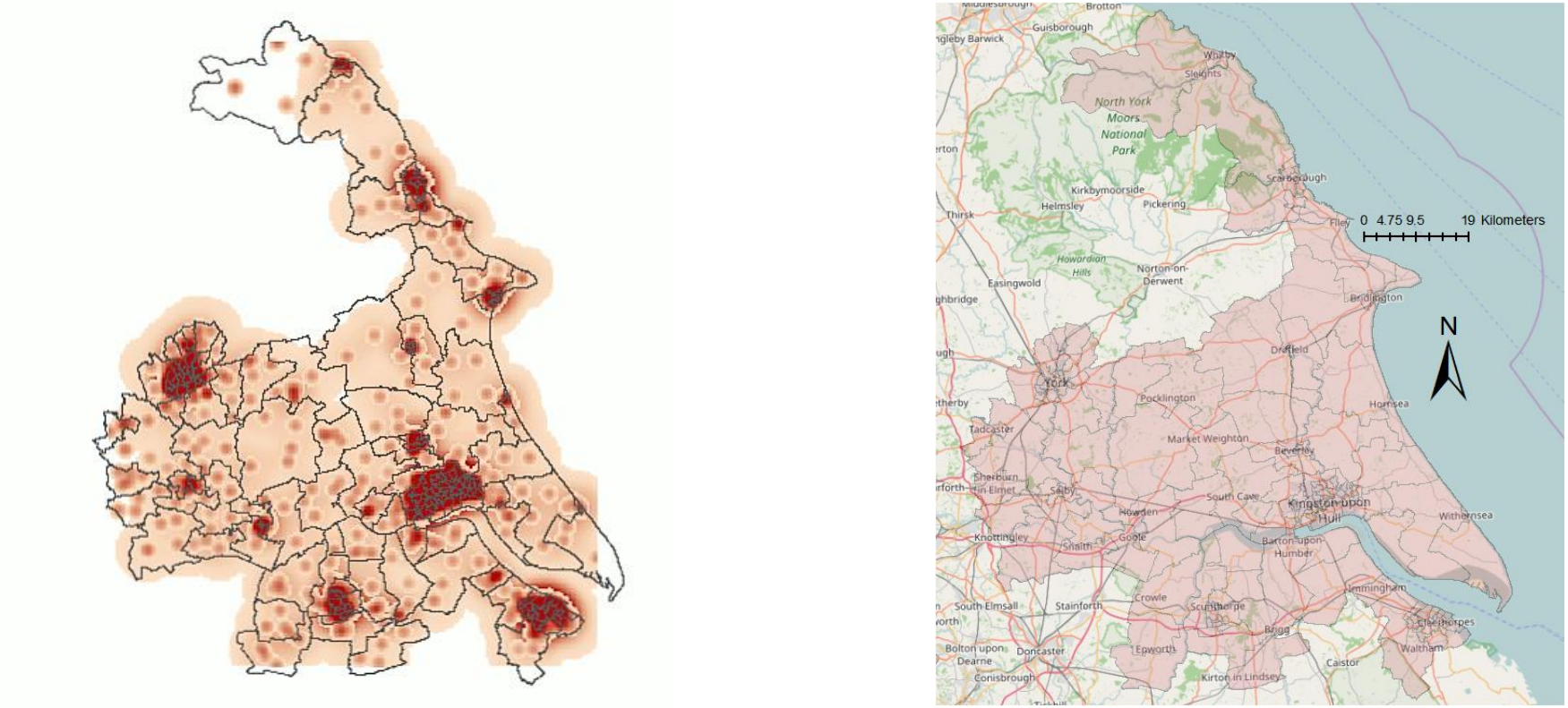


Figure 64 Kernel density estimation showing the most dense areas of GOC subjects – Grimsby, Scunthorpe, Selby, York, Bridlington, , Filey, Scarborough, Whitby, Driffield, Hull, Beverley, Goole and Pocklington.

7.4.4 K Means Cluster analysis – characterising the catchment area to reflect potential GOC Incidence

Classification of areas require unique variables which are measured in a range of different ways. This thesis has presented a case to identify that GOC is more prevalent in those who are smokers, who consume alcohol, in aged populations, and (in some cases), with lower socioeconomic groupings. These attributes have all been linked with increased risk of GOC (Launay et al., 2012, Bus et al., 2012, Mao et al., 2011, Brewster et al., 2000, Sharpe et al., 2012).

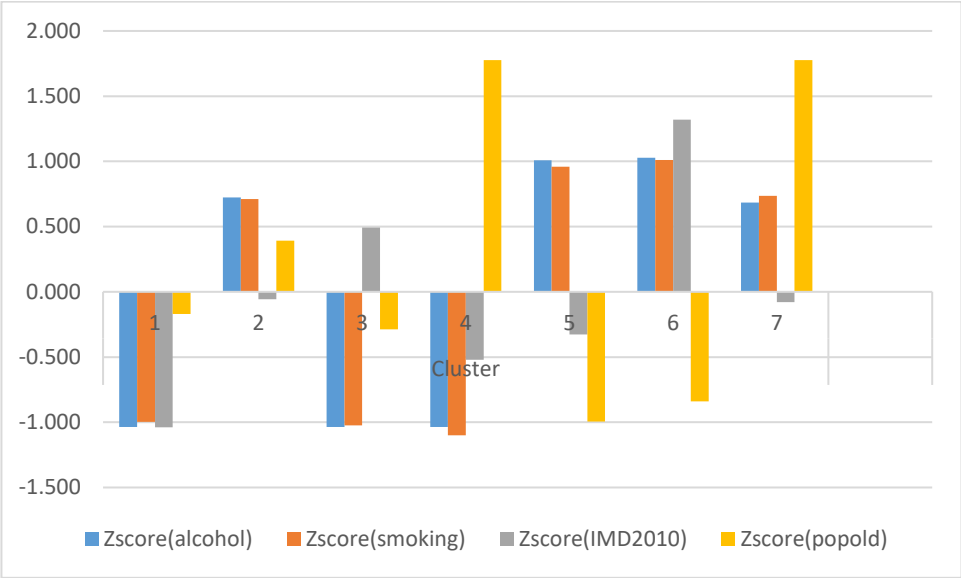
All data relating to these attributes were placed in order of priority and weighted according to the findings presented through the literature on attributable factors. Each LSOA population was weighted to create a variable whereby populations could be considered as 'older age' and where IMD status was lower. The rationale for applying IMD status is because this is a measurement of overall deprivation and this involves a composite score based on a total 38 indicators, for example, access to services, air quality and lifestyle data.

So, one variable did not inadvertently dominate the characterisation process, Data were normalised to Z scoring system and groups prioritised to fit alongside the weighted percentages of attributable factors identified in Table 10. Each input variable was apportioned with a mean of 0 and standard deviation of 1 to each MSOA.

For the purposes of this classification, positive and negative associations were considered, and these were weighted on a scoring system. Good access to services, for example, was identified as a positive, (Z scored + for areas with best access) whereas

high levels of smoking in the region was negative (Z scored - for those areas with worse scores).

Figure 65 K means clustering algorithm



After standardisation of the variables to Z classification scores, a K Means clustering algorithm was applied to each MSOA to determine best fit across all groups. Iterations of nine groups, five groups, then finally seven groups were developed. Grouped characterisation appeared to reflect the best fit alongside data and attributes. From these data, the following areas were determined, with risk stratified accordingly.

Table 25 risk stratified LSOAs

Group descriptors	Risk profile
1 Not deprived, low smoking and alcohol, normal population profile	Extremely Low
2 Not deprived, high smoking and alcohol, young population	Low
3 Not deprived, high levels of smoking and alcohol mid population ageing profile	Higher
4 Older population low deprivation, low smoking and alcohol	Expected
5 Not deprived, young population high smoking and alcohol	Low
6 High deprivation, young population, high levels of smoking and alcohol	High
7 Older population, high smoking and alcohol low deprivation	Extremely high

The mean diagnoses per LSOAs were spread across the groups are illustrated in the following figure.

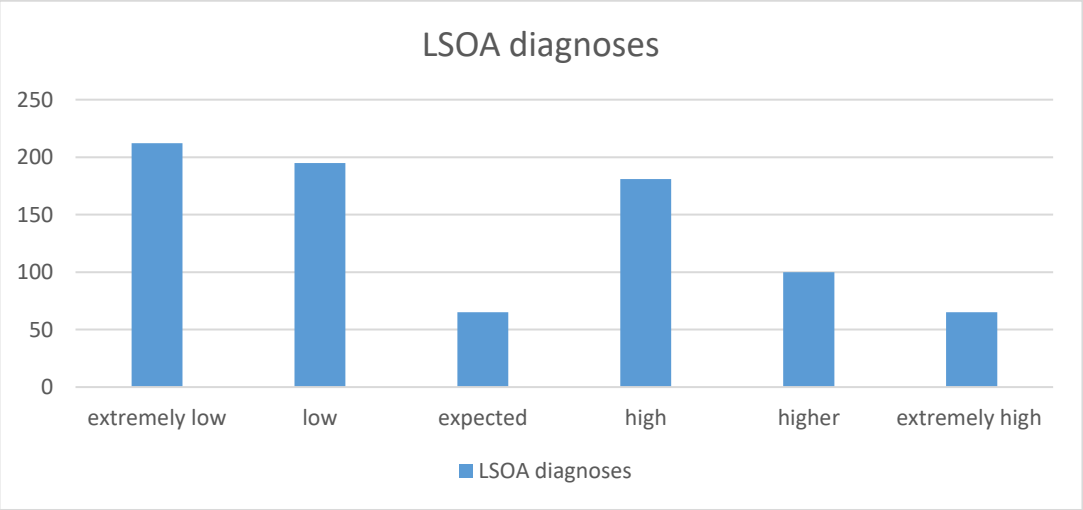


Figure 66 GOC diagnoses/LSOA per characterisation group.

Data are illustrated against clustering algorithms and observed versus expected rates in the following maps.

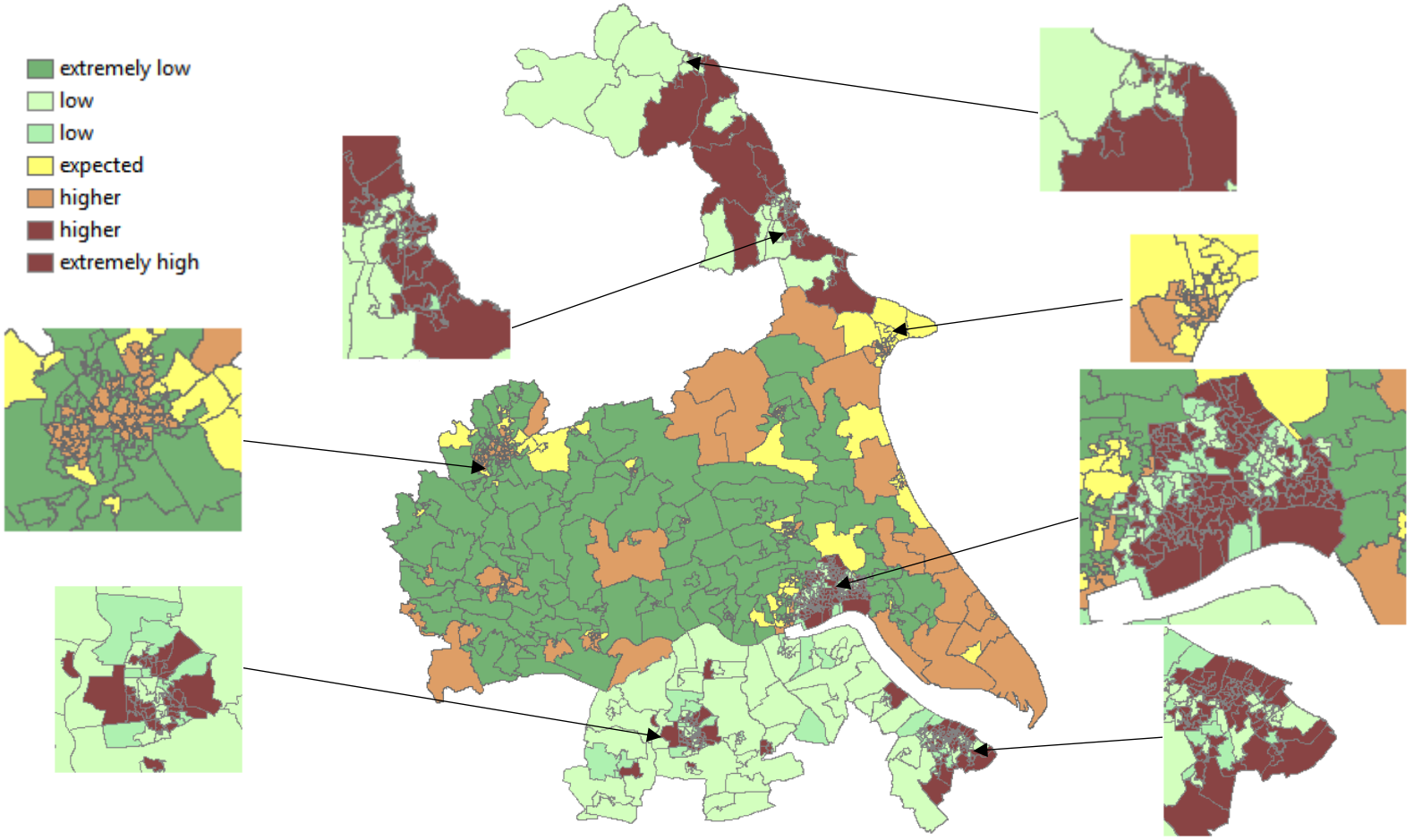


Figure 67 mapped K-Means clusters

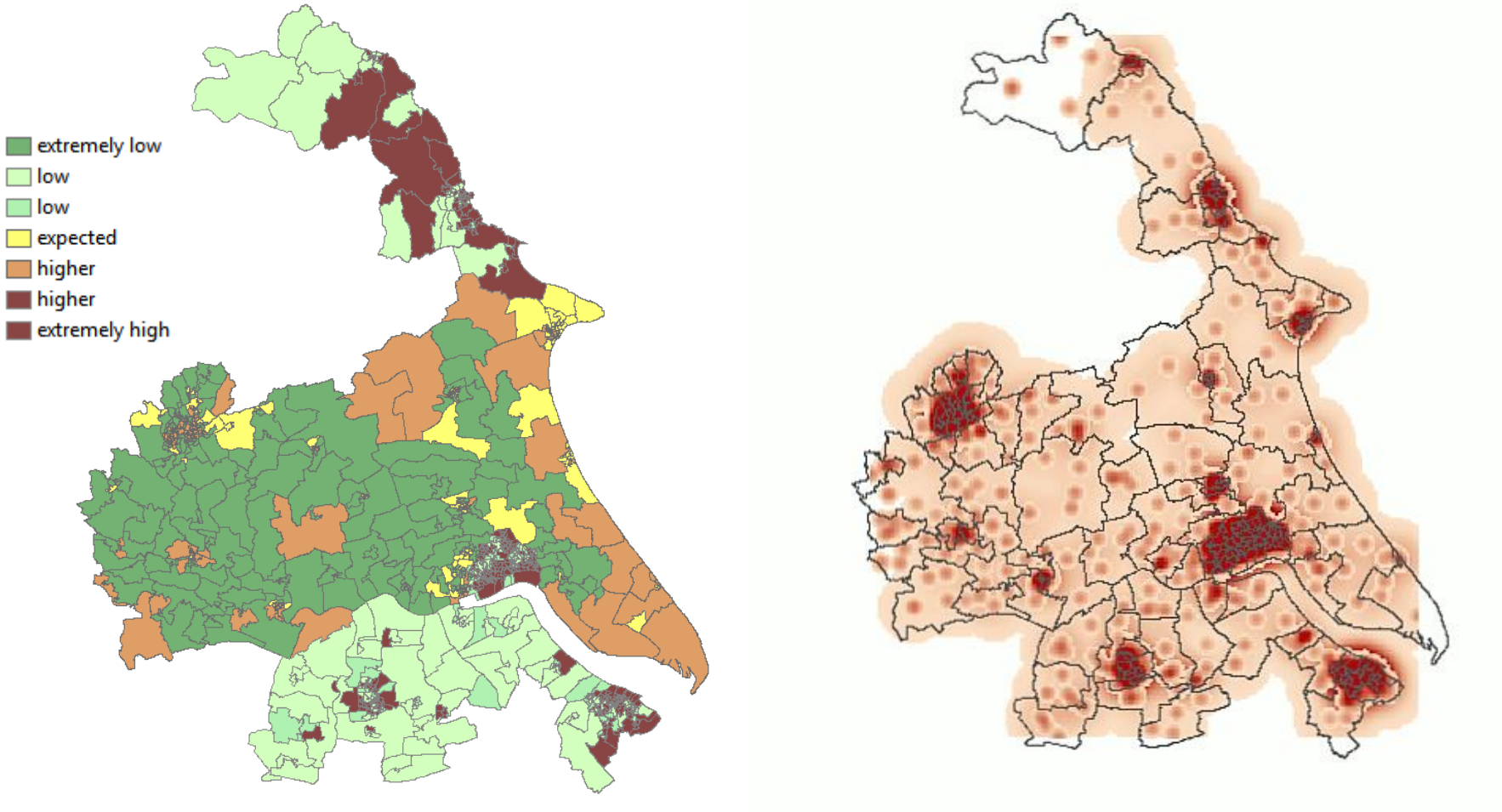
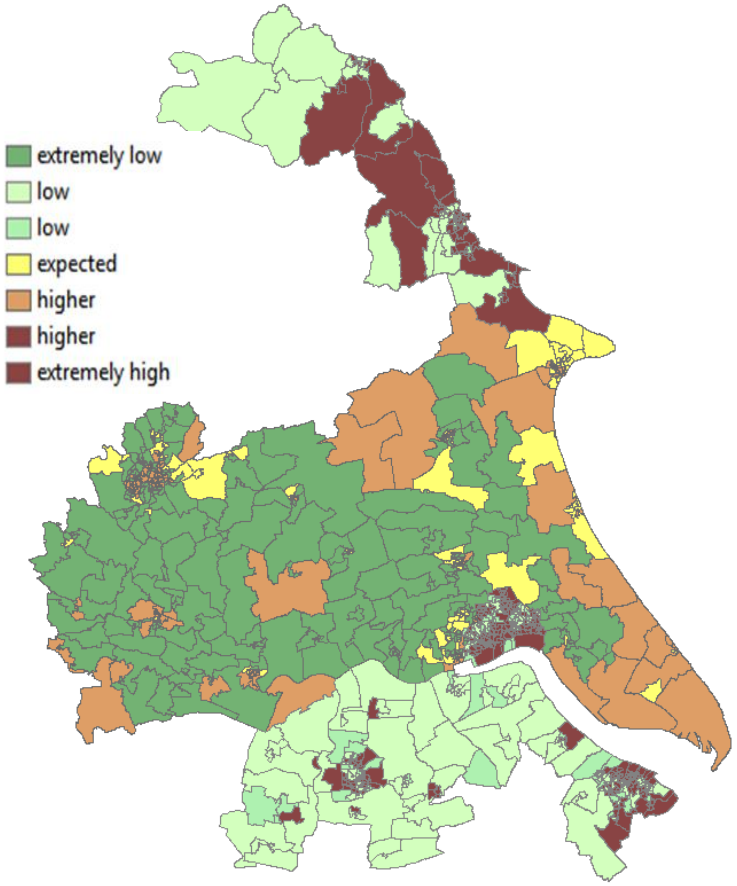


Figure 68 Characterisation against all GOC diagnoses 2000-2013



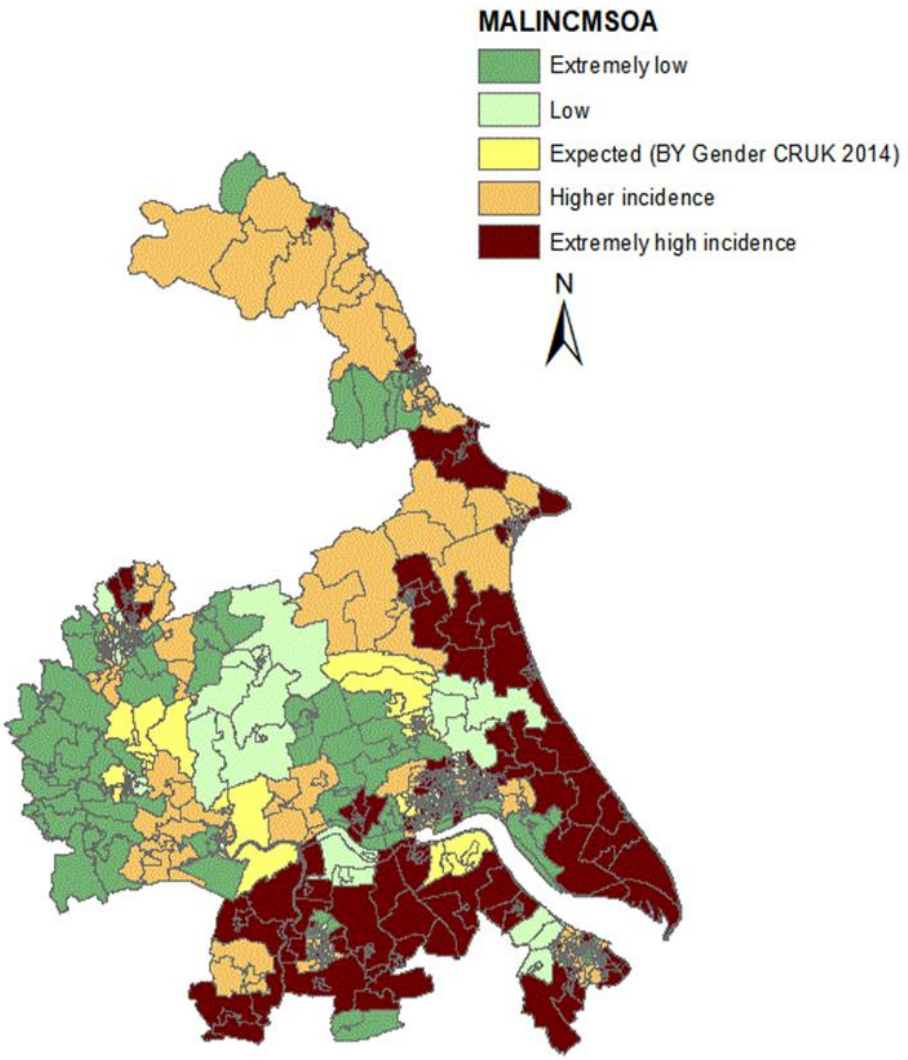


Figure 69 area characterisation tool – compared to CRUK incidence rates (higher or lower than anticipated)

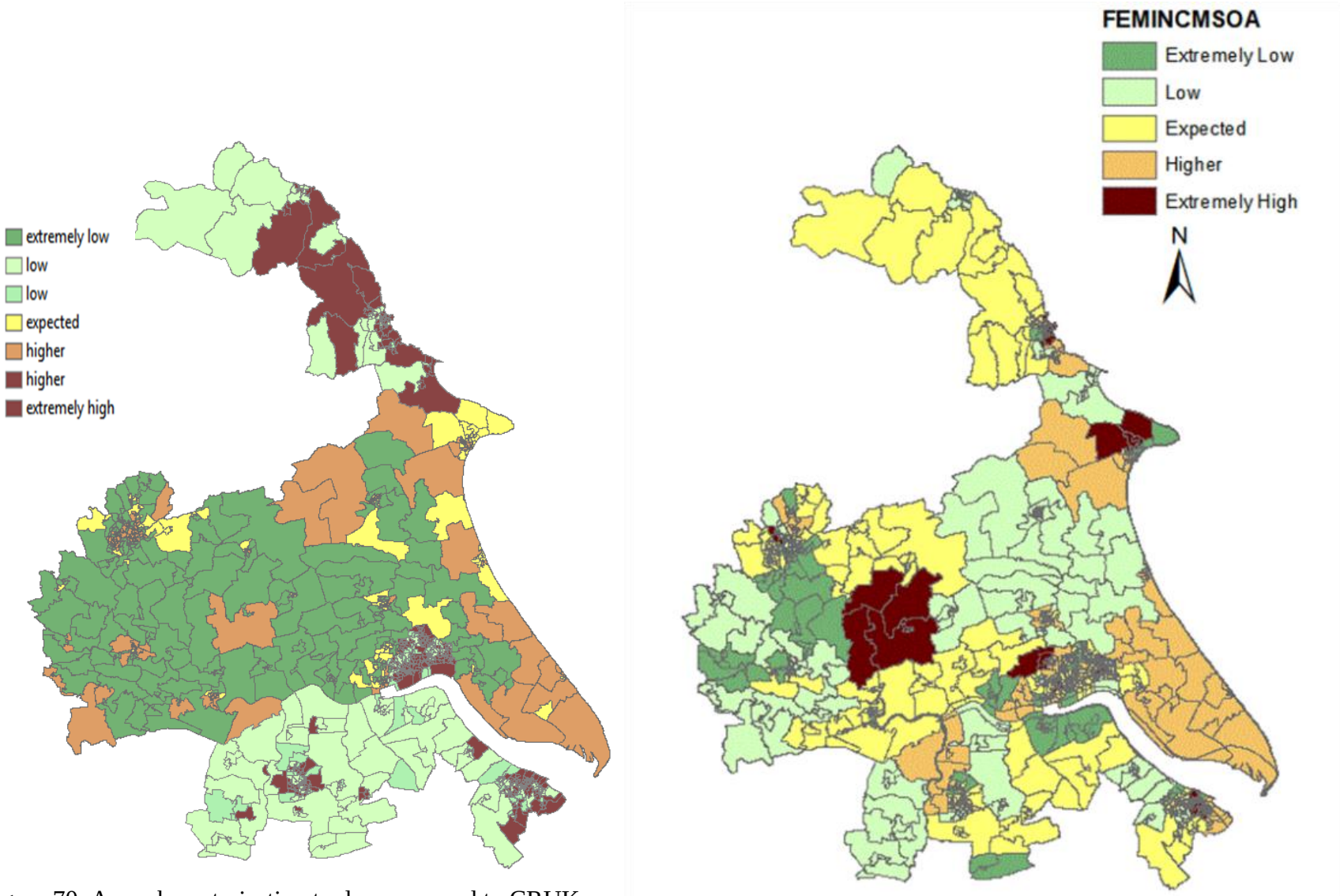


Figure 70 Area characterisation tool – compared to CRUK FEMALE incidence rates (higher or lower than anticipated

Figure 68 suggests the characterised areas are not necessarily reflected in population standardised (historical) incidence. High risk areas were identified across Scarborough and the North York Moors, but the characterisation tool related this to the aging population profile and the higher levels of smoking and alcohol related incidences in these areas.

Several upper mid risk stratified areas included areas around Driffield, Withernsea and North of the Humber Estuary, with Goole, Nottingley and Selby as potentially high. Smaller scale analysis in more densely populated areas were reflected in incidence density (Figure 68).

Figures 69 & 70 offer a comparison between the age standardised 'observed versus expected' incidence by gender across the region. The maps are populated using age stratified population denominators, so those areas which identify higher (or lower) than average expected rates should be reflected in the characterisation tool.

