Spatial and social variations in colorectal cancer within England and Wales

Charlotte Elizabeth Sturley

Submitted in accordance with the requirements for the degree of PhD

The University of Leeds

School of Medicine

&

School of Geography

November 2021

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would like to express my sincere thanks to my supervisors Amy Downing, Michelle Morris and Paul Norman for their guidance and encouragement throughout. The patience and support they have shown is greatly appreciated, as is their knowledge and expertise, from which I have learnt so much. I feel extremely fortunate to have worked with you all. I am also grateful to Richard Feltbower and Graham Clarke for their input as transfer viva assessors and for providing advice and constructive feedback on my work.

Thank you to the Cancer Epidemiology Group for welcoming me into your team and for your patience and time in answering my many questions. Particular thanks to Rebecca Birch for the advice and encouragement with the analysis. I also appreciate the support of members of the Centre for Spatial Analysis and Policy research group.

My thanks to colleagues at the GeoHealth Laboratory at the University of Canterbury for hosting us during our research visit to New Zealand, which was a highlight of my PhD experience.

This studentship was supported by Cancer Research UK (C23434/A24939), which underpinned data access via the UK Colorectal Cancer Intelligence Hub. This project involves data that have been provided by, or derived from, patients and collected by the NHS as part of their care and support.

The permission of the Office for National Statistics to use the Longitudinal Study is gratefully acknowledged, as is the help provided by staff of the Centre for Longitudinal Study Information & User Support (CeLSIUS), particularly Aly Sizer and Rachel Stuchbury. CeLSIUS is supported by the ESRC under project ES/V003488/1.The authors alone are responsible for the interpretation of the data.

This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

There are many people who have contributed to my PhD experience and I would like to thank them all for their support and encouragement.

The friendship and support of my fellow students, in particular Amanda Otley, Fran Pontin, Rachel Oldroyd and Victoria Jenneson, has been invaluable. The ‘Shut Up and Write’ sessions and tea breaks have kept me going, particularly over the past 18 months.

I would like to thank my parents for their love, support and encouragement in everything I do. Thanks also to my brother, sister-in-law and nieces for always lifting my spirits and providing a welcome distraction from my PhD when needed. Finally, thank you to Richard for always believing in me and keeping me going during the final few weeks. I could not have done it without your support.

Abstract

Each year in the UK, more than 42,000 cases of colorectal cancer (CRC) are diagnosed and there are over 16,000 deaths. Despite efforts to reduce inequalities in cancer outcomes, disparities have persisted. This thesis investigates spatial and social variations in CRC incidence, survival and mortality in England and Wales.

A range of data sources were used to obtain information about CRC patients: traditional mortality records; data from the Office for National Statistics Longitudinal Study (LS); and the COloRECTal cancer data Repository (CORECT-R), a contemporary database of CRC data.

Standardised CRC mortality rates were calculated by Local Authority (LA) and area deprivation over a 20-year period, and compared to those for all cancers combined. Time-to-event analysis examined associations between individual-level indicators of socio-economic status and CRC incidence and survival in the LS. Small-area data linkage to CORECT-R enabled investigation of associations between local area characteristics and CRC incidence, using a bespoke risk index.

Spatial variation in colorectal cancer mortality was observed, but there was no clear pattern by LA or area deprivation, in contrast to that for all cancers combined. A stronger association was found between individual socio-economic attributes (educational attainment, social class and housing tenure) and both CRC incidence and survival. The CRC risk index revealed greater distance to health services and green space was associated with worse CRC outcomes and surprisingly closer proximity to retail outlets was associated with better CRC outcomes. No associations of the risk index and stage at diagnosis were found.

Understanding spatial and social variations in CRC is important to inform policies and target interventions to reduce inequalities. Access to up-to-date data is essential to monitor current health outcomes, while combining more traditional datasets with novel data offers potential to examine these relationships in more detail.