There were a number of LSOAs undetected as high-risk areas when EU age standardised data were applied, meaning the tool lacked specificity in predicting outcomes. This is discussed further in the next chapter. It should be noted that this area characterisation tool was developed following several different iterations to group LSOAs into characterised variables.

Chapter 8 - Discussion

This chapter presents a discussion of the findings in relation to the literature. This exploratory case study was undertaken to review whether there were any geographical patterns in incidence, presentation and survival in gastroesophageal cancer. The substantive preposition of this thesis is based on a wealth of evidence used by social marketing, whereby customer attributes and location can provide intelligence on the best appropriate areas for provision of services.

The rationale for this study was to seek novel ways to reveal geographical areas with higher density of people at higher risk of gastroesophageal cancer, to reveal potential areas with worse survival or incidence rates, and to evaluate whether later stage presenters 'cluster' over particular geographies. This is the first study examining patient postcodes in relation to presentation and survival in gastroesophageal cancer. As such, there were no previous theoretical prepositions to provide a basis for the study.

Following significant deliberations on theoretical frameworks appropriate for the research, case study offered an appropriate methodology to capture the disparate elements regarding the geography of presentation, survival and incidence in gastroesophageal cancer. It also allowed the researcher to undertake a range of methodologies derived from geography, and epidemiology to describe the cohort fully, its geography and its demographic catchment area.

The initial phase of any case study is to develop the underlying theory and this is undertaken through exploration of existing research, gathering information and identifying ways to answer the research objectives and therefore, the research question.

The background chapter revealed the stark geographic affiliation to incidence of gastroesophageal cancer and revealed that this cancer has a specific patient demography. This information supported the need for small scale analysis of gastroesophageal cancer and further study into incidence, presentation and survival in relation to encouraging earlier presentation and rapid transition to treatment in gastroesophageal cancer.

The literature review identified the most crucial time period for encouraging this earlier presentation. It narrowed the focus to the 'patient interval' as the most relevant for further study and supported a rationale to use patients, rather than service transitions as a way of studying presentation and survival. A methodology derived from social marketing, geography and epidemiology was described in a case study format, to structure the research.

Yin's (2003) case study methodology and four objectives were presented as a methodology to reveal whether there are any patterns in incidence, presentation and survival in gastroesophageal cancer patients' neighbourhoods. Further geographical analysis techniques were applied to explore the potential to use these factors as a way to inform targeted interventions. A small regional referral centre was chosen as an appropriate location for the study and a retrospective cohort was identified as means to analyse past events, therefore, identify whether future events may be predicted at local levels.

8.1 Key findings from this research

This research hypothesised that the larger scale geographic studies which have linked a geographical affiliation to GOC, may potentially be reflected at smaller, more local scales. This exploratory study applied spatial scan statistics to reveal patterns in

incidence, presentation and survival in gastroesophageal cancer patients' neighbourhoods and consider whether this intelligence had a potential to inform targeted interventions which would promote earlier presentation and improve survival in GOC.

A case study within a regional referral centre presented a retrospective analysis of 14 years of GOC subjects, analysed against the underlying population profiles. The case study evaluated demographic and histological factors associated with presentation and survival in the GOC cohort and then presented a neighbourhood description of the catchment area using Census data from that period. The specific aim of this research was to review whether GOC attributes in a given population, had the potential to predict geographical areas of higher need, thus, to inform health service delivery.

This research has found the following in relation to the 4 objectives:

Population and attributes in GOC

- Diagnosis of GOC is more common in males and in those over 65.
- Main histological subtypes were adenocarcinoma and squamous cell carcinoma.
- There is a wealth of missing data in gastroesophageal cancer – this relates to TNM staging, site of cancer and how patients present for diagnosis.

Presentation factors

- Most GOC patients present as 'inpatients' and through GP referrals, and those with worse outcomes present as emergency admissions.

Survival in GOC

- Advancing age impairs survival outcome, but survival has significantly increased since 2006. There are no gender differences in survival outcomes.
- Surgery increases mean survival in GOC.
- GOC diagnosed patients under the age of 54 are more likely to be from lower socioeconomic groups.
- The usual ‘one year’ survival grouping does not appropriately represent survival in GOC. In this cohort, over 39% died within 6 months of diagnosis.
- Survival in lower socioeconomic groups is marginally lower than survival in those patients from the least deprived areas.
- Emergency presentation is linked with reduced survival in this cohort.
- More advanced TNM stage presentation is linked to earlier death.

The Neighbourhood, lifestyle and demography of the catchment area.

- MSOA levels of enumeration are the most effective scales for longitudinal analysis of these types of GOC cancer data.
- Population attributes vary across the region, and there are areas with higher levels of smoking, alcohol consumption and socioeconomic deprivation.
- The research revealed MSOAs with high and extremely high incidences of GOC in the region, when compared to EU age standardised incidence rates.

Clustering, prediction and informing interventions.

- Incidence data placed into a mapped format requires appropriate population weighting.
- When appropriately weighted for populations, there are clusters in incidence of GOC over the full timescale of the cohort, however, an area characterisation

tool with existing available datasets lacks clinical specificity to inform practice.

8.2 Common attributes in gastroesophageal cancer (Objective 1)

The GOC cohort within this research displayed very similar attributes when compared with UK level studies (Coupland et al., 2012, NCIN, 2010, NHS information centre, 2008). Age, gender and clinical interventions were also similar to other UK regions. The wider literature suggests 56% of newly diagnosed GOC cases are in the over 70 age group (CRUK, 2017). This research identified the average age at presentation to be 70 years. This concurs with the wealth of evidence associating gastroesophageal cancers with older age group presentations. Out of 2785 diagnoses in the region, only 43 cases were identified in subjects aged below 44. Many studies identify that men tend to present at a lower age group than women (males have a mean presentation age of 65 and women 70 years and this was reflected in the cohort, where the mean male age was 67 years and females, 71).

The gender ratio of 5 males to every 2 females found in this cohort is supported in several studies on GOC. Many studies apply ‘rounding’ of ratios, offering male to female counts of 3:1, or 2:1.(Coupland et al., 2012, Belgian Cancer Registry, 2013, DeAngelis et al., 2014, Gavin et al., 2012, Karim-Kos et al., 2008, NHS information centre, 2010, Levi et al., 2013, Wei et al., 2011).

Only 19% of the GOC cohort received surgery. However, it must be noted that clinical improvements and developments in surgical interventions are improving over time, so these lower rates may reflect the timespan 2000-2013. More recent UK studies identify that 38% of patients received surgery between the years 2007 and 2009 and 52% received surgery between 2013-2015 (NHS digital information centre, 2016), perhaps

indicating a rise in interventions with curative intent. UK surgical intervention is higher than the worldwide average which is cited at 29.8 % (Stordeur et al., 2015), yet this may represent several confounding phenomena in healthcare delivery. For example, the free NHS system of the UK means that patients have increased access to healthcare. Likewise, citing a worldwide average will also include the third world countries, where diagnoses and treatment options are significantly different.

The main histological type in this cohort was adenocarcinoma (ADC) then squamous cell carcinoma (SCC). This is reflected in gastroesophageal cancers across the world (Stordeur et al., 2015). The site of the tumour was most commonly identified as junctional and this is, again, reflected in worldwide studies in gastroesophageal cancers (NHS digital Information centre, 2016, Gavin, 2012, Cronin-Fenton, 2008).

Socioeconomic groups identified within the cohort were similar to previous studies (Levi et al., 2003, Ljung et al., 2013a, Coupland, 2012, Ellis et al., 2012). This has several implications. As discussed in chapter 2, socioeconomic grouping offers a means to describe a variety of factors contributable to health and wellbeing. The UK Indices of deprivation (IMD) are a constellation of variables, each weighted in significance to overall 'general health and wellbeing'. These variables draw from aggregated data on lifestyle factors, access to healthcare housing and education, income and employment, living environment and crime statistics (or those variables assumed meaningful to deprivation. IMD scores are apportioned to set geographical areas (LSOA) and so are limited in that they can only represent the 'average' presentation per geographical area across a 5-point scale ranging least to most deprived status. As factors meaningful to deprivation change over the years, and communities change to adapt to economic and social forces, IMD scores are subject to several iterations. There are alternatives, such as the Townsend index (Townsend et

al, 1998), however, IMD is most commonly presented by the Office of National Statistics and Public Health England. For the purposes of this research, IMD was chosen as it offered temporally appropriate data at the right level of resolution for this study. The Office of National Statistics and CASWEB release IMD data at LSOA level to underpin research of this type. Additionally, IMD scores were available via Pubic Health England data, captured at the time of presentation and diagnosis and recorded in cohort data. To determine whether there were any significant temporal changes in IMD iterations which could be present in the cohort, IMD 2004, 2007 and 2010 were compared and no major differences were identified across the iterations.

Several worldwide studies on factors attributable to GOC were identified to support objective 1. There are many tools to assess risks in a range of cancers, resulting in particular algorithms suggested as tools to alert the diagnostic and referral processes in cancer care. However, cancer has a long latency period, there are multiple risk factors with relatively small impact and the causal pathways to developing cancers are yet undefined. Diagnoses and clinical decision-making relies on several attributable factors, signs and symptoms, and apportioning risk, or high-level risk behaviours to subsequent development of cancer is fraught with methodological difficulties. Each variable must be accounted for, confounding variables must be apportioned and weighted according to their effects on subsequent development of the disease and then compared with an infinite number of other factors, such as diet, lifestyle and underlying health status.

Hence this thesis drew from existing meta-analyses of studies identifying those factors common to patients diagnosed with GOC. These studies formed the basis for the K means cluster analyses in objective 3.

8.3 Survival and presentation in gastroesophageal cancer (Objective 2)

As described in chapters 2 and 3, there is a vast array of contesting views on nomenclature of survival, delay and presentation stage in gastroesophageal cancers. Originally, the research intended to base survival and presentation analysis based on its TNM classification. This describes the stage at presentation through assessing the extent of **T**umour invasion, lymph **N**odal involvement and whether there are any **M**etasteses present. Numeric coding is apportioned to denote the severity of anatomical invasion. This classification is regularly updated and reviewed by the union for International Cancer Control and the American Joint Committee on Cancer (AJCC) and is the internationally accepted classification system. The AJCC release staging manuals to describe processes for staging. These are in line with the AJCC cancer staging manual (currently both on their 8th published editions following several iterations between 1968 and 2016). These iterations ensure clinical relevance and information is updated with emerging evidence and clinical improvements.

In this cohort, complete TNM data were not available for all subjects. Data were absent for 1547 of 2785 subjects (55%), and the Public Health England 'stagebest' was applied as a variable to reveal stage at presentation. This variable is developed via histological confirmation, at post-mortem, through clinician assessment, or during surgical intervention. Rather than rely on missing data, groups for presentation were developed from time at histological diagnosis, to time of death. This approach was developed from the Aarhus statement (Weller et al, 2012) revealing key time intervals for early diagnosis cancer research (Table 2).

Statistical analysis of survival data requires a movement from one status, to another (diagnosis – to death, or a censored 'end-point' in data). The end point in these data related to either death, or a 5-year point which was identified to denote 'survival after diagnosis'. To identify the most common groups of survival in days, percentiles were calculated. The most common elements were then derived through Tukeys Hinges calculation, comparing all possible means, to provide a standardised range distribution (Bryman & Cramer, 2011). This allows for a more conservative estimation where group sizes are different. The mean survival time in days within this research was commensurate with existing UK literature on survival times in gastroesophageal cancer (Cancer Research UK, 2017, Coupland et al., 2011). However, this is the first application of the 2012 Aarhus key time intervals to apply groups to cancer survival. These groups are based on N = 2215 subjects to allow for the 5 year censoring, so it must be noted that survival analysis only relates to the data on GOC presentations with histologically confirmed diagnosis, between the years 2000 and (June) 2011. It also did not include post mortem diagnoses.

The survival groups allow for an analysis of death within 6 months of histological confirmation and this related to the largest group in the survival cohort. 39% of these deaths occurred before 6 months, and only a further 21% died before the 1-year survival statistic. Generally, results are reported as one year cut point, and this would have related to a total 60% deaths within a year. However, noting the 6-month statistic separates and clarifies the extreme late presentation phase in gastroesophageal cancers. A biologically plausible explanation for this would be to suggest those subjects who died within the first 6 months, either had other comorbidities, presented at a very late stage, or had extremely aggressive tumour forms.

All subjects presenting between 2006 – June 2013 (n = 1097), had a recorded 'route to diagnosis'. These data were useful to compare with the survival statistics. Most referrals were made via the 'two-week wait' (TWW) rule (42% of the cohort). The TWW rule was instigated in 2000 by the UK Department of Health to reduce waiting times for suspected cancer (NICE, 2015). The Department of Health set standards and guidelines for referral, structured care delivery pathways, and the 'two week wait' standards. However, there remains a regional variation in referrals (Meecham et al., 2012, Vedstead & Olessen, 2011, Hamilton et al., 2015). The premise was that UK average referral and wait times were longer in comparison to European standards (Souhami, 2010). The most recent iteration to the referral guidelines was in 2015, when NICE (2015) published their less prescriptive standards aimed towards clinical diagnostic rationales and based on the most recent evidence of attributable factors and symptomology.

Unsurprisingly, the majority of extremely late presenters in the grouped cohort (those dying on or before 6 months from diagnosis, presented as emergencies). Most subjects were referred through this TWW pathway. Factors explaining the early demise of emergency presenters lie within the physiological status of patients during the process of gastroesophageal cancer. Emergency presentations are generally identified when subjects are at the later stage of their disease process, where the tumour has infiltrated surrounding structures, either to cause gastrointestinal bleeding or perforation, respiratory distress, or with extreme dysphagia, pain, nausea or fever (Marmo et al., 2012, Shah et al., 2010, Blackshaw et al., 2004, Bosscher et al., 2014).

This research also correlated advancing age with reduced survival, which, in some ways, could be described as biologically attributable to the ageing process, to increased comorbidities and frailty which is strongly associated with ageing. The mean

age at presentation was 70 years, so the skew to an older population in gastroesophageal cancer means the cohorts studied will be subject to the many confounding variables of ageing.

Male and female survival rates were similar, and this supports current evidence within the literature (Coupland et al., 2011, CRUK, 2017). This was important to review, as there is a difference in lifestyle behaviours between the genders in the 'baby boomer' generation. Males had higher body mass indices compared with females, and smoking and alcohol consumption were higher (Worsley, Wang & Hunter, 2012). All of these are considered as attributable factors in gastroesophageal cancer, and have been inked with impaired survival in several other studies (Coupland et al., 2012, CRUK, 2016).

Parametric and nonparametric tests revealed links between worsening deprivation scores and impaired survival. As previously discussed, socioeconomic status measured in terms of deprivation indices rely upon a host of lifestyle, economic, social, dietary and psychological parameters. The modifiable risk factors such as smoking, diet, physical activity and increased BMI more commonly identified in deprived groups, are attributed to reduced survival outcomes in several studies on cancer (Hastert et al., 2016, Danzig et al., 2014, Hagedoorn, 2016).

To investigate the relation between survival time and socioeconomic deprivation scores, the Cox Proportional Hazards model was applied. This model was chosen over the alternative Kaplan Meir survival curve (the univariate alternative). Cox Proportional Hazards model enables a multivariate parametric analysis of relative risk, against nonparametric hazard functions (IMD score). This model enables the prediction of survival differences in relation to hazard functions, measured against

time-dependent factors (Bryman & Cramer, 2011, Chan, 2004, Bland & Altman, 1998).

The morphologies within this cohort also reflected the wider literature, showing adenocarcinoma to be the most common presentation, closely followed by squamous cell carcinoma. Again, TWW referrals were more common in these groups and survival was similar between groups.

Mean days survival in the cohort with available TNM staging data showed a drastic decline as stage progressed from 1 – 1V.

8.4 Describing neighbourhoods in relation to specific cancer attributes (Objective 3).

The third objective in this research was to describe and map neighbourhoods with the attributes identified in the first two objectives, to compare against incidence, presentation and survival in gastroesophageal cancer within the cohort. These attributes included population density, social descriptors, access to services and lifestyle indicators (smoking and alcohol). Incidence data across the longitudinal timeframe was then reported against underlying population density and compared with EU standardised data to crudely identify whether there were any areas of increased, or decreased incidence across the region.

The catchment area was initially defined, and this was rather complex, as NHS boundaries and area classifications have changed over the years. These boundaries are co-terminus with output area scales, as are the variety of health and lifestyle data published through the Office for National Statistics (ONS). These are described fully in chapter 4.

Scale was an important factor to consider through this research, as it attempted to find the highest scale of resolution while maintaining patient confidentiality. This exploratory case study was undertaken as a response to the number of calls to undertake small area analysis of cancer data (NCIN, 2010, Abbas et al., 2009, Abel-Rahman, 2009, Aguilar et al., 2013, Bell et al., 2012, Candace et al., 2011, Dummer, 2008, Hahn, 2014, Hanafi Bjoyd et al., 2012, Hu et al., 2012, Lewandosky et al., 1995, Mohebbi, 2008, Parrott, 2010, Sharp & Donnely, 2014, Wang, et al. 2010). In searching for lifestyle data, it became evident that a variety of scales were applied to findings. For example, socioeconomic and some lifestyle data were available at LSOA levels, which was very useful, but smoking and alcohol related data had not been released by the Office for National Statistics at this same level. These data were enumerated to very large scales and this could affect results and interpretation. This thesis presented the issue of the modifiable areal unit problem (MAUP) (as we merge data on higher and higher scales, it becomes less meaningful) and ecological fallacy (all persons are equal, so long as they reside in the same post-coded neighbourhood) in chapter 6.

Social descriptors of the region were presented as Index of Multiple Deprivation or IMD (2010) denoting levels of deprivation across the region. The IMD is the favoured indicator for England (Noble et al., 2006). These data described populations in the region. However, deprivation scores do not necessarily reflect whether small areas change over time. IMD based on decennial census survey data cannot fully capture the issues which occur between census dates. Industrial closures, local area planning initiatives or levels of deprivation changes which may affect how populations are 'grouped' into indices of deprivation, occur rapidly and affect people frequently.

Additionally, larger enumerated ‘indices’ of deprivation do not necessarily reflect what is going on at local levels (see MAUP in the previous paragraph).

There is a strong relationship between health and unemployment and that these data are available on an annual basis (Haynes, 2006), so this could be harnessed to reveal annual differences in deprivation which is linked with health status. Ongoing work by experts such as Ajebon & Norman (2015) has led to a methodology of area classification according to employment attributes. These are published through the Small Area Health Statistic Unit of the UK Office of National Statistics (ONS,2009). These classifications were applied to the GIS to produce the map denoting social descriptors based on income status, hence, ‘countryside, disadvantaged urban etc’ (see Figure 58 Area descriptions – Norman (2010) Advantaged/ less advantaged areas. Access to health services (raised in chapter 2) was an important factor to consider for GOC presentation and this is detailed using GP and Hospital access data from the Department of Transport. Analysis of presentation over time was undertaken to reveal whether there were temporal trends in incidence. Application of mean point prevalence rates (Moon et al., 2000) revealed no significant differences between groups. 2003 showed the smallest number of GOC diagnoses (n = 167) and years after 2011 show the highest (>210 per year). The clinical relevance of this must be considered. These cohort data do not include diagnoses made at post-mortem. Clinical care, referrals and diagnostics processes have been improving since 2000. Thus, raising the potential for decreased levels of diagnoses at post-mortem.

The UK cancer statistics on age standardised incidence rates of gastroesophageal cancers was applied as a comparator against this cohort. Again, this enumeration could be subject to debate, as MSOA level data on populations were applied to this map.

There were areas of increased density identified within the cohort and these, gender specific maps detailed in Figure 45, depict geographical areas where these occur.

The initial bar chart (Figure 44) depicts all MSOAs within the geographical catchment area and identifies several MSOAs with high and extremely higher incidence of GOC diagnoses. These were applied to develop a quintile classification of observed versus expected incidence within the geographical information system. MSOA level population data were stratified by gender and applied to display this mapped version of data. This was necessary due to the disparity between male and female incidence in GOC. There were differences in observed vs expected incidences between males and females when data were presented by MSOA. Selby and York had far fewer areas of higher rates than expected for both genders. The largely industrial areas in North and North East Lincolnshire and Scarborough and East coastal regions of the catchment area had higher than average levels of male incidence GOC.

These are crude maps and any causal links cannot be drawn from these data. They merely illustrate incidence over time within the catchment area, compared with the EU expected rates. However, these maps remain informative. Incidence over the years is something which requires close observation. GOC has been linked with several climate factors, nutrition and diet status in countries such as China and Iran (li et al., 2014, Smith et al., 2008, Tran et al., 2005, Wu et al., 2007, Wang et al., 2006, Wei et al., 2011, Abedi Ardekani et al., 2011, Ahari et al., 2013, Islami 2004), however, determining causality is very complex. The next phase of analysing presentation and survival in gastroesophageal cancer within this cohort, is to triangulate all findings from objectives 1, 2 and 3.

8.5 Evaluating clusters and patterns in the geography of presentation and survival in gastroesophageal cancer (Objective 4)

This objective is underpinned through the development of a geographical information system to hold all data previously presented within this research. A mapped display of male versus female incidence merely pinpoints all cases to the map, using patient postcodes as the pin. This research undertook a range of tools for identification of clusters. Spatial and aspatial techniques to identify clustering have been applied in many studies previously. They may offer an historical view on incidence data, and some have revealed distinct patterns in the past (Sellafield and childhood leukaemia for example). However, this research draws no assumptions on the clustering, other than to identify patterns in incidence and presentation of GOC across the area.

A comparison with EU standardised incidence rates illustrated areas where incidence was higher than anticipated, and on reflection, this was found to be the most clinically relevant tool. Observed versus expected rates, presented in a mapped format, offered a useful illustration and historical account of where diagnoses fell below or above anticipated levels.

The observed versus expected map was compared with findings from the K-Means cluster analysis groups, as a way to determine the accuracy of the final (K-Means) characterisation tool. Both maps showed areas of homogeneity, however, there were also significant differences to note, between population profiles (K-Means characterisation tool) and historical incidence (Observed versus expected).

There were clusters evident across the catchment area which are unexplained by the underlying population profile and this requires further review. However, this is not uncommon. There is a distinct lack of lower output area data on lifestyles and a current drive to develop required datasets is underway across several universities and

institutions. It is left to determine whether more refined characterisation tools, based on more accurate datasets, can to predict service requirements with more specificity. However, the many confounding factors attributable to a cancer diagnosis, and a later staged presentation will perhaps always have an effect to limit area characterisation tools in their ability to predict areas of need.

In the UK, data on health and lifestyle remains limited, and data on dietary and nutritional status were not available to underpin the K Means cluster analysis. It is widely known that diets high in fruit and vegetables can reduce risks of cancer (NCIN, 2012), yet datasets are not routinely available through the Office of National Statistics at this present time. Similarly, postcode level data on smoking and alcohol were unavailable and these would have assisted the descriptions of neighbourhoods within the thesis. Despite many calls for smaller area studies into diseases, published evidence is limited through the requirement for anonymity.

8.6 Application of GIS in cancer research

Gathering information to reveal any potential relationships between location and cancer incidence, presentation and survival is crucial to clinical interventions and planning of services. Whether previous incidence has the potential to predict future incidence remains to be shown. Geographical Information Science is developing rapidly to allow a new lens to be applied to large datasets, and as we collate increasingly complex datasets, there is a need to manage them appropriately, to analyse them and evaluate findings which will be truly meaningful in clinical practice.

Studies which examine the geographic variations in presentation, survival and incidence of cancers have a potential to identify important (geographically based)

prognostic factors. They can support the development of hypotheses for further investigation (Henry, 2009). Furthermore, they can develop a healthcare intelligence which supports policymakers to maximise resources, make services more 'geographically targeted and therefore, improve efficiency (Mobley, 2008). However, they are hampered by the requirement to maintain patient anonymity. Small scale analysis of cancer data is complex. Datasets must be fixed to residencies, presentation of findings can be interpreted in many ways. For example, in identifying areas with higher levels of advanced stage presentation, one could inadvertently identify referral services as inadequate. Similarly, individual hospitals, healthcare centres (and even clinicians) may be detectable through analysis of events at postcode levels, so the research must be presented with significant caveats.

During the process of developing the geographical information system and plotting patient postcodes and incidence, the issue of chloropleth map illustration had to be considered. Where data are displayed in certain colours on a map, they can be interpreted by users in a different way to that originally intended. Symbology applied to data presentation is a significant factor in geographical research and findings had to be interpreted accordingly.

8.7 Use of case study methodology to evaluate presentation and survival in gastroesophageal cancer

Yin's (2014) approach to case study research offers a framework to consider real life situations. Its credibility lies in the case study protocol and this leads the researcher to consider a variety of ways to approach a research question. As a descriptive study of presentation and survival in gastroesophageal cancer, the framework offered a way to manage each 'case' as an objective, then to triangulate findings which were drawn through quantitative data and geodemographic analysis through a geographical

information system. One criticism of case study methodology lies in its lack of generalisability. However, this was not the case within this research. Rather, the aim was to describe a geography of presentation and survival in gastroesophageal cancer in a cohort of patients within a regional referral centre. Using a patient survey and enhancing the research through patient mediated data was considered, however, this moved the focus from the original research question.

Chapter 9 – Conclusion and Recommendations

This chapter addresses how and why the results of this case study are important. It identifies strengths and limitations of the study raised in the discussion chapter and concludes with suggestions for future research into presentation and survival in gastroesophageal cancer.

9.1 The importance of these results.

This research was undertaken to explore presentation and survival in gastroesophageal cancer and the geography of incidence. The research question was derived through the initial 'theory generation' element of case study methodology. The case derived from the geographical affiliation of GOC and took this to a spatial scale previously un-investigated.

Identifying areas of increased density in diagnoses of GOC can offer an intelligence to inform agencies responsible for healthcare delivery. When data are presented in a mapped format, these agencies can readily interpret findings and assimilate information to inform practice. However, that information must be clinically relevant and specific to the disease processes.

The UK Government targets to reduce later stage presentation and improve cancer outcomes need to consider a range of options on how to inform that practice and this research offers a novel, 'geographical lens' as an option. Geographical information systems and science can be harnessed to inform health service delivery, but it is generally applied at very large scales of enumeration when presented to the public. However, this research presented analysis at a more local level, with the aim of offering intelligence which was of clinical relevance.

The case study protocol could be adapted to suit a range of disease processes and a variety of disease specific attributes. The cluster analysis techniques used in this

research were based on spatial and aspatial scan statistics and this is necessary so that the data are not artificially constrained to boundaries such as MSOA, or LSOA.

This large cohort of subjects has provided a basis for analysis of survival, presentation and incidence in a geographical catchment area. Through a retrospective analysis of GOC across the regional referral centre, the research has provided a range of mapped data illustrations with a potential to inform future healthcare delivery.

The research revealed (and challenged) the usual '1-year' survival statistic and presented an alternative which captures the 39% of the population who die within the first 6 months. These could also offer an alternative to the diverse nomenclature associated with cancer 'delays'.