Table of contents

Table of figures

Table of tables

Abbreviations

ASR	Age-standardised rate
BCSP	Bowel Cancer Screening Programme
CCG	Clinical Commissioning Group
CDRC	Consumer Data Research Centre
CeLSIUS	Centre for Longitudinal Study Information and User Support
CI	Confidence Interval
ESP	European Standard Population
HES	Hospital Episode Statistics
HR	Hazard Ratio
IMD	Index of Multiple Deprivation
IRR	Incidence Rate Ratio
LA	Local Authority
LS	Longitudinal Study
LSOA	Lower Layer Super Output Area
MSOA	Middle Layer Super Output Area
NHS	National Health Service
NHSIC	National Health Service Information Centre
OA	Output Area
ONS	Office for National Statistics
SES	Socio-economic status
TRE	Trusted Research Environment
UKDS	UK Data Service
WCRF	World Cancer Research Fund
WHO	World Health Organisation

Chapter 1
Introduction

1.1 Overview

This chapter sets out the research problem that the thesis seeks to address and provides some background information. The conceptual framework that will be used throughout the thesis is introduced. The research aims and objectives are described and the overall structure of the thesis is outlined.

1.2 Colorectal cancer

Colorectal cancer refers to cancer that starts in the colon (large bowel) or rectum. It is the fourth most common cancer in the UK but the second most common cause of cancer death. Around 41,000 people are diagnosed with colorectal cancer each year in the UK and there are around 16,000 colorectal cancer deaths (Cancer Research UK, 2018). It therefore represents a major public health concern.

Colorectal cancer incidence is strongly linked to age, with higher incidence rates in older people. Over ninety percent of cases in England diagnosed in people aged over 50 (Office for National Statistics, 2018). Colorectal cancer is also more common in men than women: fifty-five per cent of colorectal cancer cases in the UK occur in males (Cancer Research UK, 2018).

Overall, seventy-eight per cent of people diagnosed with colorectal cancer in England survive for one year or more. This falls to 58% of people surviving five years or more (Cancer Research UK, 2018). Early diagnosis is key to colorectal cancer survival. If the cancer is detected early, treatment is more likely to be successful and there is a better chance of survival. When diagnosed at the earliest stage (Stage 1), nearly all (98%) people with colorectal cancer will survive for at least one year. This figure is much lower (44%) when the disease is detected at the most advanced stage (Stage 4) (Office for National Statistics, 2016). Currently, over half of colorectal cancer cases are diagnosed at a late stage (Stage 3 or 4) (National Cancer Registration and Analysis Service, 2016), so improving early detection is a priority. Cancer is more likely to be detected at an early stage through screening or via an urgent (2 week wait) GP referral when a patient presents with symptoms. It is more likely to be diagnosed at a

late stage when it is diagnosed as a result of an emergency presentation and 23% of colorectal cancer cases were detected via this route in 2018 (CRUK Early Diagnosis Tool).

Surgery is the most common treatment for colorectal cancer. Sixty-six per cent of colon cancer patients and sixty-three per cent of rectal cancer patients have surgery to remove the tumour as part of their treatment (NCRAS, 2017). Radiotherapy is also commonly used to treat rectal cancer patients, either alone or together with surgery or chemotherapy, but it not usually used to treat colon cancer. Both colon and rectal cancer patients receive chemotherapy as part of their treatment.

Survival for colorectal cancer has improved over recent decades, however, there is concern that outcomes from the disease in England lag behind many other comparable countries. The International Cancer Benchmarking Partnership reported persistently lower colorectal cancer survival in England, compared to Australia, Canada, Norway and Sweden (Coleman et al., 2011), particularly in the first year after diagnosis. Later stage at diagnosis or differences in diagnostic practices and treatment for colorectal cancer were given as possible reasons for the disparities in survival rates.