Analysing where patients live, in terms of how they present for treatment and diagnosis and how they survive, is a novel approach to these data. Findings relating to the geography of presentation were inconclusive, but when larger longitudinal data are available, then this may change. Findings relating to incidence across the catchment area revealed areas with higher density poor survival and higher density incidence when weighted for populations. The characterisation tool was hampered through lack of specific data and attributable information on causation of cancers. It lacked a specificity to determine and inform health service interventions, however, the process of identifying clusters which were not attributable to population factors was clinically relevant. Results have been discussed with clinical consultants across the regional referral areas and all have identified a range of attributable causative factors. However, these all require intense scrutiny and a sound and thorough research process for further investigation.

9.2 Strengths and Limitations of the study

9.2.1 Data

A large cohort (N = 2785) was attainable because of the longitudinal timeframe. Accordingly, incidence data were higher, which meant patient confidentiality could be secured at lower scales of resolution (MSOA and LSOA levels). Sub cohorts could be drawn from the larger cohort and were of an appropriate size to validate survival analysis and a separate analysis of route to diagnosis where data were available.

Although this research relies on secondary data analysis, all sources of data are extremely accurate. Public Health England data and ONS data are collated regularly and subject to a range of analyses. Data were not captured at the time of events for this type of analysis, which reduces the potential for bias in the research.

All data retrieval processes required strict ethics approvals, data release and data management policies. These processes meant that the study was presented to a wide audience on many different occasions. All feedback generated at these events was implemented in this thesis.

There were many missing datasets in relation to TNM staging and type of presentation, and as with many datasets in social sciences, presence of outliers had to be considered in relation to its impact on findings. To account for missing data, presentation groups were developed and quantified in the results section. A sub cohort of subjects with 'route to diagnosis' data (post 2016) was derived from the original dataset to enable further analysis. To manage the outliers, a '5-year survival' cut off point was attributed to the dataset. This may have affected the results of survival analysis in this study.

The diverse nature of this research meant that data storage, retrieval and transfer between software packages such as SPSS, ArcGIS and Excel had to be accurate. At each iteration, datasets were assessed for accuracy in transfer and this was based on the overall cohort N, represented through unique identifiers. SPSS syntax is identified in Appendix 6.

9.2.2 Theoretical preposition

This thesis sits on a theoretical assumption that where patients live may impact on their GOC diagnosis, presentation and survival. This can be challenged in several ways.

GOC diagnosis and place of habitation

Patient postcodes at the time of diagnosis do not necessarily reflect the subject's main place of habitation. The cohort age range mainly focussed on older patients. There is a potential that many of these patients could have been relocated to residential care settings. Thus, clusters in certain areas would be higher where there are more nursing or residential care homes. In the wider literature, most spatial cluster disease analyses are based on post coded residence at the time of diagnosis, so this it is not unusual to use these as a basis, however, the implications must be considered against findings within this research.

This research makes no claims to prove causality between the patient's environment (or place of residence) and the development of GOC. Gastroesophageal cancer has a long latency period (Tse Lap-Ah et al., 2007) and levels of exposure to potential attributable factors is very difficult to monitor. Early life risk factor exposure may be significant to developing many cancers (Han et al., 2005, Parkin et al., 2011, IARC,

2016, Lauby et al., 2016). This exposure is not necessarily reflected in the subject's place of residence at the time of data capture

GOC presentation and survival

There are many health service providers in each geographical area. Although NAEDI (2016) identify the need to refer GOC symptoms urgently via the two week wait system, this still relies on the patient managing to identify a complete history and symptoms, and the GP or referrer to attribute those symptoms to potential GOC. As with any service, there may be differences in experience, processes and structures for referrals. Indeed, the results chapter highlighted over 25% of GOC subjects were referred by their GP, but not via the 2-week wait system. Therefore, clusters identified for late presentation and reduced survival in geographical areas have a potential to reflect individual patients and their GP and Healthcare service providers.

Population profiling

The underlying explanations which constituted the geodemographic classifications relied on several assumptions. Smoking and Alcohol data were enumerated to a very large scale and this poses a limitation to the study. It is also extremely difficult to apportion relative risk to socioeconomic status, alcohol and smoking, as they are poorly defined entities (Kamangar et al., 2009). Many variables are applied to apportion socioeconomic status to an area, and artificial boundaries set to different geographical scales may mean several people who do not 'fit' with the apportioned status are left unrepresented within the 'scaled up' geographical area. Alcohol and smoking data only captured those persons whose consumption led to hospital admissions where this was recorded. Socioeconomic status labels have been through

several different alliterations through the years and many different indices have historically been applied.

Application of Census data from 2000 was felt to be the most representative of the cohort, though area classifications can change over time and this must be considered in relation to findings (Norman, 2006). Age standardised incidence rates were calculated by EU standards and it must be noted that this catchment area has a higher rate of GOC diagnosis than many European countries, in fact, the UK has higher average GOC diagnosis than many countries in Europe (Kollarova et al., 2012, Coupland et al., 2016).

9.2.3 Scales of resolution

The modifiable areal unit problem (MAUP) occurs when data are aggregated to larger geographical areas, allowing the observer to review data at large scales of resolution. When variables are plotted into a geographical catchment area, their geographical boundaries are implied by the commonest denominator. For example, census wards and output areas (LSOA and MSOA) are imposed to geographical catchment areas, where each boundary is set artificially to 'similarities' in data. This may be population size, geographic similarities, or a range of attributes within populations and land use. The MAUP suggests that these 'areas (output, LSOA or MSOA) may not necessarily be meaningful to the variables of interest. For example, if a catchment area encompassed 15 residential care settings for the elderly, then the underlying population would be artificially described with an 'older population', when compared to an area with no elderly residential care settings. Data are skewed by where these artificial boundaries are drawn.

This thesis applied presentation and survival to the lowest levels of enumeration, so that results could be clinically meaningful, but also, so that they maintained clinical confidentiality. Many studies identify that to reduce this MAUP, data should be analysed at the smallest possible scales of resolution, finding a lens to analyse at the most local levels, without artificially constrained boundaries. Analysis of events within specific geographies should be at the smallest levels of enumeration so they can capture data against attributes in the most specific way determinable.

This is relevant to the aggregation of alcohol and smoking lifestyle data within this thesis and must be considered as a potential limitation to the case study. However, the incidence, survival and outcome data mapped were specifically enumerated to patient postcodes. This provided a small area, locally specific demonstration of where subjects were at the time of diagnosis.

9.2.4 Longitudinal timeframe

The cohort chosen for this case study fell between the years 2000 and 2013. This longitudinal data has strengths and limitations. Data capture within this timeframe allowed higher numbers of GOC diagnosed patients, which allowed data to be analysed at smaller scales of resolution. This has proven effective in several other clinical studies (Wang et al., 2010, Wang & Luo, 2005a, Hu et al., 2012).

Clinical improvements in care and services have occurred through the timespan of these data. Advances in GOC treatments and outcomes are always being sought, as are implementation plans for improved early diagnosis and referral processes. This needs to be considered against survival outcomes presented in this thesis. The median days survival identified in the results chapter concur with this statement, as survival in days showed increased across the cohort timeframe. This supports several affirmations in

small area health analysis literature that analysis of temporal data must account for improvements in clinical care over time (Wang et al., 2010, Goovaerts, 2006, Norman, 2010).

9.3 Recommendations for further research

This research sought to identify potential patterns in presentation and survival in gastroesophageal cancer, to evaluate how patients presented and survived between the years 2000 and 2013. It offered a geographically based analysis of retrospective cancer incidence data, to review whether there were any geographical patterns in presentation and survival. Through analysis of clusters in incidence, it revealed geographical areas with higher and lower than average diagnoses. However, inferences and predictions could not be based on this intelligence.

Further research is required to identify factors which are associated with the causality of cancers. As data becomes richer, capacity for analysis becomes more complex, and so does the ability to identify individuals through that analysis. Geographical information systems require significant computing power to manage, and databases must be adequately anonymised to protect the public.

Harnessing sociodemographic information to predict geographical areas with a potential to cancer incidence is a goal for the future. However, availability of datasets containing variables attributed to these cancers must be improved. Household level data on smoking and alcohol, on obesity and dietary habits could lead to improved research at small area levels.

There is a wealth of missing data in gastroesophageal cancer – this relates to TNM staging, site of cancer and how patients present for diagnosis. This absence of data is

being addressed at worldwide scale, but it must be recognised that individual patients and their choices will always impact on ability to collect 100% of data effectively. Not all patients will wish to undergo invasive diagnostic procedures or treatment and so missing data are intrinsic to any research in humans. Further research is required to support non-invasive cancer diagnostic techniques and hopefully this will occur alongside clinical improvements in cancer care.

The fact that most GOC patients present as ‘inpatients’ and through GP referrals and those with worse outcome, present as emergency is essential to informing cancer care provision. Targeting and reducing emergency presentations in gastroesophageal cancer has to be a priority in future research.

This research identified that GOC diagnosed patients under the age of 54 are more likely to be from lower socioeconomic groups, which requires further investigation. An analysis of socioeconomics, ageing and survival across countries is required, to determine whether this is an issue on a larger scale.

This thesis also highlighted the significant mortality in GOC before 6 months. Further review of the classifications of ‘1-year survival’ should be undertaken, as it did not capture over 39% of the subjects who died before 6 months. This research proposes that clearly defined survival times and ways to calculate survival should be investigated further.

References

Aansen, R.P., Vested, P., Sokolowski, I., Sondergaard, J., Olessen, F. (2011) Time intervals from first symptom to treatment of cancer: a cohort study of 2,212 newly diagnosed cancer patients. *BMC Health Services Research,* 11 284-291.

Abbas, J., Ojo, A., Orange, S (2009) Geodemographics - a tool for health intelligence *Public Health* 123, e35-e39.

Abdalla, A., Gunst, M., Ghaemmaghami, V., Gruszecki, A., C., Urban, J., Barber, R., C., Gentilello, L., M. & Shafi, S. (2012) Spatial analysis of injury-related deaths in Dallas County using a geographic information system. *Baylor University Medical Center Proceedings*, 25 (3), 208-213.

Abdel-Rahman, M., Stockton, D., Rachet, B., Hakulinen, T. & Coleman, M.P. (2009) What if cancer survival in Britain were the same as in Europe: how many deaths are avoidable? *British Journal of Cancer,* 101 Suppl 2 S115-24.

Abdullah, M., Karim, A.A. & Goh, K. (2010) Late presentation of esophageal cancer: observations in a multiracial South-East Asian population. *Journal of Digestive Diseases,* 11 (1), 28-33.

Abedi-Ardekani, B., Kamangar, F., Sotoudeh, M., Villar, S., Islami, F., Aghcheli, K., Nasrollahzadeh, D., Taghavi, N., Dawsey, S.M., Abnet, C.C., Hewitt, S.M., Fahimi, S., Saidi, F., Brennan, P., Boffetta, P., Malekzadeh, R. & Hainaut, P. (2011) Extremely high Tp53 mutation load in esophageal squamous cell carcinoma in Golestan Province, Iran. *PloS One,* 6 (12), e29488.

Abnet CC, Kamangar F, Islami F, Nasrollahzadeh D, Brennan P, Aghcheli K, Merat S, Pourshams A, Marjani HA, Ebadati A (2008a) Tooth loss and lack of regular oral hygiene are associated with higher risk of esophageal squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* 17(11): 3062–3068.

Abrams, J.A., Gonsalves, L. & Neugut, A.I. (2013) Diverging trends in the incidence of reflux-related and Helicobacter pylori-related gastric cardia cancer. *Journal of Clinical Gastroenterology*, 47 (4), 322-327.

Adair, T., Hoy, D., Dettrick, Z. & Lopez, A., D. (2011) Trends in oral, pharyngeal and oesophageal cancer mortality in Australia: the comparative importance of tobacco, alcohol and other risk factors. *Australian & New Zealand Journal of Public Health,* 35 (3), 212-219.

Adams A, Buckingham CD, Lindenmeyer A, (2008). The influence of patient and doctor gender on diagnosing coronary heart disease. *Sociology of Health and Illness* 30, 1–18.

Afzelius, P., Zedeler, K., Sommer, H., Mouridsen, H.T., Blichert-Toft, M. (1994) Patient's and doctor's delay in primary breast cancer. Prognostic implications. Acta Oncol. 33, 345-351.

Aguilar, I., Compés, L., Feja, C., Rabanaque, M.J. & Martos, C. (2013) Gastric cancer incidence and geographical variations: the influence of gender and rural and

socioeconomic factors, Zaragoza (Spain). *Gastric Cancer : Official Journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association,* 16 (2), 245-253.

Ahari, S.S., Agdam, F.B., Amani, F., Yazdanbod, A. & Akhghari, L. (2013) Analysis of the relationships between esophageal cancer cases and climatic factors using a Geographic Information System (GIS): a case study of Ardabil province in Iran. *Asian Pacific Journal of Cancer Prevention* 14 (3), 2071-2077.

Ajebon, M.O. and Norman, P., (2015). Can administrative data be used to create a geodemographic classification? Geographic Information Systems Registry UK Proceedings. pp.20-32.

Akram, M., Siddiqui, S.A. & Karimi, A.M. (2014) Patient Related Factors Associated with Delayed Reporting in Oral Cavity and Oropharyngeal Cancer. *International Journal of Preventive Medicine,* 5 (7), 915-919.

Alimoghaddam, K., Jalali, A., Aliabadi, L.S., Ghaffari, F., Maheri, R., Eini, E., Mashhadireza, M., Mousavi, S.A., Bahar, B., Jahani, M. & Ghavamzadeh, A. (2014) The outcomes of esophageal and gastric cancer treatments in a retrospective study, single center experience. *International Journal of Haematology-Oncology and Stem Cell Research,* 8 (2), 9-13.

Allum, W.H., Griffith, S.M., Watson, A. & Colin-jones, D. (2002) Guidelines for the management of oesophageal and gastric cancer. *Gut* 50 (5), v1-v23.

Allum, W.H., Blazeby, J.M., Griffin, S.M., Cunningham, D., Jankowski, J.A., Wong, R. & Association of Upper Gastrointestinal Surgeons of Great Britain and Ireland, the British Society of Gastroenterology and the British Association of Surgical Oncology (2011) Guidelines for the management of oesophageal and gastric cancer. *Gut,* 60 (11), 1449-1472.

Altoriki, N., Harrison, S. (2017). What is the role of neoadjuvant chemotherapy, radiation, and adjuvant treatment in resectable esophageal cancer? *Annals of Cardiothoracic surgery* 6(2) 749-756.

American Joint Committee on Cancer. *AJCC Cancer Staging Manual* available at https://cancerstaging.org/references-tools/deskreferences/Pages/ [accessed 12th Sept 2017].

Aragonés, N., Ramis, R., Pollán, M., Pérez-Gómez, B., Gómez-Barroso, D., Lope, V., Boldo, E.I., García-Pérez, J. & López-Abente, G. (2007) Oesophageal cancer mortality in Spain: a spatial analysis. *BMC Cancer,* 7 (1), 3-3.

Aronson, R.E., Wallis, A.B., O'Campo, P. & Schafer, P. (2007) Neighbourhood mapping and evaluation: a methodology for participatory community health initiatives. *Maternal & Child Health Journal,* 11 (4), 373-383.

Badreddine, R.J. & Wang, K.K. (2008) Barrett's esophagus: pathogenesis, treatment, and prevention. *Gastrointestinal Endoscopy Clinics of North America,* 18 (3), 495-512.

Bartsch, H., Ohshima, H., Pignatelli, B. & Calmels, S. (1992) Endogenously formed N-nitroso compounds and nitrosating agents in human cancer aetiology. *Pharmacogenetics,* 2 (6), 272-277.

Baughan, P., O'Neill, B. & Fletcher, E. (2009) Auditing the diagnosis of cancer in primary care: the experience in Scotland. *British Journal of Cancer,* 101 S87-S91.

Baxter, M.S.,MacKenzie, A.B., East, B.W. (1996) Natural decay series radionuclides in and around a large metal refinery. *Journal of Environmental Radioactivity*. 32, (1,2) 115–133

Bazemore, A., Phillips, R., Miyoshi, T. (2010) Harnessing Geographic Information systems to enable community oriented primary care *Journal of the American Family Based Medicine* (23)1. 22-30

Becker, U., Deis, A., Sorensen, T.I., Gronbaek, M., Borch-Johnsen, K., Muller, C.F., Schnohr, P. & Jensen, G. (1996) Prediction of risk of liver disease by alcohol intake, sex, and age: A prospective population study. *Hepatology,* 23 (5), 1025-1029.

Bektas, A., Yasa, M.H., Kuzu, I., Dogan, I., Unal, S. & Ormeci, N. (2001) Flow cytometric DNA analysis, and immunohistochemical p53, PCNA and histopathologic study in primary achalasia: preliminary results. *Hepato-Gastroenterology,* 48 (38), 408-412.

Bell, B.S., Hoskins, R., Pickle, L. & Wartenberg, D. (2006) Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. *International Journal of Health Geographics,* 5 (1), 49.

Bell, S., Wilson, K., Shah, T.I., Gersher, S. & Elliott, T. (2012) Investigating impacts of positional error on potential health care accessibility. *Spatial and Spatio-Temporal Epidemiology,* 3 (1), 17-29.

Beyer, K.M., Comstock, S. & Seagren, R. (2010) Disease maps as context for community mapping: a methodological approach for linking confidential health information with local geographical knowledge for community health research. *Journal of Community Health,* 35 (6), 635-644.

Beyer, K.M. & Rushton, G. (2009) Mapping cancer for community engagement. *Preventing Chronic Disease,* 6 (1), A03-A03.

Bhat, S., Coleman, H.G., Yousef, F., Johnston, B.T., McManus, D.T., Gavin, A.T., Murray, L.J. (2011) Risk of malignant progression in Barrett's esophagus patients: results from a large population-based study. *Journal of the National Cancer Institute,* 103 (13), 1049-1057.

Bhatt, V. & Tiwari, N. (2013) A spatial scan statistic for survival data based on Weibull distribution. *Statistics in Medicine,*21 (4) 67-679.

Bhopal, R.S. (2008) *Concepts of epidemiology: Integrating the ideas, theories, principles, and methods of epidemiology* Oxford: Oxford University Press.

Bithell., J, F & Stone,, R, A (1989) On statistical methods for analysing the geographical distribution of cancer cases near nuclear installations. *Journal of Epidemiology & Community Health,* 43 (1), 79-85.

Blakely, T., Barendregt, J.J., Foster, R.H., Hill, S., Atkinson, J., Sarfati, D. & Edwards, R. (2013a) The association of active smoking with multiple cancers: national census-cancer registry cohorts with quantitative bias analysis. *Cancer Causes & Control: CCC,* 24 (6), 1243-1255.

Blakely, T., Barendregt, J.J., Foster, R.H., Hill, S., Atkinson, J., Sarfati, D. & Edwards, R. (2013b) The association of active smoking with multiple cancers: national census-cancer registry cohorts with quantitative bias analysis. *Cancer Causes & Control : CCC,* 24 (6), 1243-1255.

Blackshaw, G.R., Stephens, M.R., Lewis, W.G., et al. (2004) Prognostic significance of acute presentation with emergency complications of gastric cancer. *Gastric Cancer*. 7(2):91–6.

Blom, R.L.G.M., Lagarde, S.M., Klinkenbijl, J.H.G., Busch, O.R.C. & van Berge Henegouwen, M.I. (2012) A high body mass index in esophageal cancer patients does not influence postoperative outcome or long-term survival. *Annals of Surgical Oncology,* 19 (3), 766-771.

Blot, W.J. & Li, J.Y. (1985) Some considerations in the design of a nutrition intervention trial in Linxian, People's Republic of China. *National Cancer Institute Monograph,* 69 29-34.

Blot, W.J., Li, J.Y., Taylor, P.R., Guo, W., Dawsey, S., Wang, G.Q., Yang, C.S., Zheng, S.F., Gail, M. & Li, G.Y. (1993) Nutrition intervention trials in Linxian, China: supplementation with specific vitamin/mineral combinations, cancer incidence, and disease-specific mortality in the general population. *Journal of the National Cancer Institute,* 85 (18), 1483-1492.

Booth, C., L. & Thompson, K., S. (2012) Barrett's esophagus: A review of diagnostic criteria, clinical surveillance practices and new developments. *Journal of Gastrointestinal Oncology,* 3 (3), 232-242.

Bosscher, M.R., van Leeuwen, B.L., Hoekstra, H.J. (2014) Surgical emergencies in oncology. *Cancer Treat Rev*. 40(8):1028–36.

Boulos, D.N.K., Ghali, R.R., Ibrahim, E.M., Boulos, M.N.K. & AbdelMalik, P. (2011) An eight-year snapshot of geospatial cancer research (2002-2009): clinico-epidemiological and methodological findings and trends. *Medical Oncology (Northwood, London, England),* 28 (4), 1145-1162.

Braun, V., Clarke, V. (2006) *Using thematic analysis in psychology. Qualitative Research in Psychology,* 3 (2). pp. 77-101. ISSN 1478-0887 Available at: http://eprints.uwe.ac.uk/11735 [accessed June 2015]

Brennan, J.A., Boyle, J.O., Koch, W.M., Goodman, S.N., Hruban, R.H., Eby, Y.J., Couch, M.J., Forastiere, A.A. & Sidransky, D. (1995) Association between Cigarette Smoking and Mutation of the p53 Gene in Squamous-Cell Carcinoma of the Head and Neck. *New England Journal of Medicine,* 332 (11), 712-717.

Brewer, C. (2006) Basic mapping principles for visualizing cancer data using geographic information systems (GIS). *American Journal of Preventive Medicine,* 30 S25-S36.

Brewster, D., Fraser, L. & McKinney, P.A. (2000) Socioeconomic status and risk of adenocarcinoma of the oesophagus and cancer of the gastic cardia in Scotland. *British Journal of Cancer,* 83 (3), 387-390.

Brierley, J.D.; Gospodarowicz, M.K.; Wittekind, Ch., eds. (2017). *TNM classification of malignant tumours* (8th ed.). Chichester, West Sussex, UK: Wiley-Blackwell

Brocklehurst, P., Kujan, O., O'Malley, L.A., Ogden, G., Shepherd, S. & Glenny, A.M. (2013) Screening programmes for the early detection and prevention of oral cancer. *The Cochrane Database of Systematic Reviews*, 11 CD004150.

Broker, L.E., HurenKamp, G.B., Riet, G.T., Schellevis, F.G., Grundmeijer, H.G. & Van Weert, H.C. (2009) Upper Gastrointestinal symptoms, psychosocial comorbidity nad health care seeking in general practice : population based case control study. *BMC Family Practice,* 10 (63),.

Bromley, D.B. (1990) Academic contributions to psychological counselling 1. A philosophy of science for the study of individual cases. *Counselling psychology quarterly* 3 (1) 299-307.

Brown LM, Hoover R, Silverman D, Baris D, Hayes R, Swanson GM et al. (2001) Excess incidence of squamous cell esophageal cancer among US Black men: role of social class and other risk factors. *American Journal of Epidemiology* 153:114–122

Brown, L.M., Devesa, S.S. & Chow, W.-. (2008) Incidence of adenocarcinoma of the esophagus among white Americans by sex, stage, and age. *Journal of the National Cancer Institute,* 100 (16), 1184-1187.

Bryere, J., Launay, L., Guittet, L., Launoy, G., Dejardin, O., Bouvier, V., Colonna, M., Guizard, A., Troussard, X., Pornet, C., Galateau-Salle, F. & Bara, S. (2014) Socioeconomic environment and cancer incidence: a French population-based study in Normandy. *BMC Cancer,* 14 (1), 87-87.

Bryman, A., Cramer, D. (2011) *Quantitative Data Analysis with IBM SPSS 17, 18 and 19. A guide for social scientists.* New York: Routledge.

Buas, M.F. & Vaughan, T.L. (2013) Epidemiology and risk factors for gastroesophageal junction tumors: understanding the rising incidence of this disease. *Seminars in Radiation Oncology,* 23 (1), 3-9.

Bus, P., Aarts, M.J., Lemmens, V.E., van Oijen, M.G., Creemers, G.J., Nieuwenhuijzen, G.A., van Baal, J.W. & Siersema, P.D. (2012) The Effect of Socioeconomic Status on Staging and Treatment Decisions in Esophageal Cancer. *Journal of Clinical Gastroenterology,* 156 (3) 526-532.

Bus, P., Lemmens, V.E., van Oijen, M.G., Creemers, G.J., Nieuwenhuijzen, G.A., van Baal, J.W. & Siersema, P.D. (2014) Prognostic factors for medium- and long-term survival of esophageal cancer patients in the Netherlands. *Journal of Surgical Oncology,* 109 (5), 465-471.

Cancer Research UK, (2010), *Improving Cancer Outcomes: An analysis of implementation of the UK's cancer strategies 2006-2010.* London, Cancer Research UK.

Cancer Research UK (2014) Cancer Diagnosis and treatment statistics. Available at http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer [accessed June 2015]

Cancer research UK (2015) Cancer statistics, oesophageal cancer available at; http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer [accessed Oct 2017]

Cancer Research UK (2016) *Cancer statistics oesophageal cancer* available at http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer#heading-Zero [accessed Nov 2017]

Cancer Research UK (2016). Cancer diagnosis and treatment statistics. 2016 http://www.cancerresearchuk.org/health-professional/cancer-statistics/diagnosis-and-treatment#heading-Zero [accessed April 2018]

Cancer research UK (2017) Oesophageal cancer survival statistics available at http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/survival#heading-Zero [accessed 092017].

Cancer Research UK (2017). Accelerate, Coordinate, Evaluate (ACE) programme. Available at; http://www.cancerresearchuk.org/health-professional/early-diagnosis-activities/ace-programme [accessed May 2018].

Candace, I.J., Nykiforuk, J., Flaman, L. (2011) Geographic information systems for health promotion and public health: a review *Health Promotion practice* 12 (1) 63-73.

Caprarelli, G. & Fletcher, S. (2014) *A brief review of spatial analysis concepts and tools used for mapping, containment and risk modelling of infectious diseases and other illnesses. England*: Cambridge University Press.

Carr, J.S., Zafar, S.F., Saba, N., Khuri, F.R. & El-Rayes, B.F. (2013) Risk factors for rising incidence of esophageal and gastric cardia adenocarcinoma. *Journal of Gastrointestinal Cancer,* 44 (2), 143-151.

Cassetti, T., La Rosa, F., Rossi, L., D'Alo, D. & Stracci, F. (2008) Cancer incidence in men: a cluster analysis of spatial patterns. *BMC Cancer,* 8 (1), 344.

Cassell, C., Symon, G. (1994) *Qualitative methods in organizational research: A practical guide.* 209–229. London: Sage

Caygill, C.P., Royston, C., Charlett, A., Wall, C.M., Gatenby, P.A., Ramus, J.R., Watson, A., Winslet, M., Hourigan, C.S. & Dev Bardhan, K. (2011) Barrett's, blood groups and progression to oesophageal cancer: is nitric oxide the link? *European Journal of Gastroenterology & Hepatology,* 23 (9), 801-806.

Chadwick, G., Groene, O., Cromwell, D., Hardwick, R.H., Riley, S., Crosby, T.D.L. and Greenaway, K., (2013), *The National Oesophago-Gastric Cancer Audit* London, Health and Social Care Information Centre.

Chan, Y.V. (2004) Biostatistics 203. Survival analysis *Singapore medical journal* 45 (6) 249-256.

Chen, D., Zheng, X.F., Yang, Z.Y., Liu, D.X., Zhang, G.Y., Jiao, X.L. & Zhao, H. (2012) S100A4 silencing blocks invasive ability of esophageal squamous cell carcinoma cells. *World Journal of Gastroenterology* 18 (9), 915-922.

Chen, W.Q., He, Y.T., Sun, X.B., Wen, D.G., Chen, Z.F. & Zhao, D.L. (2011) Analysis of risk factors for upper gastrointestinal cancer in China: a multicentric population-based case-control study. *Chinese Journal of Preventive Medicine,* 45 (3), 244-248.

Chen, W., He, Y., Zheng, R., Zhang, S., Zeng, H., Zou, X. & He, J. (2013) Esophageal cancer incidence and mortality in China, 2009. *Journal of Thoracic Disease,* 5 (1), 19-26.

Cheng, K.K. & Day, N.E. (1996) Nutrition and esophageal cancer. *Cancer Causes and Control,* 7 33-40.

Chi, W., Wang, J., Li, X., Zheng, X. & Liao, Y. (2008) Analysis of geographical clustering of birth defects in Heshun county, Shanxi province. *International Journal of Environmental Health Research,* 18 (4), 243-252.

Choi, K.W., Wong, N.S., Lee, L.Y. & Lee, S.S. (2010) Surveillance of febrile patients in a district and evaluation of their spatiotemporal associations: a pilot study. *BMC Public Health,* 10 84-84.

Choi, S., E. & Hur, C. (2012) Screening and surveillance for Barrett's esophagus: current issues and future directions. *Current Opinion in Gastroenterology,* 28 (4), 377-381.

Cohen, J.W. (1988) *Statistical Power Analysis for the behavioural sciences* (2nd Edition) Hillside, NJ; Lawrence Erlbaum Associates.

Coleman, M.P., Babb, P., Damieki, P., GrosClaude, P.C., Honjo, S. and Jones, J., (1999), *Cancer survival trends in England and Wales 1971 – 1995. Deprivation and NHS region: studies on medical and population subjects no 61.* London, HMSO.

Coleman, H.G., Bhat, S., Johnston, B.T., McManus, D., Gavin, A.T. & Murray, L.J. (2012) Tobacco smoking increases the risk of high-grade dysplasia and cancer among patients with Barrett's esophagus. *Gastroenterology,* 142 (2), 233-240.

Collins, S.E., Haining, R.P., Bowns, I.R., Crofts, D.J., Williams, T.S., Rigby, A.S. & Hall, D.M. (1998) Errors in postcode to enumeration district mapping and their effect on small area analyses of health data. *Journal of Public Health Medicine,* 20 (3), 325-330.

Congdon, P. (2012) Spatial Health Factors with Selection among Multiple Causes: Lung Cancer in U.S. Counties. *Communications in Statistics - Theory and Methods*, 41 (11), 1933-1953.

Corley, D.A. (2008) Cancer incidence in Barrett's esophagus: does it really matter, and who's counting anyway? *Gastrointestinal Endoscopy,* 67 (3), 399-401.

Coupland, V.H., Allum, W., Blazeby, J.M., Mendall, M.A., Hardwick, R.H., Linklater, K.M., Moller, H. & Davies, E.A. (2012) Incidence and survival of oesophageal and gastric cancer in England between 1998 and 2007, a population-based study. *BMC Cancer,* 12 11.

Courtney, K.E., Ashenhurst, J., Bacio, G., Moallem, N., Bujarski, S., Hartwell, E. & Ray, L.A. (2013) Craving and subjective responses to alcohol administration: validation of the desires for alcohol questionnaire in the human laboratory. *Journal of Studies on Alcohol and Drugs,* 74 (5), 797-802.

Crawford, S., Davis, J., Siddiqqui, N., deCastecker, L., Gillis, C., Penney, C. (2002) The waiting time paradox, Population based retrospective study of treatment delay and survival of women with endometrial cancer in Scotland *British Medical Journal* 325: 196.

Cronk, C.E., Gangnon, R., Cossette, S., McElroy, J.A. & Pelech, A.N. (2011) Modeling geographic risk of complex congenital heart defects in Eastern Wisconsin. *Birth Defects Research. Part A, Clinical and Molecular Teratology,* 91 (7), 631-641.

Cronin-Fenton, D.P., Sharp, A.E., Carsin, A., Comber, H. (2007) Patterns of care and effects on mortality for cancers of the oesophagus and gastric cardia: a population based study *European journal of cancer* 43 565-575

Curado, M.P., Edwards, B., Shin, H.R., Torm, H., Ferlay, J., Heanue, M. and Boyle, P., (2009), *Cancer incidence in five continents* 160, Lyons, France, IARC Scientific Publication.

Danesh, J., Gault, S., Semmence, J., Appleby, P. & Peto, R. (1999) Postcodes as useful markers of social class: population based study in 26000 British households. *British Medical Journal (Clinical Research Ed.),* 318 (7187), 843-844.

Danzig, M.R., Weinberg, A.C., Ghandour, R.A., Kotamarti, S., McKiernan, J.M., Badani, K.K. (2014) The Association Between Socioeconomic Status, Renal Cancer Presentation, and Survival in the United States: A Survival, Epidemiology, and End Results Analysis *Urology* 84, (3) 583-589

Dar, N.A., Islami, F., Bhat, G.A., Shah, I.A., Makhdoomi, M.A., Iqbal, B., Rafiq, R., Lone, M.M., Abnet, C.C., Boffetta, P., (2013a) Poor oral hygiene and risk of esophageal squamous cell carcinoma in Kashmir *British Journal of Cancer* 109, 1367–1372

Dar, N.A., Shah, I.A., Bhat, G.A., Makhdoomi, M.A., Iqbal, B., Rafiq, R. (2013b) Socioeconomic status and esophageal squamous cell carcinoma risk in Kashmir, India. *Cancer Science* 104:1231–1236

De Angelis, R., Sant, M., Coleman, M.P. (2014) Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE--5-a population-based study. *Lancet Oncol.* 15:23–34.

Defoe, S.G., Pennathur, A., Flickinger, J.C., Heron, D.E., Gibson, M.K., Luketich, J.D. & Greenberger, J.S. (2011) Retrospective review of patients with locally advanced esophageal cancer treated at the University of Pittsburgh. *American Journal of Clinical Oncology,* 34 (6), 587-592.

Department of Health (UK) (2012) *Health Profile - East Riding of Yorkshire*. DoH, crown copyright.

Department of Health (2014) The National Cancer Strategy; 4th Annual report available online https://www.gov.uk/government/publications/the-national-cancer-strategy-4th-annual-report [accessed March 2018].

Dhonapal, R., Saraswathi, T., Govinel, R. (2010) Cancer cahexia. *Journal of maxillofacial pathology.* 15(3), 257-260.

Di Pietro, M., Fitzgerald, R.C. (2013a) Screening and risk stratification for Barrett's esophagus: how to limit the clinical impact of the increasing incidence of esophageal adenocarcinoma. *Gastroenterology Clinics of North America,* 42 (1), 155-173.

Diederich, S. (2007) Staging of oesophageal cancer. *Cancer Imaging.* 7. 42-57.

Dixon-Woods, M., Bonas, S., Booth, A., Jones, D.R., Miller, T., Sutton, A.J., Shaw, R.L., Smith, J.A., Young, B. (2006a) How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research* 6 (1) 27-44

Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Katbamna, S., Olsen, R., Smith, L., Riley, R., Sutton, A.J. (2006b) Conducting a

critical interpretative synthesis of the literature on access to healthcare by vulnerable groups. *Medical Research Methodology* BMC

Dolan, K., Sutton, R., Walker, S.J., Morris, A.I., Campbell, F. & Williams, E.M. (1999) New classification of oesophageal and gastric carcinomas derived from changing patterns in epidemiology. *British Journal of Cancer,* 80 (5-6), 834-842.

Drewnowski, A., Aggarwal, A., Hurvitz, P., M., Monsivais, P. & Moudon, A., V. (2012) Obesity and Supermarket Access: Proximity or Price? *American Journal of Public Health,* 102 (8), e74-80.

Dubecz, A., Gall, I., Solymosi, N., Schweigert, M., Peters, J.H., Feith, M. & Stein, H.J. (2012) Temporal trends in long-term survival and cure rates in esophageal cancer: a SEER database analysis. *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer,* 7 (2), 443-447.

Dulin, M.F., Luddden, T., Tapp, H., Blackwell, J., Urquettia de Hernandez, B., Smith, H., Furuseth, U. (2010) Using geographic Information systems (GIS) to understand a community's primary care needs *Journal of American Family Based Medicine* 23 (1) 13-21 available at www.jabfm.2010.01.090135 [accessed 062015].

Dummer, T (2008) Health geography; supporting public health policy and planning *Canadian Medical Association Journal* 178 (9) 1177-1180

Dutta, S., Going, J.J., Crumley, A.B., Mohammed, Z., Orange, C., Edwards, J., Fullarton, G.M., Horgan, P.G. & McMillan, D.C. (2012) The relationship between tumour necrosis, tumour proliferation, local and systemic inflammation, microvessel density and survival in patients undergoing potentially curative resection of oesophageal adenocarcinoma. *British Journal of Cancer,* 106 (4), 702-710.

Dwivedi, A.K., Dwivedi, S.N., Deo, S., Shukla, R., Pandey, A. & Dwivedi, D.K. (2012) An epidemiological study on delay in treatment initiation of cancer patients. *Health* 4 66-79.

Elebead, F.M., Hamid, A., Hilmi, H.S. & Galal, H. (2012) Mapping Cancer Disease Using Geographical Information System (GIS) in Gezira State-Sudan. *Journal of Community Health,.*

Elferink, M.A., Pukkala, E., Klaase, J.M. & Siesling, S. (2012) Spatial variation in stage distribution in colorectal cancer in the Netherlands. *European Journal of Cancer,* 48 (8), 1119-1125.

Elit, L., O'Leary, E., Pond, G., Smow, H. (2013) Impact of wait times on survival for women with uterine cancer *Journal of clinical oncology* 31: 67.

Elliott, P., Wakefield, J., Best, N. & Briggs, D. (eds) (2009) *Spatial Epidemiology: Methods and Applications.* Oxford Brookes.

Elliott, P. & Savitz, D.A. (2008) Design issues in small-area studies of environment and health. *Environmental Health Perspectives,* 116 (8), 1098-1104.

Ellis, L., Coleman, M.P. & Rachet, B. (2012) How many deaths would be avoidable if socioeconomic inequalities in cancer survival in England were eliminated? A national population based study 1996-2006. *European Journal of Cancer,* 48 270-278.

Elliss-Brookes, L., McPhail, S., Ives, A., Greenslade, M., Shelton, J., Hiom, S. & Richards, M. (2012) Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *British Journal of Cancer,* 107 (8), 1220-1226.

Fagan, K.J., Irvine, K.M., Kumar, S., Bates, A., Horsfall, L.U., Feeney, G.F. & Powell, E.E. (2013) Assessment of alcohol histories obtained from patients with liver disease: opportunities to improve early intervention. *Internal Medicine Journal,* 43 (10), 1096-1102.

Fajans, P., Simmons, R., Ghiron, L (2006) Helping public sector health systems innovate: The strategic approach for strengthening reproductive policies and programs. *American Journal of Public Health* 96 435-440.

Feigin, R., Oram, A., Sjeburg, G. (1991) *A case for case study*. Chapel Hill; University of North Carolina Press.

Flyubjerg, F (2006) Five Misunderstandings about case study research. *Qualitative inquiry* 12(2) 219-245.

Forbes, L.J.L., Warburton, F., Richards, M.A. & Ramirez, A.J. (2014) Risk factors for delay in symptomatic presentation: a survey of cancer patients. *British Journal of Cancer,* 111 (3), 581-588.

Foster, D.W., Yeung, N. & Neighbours, C. (2013) I think I can't: Drink refusal self-efficacy as a mediator of the relationship between self-reported drinking identity and alcohol use. *Addictive Behaviours,* 14 (4) 105-9.

Foulc, P., Nguyen, J.M., Dreno, B. (2003) Prognostic factors in Sezary syndrome - a study of 28 patients *British Journal of Dermatology.* 149 1152-1158.

Fransen, G.A.J., Janssen, M.J.R., Muris, J.W., Laheij, R.J.F. & Jansen J.B.M.J. (2004) Meta analysis : The diagnostic value of alarm symptoms for upper gastrointestinal malignancy. *Alimentary Pharmacology & Therapeutics,* 20 1045-1052.

Frye, C. (2001) Making maps that communicate. *ArcUser,* 4 38-43.

Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. 2015. Available from http://globocan.iarc.fr [accessed 0417].

Gardner MJ, Snee MP; Hall AJ; Powell CA; Downes S; Terrell JD (1990) Results of case-control study of leukaemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. *British Medical Journal*. 1990; 300:423–429.

Galdas PM, Cheater F, Marshall P. (2005) Men and health help-seeking behaviour: literature review. *Journal of Advanced Nursing*. 49:616–23

Gatenby, P.A., Hainsworth, A., Caygill, C., Watson, A. & Winslet, M. (2011) Projections for oesophageal cancer incidence in England to 2033. *European Journal of Cancer Prevention : The Official Journal of the European Cancer Prevention Organisation (ECP),* 20 (4), 283-286.

Gavin, A.T., Francisci, S., Foschi, R., Donnelly, D.W., Lemmens, V., Brenner, H. & Anderson, L.A. (2012) Oesophageal cancer survival in Europe: A EUROCARE-4 study. *Cancer Epidemiology,* 36 (6), 505-512.

Gentil, J., Dabakuyo, T., Sandrine, Ouedraogo, S., Poillot, M., Dejardin, O. & Arveux, P. (2012) For patients with breast cancer, geographic and social disparities are independent determinants of access to specialized surgeons. A eleven-year population-based multilevel analysis. *BMC Cancer,* 12 351-351.

Gerritsen AA, Deville WL (2009). Gender differences in health and health care utilisation in various ethnic groups in the Netherlands: a cross-sectional study. *BMC Public Health* 9, 109.

Gibbs, J.F., Rajput, A., Chadha, K.S., Douglas, W.G., Hill, H., Nwogu, C., Nava, H.R. & Sabel, M.S. (2007) The changing profile of esophageal cancer presentation and its implication for diagnosis. *Journal of the National Medical Association,* 99 (6), 620-626.

Goli, A., Oroei, M., Jalalpour, M., Faramarzi, H. & Askarian, M. (2013a) The spatial distribution of cancer incidence in fars province: A GIS-based analysis of cancer registry data. *International Journal of Preventive Medicine,* 4 (10), 1122-1130.

Goli, A., Oroei, M., Jalalpour, M., Faramarzi, H. & Askarian, M. (2013b) The Spatial Distribution of Cancer Incidence in Fars Province: A GIS-Based Analysis of Cancer Registry Data. *International Journal of Preventive Medicine,* 4 (10), 1122.

González Ortiz, D.,I. & Toro, D.H. (2009) Esophageal cancer subtypes and survival rates at the VA Caribbean Healthcare System: a 10-year experience. *Boletín De La Asociación Médica De Puerto Rico,* 101 (3), 14-17.

Goodman, M.S. (2010) Comparison of small-area analysis techniques for estimating prevalence by race. *Preventing Chronic Disease,* 7 (2), A33-A33.

Goovaerts, P. & Xiao, H. (2012) The impact of place and time on the proportion of late-stage diagnosis: the case of prostate cancer in Florida, 1981-2007. *Spatial and Spatio-Temporal Epidemiology,* 3 (3), 243-253.

Goovaerts, P. & Xiao, H. (2011) Geographical, temporal and racial disparities in late-stage prostate cancer incidence across Florida: a multiscale joinpoint regression analysis. *International Journal of Health Geographics,* 10 63-63.

Gopal, D.V., Reichelderfer, M., Gaumnitz, E.A., Harter, J. & Jobe, B.A. (2004) Barrett's esophagus: is screening and surveillance justified? *Disease Management & Health Outcomes,* 12 (6), 353-361.

Green, C., Yu, B., Nancy & Marrie, R., Ann (2013) Exploring the Implications of Small-Area Variation in the Incidence of Multiple Sclerosis. *American Journal of Epidemiology,* 178 (7), 1059-1066.

Gregorio, D.I., DeChello, L.M., Samociuk, H. & Kulldorff, M. (2005) Lumping or splitting: seeking the preferred areal unit for health geography studies. *International Journal of Health Geographics,* 4 6-10.

Griffiths, H. (2011) Management of Barrett's oesophagus. *Gastrointestinal Nursing,* 9 (2), 34-35.

Grotenhuis, B.A., Van Hagen, P., Wijnhoven, B.P.L., Spaander, M.C.W., Tilanus, H.W. & Van Lanschot, Jan J. B. (2010) Delay in Diagnostic Workup and Treatment of Esophageal Cancer. *Journal of Gastrointestinal Surgery,* 14 (3), 476-483.

Grubesic, T.H., Miller, J.A., Murray, A.T. (2014) Geospatial and geodemographic insights for diabetes in the United States. *Applied Geography* 55, 117-126

Guo, F., Liu, Y., Wang, X., He, Z., Weiss, N.S., Madeleine, M.M., Liu, F., Tian, X., Song, Y., Pan, Y., Ning, T., Yang, H., Shi, X., Lu, C., Cai, H. & Ke, Y. (2012) Human Papillomavirus Infection and Esophageal Squamous Cell Carcinoma: A Case-Control Study. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology, 5 (8) 1194-1205*

Guzman-Laura, K., Bolibar, R.I., Alepuz, M., Gonzalez, M., Martin, M. (2011) Impact on the care and time to tumour stage of a program of rapid diagnosis and treatment of colorectal cancer *rev espania enfermerie diagnosis* 103 13-19.

Hartley, J. F. (1994) Case studies in organizational research. In Cassell C., Symon, G., (1994) Qualitative methods in organizational research: A practical guide. 209–229. London: Sage

Hagedoorn, P., Vandenheede, H., Vanthomme, K., Willaert, D., Gadeyne, S. (2016) A cohort study into head and neck cancer mortality in Belgium (2001–11): Are individual socioeconomic differences conditional on area deprivation? *Oral Oncology.* 61 ; 76-82

Hahn, K.A. (2014) Book review - Epidemiology Matters. *American Journal of Epidemiology.* 180 (11) 1126-1127.

Hamel, J. (1993) C*ase Study Methods Qualitative Research Method*s (32) London: Sage.

Hamilton, W., Hajioff, S., Graham, J., Schmidt-Hansen, M. (2015) Suspected cancer (part 2-adults): reference tables from updated NICE guidance. *British Medical Journal* 350(h) 30-44.

Hammoud, G.M., Hammad, H. & Ibdah, J.A. (2014) Endoscopic assessment and management of early esophageal adenocarcinoma. *World Journal of Gastrointestinal Oncology,* 6 (8), 275-288.

Hanafi-Bojd, A., Vatandoost, H., Oshaghi, M.A., Charrahy, Z., Haghdoost, A.A., Zamani, G., Abedi, F., Sedaghat, M.M., Soltani, M., Shahi, M. & Raeisi, A. (2012) Spatial analysis and mapping of malaria risk in an endemic area, south of Iran: a GIS based decision making for planning of control. *Acta Tropica,* 122 (1), 132-137.

Hanchette, C.L. & Schwartz, G.G. (1992) Geographic patterns of prostate cancer mortality. Evidence for a protective effect of ultraviolet radiation. *Cancer,* 70 (12), 2861-2869.

Hansen, R., Vedsted, P., Sokolowski, I., Sondergaard, J. & Olessen, F. (2011) Time intervals from first symptom to treatment of cancer: a cohort study of 2,212 newly diagnosed cancer patients.. *BMC Health Services Research,* 11 284-291.

Hardwick, R.H., Barham, C.P., Ozua, P., Newcomb, P.V., Savage, P., Powell, R., Rahamin, J. & Alderson, D. (1997) Immunohistochemical detection of p53 and c-erbB-2 in oesophageal carcinoma; no correlation with prognosis. *European Journal of Surgical Oncology: The Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology,* 23 (1), 30-35.

Harris, R (2016) *Quantitative Geography - The Basics.* London: Sage.

Hashemi, N., Loren, D., DiMarino, A.J. & Cohen, S. (2009) Presentation and prognosis of esophageal adenocarcinoma in patients below age 50. *Digestive Diseases and Sciences,* 54 (8), 1708-1712.

Hastert, T.A., Ruterbusch, J.J., Beresford, S.A., Sheppard, L., Whit, E., (2016) Contribution of health behaviors to the association between area-level socioeconomic status and cancer mortality. *Social Science & Medicine (*148), 52-58

Hawker, S., Payne, S., Kerr, C., Hardey, M. & Powell, J. (2002) Appraising the evidence: reviewing disparate data systematically. *Qualitative Health Research,* 12 (9), 1284-1299.

Hayes, K., Newton, S. (2003) *Promoting learning in organizations through embedded case studies*. Available at; http://notebook.lausd.net/pls/ptl/docs/page/ca_lausd/fldr_organizations/fldr_plcy_res_dev/par_division_main/research_unit/publications/conference_presentations/embedded%20case%20studies,%20aea%202003.pdf [accessed 030317].

Haynes, R., Gale, S., Lovett, A., & Bentham, G. (1996). Unemployment rate as an updatable health needs indicator for small areas. Journal of Public Health Medicine, 18(1), 27–32.

Healthcare Quality Improvement Partnership (2016) Report- National Oesophagogastric Cancer Audit available at https://www.hqip.org.uk/resources/national-oesophago-gastric-cancer-audit-2016/ [accessed March 2018]

Healthcare Quality Improvement Partnership (2015) Report- National Oesophagogastric Cancer Audit available at https://www.hqip.org.uk/resources/national-oesophago-gastric-cancer-audit-report-2015/ [accessed March 2018].

Hendryx, M., Fedorko, E. & Anesetti-Rothermel, A. (2010a) A geographical information system-based analysis of cancer mortality and population exposure to coal mining activities in West Virginia, United States of America. *Geospatial Health*, 4 (2), 243-256.

Hendryx, M., Fedorko, E. & Anesetti-Rothermel, A. (2010b) A geographical information system-based analysis of cancer mortality and population exposure to coal mining activities in west Virginia, united states of America. *Geospatial Health*, 4 (2), 243-256.

Hewapathirana, R. & Wijayarathna, G. (2010) Spatiotemporal antibiotic resistance pattern monitoring using geographical information system based hierarchical cluster analysis... MEDINFO 2010: Proceedings of the 13th World Congress on Medical Informatics, Part 1. *Studies in Health Technology & Informatics*, 160 501-504.

Hexi, Z., Zhichao, C., Jinpeng, C., Xiaoling, Z., Weiren, G., Andong, H., Yukai, D., Yikai, Z., Youjie, W. (2010) The high inciodence of esophageal cancer in parts of china may result primarily from genetic rather than environmental factors *Diseases of the oesophagus 23 392-397.*

Higgs, G., Gould, M. Is there a role for GIS in the 'new NHS'? *Health and Place* 7 247-259

Hiom, S.C. (2015) Diagnosing cancer earlier: reviewing the evidence for improving cancer survival. *British Journal of Cancer*. Mar 31;112 S1-5.

Holmang, S., Johansson, S.L. (2006) Impact of diagnostic and treatment delay on survival in patients with renal pelvic and ureteral cancer *Scandinavian Journal of Urology and Nephrology* 40 479-494.

Holscher, A.H., Bollschweiler, E., Schroder, W., Metzger, R., Gutschow, C. & Drebber, U. (2011) Prognostic impact of upper, middle, and lower third mucosal or submucosal infiltration in early esophageal cancer. *Annals of Surgery*, 254 (5), 802-7; discussion 807-8.

Homs, M.Y., van Oijen, M., Wijnhoven, B.P., van Hillegersberg, R., de Boer-Dennert, M. & Siersema, P.D. (2012) Changes in diagnostic and treatment strategies

of oesophageal cancer in the period from 2001 to 2009: a survey in the Netherlands. *European Journal of Gastroenterology & Hepatology*, 24 (2), 126-133.

Hong, L., Zhang, H., Zhao, Q., Han, Y., Yang, J. & Brain, L. (2013) Relation of excess body weight and survival in patients with esophageal adenocarcinoma: a meta-analysis. *Diseases of the Esophagus*, 26 (6), 623-627.

Hoyo, C., Cook, M.B., Kamangar, F., Freedman, N.D., Whiteman, D.C., Bernstein, L., Brown, L.M., Risch, H.A., Ye, W., Sharp, L., Wu, A.H., Ward, M.H., Casson, A.G., Murray, L.J., Corley, D.A., Nyren, O., Pandeya, N., Vaughan, T.L., Chow, W.H. & Gammon, M.D. (2012) Body mass index in relation to oesophageal and oesophagogastric junction adenocarcinomas: a pooled analysis from the International BEACON Consortium. *International Journal of Epidemiology*, 41 (6), 1706-1718.

Hu, J., Wang, F., Sun, J., Sorrentino, R. & Ebadollahi, S. (2012) A healthcare utilization analysis framework for hot spotting and contextual anomaly detection. *AMIA...Annual Symposium Proceedings/AMIA Symposium.AMIA Symposium,* 2012 360-369.

Huang, L., Pickle, L.W. & Das, B. (2008) Evaluating spatial methods for investigating global clustering and cluster detection of cancer cases. *Stat Med,* 27 5111-5142.

Hudson, C.G. & Abbott, M.W. (2013) Modeling the geographic distribution of serious mental illness in New Zealand. *Social Psychiatry & Psychiatric Epidemiology*, 48 (1), 25-36.

Hunt, R.H., Sumanac, K. & Huang, J.Q. (2001) Review article: should we kill or should we save Helicobacter pylori? *Alimentary Pharmacology & Therapeutics,* 15 Suppl 1 51-59.

International Agency for Research on Cancer (2017). List of Classifications by cancer sites with sufficient or limited evidence in humans, Volumes 1 to 116* (link is external). [Accessed 04/2017].

International Agency for Research on Cancer (IARC) (2015), CancerBase No. 10 [Internet]. Lyon, France: International Agency for Research on Cancer; 2012. Available from: http://globocan.iarc.fr, [accessed November 2015]

Islami, F., Kamangar, F., Aghcheli, K., Fahimi, S., Semnani, S., Taghavi, N., Marjani, H.A., Merat, S., Nasseri-Moghaddam, S., Pourshams, A., Nouraie, M., Khatibian, M., Abedi, B., Brazandeh, M.H., Ghaziani, R., Sotoudeh, M., Dawsey, S.M., Abnet, C.C., Taylor, P.R., Malekzadeh, R. (2004) Epidemiologic features of upper gastrointestinal tract cancers in Northeastern Iran. *British Journal of Cancer* **90**(7): 1402–1406

Islami, F., Kamangar, F., Nasrollahzadeh, D., Aghcheli, K., Sotoudeh, M., Abedi-Ardekani, B. (2009) Socio-economic status and oesophageal cancer: results from a population-based case-control study in a high-risk area. *International Journal of Epidemiology* 38:978–988

Jacquez, G.M. (2004) Current practices in the spatial analysis of cancer: flies in the ointment. *International Journal of Health Geography,* 3 22.