Significant efforts have been made to improve cancer outcomes across the NHS (Department of Health, 2007) and consequently reduce these disparities. Major initiatives have been introduced over the past 20 years aimed at increasing survival from colorectal cancer. Specialist multidisciplinary teams have been established to improve the quality of care delivered in NHS hospitals. The Bowel Cancer Screening Programme was introduced in 2006, aiming to detect cancer at an early stage of the disease when treatment is more effective and therefore reduce mortality among those invited for screening. It is currently offered to all adults in England aged 60-74 but the age at which screening is offered is to be lowered to start at age 50 (UK National Screening Committee, 2018). Uptake of bowel cancer screening is below that of for breast and cervical screening at 54% overall, and there is considerable variation between areas and demographic groups.

It is estimated that over half of colorectal cancer cases in the UK are linked to lifestyle risk factors (Brown et al. 2018). These include consumption of red and processed meat, being overweight or obese, alcohol consumption and smoking.

Conversely, physical activity and fibre consumption may protect against colorectal cancer (International Agency for Research on Cancer and World Cancer Research Fund). Changing population-level exposure to modifiable risk factors is a key driver of changing cancer incidence (Brown et al. 2018). Furthermore, prevalence of exposure to risk factors varies by geographic area and socio-demographic group.

In addition to substantial morbidity and mortality as a result of colorectal cancer, it is estimated that detecting and managing the disease costs the National Health Service (NHS) in excess of £1.1 billion annually (Incisive Health and Cancer Research UK, 2015; Laudicella et al., 2016). Despite this, major variations in diagnosis, treatment and outcomes remain. Therefore, more research is required to inform policies to address this.

1.3 Determinants of health

Inequalities in health exist between subgroups of the population and from one place to another, at national, regional and local level. In order to understand why such inequalities occur, there is a need to understand the factors that determine good and bad health.

It is well established that there is a social gradient in health: the lower and individual's social position, the worse their health is likely to be (Marmot, 2010). This has led to a growing awareness of social factors (as opposed to biological or genetic factors) in determining health. Economic, environmental and social inequalities can affect people's risk of contracting disease, their ability to prevent serious illness and their access to timely and effective treatments.

There is also an ongoing debate about the role of individual and area effects on health outcomes (Roux, 2001; van Ham and Manley, 2012). While geographical variations in health have been examined historically, the importance given to the study of area differences when investigating the causes of disease has varied over time. The focus on individual-level risk factors for disease in epidemiology over the past few decades led to less interest in the role of area attributes as possible causes of disease. The concept of place as an important determinant of health remerged in the 1980s and 1990s (Carstairs and Morris, 1989; Humphreys and Carr-Hill, 1991) and since then there has been a marked increase in interest. A large body of literature examining neighbourhood effects

has been published in epidemiology and Public Health journals, across a range of health outcomes and health behaviours, including cardiovascular disease (Smith et al., 1998), self-reported illness (Shouls et al., 1996), smoking behaviour (Duncan et al., 1999) and mortality (Sloggett and Joshi, 1994).

The upsurge in interest in area effects coincided with an increased interest in the social determinants of health. The type of area in which people live is one way in which social factors can influence health. At the same time there was an emerging debate about the use of ecological variables in epidemiology (Diez-Roux, 1998) and the notion that neighbourhood context may have an effect on a person's health independently of their individual attributes. This led to the emergence of new methodological approaches, such as multi-level modelling, to try and understand the role of context and composition in area variation in health outcomes (Duncan et al., 1998).

Despite the volume of research on neighbourhood effects, enough is still not known about the causal mechanisms which produce them and how important they are in shaping an individual's life chances (van Ham et al., 2012). A number of conceptual and methodological challenges remain in order to better understand neighbourhood effects on health outcomes (Roux, 2001; van Ham and Manley, 2012). These include how to better define neighbourhoods, how to conceptualise the complex relationship between neighbourhood characteristics and individual-level socio-economic position and understanding the scale and temporality of neighbourhood exposures. Furthermore, incorporating life course and longitudinal dimensions in work on health and place is an emerging research area (Pearce, 2018). Researchers must also overcome data availability issues (Jivraj et al., 2020). Finally, combining quantitative and qualitative approaches into one research design may enable better understanding of the processes that underlie the causal mechanisms (van Ham and Manley, 2012).