Jansen L, Eberle A, Emrich K, Gondos A, Holleczek B, Kajuter H et al. (2014) Socioeconomic deprivation and cancer survival in Germany: an ecological analysis in 200 districts in Germany. *International Journal of Cancer* 134:2951–2960

Jayasekera, C.S., Desmond, P.V., Holmes, J.A., Kitson, M., Taylor, A.C. (2012) Cluster of 4 cases of esophageal squamous cell cancer developing in adults with surgically corrected esophageal atresia--time for screening to start. *Journal of Pediatric Surgery* 47(4). 646-51

Jensen, A.R., Nellemann, H.M., Overgaard, J. (2007) Tumor progression in waiting time for radiotherapy in head and neck cancer. *Radiother Oncol.* 84: 5-10. 10.1016/j.radonc.2007.04.001

Jeurnink, S.M., Buchner, F.L., Bueno-de-Mesquita, H.B., Siersema, P.D., Boshuizen, H.C., Numans, M.E., Dahm, C.C., Overvad, K., Tjonneland, A., Roswall, N., Clavel-Chapelon, F., Boutron-Ruault, M.C., Morois, S., Kaaks, R., Teucher, B., Boeing, H., Buijsse, B., Trichopoulou, A., Benetou, V., Zylis, D., Palli, D., Sieri, S., Vineis, P., Tumino, R., Panico, S., Ocke, M.C., Peeters, P.H., Skeie, G., Brustad, M., Lund, E., Sanchez-Cantalejo, E., Navarro, C., Amiano, P., Ardanaz, E., Ramon Quiros, J., Hallmans, G., Johansson, I., Lindkvist, B., Regner, S., Khaw, K.T., Wareham, N., Key, T.J., Slimani, N., Norat, T., Vergnaud, A.C., Romaguera, D. & Gonzalez, C.A. (2012a) Variety in vegetable and fruit consumption and the risk of gastric and esophageal cancer in the European prospective investigation into cancer and nutrition. *International Journal of Cancer.Journal International Du Cancer,.*

Jeurnink, S.M., Büchner, F.L., Bueno-de-Mesquita, H.B., Siersema, P.D., Boshuizen, H.C., Numans, M.E., Dahm, C.C., Overvad, K., Tjønneland, A., Roswall, N., Clavel-Chapelon, F., Boutron-Ruault, M.C., Morois, S., Kaaks, R., Teucher, B., Boeing, H., Buijsse, B., Trichopoulou, A., Benetou, V., Zylis, D., Palli, D., Sieri, S., Vineis, P., Tumino, R., Panico, S., Ocké, M.C., Peeters, P.H.M., Skeie, G., Brustad, M., Lund, E., Sánchez-Cantalejo, E., Navarro, C., Amiano, P., Ardanaz, E., Ramón Quirós, J., Hallmans, G., Johansson, I., Lindkvist, B., Regnér, S., Khaw, K.T., Wareham, N., Key, T.J., Slimani, N., Norat, T., Vergnaud, A.C., Romaguera, D. & Gonzalez, C.A. (2012b) Variety in vegetable and fruit consumption and the risk of gastric and esophageal cancer in the European prospective investigation into cancer and nutrition. *International Journal of Cancer,* 131 (6), E963-E973.

Johnson, G.D. (2004) Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modelling. *International Journal of Health Geographics,* 3 29.

Jones, R., Latinovic, R., Charlton, J. & Gulliford, M.C. (2007) Alarm symptoms in early diagnosis of cancer in primary care: cohort study using General Practice Research Database. *BMJ: British Medical Journal (International Edition),* 334 (7602), 1040-1044.

Ka, H., Nm, S. & Ay, K. (2013) The role of geography in low mammography screening rates and late-stage breast cancer diagnosis in utah. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology,* 22 (3), 476-477.

Kabir M, Khoo D (2016) Impact of The 'Be Clear On Cancer' National Oesophago-Gastric Cancer Awareness Campaign on Endoscopy Services and Cancer Diagnosis Rates and Outcomes *Gut* **65:**A207-A208.

Kamangar, F., Dores, G.M. & Anderson, W.F. (2006a) Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *Journal of Clinical Oncology,* 24 (14), 2137-2150.

Kamangar, F., Qiao, Y.L., Schiller, J.T., Dawsey, S.M., Fears, T., Sun, X.D., Abnet, C.C., Zhao, P., Taylor, P.R. & Mark, S.D. (2006b) Human papillomavirus serology and the risk of esophageal and gastric cancers: results from a cohort in a high-risk region in China. *International Journal of Cancer.Journal International Du Cancer,* 119 (3), 579-584.

Kamangar, F., Chow, W., Abnet, C.C. & Dawsey, S.M. (2009) Environmental causes of esophageal cancer. *Gastroenterology Clinics of North America,* 38 (1), 27.

Kang, S.Y., McGree, J. & Mengersen, K. (2013) The impact of spatial scales and spatial smoothing on the outcome of bayesian spatial model. *PloS One,* 8 (10), e75957.

Kato, H., Yoshihara, M., Miyazacki, T., Nakajima, M., Fukai, Y., Tajima, K., Norihiro, T., Katsuhiko, M., Fukuda, T., Nakajima, T. & Kuwano, H. (2001) Expression of P53 protien related to smoking and alcoholic beverage drinking habits in patients with esophageal cancers. *Cancer Letters,* 167 (1), 65-72.

Kayani, B., Zacharakis, E., Ahmed, K. & Hanna, G.B. (2011) Lymph node metastases and prognosis in oesophageal carcinoma--a systematic review. *European Journal of Surgical Oncology : The Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology,* 37 (9), 747-753.

Khomichuk, A.L., Sharafetdinov, K., Vvoznyi, E.K., Shakhovskaia, A.K., Plotnikova, O.A. & Rusakova, D.S. (2011) Nutrition management of cancer patients after surgery on the esophagus and stomach: modern view on problem. *Voprosy Pitaniia,* 80 (5), 71-77.

Kitchin, R., Lauriault, T.P., Wilson, M.W (2017) *Understanding Spatial media* London; Sage.

Kmietowicz, Z. (2014) Many oesophageal and gastric cancers are detected too late. *British Medical Journal,* 349 (7896), g7340-2.

Kötz, B.S., Croft, S. & Ferry, D.R. (2006) Do delays between diagnosis and surgery in resectable oesophageal cancer affect survival? a study based on West Midlands cancer registration data. *British Journal of Cancer,* 95 (7), 835-840.

Kötz, B.S., Croft, S. & Ferry, D.R. (2006) Do delays between diagnosis and surgery in resectable oesophageal cancer affect survival? A study based on West Midlands Cancer registray Data. *British Journal of Cancer,* 95 835-890.

Krishnatreya, M., Saikia, A., Kataki, A., Sharma, J. & Baruah, M. (2014) Variations in the spatial distribution of gall bladder cancer: aA call for collaborative action. *Annals of Medical and Health Sciences Research,* 4 (9), 329.

Kubo, A., Cook, M.B., Shaheen, N.J., Vaughan, T.L., Whiteman, D.C., Murray, L. & Corley, D.A. (2013) Sex-specific associations between body mass index, waist circumference and the risk of Barrett's oesophagus: a pooled analysis from the international BEACON consortium. *Gut,* 62 (12), 1684-1691.

Kuwano, H., Kato, H., Miyazacki, T., Fukuchi, M., Masuda, N., Nakajima, M., Fukai, Y., Sohda, M., Kimura, H. & Faried, A. (2005) Genetic alterations in esophageal cancer. *Surgery Today,* 35 7-18.

Lagergren, J., Mattsson, F. & Nyren, O. (2013) Gastroesophageal Reflux Does Not Alter Effects of Body Mass Index on Risk of Esophageal Adenocarcinoma. *Clinical Gastroenterology and Hepatology : The Official Clinical Practice Journal of the American Gastroenterological Association,.*

Lambert, R., Saito, H., Lucas, E., Sankaranarayanan, R. (2012) Survival from digestive cancer in emerging countries in Asia and Africa. *European Journal of Gastroenterology and Hepatology.*24(6) 605-12

Lang, G.D. & Konda, V.J. (2013) Early diagnosis and management of esophageal and gastric cancer. *Minerva Gastroenterologica e Dietologica,* 59 (4), 357-376.

Lao-Sirieix, P. & Fitzgerald, R.C. (2012) Screening for oesophageal cancer. *Nature Reviews.Clinical Oncology,* 9 (5), 278-287.

Laszkiewicz, E., Dong, G., Harris, R. (2014) The effects of omitting spatial effects on social dependence on the modelling of household expenditure for fruits and vegetables. *Comparative economic research*17 (4) 155-72.

Lauby-Secretan B, Scoccianti C, Loomis D, et al. Body Fatness and Cancer--Viewpoint of the IARC Working Group (link is external). N Engl J Med. 2016 Aug 25;375(8):794-8.

Launay, L., Dejardin, O., Pornet, C., Morlais, F., Guittett, L., Launoy, G. & Bouvier, V. (2012) Influence of socioeconomic environment on survival in patients diagnosed with esophageal cancer : a population based study. *Diseases of the Esophagus,.*

Lawson, A. (2013) *Bayesian disease mapping* Boca Raton: Taylor & Francis.

Leeman, R.F., Beseler, C.L., Helms, C.M., Patock-Peckham, J.A., Wakeling, V.A. & Kahler, C.W. (2013) A Brief, Critical Review of Research on Impaired Control Over Alcohol Use and Suggestions for Future Studies. *Alcoholism, Clinical and Experimental Research*,.

Leeuwenburgh, I., Gerrits, M.M., Capello, A., van, d.B., van Dekken, H., Steyerberg, E.W., Siersema, P.D. & Kuipers, E.J. (2010) Expression of p53 as predictor for the development of esophageal cancer in achalasia patients. *Diseases of the Esophagus: Official Journal of the International Society for Diseases of the Esophagus/I.S.D.E*, 23 (6), 506-511.

LeSage, J., Pace, R.K., Campanella, R., Lam, N., Liu, X.(2011) Do What the neighbours do: Reopening businessses after Hurricane Katrina. *Significance* 8 (4) 160-3.

Levi, Z., Kark, J.D., Shamiss, A., Derazne, E., Tzur, D., Keinan-Boker, L., Liphshitz, I., Niv, Y., Furman, M. & Afek, A. (2013) Body mass index and socioeconomic status measured in adolescence, country of origin, and the incidence of gastroesophageal adenocarcinoma in a cohort of 1 million men. *Cancer*, 119 (23), 4086-4093.

Lewandowsky, S., Behrens, J.T., Pickle, L.W., Herrmann, D.J. & White, A.A. (1995) Perception of clusters in mortality maps: representing magnitude and statistical reliability. *Annals of geographic inform*. 107-132.

Lewis, R. (2017) NHS England, Cancer Research UK and Macmillan Cancer Support. Improving Diagnostic Pathways for Patients with Vague Symptoms available at . https://www.macmillan.org.uk/_images/ace-vague-symptoms-exec-summary_tcm9-312470.pdf [accessed 04 2018]

Li, J., Wang, K., Chen, X., Meng, H., Song, M., Wang, Y., Xu, X. & Bai, Y. (2012) Transcriptional activation of microRNA-34a by NF-kappa B in human esophageal cancer cells. *BMC Molecular Biology*, 13 4.

Li, X., Wang, L., Zhang, H., Jiang, S., Fang, Q., Chen, J. & Zhou, X. (2014) Spatial variations of pulmonary tuberculosis prevalence co-impacted by socio-economic and geographic factors in People's Republic of China, 2010. *BMC Public Health*, 14 (1), 257-257.

Lian, M., Struthers, J. & Schootman, M. (2012) Comparing GIS-Based Measures in Access to Mammography and Their Validity in Predicting Neighbourhood Risk of Late-Stage Breast Cancer. *PLoS ONE*, 7 (8)

Lian, M., Struthers, J. & Schootman, M. (2012) Comparing GIS-based measures in access to mammography and their validity in predicting neighborhood risk of late-stage breast cancer. *PLoS ONE [Electronic Resource]*, 7 (8), e43000.

Liao, Y., Wang, J., Wu, J., Driskell, L., Wang, W., Zhang, T., Xue, G. & Zheng, X. (2010) Spatial analysis of neural tube defects in a rural coal mining area. *International Journal of Environmental Health Research*, 20 (6), 439-450.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J. & Moher, D. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology,* 62 (10), e1-34.

Ljung, R., Drefahl, S., Andersson, G. & Lagergren, J. (2013a) Socio-Demographic and Geographical Factors in Esophageal and Gastric Cancer Mortality in Sweden. *PLoS ONE,* 8 (4), 1-6.

Ljung, R., Drefahl, S., Andersson, G. & Lagergren, J. (2013b) Socio-Demographic and Geographical Factors in Esophageal and Gastric Cancer Mortality in Sweden. *PLoS ONE,* 8 (4), 1-6.

Longley, P. (2010) *Geographic Information Systems and Science (3rd ed*) London: Wiley

Lordick F, Holscher AH, Haustermans K, Wittekind C (2012) Multimodal treatment of esophageal cancer. Langenbecks Arch Surg 10, 22-32

Lovett, D.A., Poots, A.J., Clements, J.T.C., Green, S.A., Samarasundera, E. & Bell, D. (2014) Using geographical information systems and cartograms as a health service quality improvement tool. *Spatial and Spatio-Temporal Epidemiology*, 10 67-74.

Lubin, J.H., Cook, M.B., Pandeya, N., Vaughan, T.L., Abnet, C.C., Giffen, C., Webb, P.M., Murray, L.J., Casson, A.G., Risch, H.A., Ye, W., Kamangar, F., Bernstein, L., Sharp, L., Nyren, O., Gammon, M.D., Corley, D.A., Wu, A.H., Brown, L.M., Chow, W.H., Ward, M.H., Freedman, N.D. & Whiteman, D.C. (2012) The importance of exposure rate on odds ratios by cigarette smoking and alcohol consumption for esophageal adenocarcinoma and squamous cell carcinoma in the Barrett's Esophagus and Esophageal Adenocarcinoma Consortium. *Cancer Epidemiology,*.

Luo, L. (2012) *Socio-spatial inequalities in late-stage cancer diagnosis in Illinois: Spatiotemporal trends and methodological challenges.* PHD Thesis available at http://hdl.handle.net/2142/26354.

Lurie, G., Wilkens, L.R., Thompson, P.j., Matsuno, R., Carney, M.E., Goodman, M. (2010) Symptom presentation in invasive ovarian carcinoma by tumour histological type and grade in a multi-ethnic population a case analysis: *Gynaecological Oncology* 119 278-284.

Lycett, M. & Marshan, A. (2016). Capturing Sensemaking Pattern during Data Analysis: A Conceptual Framework. In J. Gołuchowski, M. Pańkowska, C. Barry, M. Lang, H. Linger, & C. Schneider (Eds.), *Information Systems Development: Complexity in Information Systems Development (ISD2016 Proceedings).* Katowice, Poland: University of Economics in Katowice.

Lydiatt, W., M., Patel, S. G., O'Sullivan, B., Brandwein, M.S., Ridge, J.A., Migliacci, J.C., Loomis, A.M., Shah, J. P (2017) Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual *CA: A Cancer Journal for Clinicians.* 67 (2): 122–137.

Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA. (2012) Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncology* 13. 353-65

Lyratzopoulos G, Abel GA, McPhail S, Neal RD, Rubin GP. (2013) Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. *British Journal of Cancer.* 108:686-90

Lyseen, A.K., Nøhr, C., Sørensen, ,E.M., Gudes, O., Geraghty, E.M., Shaw, N.T. & Bivona-Tellez, C. (2014) A Review and Framework for Categorizing Current Research and Development in Health Related Geographical Information Systems (GIS) Studies. *Yearbook of Medical Informatics,* 9 (1), 110-124.

Macdonald, S., Macleod, U., Campbell, N.C., Weller, D. & Mitchell, E. (2006) Systematic review of factors influencing patient and practitioner delay in diagnosis of upper gastrointestinal cancer. *British Journal of Cancer,* 94 (9), 1272-1280.

Macleod, U., Mitchell, E.D., Burgess, C., Macdonald, S. & Ramirez, A.J. (2009) Risk factors for delayed presentation and referral of symptomatic cancer: evidence for common cancers. *British Journal of Cancer,* 101 Suppl 2 S92-S101.

Mahon, S.M. (2009) Prevention and screening of gastrointestinal cancers. *Seminars in Oncology Nursing,* 25 (1), 15-31.

Mao, W.M., Zheng, W.H. & Ling, Z.Q. (2011) Epidemiologic risk factors for esophageal cancer development. *Asian Pacific Journal of Cancer Prevention : APJCP,* 12 (10), 2461-2466.

Marchet, A., Mocellin, S., Ambrosi, A., Morgagni, P., Vittimberga, G., Roviello, F., Marrelli, D., de Manzoni, G., Minicozzi, A., Coniglio, A., Tiberio, G., Pacelli, F., Rosa, F. & Nitti, D. (2011) Validation of the new AJCC TNM staging system for gastric cancer in a large cohort of patients (N = 2,155): focus on the T category. *European Journal of Surgical Oncology: The Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology,* 37 (9), 779-785.

Mariette. C., Piessen, G., Triboulet, J.P. (2007) Therapeutic strategies in oesophageal carcinoma: role of surgery and other modalities. *Lancet Oncol* 8:545–553

Masaru Morita, M., Otsu, H., Kawano, H., Kasagi,Y., Kimura, Y., Saeki, H., Ando, K., Ida, S., Oki, S., Tokunaga, E., Ikeda, T., Kusumoto, T., Maehara, Y. (2013) Gender differences in prognosis after esophagectomy for esophageal cancer *Surgery today* 44 (3) 505-521

Marmo, R., Del Piano, M., Rotondano, G., et al. (2012) Mortality from nonulcer bleeding is similar to that of ulcer bleeding in high-risk patients with nonvariceal hemorrhage: a prospective database study in Italy. Gastrointest Endosc.;75(2):263–72, 272.e1

Mayor S (2014) People with oesophageal and oropharyngeal cancers delay seeing their GP, audit shows. *British Medical Journal* BMJ 348:g1324

McCandless, D (2012) *Information is beautiful* London: Harper Collins.

McCandless, D. (2014) *Knowledge is beautiful* London: Harper Collins.

McPhail S., Elliss-Brookes, L., Shelton, J. (2013). Emergency presentation of cancer and short-term mortality. B*ritish Journal of Cancer.* 109: 2027–34

McGhan, L.J., Pockaj, B.A., Gray, R.J., Bagaria, S.P. & Wasif, N. (2012) Validation of the updated 7th edition AJCC TNM staging criteria for gastric adenocarcinoma. *Journal of Gastrointestinal Surgery: Official Journal of the Society for Surgery of the Alimentary Tract,* 16 (1), 53-61.

Medical research Council (2002) Surgery with or without preoperative chemotherapy in oesophageal cancer; a randomised controlled trial. *Lancet,* 359 1727-1733.

Meecham. D., Gildea, C., Hollingworth, L., Richards, M.A., Riley, D., Rubin, G. (2012) Variation in use of the 2-week referral pathway for suspected cancer: a cross sectional analysis. *British Journal of General Practice* 62 : e590-7.

Merriman, S. (2009) *Qualitative research A guide to design and implementation.* California : Wiley & Sons.

Meyer. C., (2001) A Case in Case study methodology Field methods 13 (4) available at. http://journals.sagepub.com/doi/pdf/10.1177/1525822X0101300402 [accessed Sept 2015].

Mobley, L., R., Kuo, T., May, Watson, L. Brown,, G. (2012) Geographic disparities in late-stage cancer diagnosis: Multilevel factors and spatial interactions. *Health & Place,* 18 (5), 978-990.

Mohebbi, M., Mahmoodi, M., Wolfe, R., Nourijelyani, K., Mohammad, K., Zeraati, H. & Fotouhi, A. (2008) Geographical spread of gastrointestinal tract cancer incidence in the Caspian Sea region of Iran: Spatial analysis of cancer registry data. *BMC Cancer,* 8 1-12.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. (2009) Preferred reporting items for systematic reviews and meta analyses : the PRISMA statement. *PLoS Med,* 6 (6) 262-278.

Moher, D., Altman, D.G., Liberati, A. & Tetzlaff, J. (2011) PRISMA statement author reply 128. *Epidemiology (Cambridge, Mass.),* 22 (1), 128

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. & PRISMA Group (2010) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International Journal of Surgery (London, England),* 8 (5), 336-341.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A. & PRISMA-P Group (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews,* 4 (1), 1-4053-4-1.

Monnet, E., Ramée, C., Minello, A., Jooste, V., Carel, D. & Di Martino, V. (2008) Socioeconomic context, distance to primary care and detection of hepatitis C: A French population-based study. *Social Science & Medicine,* 66 (5), 1046-1056.

Moscow, J.A. & Cowan, K.H. (2011) Biology of Cancer. In L. Goldman and A. Schafer (eds) *Cecil medicine.* Philadelphia: Saunders Elsevier,.

Mulholland, H.G., Murray, L.J., Anderson, L.A., Cantwell, M.M. & FINBAR study group (2011) Vitamin D, calcium and dairy intake, and risk of oesophageal adenocarcinoma and its precursor conditions. *The British Journal of Nutrition,* 106 (5), 732-741.

Muthusamy, V.R. & Sharma, P. (2011) Diagnosis and management of Barrett's esophagus: What's next? *Gastrointestinal Endoscopy Clinics of North America*, 21 (1), 171-181.

Nagtegaal, I.D., Tot, T., Jayne, D.G., McShane, P., Nihlberg, A., Marshall, H.C., Påhlman, L., Brown, J.M., Guillou, P.J. & Quirke, P. (2011) Lymph nodes, tumor deposits, and TNM: are we getting better? *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 29 (18), 2487-2492.

Naimi, T.S., Xuan, Z., Brown, D.W. & Saitz, R. (2013) Confounding and studies of moderate alcohol consumption: the case of drinking frequency and implications for low-risk drinking guidelines. *Addiction,* 108 (9), 1534-1543.

Najafabadi, A.T. & Pourhassan, M. (2011) Integrating the geographic information system into cancer research. *Indian Journal of Cancer,* 48 (1), 105-109.

National Audit Office, (2010), *Report by the Comptroller and Auditor General DoH Delivering the cancer reform strategy.* London, DoH.

National Awareness and Early Diagnosis Initiative (NAEDI) (2016) *briefing paper* available at; http://www.cancerresearchuk.org/sites/default/files/health_professional_naedi_briefing_sheet.pdf [accessed Oct 2017]

National Cancer Intelligence Network (NCIN) (2010) Geographic variation in cancer of the lower oesophagus. NCIN Data Briefing available at; http://www.ncin.org.uk [accessed 062016].

National Cancer Intelligence Network (2012) Incidence of Oesophageal cancer in England 1998-2007 NCIN Data Briefing. available at http://www.ncin.org.uk/search/oesophageal+cancer [accessed Nov 2017]

National Cancer Intelligence Network (2014) One-year relative survival rates for patients diagnosed with cancer of the oesophagus, stomach, primary liver, gallbladder, biliary tract and pancreas in England, 1985-2009 NCIN Data Briefing. available at http://www.ncin.org.uk/search/oesophageal+cancer [accessed Jan 2015]

National Cancer Intelligence Network (2015) NCIN Data Briefing One Year relative survival rates for oesophageal cancer in Great Britain. available at http://www.ncin.org.uk/search/oesophageal+cancer [accessed Jan 2017]

National Cancer Intelligence Network, (2008), *Cancer Incidence by Deprivation, England 1995-2004.* London, NCIN.

National Cancer Institute (2017) *Cancer staging* available at https://www.cancer.gov/about-cancer/diagnosis-staging/staging. [accessed 12 Sept 2017]

National Institute for Health and Clinical Excellence, (2005) *Referral Guidelines for suspected cancer.* Clinical Guideline 27, UK, NIHCE.

National Institute for Health and Clinical Excellence (NICE). (2015) *Suspected cancer: recognition and referral*. available at www.nice.org.uk/guidance/ng12. [accessed sept 2017].

Naureckas, E.T. & Thomas, S. (2007) Are we closing the disparities gap? Small-area analysis of asthma in Chicago. *Chest*, 132 (5), 858S-865.

Neal, R. (2009) Do diagnostic delays in cancer matter? *British Journal of Cancer* 101 S9-12

Neal, R., Tharmanthan, P., France, B., Din, N., Cotton, N., Fergussen, J., Hamilton, W., Hendry, A., Hendry, M., Lewis, R., Macleod, U., Mitchell, M., Pickett, M., Rai, M., Shaw, M., Stuart, N., Tørring, M.L., Wilkinson, C., Williams, B., Emery, J. (2015) Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review *British Journal of Cancer* 112 5s2-s107

NHS The Information Centre, (2008), *National Oesophagogastric Cancer Audit.* First Annual report, The NHS Information centre.

NHS Digital Information Centre (2016) National Oesophagogastric cancer audit Available at **https://www.digital.nhs.uk/og** [accessed April 2017].

Noble, M., Wright, G., Smith, G., & Dibben, C. (2006). Measuring multiple deprivation at the small-area level. *Environment & Planning A, 38*, 168–185.

Noble, D., Smith, D., Mathur, R., Robson, J. & Greenhalgh, T. (2012) Feasibility study of geospatial mapping of chronic disease risk to inform public health commissioning. *BMJ Open,* 2 (1), e000711.

Norman, P. (2006). Sociodemographic spatial change in the UK: data and computational issues and solutions. *GIS Development special issue Maps and Census, 10*(12), 30–34.

Norman, P.(2010) Identifying change over time in Small Area Socioeconomic Deprivation *Applied spatial analysis and policy* 3 (2-3) 107-138.

Northern Ireland Cancer Registry (2010) *Cancer Incidence and Mortality* Available online: http://www.qub.ac.uk/research-centres/nicr/Publications/ [Accessed May 2012.

Nuting, C.M., Robinson, M. & Birchall, M. (2008) Survival from Laryngeal cancer in England and Wales up to 2001. *British Journal of Cancer*, 99 (1), s38-s39.

Nykiforuk, C., Flaman, L. (2011) Geographic information systems for health promotion and public health: a review *Health Promotion practice* 12 (1) 63-73.

O'Connor, A. & O'Moráin, C. (2013) Helicobacter pylori infection in Europe: current perspectives. *Expert Review of Gastroenterology & Hepatology*, 7 (6), 541-548.

O'Doherty, M.G., Freedman, N.D., Hollenbeck, A.R., Schatzkin, A., Murray, L.J., Cantwell, M.M. & Abnet, C.C. (2011) Association of dietary fat intakes with risk of esophageal and gastric cancer in the NIH-AARP diet and health study. *International Journal of Cancer* 131 (6) Available at; https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.27366 [accessed April 2018]

Office for National Statistics, (2005), *Cancer Statistics Registrations: Registrations of cancer diagnosed in 2005. Statistical Bulletin*, London, Office for National Statistics.

Office of National Statistics (2009) Understanding Patterns of deprivation *regional trends* (2009) 41(1) 93-114

Office for National Statistics (2011): 2001 Census aggregate data (Edition: May 2011). UK Data Service. DOI: http://dx.doi.org/10.5257/census/aggregate-2001-2 [accessed May 2016]

Office for National Statistics, (2012), *Cancer survival by cancer network, England: Patients diagnosed 1997-2010 and followed up to 2011.* statistical Bulletin, London, Office for National Statistics.

Office for National Statistics (2014) Cancer survival by NHS England Area Team: Adults diagnosed 1997-2012, followed up to 2013 | 16 December 2014 available at **https://www.ons.gov.uk** [accessed 0616].

Office for National Statistics (2014) Cancer survival in England ; Patients diagnosed between 2010 and 2014 and followed up to 2015 Statistical bulletin available at ; https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalinenglandadultsdiagnosed/2010and2014andfollowedupto2015 [accessed Nov 2017]

Office for National Statistics (2016); National Records of UK Statistics and Research Agency: 2011 Census aggregate data. UK Data Service (Edition: June 2016). DOI: http://dx.doi.org/10.5257/census/aggregate-2011-1 [accessed May 2016]

Office for National Statistics, Public Health England (2016) Cancer survival by stage at diagnosis for England (experimental statistics): 2012 to 2014. 10 Jun 2016. available at https://www.ons.gov.uk/releases/cancersurvivalbystageatdiagnosisexperimentalstatistics2012to2014. [accessed 022018]

Ojala, K., Sorri, M., Jokinen, K. & Kairaluoma, M. (1982) Symptoms of carcinoma of the oesophagus. *The Medical Journal of Australia*, 1 (9), 384-385.

Olsson, J.K., Schultz, E.M., Gould, M.K. (2009) Timeliness of care in patients with lung cancer: a systematic review. *Thorax*.64:749–756

Olessen, F., Hansen, R. & Vested, P. (2009) Delay in diagnosis; the experience in Denmark. *British Journal of Cancer*, 101 (suppl2), s1-4.

Openshaw, S., Charlton, M., Wymer, C., and Craft, A., (1987) A mark I geographical analysis machine for the automated analysis of point data sets, *International Journal of Geographical Information Systems*, 1, 335-358.

Openshaw, S., Charlton, M., Craft, A. and Birch, J. (1988) 'An investigation of leukaemia clusters by the use of a geographical analysis machine', *The Lancet*, Feb 6th, 272-273.

Openshaw, S., and Craft, A., (1991) 'Using geographical analysis machines to search for evidence of cluster and clustering in childhood leukaemia and non-Hodgkin Lymphomas in Britain. In G. Draper (ed) '*The Geographical Epidemiology of Childhood Leukaemia and non-Hodgkin Lymphomas in Great Britain 1966-83' Studies in Medical and Population Subjects No 53*, OPCS, London, HMSO

Orengo, M.A., Casella, C., Fontana, V., Filiberti, R., Conio, M., Rosso, S., Tumino, R., Crosignani, P., De Lisi, V., Falcini, F. & Vercelli, M. (2006) Trends in incidence rates of oesophagus and gastric cancer in Italy by subsite and histology, 1986-1997. *European Journal of Gastroenterology & Hepatology*, 18 (7), 739-746.

Otterstatter, M.C., Brierley, J.D., De, P., Ellison, L.F., Macintyre, M., Marrett, L.D., Semenciw, R. & Weir, H.K. (2012) Esophageal cancer in Canada: trends according to morphology and anatomical location. *Canadian Journal of Gastroenterology = Journal Canadien De Gastroenterologie*, 26 (10), 723-727.

Parkin, D.M., Boyd, L., Walker, L.C. (2010) The fraction of cancer attributes to lifestyle and environmental factors in the UK in 2010. *British Journal of Cancer* 105(S2) s77-81.

Parrott, R., Hopfer, S., Ghetian, C. & Lengerich, E. (2007) Mapping as a visual health communication tool: promises and dilemmas. *Health Communication,* 22 (1), 13-24.

Parrott, R., Volkman, J., Lengeritch, E., Ghetian, C., Chadwick, A., Hopfer, S. (2010) Using geographic information systems to promote community involvement in comprehensive cancer control *Health communication* 25. 276-285.

Parsons, S. (2010) Are we missing Gastro-Oesophageal Cancer at Endoscopy? *Annals of the Royal College of Surgeons of England,*.

Peng, W., Chen, Y., Jiang, Q., Zheng, Y. (2010) Spatial analysis of hepatocellular carcinoma and socioeconomic status in China from a population-based cancer registry. C*ancer Epidemiology* 34:29–33

Peter, A. (2013) Connection between cancer- and alcohol-related mortality in a rural practice of a South-Hungarian village. *Orvosi Hetilap,* 154 (18), 700-706.

Peterson, W., Barnett, C., Smith, J., Allen, M.H., Corbett, D. (1981) Routine early endoscopy in upper gastrointestinal tract bleeding an RCT *New England Journal of Medicine.* 304(16), 925-929.

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

Public Health England (2016) Public Health Profiles (smoking) available at http://fingertips.phe.org.uk/search/smoking#page/0/gid/1/pat/152/par/E38000085/ati/7/are/B81085 [accessed Feb 2016]

Public Health England (2016) Public Health Profiles (alcohol) available at http://fingertips.phe.org.uk/search/alcohol#pat/152/ati/7/par/E38000085 [accessed Feb 2016]

Qiao, Y.L., Dawsey, S.M., Kamangar, F., Fan, J.H., Abnet, C.C., Sun, X.D., Johnson, L.L., Gail, M.H., Dong, Z.W., Yu, B., Mark, S.D. & Taylor, P.R. (2009) Total and cancer mortality after supplementation with vitamins and minerals: follow-up of the Linxian General Population Nutrition Intervention Trial. *Journal of the National Cancer Institute,* 101 (7), 507-518.

Qiu, D. & Kaneko, S. (2005) Comparison of esophageal cancer mortality in five countries: France, Italy, Japan, UK and USA from the WHO mortality database (1960-2000). *Japanese Journal of Clinical Oncology,* 35 (9), 564-567.

Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ. (1999) Influence of delay on survival in patients with breast cancer: a systematic review. *Lance*t;353:1119-26.

Richards, M.A., Grob, JJ., Avril, M.F., Delaunay, M., Thirion, M., Wolkenstein, P., Southeryand, P., Dreno, P., Aubyn, F., Guilliot, B., Lok, C., Chemaly, P (1999)

Melanoma and tumour thickness challenges of early diagnosis *Archives dermatology* 135 269-74.

Richards, M.A. (2009a) The National Awareness and Early Diagnosis Initiative in England: assembling the evidence. *British Journal of Cancer,* 101 Suppl 2 S1-4.

Richards, M.A. (2009b) The size of the prize for earlier diagnosis of cancer in England. *British Journal of Cancer,* 101 Suppl 2 S125-9.

Richards, T.B., Croner, C.M., Rushton, G., Brown, C.K. & Fowler, L. (1999) Geographic information systems and public health: mapping the future. *Public Health Reports,* 114 359-360.

Risser, D.R. & Miller, E.A. (2012) Cancer in relation to socioeconomic status: stage at diagnosis in Texas, 2004-2008. *Southern Medical Journal,* 105 (10), 508-512.

Robertson, E.V., Derakhshan, M.H., Wirz, A.A., Lee, Y.Y., Seenan, J.P., Ballantyne, S.A., Hanvey, S.L., Kelman, A.W., Going, J.J. & McColl, K.E. (2013) Central obesity in asymptomatic volunteers is associated with increased intrasphincteric acid reflux and lengthening of the cardiac mucosa. *Gastroenterology,* 145 (4), 730-739.

Roche, L.M., Skinner, R. & Weinstein, R.B. (2002) Use of a geographic information system to identify and characterize areas with high proportions of distant stage breast cancer. *Journal of Public Health Management & Practice,* 8 (2), 26-32.

Roshandel, G., Norouzi, A., Pourshams, A., Amiriani, T., Semnani, S., Merat, S. & Khoshnia, M. (2013) Endoscopic Screening for Esophageal Squamous Cell Carcinoma. *Archives of Iranian Medicine (AIM),* 16 (6), 351-357.

Rothwell, J.F., Feehan, E., Reid, I., Walsh, T.N. & Hennessy, T.P. (1997) Delay in treatment for oesophageal cancer. *The British Journal of Surgery,* 84 (5), 690-693.

Royal College of General Practitioners. National audit of cancer diagnosis in primary care. (2011). available at http//www.rcgp.org.uk/news/2011/november/~/media/Files/News/National_Audit_of_Cancer_Diagnosis_in_Primary-Care.ashx. [accessed Feb 2018].

Royal College of Surgeons, (2014), *National Oesophagogastric Cancer Audit progress report 2014.* UK, Royal College of Surgeons of England.

Rushton, G., Armstrong, M.P., Gittler, J., Greene, B.R., Pavlik, C.E., West, M.M. & Zimmerman, D.L. (2006) Geocoding in cancer research: a review. *American Journal of Preventative Medicine,* 30 (2), s16-s24.

Sakai, H., Muramatsu, K. & Matsuda, S. (2013) DPC data based situation analysis of Regional Core Hospital: an application of the GIS Methodology for Regional Health Care Plans]. *Journal of UOEH,* 35 (1), 39-49.

Salaspuro, M. (2011) Acetaldehyde and gastric cancer. *Journal of Digestive Diseases,* 12 (2), 51-59.

Schlansky, B., Dimarino, A.J., J., Loren, D., Infantolino, A., Kowalski, T. & Cohen, S. (2006a) A survey of oesophageal cancer: pathology, stage and clinical presentation. *Alimentary Pharmacology & Therapeutics*, 23 (5), 587-593.

Scholz, R.W., Tietje, O. (2002) *Embedded case study methods: integrating quantitative and qualitative knowledge*. 1st ed. Thousand Oaks: SAGE Publications.

Schwartz, G.G. & Hanchette, C.L. (2006) UV, latitude, and spatial trends in prostate cancer mortality: all sunlight is not the same (United States). *Cancer Causes & Control:* 17 (8), 1091-1101.

Shammas, M.A. (2011) Repetitive sequences, genomic instability and Barrett's esophageal adenocarcinoma. *Mobile Genetic Elements*, 1 (3), 208-212.

Shah, M.B., Schnoll-Sussman, F. (2010) Novel use of cryotherapy to control bleeding in advanced esophageal cancer. *Endoscopy*.42 Suppl 2:9.

Stake, R. E. (1995). *The art of case study research*. London: Sage.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A. & PRISMA-P Group (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ (Clinical Research Ed.)*, 349 g7647.

Sharp, L., Donnelly, D., Hegarty, A., Carsin, A., Deady, S., McCluskey, N., Gavin, A. & Comber, H. (2014a) Risk of several cancers is higher in urban areas after adjusting for socioeconomic status. Results from a two-country population-based study of 18 common cancers. *Journal of Urban Health : Bulletin of the New York Academy of Medicine*, 91 (3), 510-525.

Sharp, L., Donnelly, D., Hegarty, A., Carsin, A., Deady, S., McCluskey, N., Gavin, A. & Comber, H. (2014b) Risk of Several Cancers is Higher in Urban Areas after Adjusting for Socioeconomic Status. Results from a Two-Country Population-Based Study of 18 Common Cancers. *Journal of Urban Health*, 91 (3), 510-525.

Sharpe, K.H., McMahon, A.D., McClements, P., Watling, C., Brewster, D.H. & Conway, D.I. (2012) Socioeconomic inequalities in incidence of lung and upper aero dogestive tract cancer by age, tumour subtype and sex: A population based study in Scotland (2000 - 2007). *Cancer Epidemiology*, in press.

Sharpe, D., Williams, R.N., Ubhi, S.S., Sutton, C.D., Bowrey, D.J. (2010) The "two-week wait" referral pathway allows prompt treatment but does not improve outcome for patients with oesophago-gastric cancer. *European Journal of Surgical Oncology.* 36 (10), 977- 81.

Sheppard, S.C., Forsyth, J.P., Earleywine, M., Hickling, E.J. & Lehrbach, M.P. (2013) Improving base rate estimation of alcohol misuse in the military: a preliminary report. *Journal of Studies on Alcohol and Drugs*, 74 (6), 917-922.

Shinagare, A.B., Zukotynski, K.A., Krajewski, K.M., Jagannathan, J.P., Butrynski, J., Hornick, J.L. & Ramaiya, N.H. (2012) Esophageal gastrointestinal stromal tumor:

report of 7 patients. *Cancer Imaging : The Official Publication of the International Cancer Imaging Society,* 12 100-108.

Siau K, Yew AC, Hingley S, Rees, J., Trudgill, N., Veitch, A., Fisher, N. (2017) The 2015 upper gastrointestinal "Be Clear on Cancer" campaign: its impact on gastroenterology services and malignant and premalignant diagnoses. *Frontline Gastroenterology* **8:**284-289.

Singhal, S., Quiñonez, C.R. & Jha, P. (2013) An observational study to assess changes in social inequality in smoking-attributable upper aero digestive tract cancer mortality among Canadian males between 1986 and 2001. *BMC Public Health,* 13 (1), 1-6.

Sitas, F., Egger, S., Urban, M.I., Taylor, P.R., Abnet, C.C., Boffetta, P., O'Connell, D.L., Whiteman, D.C., Brennan, P., Malekzadeh, R., Pawlita, M., Dawsey, S.M., Waterboer, T. & InterSCOPE Collaboration (2012) InterSCOPE study: Associations between esophageal squamous cell carcinoma and human papillomavirus serological markers. *Journal of the National Cancer Institute,* 104 (2), 147-158.

Sloggett, A., Young, H. & Grundy, E. (2007) The association of cancer survival with four socioeconomic indicators : a longitudinal study of the older population of England and Wales 1981 - 2000. *BMC Cancer,* 7. 20.

Smith, A.P. & Gazin-Schwartz, A. (2008) *Landscapes of clearance: Archaeological and anthropological perspectives* Walnut Creek, Calif.: Left Coast Press.

Smith, M., Zhou, M., Whitlock, G., Yang, G., Offer, A., Hui, G., Peto, R., Huang, Z. & Chen, Z. (2008) Esophageal cancer and body mass index: results from a prospective study of 220,000 men in China and a meta-analysis of published studies. *International Journal of Cancer.Journal International Du Cancer,* 122 (7), 1604-1610.

Smithers, B.M., Fahey, P.P., Corish, T., Gotley, D.C., Falk, G.L., Smith, G.S., Kiroff, G.K., Clouston, A.D., Watson, D.I. & Whiteman, D.C. (2010) Symptoms, investigations and management of patients with cancer of the oesophagus and gastro-oesophageal junction in Australia. *The Medical Journal of Australia,* 193 (10), 572-577.

Sobin, L.H.; Gospodarowicz, M.K.; Wittekind, Ch., eds. *(2010). TNM classification of malignant tumours* (7th ed.). Chichester, West Sussex, UK: Wiley-Blackwell.

Souhami, R. (2010) Are UK cancer cure rates worse than in most other European countries? *British Journal of General Practice* 60:81-2.

Stake, R.E. (2000) Case studies. In: Denzin NK, Lincoln YS, (eds). *The handbook of qualitative research.* 2nd ed. Thousand Oaks: SAGE Publications; 2000. p. 435–54.

Steevens, J., Schouten, L., Goldbohm, R.A., Van Der Brant, P.A. (2009) Alcohol consumption, cigarette smoking and risk of subtypes of oesophageal and gastric cancer; a prospective cohort study. *GUT,* 59 39-48

Stordeur, S., Vlayen, J., Vrijens, F., Camberlin, C., Gendt, G., Van Eycken, E., Lerut, T. (2015) Quality indicators for oesophageal and gastric cancer : A population based study in Belgium 2004-08 *European Journal of Cancer care.* 24 (3) 376-386

Subasinghe, D. & Samarasekera, D.N. (2010) Delay in the diagnosis of esophageal carcinoma: experience of a single unit from a developing country. *Indian Journal of Cancer,* 47 (2), 151-155.

Suh, N. & Pezzuto, J.M. (2012) Strawberry fields forever? *Cancer Prevention Research (Philadelphia, Pa.),* 5 (1), 30-33.

Sui, D (2017) Geospatial big data. Cited in Kitchin, R., Lauriault, T.P., Wilson, M.W. (eds) U*nderstanding Spatial media* London; Sage.

Sun, L., Zhang, H., Wu, k. (2014) Esophageal cancer ; Current options for therapeutic management *Gastrointestinal tumours* 1 105-113

Tachibana, M., Kinugasa, S., Shibakita, M., Tonomoto, Y., Hattori, S. & Hyakudomi, R. (2006) Surgical treatment of superficial esophageal cancer. *Langenbecks Archives of Surgery,* 391 (4), 304-321.

Tanaka, H. (2014) Advances in cancer epidemiology in Japan. *International Journal of Cancer.* 134 (4), 747-754.

Tannenbaum, S.L., Hernandez, M., Zheng, D.D., Sussman, D.A. & Lee, D.J. (2014) Individual- and Neighborhood-Level Predictors of Mortality in Florida Colorectal Cancer Patients. *PLoS ONE,* 9 (8), 1-10.

Tapp, H., Smith, H.A., Dixon, J.T., Ludden, T. & Dulin, M. (2013) Evaluating Primary Care Delivery Systems for an Uninsured Hispanic Immigrant Population. *Family & Community Health,* 36 (1), 19-33.

Tarleton, H.P., Chang, S., Park, S.L., Cai, L., Ding, B., He, N., Hussain, S.K., Jiang, Q., Mu, L., Rao, J., Wang, H., You, N.Y., Yu, S., Zhao, J. & Zhang, Z. (2014) Genetic variation at 8q24, family history of cancer, and upper gastrointestinal cancers in a Chinese population. *Familial Cancer,* 13 (1), 45-56.

Tentzeris V, Lake B, Cherian T, (2011) Poor awareness of symptoms of oesophageal Cancer. Interact Cardiovascular and Thoracic Surgery 2011;12:32

Thomson, C.A., LeWinn, K., Newton, T.R., Alberts, D.S. & Martinez, M.E. (2003) Nutrition and diet in the development of gastrointestinal cancer. *Current Oncology Reports,* 5 (3), 192-202.

Thornton, L., Reader, H., Stojkovic, S., Allgar, V., Woodcock, N. (2016) Has the 'Fast-Track' referral system affected the route of presentation and/or clinical outcomes in patients with colorectal cancer? *World Journal of Surgical Oncology.* 8, 14 (1).

Thrift, A.P., Nagle, C.M., Fahey, P.P., Smithers, B.M., Watson, D.I. & Whiteman, D.C. (2012) Predictors of survival among patients diagnosed with adenocarcinoma

of the esophagus and gastroesophageal junction. *Cancer Causes & Control,* 23 (4), 555-564.

Thrift, A.P., Kendall, B.J., Pandeya, N. & Whiteman, D.C. (2013) A Model to Determine Absolute Risk for Esophageal Adenocarcinoma. *Clinical Gastroenterology and Hepatology,* 11 (2), 138-144.e2.

Tomizawa, Y. & Wang, K.K. (2009a) Changes in screening, prognosis and therapy for esophageal adenocarcinoma in Barrett's esophagus. *Current Opinion in Gastroenterology,* 25 (4), 358-365.

Tomizawa, Y. & Wang, K.K. (2009b) Screening, surveillance, and prevention for esophageal cancer. *Gastroenterology Clinics of North America,* 38 (1), 59-73.

Tørring ML, Frydenberg M, Hamilton W, Hansen RP, Lautrup MD, Vedsted P. (2012) Diagnostic interval and mortality in colorectal cancer: U-shaped association demonstrated for three different datasets. J *Clinical Epidemiology* 65:669-78.

Tørring, m.l., Frydenberg, M., Hansen, P., Olessen, F., Hamilton, W., Vedsted, P., Time to diagnosis and mortality in colorectal cancer: A cohort study. *British Journal of Cancer* 104 934-40.

Tørring, M.l., Frydenberg, M., Hansen, R., Olessen, F., Vedsted, P. (2013) Evidence of Increasing mortality with longer diagnostic intervals for five commmon cancers: a Cohort study in promary care *European Journal of Cancer* 49, (9) 2187-2198.

Townsend, P., Phillimore, P. and Beattie, A. (1988) *Health and Deprivation: Inequality and the North.* Routledge, London.

Tran, G.D., Sun, X.D., Abnet, C.C., Fan, J.H., Dawsey, S.M., Dong, Z.W., Mark, S.D., Qiao, Y.L. & Taylor, P.R. (2005) Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *International Journal of Cancer.* 113 (3), 456-463.

Trellis, W. (1997) Introduction to Case Study *the qualitative report* 3(2)

Tsinguahua, A. (2015) Esophageal cancer project - Team Tsinguaha International Genetically Engineered Machine Competition, USA available at http://2015.igem.org/Team:Tsinghua-A [accessed 062016].

UK data Service Census support, CASWEB available at http://casweb.mimas.ac.uk/

[accessed 04/2017]

UK data Census support, InFuse. Available at http://infuse.ukdataservice.ac.uk/

[accessed 04/2017]

Uphoff, E.P., Pickett, K.E., Cabieses, B., Small, N. & Wright, J. (2013) A systematic review of the relationships between social capital and socioeconomic inequalities in health: a contribution to understanding the psychosocial pathway of health inequalities. *International Journal for Equity in Health,* 12 (1), 54.

Valkanova, N., Jorda, S. VanDe Moere, A., (2005) Public visualisation displays of citizen data: design, impact and implications *International journal of human computer studies* 81 P4-16.

Van Rensburg, S.J. (1981) Epidemiologic and dietary evidence for a specific nutritional predisposition to esophageal cancer. *J Natl Cancer Inst,* 67 (2), 243-51.

Varghese, S., Lao-Sirieix, P. & Fitzgerald, R.C. (2012) Identification and clinical implementation of biomarkers for Barrett's esophagus. *Gastroenterology,* 142 (3), 435-441.e2.

Vedsted, P., Olessen, F. (2011) Are the serious problems in cancer survival partly rooted in gatekeeper principles? An ecologic study *British Journal of General Practice* 61:e508-12.

Vieira, V., Webster, T., Weinberg, J. & Aschengrau, A. (2009) Spatial analysis of bladder, kidney, and pancreatic cancer on upper Cape Cod: an application of generalized additive models to case-control data. *Environmental Health: A Global Access Science Source,* 8 3.

Wagner, S.E., Bauer, S.E., Bayakly, A.R. & Vena, J.E. (2013) Prostate cancer incidence and tumor severity in Georgia: descriptive epidemiology, racial disparity, and geographic trends. *Cancer Causes & Control,* 24 (1), 153-166.

Wallace, M.B., Durkalski, V.L. & Vaughan, J. (2001) Age and alarm signals do not predict endoscopic findings among patients with dyspepsia : A multicentre database study. *Gut,* 49 29-34.

Walter, F., Webster, A., Scott, S. & Emery, J. (2012) The Andersen Model of Total Patient Delay: a systematic review of its application in cancer diagnosis. *Journal of Health Services Research & Policy,* 17 (2), 110-118.

Walters, S., Quaresma, M., Coleman, M.P., Gordon, E., Forman, D. & Rachet, B. (2011) Geographical variation in cancer survival in England, 1991--2006: an analysis by Cancer Network. *Journal of Epidemiology & Community Health,* 65 (11), 1044-1052.

Wan, N., Zhan, F.B., Zou, B. & Wilson, J.G. (2013) Spatial Access to Health Care Services and Disparities in Colorectal Cancer Stage at Diagnosis in Texas*. *Professional Geographer,* 65 (3), 527-541.

Wang, K. & Wongkeesong, M.B., N.S. (2005) American Gastroenterological Association technical review on the role of the gastroenterologist in the management of esophageal carcinoma. *Gastroenterology* 2005;128(5):1471-505. *Gastroenterology,* 128 (5), 1471-1505.

Wang, F., Luo, L. & McLafferty, S. (2010) Healthcare access, socioeconomic factors and late-stage cancer diagnosis: an exploratory spatial analysis and public policy implication. *International Journal of Public Policy,* 5 (2-3), 237-258.

Wang, F. & Luo, W. (2005a) Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. *Health & Place,* 11 (2), 131-146.

Wang, F. & Luo, W. (2005b) Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. *Health & Place,* 11 (2), 131-146.

Wang, F., Guo, D. & McLafferty, S. (2012) Constructing geographic areas for cancer data analysis: A case study on late-stage breast cancer risk in Illinois. *Applied Geography,* 35 (1–2), 1-11.

Wang, F., Luo, L. & McLafferty, S. (2010) Healthcare access, socioeconomic factors and late-stage cancer diagnosis: an exploratory spatial analysis and public policy implication. *International Journal of Public Policy,* 5 (2-3), 237-258.

Wang, F., McLafferty, S., Escamilla, V. & Luo, L. (2008a) Late-Stage Breast Cancer Diagnosis and Health Care Access in Illinois. *Professional Geographer,* 60 (1), 54-69.

Wang, J., Liu, F., Gao, H., Wei, W., Zhang, X., Liang, Y. & Cheng, Y. (2008b) The symptom-to-treatment delay and stage at the time of treatment in cancer of esophagus. *Japanese Journal of Clinical Oncology,* 38 (2), 87-91.

Wang, J.B., Abnet, C.C., Fan, J.H., Qiao, Y.L. & Taylor, P.R. (2013) The randomized Linxian Dysplasia Nutrition Intervention Trial after 26 years of follow-up: no effect of multivitamin supplementation on mortality. *JAMA Internal Medicine,* 173 (13), 1259-1261.

Wang, L.D., Zhou, Q., Feng, C.W., Liu, B., Qi, Y.J., Zhang, Y.R., Gao, S.S., Fan, Z.M., Zhou, Y., Yang, C.S., Wei, J.P. & Zheng, S. (2002) Intervention and follow-up on human esophageal precancerous lesions in Henan, northern China, a high-incidence area for esophageal cancer. *Gan to Kagaku Ryoho.Cancer & Chemotherapy,* 29 Suppl 1 159-172.

Wang, N., Cao, F., Liu, F., Jia, Y., Wang, J., Bao, C., Wang, X., Song, Q., Tan, B., Cheng, Y. (2015) The effect of socioeconomic status on health-care delay and treatment of esophageal cancer. *Journal of Translational Medicine* **13**:241

Wang, Y., Hunt, K., Nazareth, I., Freemantle, N., Petersen, I (2013) Do men consult less than women? An analysis of routinely collected general practice data *British Medical Journal open* available at ; http://bmjopen.bmj.com/content/3/8/e003320 [accessed April 2018].

Wang, Z., Goodman, M., Saba, N. & El-Rayes, B. (2013) Incidence and prognosis of gastroesophageal cancer in rural, urban, and metropolitan areas of the United States. *Cancer (0008543X)*, 119 (22), 4020-4027.

Ward, M.H., Cross, A.J., Abnet, C.C., Sinha, R., Markin, R.S. & Weisenburger, D.D. (2012) Heme iron from meat and risk of adenocarcinoma of the esophagus and stomach. *European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation (ECP)*, 21 (2), 134-138.

Warden, C., Cudnik, M., T., Sasson, C., Schwartz, G. & Semple, H. (2012) Poisson cluster analysis of cardiac arrest incidence in Columbus, Oohio. *Prehospital Emergency Care*, 16 (3), 338-346.