Several conceptual models have been developed to illustrate the social and environmental factors which contribute to health, for example social ecological models (McLeroy et al., 1988), the Health Belief Model (Janz and Becker, 1984) and the triangle of human ecology (Meade and Earickson, 2000). Perhaps the most widely used of these is Dahlgren and Whitehead's 'rainbow' model of health determinants which maps the relationship between the individual, their

environment and health (Dahlgren and Whitehead, 1991). Individuals are placed at the centre of the model along with personal characteristics such as sex, age, ethnic group and hereditary factors. Surrounding them are the various layers that influence health (Figure 1.1). The first layer comprises individual lifestyle factors such as smoking, alcohol use and physical activity. This is followed by social and community influences which include family and wider social circles. The next layer encompasses living and working conditions and access to essential goods and services. The outermost layer includes macroeconomic, cultural and environmental conditions.

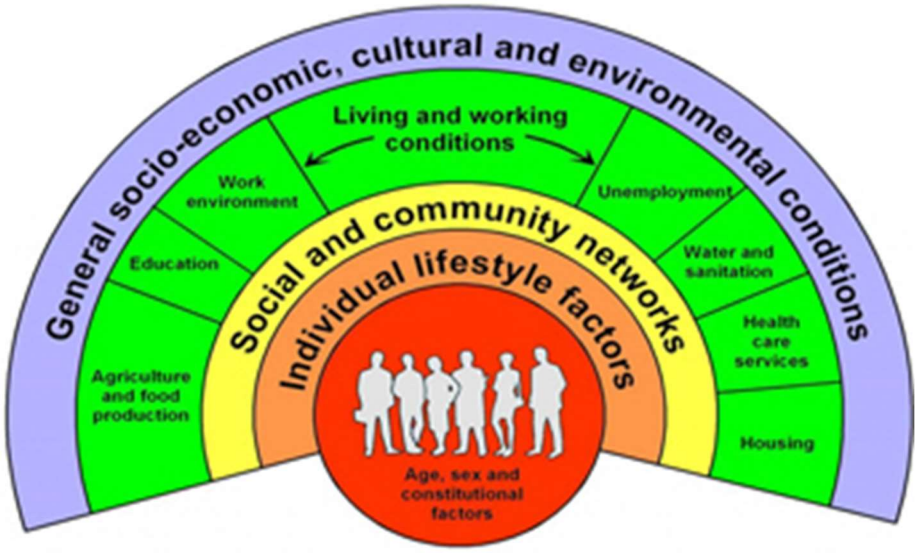


Figure 1.1 Dahlgren and Whitehead model of health determinants

Social models of health determinants such as these are relevant to this PhD as it is concerned with investigating social and spatial variation in colorectal cancer risk and outcomes. Of particular interest is how geographic location and environment contribute to inequalities in colorectal cancer risk and outcomes both independently or via social factors. Better understanding of how geographic factors interact with socio-economic position and lifestyle factors could help to explain the mechanisms behind these disparities and thereafter be used for appropriate targeting for policy whether aimed at individuals or at areas; whether specific places or types of spaces.

The core concepts and elements of the Dahlgren and Whitehead model will be used to inform this PhD.

1.4 Framework

Figure 1.2 shows an adapted version of the Dahlgren and Whitehead model, relevant to this PhD. The model has been simplified to 3 layers: the individual, lifestyle factors and socio-economic conditions. An additional spatial component has been added, intersecting the lifestyle and socio-economic factors. This model will be used as a framework throughout the PhD.

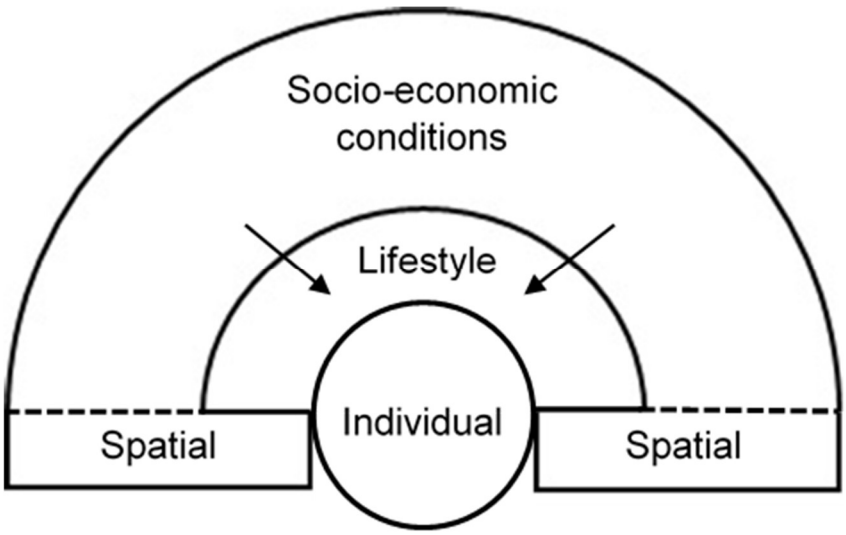


Figure 1.2 Thesis framework

The spatial component encompasses factors relating to the geographic location (environment context) of a place. This can be considered at various different scales, such as nationally, regionally and locally. Traditionally, census geographies in England have been used in research to define area boundaries. Spatial factors also include differences between urban and rural areas, which could influence access to health care services and health-related behaviours due to aspects of the physical environment. Proximity to features of the built and physical environment which may influence health-related behaviours will also be considered.

While the spatial component relates to physical location, the social component relates to the demographic characteristics and the composition of the area in which people live. This includes both individual-level socio-economic position, measured by indicators such as educational attainment, unemployment,

occupation and housing and area-level indicators, such as area deprivation. In many cases area-level indicators are used as a proxy for individual-level measures where these are not available, however, area characteristics may influence health regardless of individual socio-economic position.

Social and spatial factors are closely intertwined. By separating out the spatial component, its influence on health can be considered both independently and via the social composition of areas.

Lifestyle factors and behaviours that influence health, and specifically, risk of colorectal cancer will be considered. Prevalence of these risk factors varies both socially and spatially. While not a primary focus of this thesis, individual characteristics at the centre of the model, such as age, gender, ethnicity and hereditary factors are important determinants of colorectal cancer risk and outcomes.

1.5 Available data

In order to investigate the social and spatial variations in colorectal cancer incidence and outcomes a range of social and spatial information over time is required.

A national census is carried out every ten years in England and Wales by the Office for National Statistics (ONS). Detailed socio-demographic information is collected for all people and households resident in the country at the time of the census. These data are then collated and disseminated to users at a range of census geographies.

Annual mortality statistics are also collated by the ONS. These data contain all deaths in England and Wales by sex, five-year age group and cause of death for a selection of geographies.

National cancer registries collect information about cancer diagnoses in the UK. The cancer registry contains detailed information about cancer patients diagnosis, treatment and outcomes, however, only limited demographic information is collected which does not include details about an individual's socio-economic position. Most research using cancer registry data uses area-based measures of SES, either census-based (Carstairs or Townsend Index) or the Index of Multiple Deprivation (IMD).

Longitudinal studies provide a means of linking socio-demographic and health information over time for a sample of the population. There are a number of national longitudinal cohort studies in the UK that are widely used for health research (Centre for Longitudinal Studies, 2021).