Warneke, V.S., Behrens, H., Hartmann, J., Held, H., Becker, T., Schwarz, N.T. & Röcken, C. (2011) Cohort study based on the seventh edition of the TNM classification for gastric cancer: proposal of a new staging system. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 29 (17), 2364-2371.

Wei, W., Yang, J., Zhang, S., Chen, W. & Qiao, Y. (2011) Esophageal cancer mortality trends during the last 30 years in high risk areas in China: comparison of results from national death surveys conducted in the 1970's, 1990's and 2004-2005. *Asian Pacific Journal of Cancer Prevention: APJCP*, 12 (7), 1821-1826.

Weller, D., Vedsted, P., Rubin, G., Walter, F.M., Emery, J., Scott, S., Campbell, C., Andersen, R.S., Hamilton, W., Olessen, F., Rose, P., Nafees, S., van Rijswijk, E., Hiom, S., Muth, C., Beyer, M. & Neal, R.D. (2012) The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *British Journal of Cancer*, 106 (7), 1262-1267.

Whitehead, W., Trivedi, J, Bond, E., van Berkel, V., Fox, M. (2018) Optimal Therapy in Locally Advanced Esophageal Cancer: A National Cancer Database Analysis Journal of gastrointestinal surgery 22 (2) available online via https://link.springer.com/article/10.1007/s11605-017-3548-1[accessed Feb, 2018].

Whiteman, D.C., Parmar, P., Fahey, P., Moore, S.P., Stark, M., Zhao, Z.Z., Montgomery, G.W., Green, A.C., Hayward, N.K. & Webb, P.M. (2010) Association of Helicobacter pylori infection with reduced risk for esophageal cancer is independent of environmental and genetic modifiers. *Gastroenterology*, 139 (1), 73.

Wiringa, A.E., Shutt, K.A., Marsh, J.W., Cohn, A.C., Messonnier, N.E., Zansky, S.M., Petit, S., Farley, M.M., Gershman, K., Lynfield, R., Reingold, A., Schaffner, W., Thompson, J., Brown, S.T., Lee, B.Y. & Harrison, L.H. (2013) Geotemporal Analysis of Neisseria meningitides Clones in the United States: 2000-2005. *PloS One*, 8 (12), e82048.

Wolf, M., Zehentmayr, F., Schmidt, M., Holzel, D. & Belka, C. (2012) Treatment strategies for oesophageal cancer - time-trends and long term outcome data from a large tertiary referral centre. *Radiation Oncology (London, England)*, 7 (1), 60.

World Cancer Research Fund/American Institute for Cancer Research, Food, Nutrition, physical Activity, and the Prevention of Cancer: a Global Perspective. Available from wcrf.org/int/research-we-fund/continuous-update-project-cup/second-expert-report. 2007 [accessed 042017].

World Cancer Research Fund/American Institute for Cancer Research. *Continuous Update Project Findings & Reports.* available at http://www.wcrf.org/int/research-we-fund/continuous-update-project-findings-reports [accessed 042017].

World Health Organization (2011) *European age-standardised rates calculated by the Statistical Information Team at Cancer Research UK, 2011 using data from GLOBOCAN, IARC, version 1.2.* Available online: http://globocan.iarc.fr May 2012].

World Health Organisation ICD10 version 2014 Available online http://apps.who.int/classifications/icd10/browse/2016/en [accessed Feb 2015]

Worsley, A., Wang, A., Hunter, W. (2011) The relationship between eating habits, smoking and alcohol consumption and body mass index among baby boomers *Appetite* 58 (1) 74-80

Wu, A.H., Crabtree, J.E., Bernstein, L., Hawtin, P., Cockburn, M., Tseng, C. & Forman, D. (2003) Role of Helicobacter pylori CagA+ strains and risk of adenocarcinoma of the stomach and esophagus. *International Journal of Cancer.Journal International Du Cancer,* 103 (6), 815-821.

Wu, K.S., Huo, X., Zhu, G.H.. (2007) Relationships between esophageal cancer and spatial environment factors by using Geographic Information System *Science of total environment* 393 219-225.

Wu, K., li, k. (2007) Association between esophageal cancer and drought in China by using Geographic information system *Environment international* 33 603-608.

Wu, Y., Fan, Y., Jiang, Y., Wang, Y., Liu, H. & Wei, M. (2013) Analysis of risk factors associated with precancerous lesion of gastric cancer in patients from eastern China: a comparative study. *Journal of Cancer Research and Therapeutics,* 9 (2), 205-209.

Xiao, H., Tan, F. & Goovaerts, P. (2011) Racial and geographic disparities in late-stage prostate cancer diagnosis in Florida. *Journal of Health Care for the Poor and Underserved,* 22 (4), 187-199.

Xie, F., Zhang, Y., Zheng, Q., Jin, H., Wang, F., Chen, M., Shao, L., Zou, D., Yu, X. & Mao, W. (2013) Helicobacter pylori infection and esophageal cancer risk: an updated meta-analysis. *World Journal of Gastroenterology: WJG,* 19 (36), 6098-6107.

Yang, S., Huang, Y., Shao, Y., Chen, X.Y., Xian, L., Zheng, J., Wen, Y., Chen, X., Li, H. and Yang, C., (2012), *Screening for oesophageal cancer (Cochrane Review).* Issue 12, London, Wiley.

Yang, T., Shoff, C. & Noah, A.J. (2013) Spatializing health research: what we know and where we are heading. *Geospatial Health,* 7 (2), 161-168.

Yin, F., Feng, Z. & Li, X. (2012) Spatial analysis of county-based gonorrhoea incidence in mainland China, from 2004 to 2009. *Sexual Health (14485028),* 9 (3), 227-232.

Yin, R. K. (1989). *Case study research: Design and methods. Applied Social Research Series,* Vol. 5. London: Sage.

Yin, R. K. (1993). *Applications of case study research. Applied Social Research Series, Vol. 34.* London: Sage

Yin, R.K. (2013) *Case study research: Design and methods* London; Los Angeles, California: Sage.

Yin, R.K. (2014) *Case study research: design and methods.* 5th ed. Thousand Oaks: SAGE Publications.

Zafar, A., Mak, T., Whinnie, S., Chapman, A. (2012) The 2-week wait referral system does not improve 5-year colorectal cancer survival. Colorectal Dis. 14(4):e177-80.

ZHANG Xueyan ZHUANG Dafang MA Xin JIANG Dong (2014) Esophageal cancer spatial and correlation analyses： Water pollution, mortality rates, and safe buffer distances in China. Chinese Medical Journal. 24 (1), 46-58.

Zhang, D., Su, X. & Lin, P. (2008) Survival analysis of patients with stage II squamous cell carcinoma of the thoracic esophagus after esophagectomy. *Ai Zheng [Chinese Medical Journal]* 2008;27(2):113-8. *Chinese Medical Journal,* 27 (2), 113-118.

Zhao, L., Wei, W.Q., Zhao, D.L., Hao, C.Q., Lin, D.M., Pan, Q.J., Li, X.Q., Lei, F.H., Wang, J.W., Wang, G.Q., Shang, Q. & Qiao, Y.L. (2012) Population-based study of DNA image cytometry as screening method for esophageal cancer. *World Journal of Gastroenterology : WJG,* 18 (4), 375-382.

Zheng, S., Vuitton, L., Sheyhidin, I., Vuitton, D.A., Zhang, Y. & Lu, X. (2010a) Northwestern China: a place to learn more on oesophageal cancer. Part one: behavioural and environmental risk factors. *European Journal of Gastroenterology & Hepatology,* 22 (8), 917-925.

Zheng, S., Vuitton, L., Sheyhidin, I., Vuitton, D.A., Zhang, Y. & Lu, X. (2010b) Northwestern China: a place to learn more on oesophageal cancer. Part one: behavioural and environmental risk factors. *European Journal of Gastroenterology & Hepatology,* 22 (8), 917-925.

Appendix

Appendix 1 FHSC Data Management Plan

(NB: This form should be completed at the start of all projects where data are not being stored in alternative sources, e.g. Clinical Trial Data held in the NHS).

Date	**April 2014**
Researcher(s)	**Amanda Lee (PhD student)** **Professor Roger Watson (Supervisor)/Sam Khulusi (Supervisor) Dr Graham Ferrier (Supervisor) Dr Erik Gardiner (statistician)**
Project title	Spatially accountable differences in presentations of patients with Gastroesophageal Cancer
Brief description	This research uses quantitative methodology to define 'late' presentations so that all sufferers who present too late for treatment, or with particularly aggressive tumours, may be captured and reported in a mapped representation of hotspots. These 'hotspots' will facilitate targeted clinical interventions.

Section 1: Project Information

Project title: **Spatially accountable differences in presentations of patients with Gastroesophageal cancer**
Project duration - this is a 5-year part time PhD thesis
Partners (if applicable) Hull Yorkshire Cancer Collaborative – Clinical consultant Sam Khulusi
Brief description This research uses quantitative methodology to define 'late' presentations so that all sufferers who present too late for treatment, or with particularly aggressive tumours, may be captured and reported in a cartographic representation of hotspots which will facilitate targeted clinical intervention.
Faculty or University requirements for data management Data will be stored on the secure password controlled University X drive which is personal to the main researcher. No identifying data will be available to share unless prerequisite ethical approval has been awarded.
1.6 Funding body(ies) N/A – Self funded PhD
1.7 Budget (estimate if necessary) 35000 PhD funding over 5-years.
1.8 Funding body requirements for data management University of Hull data storage – there is no charge. This is enabled through the PhD student's current role as lecturer at the University of Hull.

Section 2: Data, Materials, Resource Collection Information

2.1 Brief description of data sources Office for National Statistics (ONS) population data Demographic and attribute data from Humber and Yorkshire Cancer Collaborative Network – supplied by National Cancer Collaborative. (Quasi identifiable)
2.2 Data collection process All data are available in Excel format. These will be transferred to SPSS V20 and ARCGIS for purposes of quantitative analysis. **Data retrieved from NCRAS cancer registration datasets,** **Public Health England publications gateway number: 2015778 –** **dataset First published: April 2016© Crown copyright 2016 details available at** **www.nationalarchives.gov.uk/doc/open-government-licence/version/3/**
2.3 Will data be available in electronic format (if so then state format(s))? Yes SPSS/excel/Arc Gis. They will be stored in a password controlled area which is only accessible to the student and her academic supervisors.
2.4 Will the data be available in hard copy (if so then state format(s))? No
2.5 Will the data stand alone and be comprehensible to a third party or be accompanied by explanatory documentation?

These data are comprehensible to any third party and are therefore subject to Data Protection Act and confidentiality.
2.6 Describe quality assurance process for data management Data from OCNS have been identified as accurate to 95% CI. NCIN data from national cancer collaborative have the same. All variables created from original datasets will be scrutinized for accuracy and reported accordingly.

Section 3: Ethics, Intellectual Property

3.1 How have the ethical aspects of data storage and subsequent access been addressed?

The data were sourced several years previously by clinical gastroenterology consultant.

Ethics approval was granted as data were not considered identifiable. However, this has now changed, and data are considered as quasi identifiable. This has been highlighted when the PhD student has attempted to access updated datasets.

There will now be a process to seek ethical approval, present a research protocol to data agency and request re access to data.

As this is a very large dataset, with several participants who have died, it is impossible and impractical to contact each for consent to participate. Rather, the suggestion to maintain confidentiality in dataset and integrity of records is key to the underpinning data management plan.

The potential for several subjects to be unhappy to share their diagnosis needs to be considered in relation to the overall dataset. GOC has been linked with some contributing factors which may make certain subjects hesitant to share the diagnosis and this must be considered.

It is left to the researcher to consider whether this is in the public interest to research. There are risks to patient and public in regard to postcode level diagnostic data. However, the researcher is a trained nurse, PhD supervisors all work in the NHS system and under professional codes of conduct which demand confidentiality. Information will be securely held and no information will be disclosed without sufficient aggregation to ensure anonymity.

3.2 Will the data comply with relevant legislation such as Data Protection Act, Copyright and Intellectual Property?

The Data Protection Act.

Under section 33 of the Data Protection Act, data are going to be processed so that they are not likely to cause substantial damage or distress to subjects or their families. Published maps displaying presentation stage 'hotspots' will aggregate findings to scales at which individual subjects are not identifiable. Survival and cluster analysis techniques will also aggregate data to a non-identifiable range prior to publication or dissemination of research.

Data will be fairly processed, and follow management plan accredited by NREC and University of Hull regulations for data management.

Other legislation considered includes : Statistics and Registration Services Act (2007); Section 60 of the Health and Social Care Act (2001).; Article 8 EU convention of Human Rights duty of confidentiality of information; Section 251 of the NHS Act 2006, and subsequent Regulations, allows the Secretary of State for Health to set aside the common law duty of confidence when it is in the public interest. This allows time-limited disclosure of identifiable patient information, without patient consent, for medical research.

3.3 If several partners are involved how will compliance with 3.2 be assured?

The initial project will be assessed through NRES and appropriate ethical approval sought.

The data management plan will ensure that;

Data on sensitive and confidential information i.e. gender, age, postcode, date of death, diagnosis place of treatment and treating authority are considered identifiable, so any

resultant publications will ensure k anonymity is maintained. All data are stored in a secure, password encoded format.

Coding of potentially identifying information will be undertaken (this includes subjects postcode to house code level, dates of birth). Any published data will consider age grouped, and district level enumerated post code data.

Data are only made available to PhD supervisors and me as primary researcher.

PhD supervisors are responsible for overseeing data processing and will act to maintain the integrity during this research.

Original source data will be available only through NCIN. The geographical information system will only publish the hotspot mapping, no identifiable data will be released without prior consent.

Any information deemed appropriate for publication will be anonymised sufficiently to maintain the duty of confidence.

Electronic data entered into my PC will continue to be stored for five years post completion of project and then destroyed and not archived.

A memory stick I use as a backup of the data will be pass code encrypted and stored under lock and key which is accessible by me only.

The data will be used as reference for projects developed from the current, to further evaluate the educational practice in question.

Section 4: Access and Use of Information

4.1 Are you required, and with whom, to share the data subsequent to completion of the project? This data will inform a thesis which will be available through public access in e'thesis format. All quasi identifying data will be removed prior to publication of any materials subject to confidentiality. Publication of maps displaying de identified hotspots of this region will be sought by the principal researcher and disseminated to clinical service providers so that interventions may be undertaken. A geographical information system which encompasses GOG related attribute data will be publicly available. However, all patient data will be removed from this system.
4.2 If 'yes' to 4.1, in what format will data be shared? Publications and e thesis sites. The geographic information system will be in GIS format, available as.dbf file,.shp file and GIS determined formatted files.
4.3 Will the data have to be stored for a specific period (if so, how long)? Data relating to the GOC patients will be maintained for a total of 5-years for purposes of assessment of data integrity. These may be accessed through appropriate application processes. (ie via the original authority). These data will be destroyed after a period of 5-years.
4.4 Who may need to have access to the data? PhD supervisors will be responsible for overseeing the data. They will be granted access to oversee processes and assess data integrity.

A consulting statistician will provide support with analysis of data which are rigorously coded for de identification. This statistician is employed through the University of Hull and is therefore subject to the university policies on data management and bound under the employer's confidentiality agreements.

4.5 How do you anticipate the data being used subsequent to the project?

Cartographic displays will identify hotspots of locations where services may be targeted for intervention (ie GP surgeries/local chemists/residential homes etc could be identified as in areas of potential higher incidences. In these circumstances, clinical GOC services will engage with providers and supply information and training as required).

District population weighting of postcodes will enable further analysis of other disease processes, and subsequent targeting of health related interventions.

Section 5: Storage and Backup of Data

5.1 Where and how will the data be stored **during the lifespan of the project?** These data will be stored in a secure x drive on University of Hull drive. This is only accessible by the student Amanda Sherratt and supervisors. Some data may be required off site, to be kept on an encrypted USB drive, to facilitate the researcher to use in her home office. Such data will be kept in a secured drawer in the office and the USB drive will be password encrypted.
5.2 Where and how will the data be stored **on completion of the project?** **All quasi identifieable data must be subject to further approval and any persons requesting access to this data can go through the national cancer collaborative and OCNS.**
5.3 What provision is being made for backup of the data? Data will be backed up on a locked and pass coded USB Flash Drive, to be secured in a locked area as per protocol. Data will also be stored on the university level x drive, again, password encrypted and maintained through the University level data encryption policy.
5.4 Will different version of the data be stored? Data will be stored in Excel/SPSS and GIS formatted documents. Newly created variables will be stored in SPSS format

Section 6: Archiving and Future Proofing of Information

6.1 What is the long-term strategy for storage and availability of the data? Annonymised data will be stored and available on X drive for a 5-year period to enable users to scrutinize methodological processes. However, principles of confidentiality must be upheld for any members viewing any data which could be considered as quasi identifiable.
6.2 Will the information be kept after the life of the project, for how long and in what format? This will be in GIS/SPSS format and results published in the PhD overall thesis.
6.3 If the data include confidential or sensitive information, how will these data be managed? This data does not refer to any persons by name. However, as postcode level data are being used, these will be aggregated to enable K anonymity factor > 10. The temporal range of data are 1999 – 2010, so reflects mainly GOC of deceased subjects. However, in areas where there have been very small numbers of cases, data are aggregated to compensate for potential identification.
6.4 If meta data or explanatory information is to be stored, how will this be linked to the data? See above
6.5 How will the data be cited? All originating sources will be cited in accordance with their policies.

The Geographic Information System which is developed from this project, may be referred to from resultant publications.

E thesis will be referenced as per referencing guidelines.

Section 7: Resourcing of Data Management

7.1 List the specific staff who will have access to the data and denote who will have the responsibility for data management. Amanda Lee – PhD student – responsible for cleaning and management of data relating to this project Overseen by Professor Roger Watson– PhD supervisor/Clinical Consultant Dr Sam Khulusi – Consultant Gastroenterologist –/Dr Graham Ferrier – Geographer and PhD Supervisor/Dr Eric Gardiner – consulting statistician.
7.2 How will data management be funded? Self funded through PhD studentship with University of Hull.
7.3 How will data storage be funded? Self funded with University of Hull agreeing to provide data storage capacity through employment as lecturer

Section 8: Review of Data Management process

8.1 How will the data management plan be adhered to?

Data on sensitive and confidential information i.e. gender, age, postcode, date of death, diagnosis place of treatment and treating authority are considered identifiable, so any resultant publications will ensure k anonymity is maintained. All data are stored in a secure, password encoded format.

Coding of potentially identifying information will be undertaken (this includes subjects postcode to house code level, dates of birth). Any published data will consider age grouped, and district level enumerated post code data. Data are only made available to PhD supervisors and me as primary researcher.

PhD supervisors are responsible for overseeing data processing and will act to maintain the integrity during this research. Original source data will be available only through NCIN. The geographical information system will only publish the hotspot mapping, no identifiable data will be released without prior consent.

Any information deemed appropriate for publication will be anonymised sufficiently to maintain the duty of confidence.

Electronic data entered into my PC will continue to be stored for five years post completion of project and then destroyed and not archived.

A memory stick I use as a backup of the data will be pass code encrypted and stored under lock and key which is accessible by me only.

The data will be used as reference for projects developed from the current, to further evaluate the educational practice in question

8.2 Who will review the data management plan?

PhD supervisors Professor Watson, Dr Khulusi and Dr Ferrier. With external review through NHS information centres.

Professor Roger Watson (author of FHSC data management documentation and Professor of Nursing) Professor Kate Galvin (Dean of research – University of Hull)

Section 9: Statements and Personnel Details

9.1 Statement of agreement

I/we agree to the specific elements of the plan as outlined:

Principal investigator or PhD supervisor

Title	Dr	Professor
Designation	PhD supervisor and Clinical gastroenterology consultant	PhD supervisor
Name	Sam Khulusi	Roger Watson
Date	14/07/2014	14/07/2014
Signature		

Researcher

Title	Ms
Designation	PhD student
Name	Amanda Jayne Lee
Date	14/07/2014
Signature	

PhD supervisor

Title	Dr
Designation	Lecturer and PhD supervisor
Name	Graham Ferrier
Date	14/07/2014
Signature	

Appendix 2 NHS rec form submitted for review.

NHS R&D Form IRAS Version 5.3.1

Welcome to the Integrated Research Application System

IRAS Project Filter

The integrated dataset required for your project will be created from the answers you give to the following questions. The system will generate only those questions and sections which (a) apply to your study type and (b) are required by the bodies reviewing your study. Please ensure you answer all the questions before proceeding with your applications.

Please complete the questions in order. If you change the response to a question, please select "Save" and review all the questions as your change may have affected subsequent questions.

Please enter a short title for this project (maximum 70 characters)
factors associated with late presenting GOC

1. Is your project research?

(●) Yes ( ) No

2. Select one category from the list below:

( ) Clinical trial of an investigational medicinal product

( ) Clinical investigation or other study of a medical device

( ) Combined trial of an investigational medicinal product and an investigational medical device

( ) Other clinical trial to study a novel intervention or randomised clinical trial to compare interventions in clinical practice

( ) Basic science study involving procedures with human participants

(●) Study administering questionnaires/interviews for quantitative analysis, or using mixed quantitative/qualitative methodology

( ) Study involving qualitative methods only

( ) Study limited to working with human tissue samples (or other human biological samples) and data (specific project only)

( ) Study limited to working with data (specific project only)

( ) Research tissue bank

( ) Research database

If your work does not fit any of these categories, select the option below:

( ) Other study

2a. Please answer the following question(s):

a) Does the study involve the use of any ionising radiation?	( ) Yes	(●) No
b) Will you be taking new human tissue samples (or other human biological samples)?	( ) Yes	(●) No
c) Will you be using existing human tissue samples (or other human biological samples)?	( ) Yes	(●) No

3. In which countries of the UK will the research sites be located? *(Tick all that apply)*

[x] England
[ ] Scotland

Appendix 3 – Ethics approval University of HULL – note researcher name change since application

UNIVERSITY OF Hull

Ms A Sherratt
Faculty of Health & Social Care
University of Hull
Hull
HU6 7RX

FACULTY OF HEALTH AND SOCIAL CARE
T: 01482 464680
E: j.dyson@hull.ac.uk

OUR REF: 156
07 September 2014

Dear Amanda

Re: Small area analysis of presentation stage gastroesophageal cancer

Thank you for your correspondence submitted in response to the points raised by the Faculty of Health and Social Care Research Ethics Committee in our letter of 04 September 2014.

Given the information you have provided, I am delighted to grant Chair's approval as per the Terms of Reference of the Faculty of Health & Social Care Research Ethics Committee.

I wish you every success with your study.

Yours sincerely

Dr Judith Dyson
Chair, Research Ethics Committee

cc: file/supervisors

Appendix 4 Health Research Authority - IRAS approval letter

NHS
Health Research Authority
NRES Committee Yorkshire & The Humber - South Yorkshire
Unit 001
Jarrow Business Centre
Rolling Mill Road
Jarrow
Tyne and Wear
NE32 3DT

Telephone: 0191 428 3565

2 September 2015

Ms Amanda Jayne Lee
Lecturer Health Professional Studies / PhD Student
Dept of Health & Social Care
University of Hull
Hull
HU6 7RX

Dear Ms Lee

Study title:	**A case study of spatiality in advanced presentation gastroesophageal cancer**
REC reference:	**15/YH/0318**
Protocol number:	**self funded**
IRAS project ID:	**161434**

Thank you for your response of 18 August 2015, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager, Ms Gillian Mayer, nrescommittee.yorkandhumber-southyorks@nhs.net. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a **Favourable** ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at **http://www.rdforum.nhs.uk**.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations

Registration of Clinical Trials
All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.
There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.
To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.
If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non-registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.
It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites
NHS sites
The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).
Approved documents
The final list of documents reviewed and approved by the Committee is as follows:

Document Version Date 04 June 2015

A Research Ethics Committee established by the Health Research Authority

Evidence of Sponsor insurance or indemnity (non NHS Sponsors only)

04 January 2015

Letter from sponsor [Letter from University of Hull] 22/3/18

Other [Supervisory Meeting - 26/2/15]

26 February 2015

Other [Supervisory Meeting - 21/11/15]

21 November 2014

Other [FHSC Approval Letter]

07 September 2014

REC Application Form [REC_Form_18082015]

18 August 2015

Research protocol or project proposal

Summary CV for Chief Investigator (CI) [CV for Amanda Lee]

Summary CV for student [CV for Amanda Lee]

Summary CV for supervisor (student research) [CV for Dr Sam Khulisi]

Summary CV for supervisor (student research) [CV for Roger Watson]

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

□ Notifying substantial amendments

□ Adding new sites and investigators

□ Notification of serious breaches of the protocol

□ Progress and safety reports

□ Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

15/YH/0318 please quote this number on all correspondence

A Research Ethics Committee established by the Health Research Authority

With the Committee's best wishes for the success of this project.

Yours sincerely

pp

Dr Ian Woollands

Chair

Email: nrescommittee.yorkandhumber-southyorks@nhs.net

Email:nrescommittee.yorkandhumber-southyorks@nhs.net

Enclosures: 'After ethical review – guidance for researchers'

Copy to:

Dr Andrew Taylor –R&D Dept, University of Hull

HRA Confidentiality Advice Team

Appendix 5 HRA research summary

Available at: **[http://www.hra.nhs.uk/news/research-summaries/spatial-evaluation-of-patient-interval-factors-in-goc/].**

Spatial evaluation of patient interval factors in GOC Full title A case study of spatiality in advanced presentation gastroesophageal cancer. Research type Research study

IRAS ID 161434

Contact Name Amanda Jayne Lee

Contact Email **a.sherratt@hull.ac.uk**

Sponsor organisation University of Hull

Additional reference number fields REF ODR_2014_162,

NCIN request for data

Research summary

Gastroesophageal cancers (GOC) are the 8th most common malignancy in the world. Patients are often over 65, male and from lower socioeconomic groups, many present at very advanced stages with their cancer, resulting in poor survival rates. These circumstances provide a unique platform to evaluate if mapping patient and population characteristics can predict geographical areas of higher incidence. This information can be used to target populations and encourage earlier presentation. This case study analyses a 10 year cohort of GOC patients, to reveal patterns in populations, behaviours and ecology which may be related to advance presenting gastroesophageal cancers. Data are geographically analysed so results are displayed as maps for easy interpretation. This research fully integrates the four pillars of geography, capturing the essence of person, place, space and time linked with the GOC cohort. This approach evaluates all elements of care, historical changes in treatments and means that geographical boundaries are not constrained to artificially produced large scale 'groupings' which frequent the public domain. The findings of this research will map areas of increased density GOC which are not explained through population descriptors. Cluster analysis techniques will identify cancer 'hot spots' to inform targeted health promotion and intervention. Results which are displayed in a map format are more accessible to the audience. This also enables cross analysis of a wide range of information related to cancer. Geographical analysis of data can be useful in prediction of disease incidence. Localised (smaller scale) analysis can capture many elements of care and has a potential to inform service providers about the populations most at risk. This research presents a case study of GOC in a geographically defined regional referral center. It applies

geography to health and population profiling, to produce an in depth analysis of how people present with GOC.

REC Name Yorkshire & The Humber - South Yorkshire Research Ethics Committee

REC Reference 15/YH/0318

REC Opinion Further Information Favourable Opinion

Further Information Favourable Opinion

Date of REC Opinion 2 September 2015

CAG approval (University level)

22 March 2016

David Richards

University of Hull

PVC office

Room 005

Ground Floor, VENN building University of Hull

HU67RX

david.richards@hull.ac.uk

Dear David

Application title: A case study of spatiality in advanced presentation gastroesophageal cancer.

CAG reference: 15/CAG/0165

IRAS project ID: 161434

REC reference: 15/YH/0318

Thank you for your research application, submitted for approval under Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002 to process patient identifiable information without consent. Approved applications enable the data controller to provide specified information to the applicant for the purposes of the relevant activity, without being in breach of the common law duty of confidentiality, although other relevant legislative provisions will still be applicable.

The role of the Confidentiality Advisory Group (CAG) is to review applications submitted under these Regulations and to provide advice to the Health Research Authority on whether an application should be approved, and if so, any relevant conditions. This application was considered at the CAG meeting held on 06 August

2015.

Health Research Authority

The Health Research Authority, having considered the advice from the Confidentiality

Advisory Group as set out below, has determined the following:

1. The application is approved, subject to compliance with the standard and specific conditions of approval.

This letter should be read in conjunction with the outcome letter dated 14 August 2015.

Context

Purpose of application

This application from University of Hull describes a case study of cancer of the gullet (gastroesophageal cancer, GOC). GOC presents an interesting case for

study, as certain attributes are very commonly associated with sufferers, such as advancing age, male gender, smokers and those with higher alcohol intake, as well as those from lower socioeconomic groups. This project will use a case study approach to map and analyse patients' environments against these common attributes, so that maps can be produced according to underlying population demographics and to analyse how the environment plays a part in stage of presentation.

A search of the disease register database held by the National Cancer Intelligence Network (NCIN) at Public Health England will be undertaken to yield all patients with histologically confirmed GOC who presented to the Queens Centre Castle Hill Hospital, during the period 1999 2013. ONS lifestyle survey data from this period will be used to develop the weighted map of the region. ONS population data from 2010 census will be used to determine age and sex of postcode level data, so that incidences may be weighted accordingly. Data from the Environmental Systems Research Institute, the geographical underpinning map datasets, will be used for the purposes of cartography and mapping the appropriate geographical catchment area.

All data will be held within the geographic information system for purposes of analysis. This will be stored in an encrypted format and only accessible to the researcher and her academic supervisors.

A recommendation for class 2 support was requested to cover access to confirmed GOC diagnosis, histological type, Interventions, GP registration, patients Date of

Birth, age, gender, date of presentation to cancer specialist services and date of death where this has occurred.

Confidential patient information requested

Access was requested to confirmed GOC diagnosis, histological type, Interventions,

GP registration, patients Date of Birth, age, gender, date of presentation to cancer specialist services and date of death.

Following the conditional approval the applicant experienced difficulty confirming Information Governance Toolkit compliance for the host institution so moved the projects data repository to the University of Leeds. It was noted that the only other change was an extension to the data ranges from 2010 to 2013 due to project delays. The chair accepted these changes and recommended final approval.

Confidentiality Advisory Group advice conclusion

The CAG agreed that the minimum criteria under the Regulations appeared to have been met and that there was a public interest in projects of this nature being conducted, and therefore advised recommending support to the Health Research Authority, subject to compliance with the specific and standard conditions of support as set out below.

Specific conditions of support

1. The applicant is requested to clarify that this is a multi-centre study. **Confirmed 18/08/2015**
2. The applicant is requested to clarify the retention of identifiers beyond 5 years post PHD and what can be done to reduce primary risks post collection.
a) What conversion into a less identifiable form can take place prior to analysis? **Confirmed 18/08/2015**

b) What does the applicant intend to do to reduce identifiability of data prior to publication? **Confirmed 18/08/2015**
3. Increased patient notification. What can the researcher do to encourage NCIN for increased notification on the website. **Confirmed 18/08/2015**
4. The applicant should address what definition of anonymous data is being used (Sweeney K and quasi anonymous data). **Confirmed 18/08/2015**
5. Favourable opinion from a Research Ethics Committee. **Confirmed 2/09/2015**
6. Confirmation from the IGT Team at the Health and Social Care Information Centre of suitable security arrangements via Information Governance Toolkit (IGT) submission. **Confirmed 17/03/2016**

As the above conditions have been accepted and/or met, this letter provides confirmation of final approval. I will arrange for the register of approved applications on the HRA website to be updated with this information.

Annual review
Please note that your approval is subject to submission of an annual review report to show how you have met the conditions or report plans, and action towards meeting them. It is also your responsibility to submit this report on the anniversary of your final approval and to report any changes such as to the purpose or design of the proposed activity, or to security and confidentiality arrangements. An annual review should be provided no later than 22/03/2017 and preferably 4 weeks before this date. If at any stage you no longer require support under the Regulations as you will cease processing confidential patient information without consent you should inform the Confidentiality Advice Team of this in writing as soon as possible.

Reviewed documents

The documents reviewed at the meeting were:	*Version*	*Date*
Document		
CAG application from (signed/authorised)	4.0.0	19 June 2015
Covering letter on headed paper	10 June 2015	
Other [IRAS XML Form]	3.5	24 June 2015
Research protocol or project proposal	2	22 June 2015

Membership of the Committee
The members of the Confidentiality Advisory Group who were present at the consideration of this item or submitted written comments are listed below.
User Feedback
The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the

service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R & D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

Yours sincerely

Ben Redclift

Email: HRA.CAG@nhs.net

Enclosures: *List of members who considered application*

Standard conditions of approval

For note – Annual review forms have been submitted via UoH research office Feb 2017

Appendix 6 SPSS Syntax for cleaning data

Reducing age

```
compute age5ya = 0 .
if( fiveyragebnd ="20- 24") age5ya = 20 .
if( fiveyragebnd ="25- 29") age5ya = 25 .
if( fiveyragebnd ="30- 34") age5ya = 30 .
if( fiveyragebnd ="35- 39") age5ya = 35 .
if( fiveyragebnd ="40- 44") age5ya = 40 .
if( fiveyragebnd ="45- 49") age5ya = 45 .
if( fiveyragebnd ="50- 54") age5ya = 50 .
if( fiveyragebnd ="55- 59") age5ya = 55 .
if( fiveyragebnd ="60- 64") age5ya = 60 .
if( fiveyragebnd ="65- 69") age5ya = 65 .
if( fiveyragebnd ="70- 74") age5ya = 70 .
if( fiveyragebnd ="75- 79") age5ya = 75 .
if( fiveyragebnd ="80- 84") age5ya = 80 .
if( fiveyragebnd ="85- 89") age5ya = 85 .
if( fiveyragebnd = "90+") age5ya = 90 .
execute .
```

Changing Vital Status numerical

```
compute recodevitalstat = 9 .
if (vitalstatus = "D") recodvitalstat = 0 .
If (vitalstatus = "A") recodvitalstat = 1 .
execute .
```

date and time wizard calculating says survival

```
compute dayssurv = 1826
if (dayssurv = ".") .
```

```
execute .

recoding IMD quintiles (04/07/10)

compute recodeimd04 = 9 .
if (quint2004 = "1 - least deprived") recodeimd04 = 1 .
If (quint2004 = "5 - most deprived") recodeimd04 = 5 .
if (quint2004 ="3") recodeimdo4 = 3 .
if (quint2004 ="2") recodeimdo4 = 2 .
if (quint2004 ="4") recodeimdo4 = 4 .
execute .


compute recodeimd07 = 9 .
if (quint2007 = "1 - least deprived") recodeimd07 = 1 .
If (quint2007 = "5 - most deprived") recodeimd07 = 5 .
if (quint2007 ="3") recodeimdo7 = 3 .
if (quint2007 ="2") recodeimdo7 = 2 .
if (quint2007 ="4") recodeimdo7 = 4 .
EXECUTE

compute recodeimd10 = 9 .
if (quint2010 = "1 - least deprived") recodeimd10 = 1 .
If (quint2010 = "5 - most deprived") recodeimd10 = 5 .
if (quint2010 ="3") recodeimd10 = 3 .
if (quint2010 ="2") recodeimd10 = 2 .
if (quint2010 ="4") recodeimd10 = 4 .
EXECUTE


Calculating descriptive stats
Age recoding
```

```
DESCRIPTIVES VARIABLES=Age
 /STATISTICS=MEAN SUM STDDEV RANGE MIN MAX KURTOSIS SKEWNESS.


RECODE Age (0 thru 44=1) (45 thru 54=2) (55 thru 64=3) (65 thru 74=4) (75 thru 84=5) (85 thru  150=6) INTO agegrpdfordescs.
VARIABLE LABELS  agegrpdfordescs 'agegrpdfordescs'.
EXECUTE.


ICD site recode
RECODE siteicd10 ('C150'=1) ('C151'=1) ('C153'=1) ('C154'=1) ('C159'=0) ('C152'=2) ('C155'=2)
  ('C160'=2) ('C161'=3) ('C162'=3) ('C163'=3) ('C164'=3) ('C165'=3) ('C166'=3) ('C167'=3) ('C168'=3)
  ('C169'=3) INTO siteofcancer.
VARIABLE LABELS  siteofcancer 'cansite'.
EXECUTE.
Remove ICD 2010 cases labelled as 3
DATASET ACTIVATE DataSet1.
USE ALL.
COMPUTE filter_$=(siteofcancer  < 3).
VARIABLE LABELS filter_$ 'siteofcancer  < 3 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.


Two stage morphology recoding


RECODE morphcoded (8000 thru 8001=1) (8010 thru 8049=2) (8050 thru 8089=3) (8140 thru 8389=4) (8440
  thru 8499=5) (8500 thru 8549=6) (8560 thru 8589=7) (8720 thru 8799=8) (8800 thru 8804=9) (8890 thru
```

```
   8929=10) (8930 thru 8999=11) INTO cancermorph.
VARIABLE LABELS  cancermorph 'cancermorph'.
EXECUTE.

RECODE cancermorph (8000 thru 8001=1) (8010 thru 8049=2) (8050 thru 8089=3)
(8140 thru 8389=4)
   (8440 thru 8499=5) (8500 thru 8999=6) INTO canmorph.
VARIABLE LABELS  canmorph 'canmorph'.
EXECUTE.

Descriptive analysis of cohort
FREQUENCIES VARIABLES=agegrpdfordescs gender
 /STATISTICS=SUM
 /ORDER=ANALYSIS.
Recode Year of diagnosis
compute recodeyodiag = 50 .
if  ( yodiag = 2011 ) recodeyodiag = 11 .
If  ( yodiag = 2010 ) recodeyodiag = 10 .
if  ( yodiag = 2009 ) recodeyodiag = 9 .
if  ( yodiag = 2008 ) recodeyodiag = 8 .
If  ( yodiag = 2007 ) recodeyodiag = 7 .
if  ( yodiag = 2006 ) recodeyodiag = 6 .
if  ( yodiag = 2005 ) recodeyodiag = 5 .
If  ( yodiag = 2004 ) recodeyodiag = 4 .
if  ( yodiag = 2003 ) recodeyodiag = 3 .
if  ( yodiag = 2002 ) recodeyodiag = 2 .
If  ( yodiag = 2001 ) recodeyodiag = 1 .
if  ( yodiag = 2000 ) recodeyodiag = 0 .
Execute
Recoding morphology to three groups
```

```
RECODE cancermorph (1=3) (2=3) (3=1) (4=2) (5=3) (6=3) INTO grpdmorph.
VARIABLE LABELS  grpdmorph 'grpdmorph'.
EXECUTE.
```

CALCULATING DISTRICT POSTCODES

DATASET ACTIVATE DataSet1.

RECODE distpostcode ('DN14'=83) ('DN15'=65) ('DN16'=67) ('DN17'=86) ('DN18'=29) ('DN19'=10)

('DN20'=55) ('DN21'=7) ('DN31'=24) ('DN32'=83) ('DN33'=43) ('DN34'=40) ('DN35'=91) ('DN36'=26)

('DN37'=44) ('DN38'=3) ('DN39'=5) ('DN40'=39) ('DN41'=8) ('DN6'=1) ('DN9'=33) ('HU1' = 3) ('HU10' =48)

('HU11' =35) ('HU12' =75) ('HU13' =31) ('HU14' =17) ('HU15' =46) ('HU16' =48) ('HU17' =87)

('HU18' =29) ('HU19' =26) ('HU2' =7) ('HU20' =2) ('HU3' =50) ('HU4' =51) ('HU5' =99) ('HU6' =82)

('HU7' =74) ('HU8' =91) ('HU9' =96) ('YO11' =68) ('YO12' =81) ('YO13' =19) ('YO14' =33)

('YO15' =44) ('YO16' =62) ('YO19' =30) ('YO21' =14) ('YO22' =24) ('YO23' =32) ('YO24' =65)

('YO25' =87) ('YO26' = 46) ('YO30' =57) ('YO31' =66) ('YO32' =77) ('YO41' =20) ('YO42' =31)

('YO43' =15) ('YO8' =75)

INTO noGOCperdist.

VARIABLE LABELS noGOCperdist 'noGOCperdist'.

EXECUTE.

2 way annova to analyse correlations between age and cancer site

DATASET ACTIVATE DataSet1.

UNIANOVA siteofcancer BY gender agegrpdfordescs

/METHOD=SSTYPE(3)

/INTERCEPT=INCLUDE

/POSTHOC=agegrpdfordescs(TUKEY)

/PLOT=PROFILE(agegrpdfordescs*gender)

```
 /PRINT=ETASQ HOMOGENEITY DESCRIPTIVE
 /CRITERIA=ALPHA(.05)
 /DESIGN=gender agegrpdfordescs gender*agegrpdfordescs.
```

Exploring relationships among variables

Socioeconomics and survival

```
compute recodeimd10 = 9 .
if (quint2010 = "1 - least deprived") recodeimd10 = 1 .
If (quint2010 = "5 - most deprived") recodeimd10 = 5 .
if (quint2010 ="3") recodeimd10 = 3 .
if (quint2010 ="2") recodeimd10 = 1 .
if (quint2010 ="4") recodeimd10 = 5 .
execute .
```

gender and age grouped survival

socioeconomic group and survival

```
 COXREG dayssurvcensored
 /STATUS=recodeimd10(1)
 /PLOT SURVIVAL HAZARDS
 /PRINT=CI(95) CORR
 /CRITERIA=PIN(.05) POUT(.10) ITERATE(20)./POSTHOC=TUKEY
ALPHA(0.05).
```

```
 COXREG dayssurvcensored
 /STATUS=recodeimd10(5)
 /PLOT SURVIVAL HAZARDS
 /PRINT=CI(95) CORR
```

/CRITERIA=PIN(.05) POUT(.10) ITERATE(20)./POSTHOC=TUKEY ALPHA(0.05).

2 way ANNOVA surgery and grouped age

DATASET ACTIVATE DataSet2.

UNIANOVA dayssurvcensored BY agegrpdfordescs surgery

/METHOD=SSTYPE(3)

/INTERCEPT=INCLUDE

/POSTHOC=agegrpdfordescs(TUKEY)

/PLOT=PROFILE(agegrpdfordescs*surgery)

/PRINT=ETASQ HOMOGENEITY DESCRIPTIVE

/CRITERIA=ALPHA(.05)

/DESIGN=agegrpdfordescs surgery agegrpdfordescs*surgery.

Recoding demographic data

RECODE airqual (.69thru .82=1) (.83thru 1.31=2) (1.32thru 1.73=3)INTO airqualcd.

VARIABLE LABELS airqual 'airqualcd'.

EXECUTE.

RECODE poorhousing (.07thru .15=1) (.16thru .36=2) (.37thru .65=3)INTO porhouscd.

VARIABLE LABELS poorhousing 'porhouscd'.

EXECUTE.

K Means Cluster analysis

RECODE accesstoGP15percent (.07thru .15=1) (.16thru .36=2) (.37thru .65=3)INTO porhouscd.

VARIABLE LABELS poorhousing 'porhouscd'.

EXECUTE.

```
DATASET ACTIVATE DataSet2.

QUICK CLUSTER Zpopold ZIMD2010 Zgpacc30 Zsmoking Zalcohol

  /MISSING=LISTWISE

  /CRITERIA=CLUSTER(5) MXITER(100) CONVERGE(0)

  /METHOD=KMEANS(NOUPDATE)

  /SAVE CLUSTER DISTANCE

  /PRINT INITIAL CLUSTER DISTAN.
```

Appendix 7 – Publications and presentations to date from this research

Systematic literature review – GOC – second and third authors undertook duties limited to initial screening and editing of final presented document.

Accepted: 9 March 2017
DOI: 10.1111/jan.13308

REVIEW PAPER

WILEY JAN

Which interval is most crucial to presentation and survival in gastroesophageal cancer: A systematic review

Amanda Lee[1] | Sam Khulusi[2] | Roger Watson[1]

[1]Faculty of Health Sciences, University of Hull, Hull, UK
[2]Queens Centre for Oncology and Haematology, Cottingham, Hull, UK

Correspondence
Amanda Lee, Faculty of Health Sciences, University of Hull, Hull, UK.
Email: a.j.lee@hull.ac.uk

Abstract

Aim: The aim of this study was to identify the most crucial interval to encourage earlier diagnosis in with gastroesophageal cancer and to identify potential factors effecting this interval.

Background: Gastroesophageal malignancy is the eighth most commonly presenting cancer with one of the worst survival rates. Identifying the most crucial period for intervention to inform earlier diagnosis is an important step towards improving survival.

Design: Mixed methods literature review.

Data Sources: CINAHL, MEDLINE and Academic search primer online databases were searched using keywords and inclusion/exclusion criteria. Empirical evidence published between 2000–2016 with a focus on gastroesophageal cancer presentation and survival was reviewed to inform this study.

Review methods: Twelve studies were extracted for further review. Selected studies were appraised and presented through Olensen's "delay interval" framework to inform the most crucial interval to survival in gastroesophageal cancer.

Results: The findings identify the patient interval as the most critical period for encouraging earlier presentation and reducing advanced stage presentation in gastroesophageal cancer. The article also highlighted some methodological limitations to cancer research, such as a lack of consensus in definitions which prevent statistical meta-analysis of cancer data, survivor bias in gastroesophageal cancer studies and a significant lack of qualitative evidence to reveal patient experience in presenting with this cancer.

Conclusion: Further research into the patient interval is required to elicit information on how and why patients present with their cancer symptoms.

KEYWORDS
advanced presentation, cancer, cancer data, cancer methodology, early presentation, GOC, interval delay, late presentation, literature review methodology, oesophagogastric cancer

1 | INTRODUCTION

Gastroesophageal cancer (GOC) is a common cancer with a high morbidity and mortality rate. This systematic literature review of empirical evidence presents a narrative synthesis of evidence to identify the most crucial stage to encouraging earlier diagnosis. It also discusses the methodological limitations in cancer research. The findings are relevant to any cancer because earlier diagnosis is essential to survival. Knowing which interval is a key in improved survival can guide further research and a more targeted clinical intervention.

Access via **http://onlinelibrary.wiley.com/doi/10.1111/jan.13308/full**

Presentation to the 6th Pan Pacific Nursing Conference, Hong Kong.

Friday, 11 March 2016

Hong Kong conference report by Amanda Lee

Earlier this month, I attended the 6th Pan- Pacific Nursing Conference and Colloquium on Chronic Illness Care at the Chinese University of Hong Kong. As one of the key nursing conferences in South-East Asia, it attracts academics from across the globe.

Prior to the conference I visited the Hong Kong Polytechnic University (HKPU) – the largest nurse education provider in Hong Kong – to meet with Professor Alex Molasiotis, the Head of the School. Professor Molasiotis is one of the most cited cancer nursing researchers in the world. We discussed the potential for collaborations linked to my research on Gastro-oesophageal cancer. I also met with Professor Loke from HKPU and Professor Pat Davidson, a founder of the advanced practice nursing movement in America, Dean of Johns Hopkins University.

As my research involves mapping incidence of cancer against socioenvironmental factors, I visited the WHO Collaborating Centre for Community Health Sciences at HKPU to meet with Professor Chen (Director of Research at HKPU). This was organised to explore potential research collaborations around cancer care in Hong Kong.

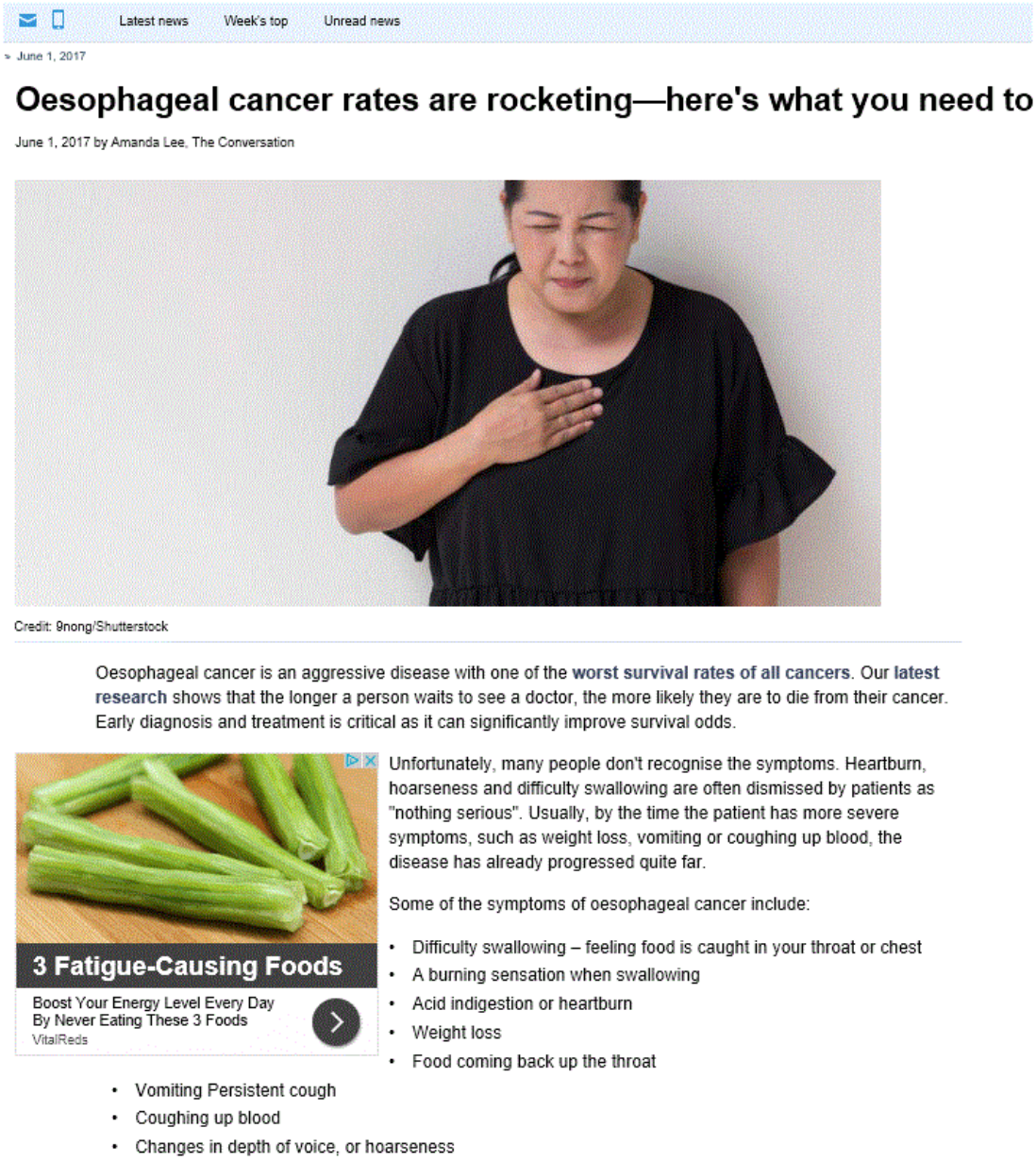
MY visit to HKPU coincided with the HKPU 6 yearly academic review and I was delighted to be invited to a working lunch with Professor Molasiotis and the Deans of Nursing from Johns Hopkins University, University of Manchester, National University of Singapore and Professor Violette Lopez, Director of Research at the National University of Singapore.

At the conference I presented an oral paper 'Using geographical research techniques to inform gastroesophageal cancer prevention: A study to predict areas of high incidence of advanced presentation', which enabled me to make links and network with other researchers in related fields. I was also able to meet Professor Yuli Zang, Deputy Dean at the Shandong University, School of Nursing to discuss research synergies and a potential visit next year. Professor Chen from HKPU attended my presentation and we had some interesting discussions on the application of my PhD research methodology on cancer patterns in Hong Kong. Professor Chen is a renowned researcher and grant holder for a number of projects across Australasia.

Posted by University of Hull Faculty of Health Sciences at 03:06

Labels: amanda lee, Amanda Sherratt, conference, Conference Presentation, hong kong

Conversation piece (received over 10000 hits in the first month after publication and shared via several sites – see below)



Latest news Week's top Unread news

» June 1, 2017

Oesophageal cancer rates are rocketing—here's what you need to

June 1, 2017 by Amanda Lee, The Conversation

Credit: 9nong/Shutterstock

Oesophageal cancer is an aggressive disease with one of the **worst survival rates of all cancers**. Our **latest research** shows that the longer a person waits to see a doctor, the more likely they are to die from their cancer. Early diagnosis and treatment is critical as it can significantly improve survival odds.

Unfortunately, many people don't recognise the symptoms. Heartburn, hoarseness and difficulty swallowing are often dismissed by patients as "nothing serious". Usually, by the time the patient has more severe symptoms, such as weight loss, vomiting or coughing up blood, the disease has already progressed quite far.

Some of the symptoms of oesophageal cancer include:

- Difficulty swallowing – feeling food is caught in your throat or chest
- A burning sensation when swallowing
- Acid indigestion or heartburn
- Weight loss
- Food coming back up the throat
- Vomiting Persistent cough
- Coughing up blood
- Changes in depth of voice, or hoarseness

Available at **http://onlinelibrary.wiley.com/doi/10.1111/jan.13308/full**

Internet sites sharing the conversation piece –

Oesophageal cancer rates are rocketing - Yahoo News UK

https://uk.news.yahoo.com/oesophageal-cancer-rates-rocketing-apos...

Oesophageal cancer rates are rocketing - Deviant World

https://www.deviantworld.com/.../oesophageal-cancer-rates-rocketing

Oesophageal cancer rates are rocketing - The UK Bulletin

theukbulletin.com/2017/06/01/oesophageal-cancer-rates-are...

Oesophageal cancer rates are rocketing—medicalxpress.com › Cancer

Oesophageal cancer rates are rocketing - econotimes.com

www.econotimes.com/Oesophageal-cancer-rates-are-rocketing-heres...

Oesophageal Cancer Rates Are Rocketing - Before It's News

beforeitsnews.com › Economy

Esophageal cancer rates are rocketing - LocalHealthGuide

localhealthguide.com/2017/06/01/esophageal-cancer-rates-rocketing...