No single dataset can provide all information required to meet the objectives of this thesis so a range of datasets will be used which will be outlined in more detail in Chapter 3, section 3.2.

1.6 Aims and objectives

Aim:

To investigate spatial and social variations in colorectal cancer incidence, mortality and survival in England and Wales.

Objectives:

Objective	Chapter
Review literature relating to colorectal cancer incidence, survival and mortality by geographic location and area deprivation.	Chapter 2, 4, 5, 6
Describe data sets and appropriate spatial and statistical methods.	Chapter 3
Describe and critique trends in colorectal cancer mortality in England and Wales at Local Authority level and by area-based socio-economic deprivation.	Chapter 4
Compare colorectal cancer mortality trends to all cancers combined.	Chapter 4
Identify whether there are statistically significant hot spots of colorectal cancer and all-cancer mortality rates by geographic area and over time.	Chapter 4
Describe the demographic and socio-economic attributes of Longitudinal Study members with a diagnoses of colorectal cancer.	Chapter 5

Investigate the association between colorectal cancer incidence and individual and area-based indicators of socio-economic status using time-to-event analysis.	Chapter 5
Investigate the association between all-cause and cause-specific survival and individual and area-based indicators of socio-economic status over a 15-year follow-up period among people with a colorectal cancer diagnosis, using time-to-event analysis.	Chapter 5
Describe individuals with a colorectal cancer diagnosis by individual attributes (age, sex, ethnicity) and characteristics of the local area in which they reside (area deprivation, population density).	Chapter 6
Develop an area-based index of the types of places with higher risk of colorectal cancer incidence, encompassing three domains: the retail environment, health services, and the natural environment.	Chapter 6
Model the relationship between colorectal cancer incidence and the colorectal cancer risk index, adjusting for demographic and socio-economic characteristics of areas.	Chapter 6
Estimate the odds of late stage diagnosis by the colorectal cancer risk index.	Chapter 6
Critique results in context of wider literature	Chapter 7
Make recommendations for future research	Chapter 7

Table 1.1 Research objectives

1.7 Funding and interdisciplinarity

This PhD project is funded by Cancer Research UK (CRUK) as part of the Colorectal Cancer Intelligence Hub research programme. The programme is a collaboration between the University of Oxford, the University of Leeds, the University of Edinburgh, the Bowel Cancer Intelligence UK Patient-Public Group and Health Data Insight. It was set up to address the challenges faced by researchers in gaining access to patient data which is vital for generating high-quality cancer intelligence. A single colorectal cancer researcher data system has been created, known as the COloRECTal cancer data Repository (CORECT-R). This novel data resource contains detailed diagnostic, management and outcome information about all individuals in England diagnosed with, or at risk of developing, colorectal cancer.

The PhD topic is an interdisciplinary subject area, encompassing Cancer Epidemiology and Geography. Interdisciplinary research is important to try to address public health concerns. In the case of this PhD, an interdisciplinary approach is needed to better understand the determinants of health at all levels and how these influence cancer risk and outcomes, specifically colorectal cancer. In particular, it is concerned with how cancer risk and outcomes vary by geographic location and socio-economic characteristics. Geographic methods and spatial analysis techniques will be applied in a health setting to try and address these questions.

The results from this thesis will provide an updated picture of social and spatial variations in colorectal cancer incidence, survival and mortality. Reducing socio-economic and geographic inequalities in cancer outcomes is a public health priority (The NHS Cancer Plan (2000), Cancer Reform Strategy (2007)). To achieve this a better understanding of cancer risk and outcomes at smaller geographic units is needed in order to develop policy and provide targeted public health initiatives. This thesis aims to address this gap in colorectal cancer research.

1.8 Thesis structure

The thesis comprises an introduction, literature review, data and methods chapter, three results chapters and a discussion. The structure is summarised in Figure 1.3. Due to the broad topic area, there is a large body of literature, therefore literature relating to each of the analysis chapters will be discussed at the start of each of those chapters.

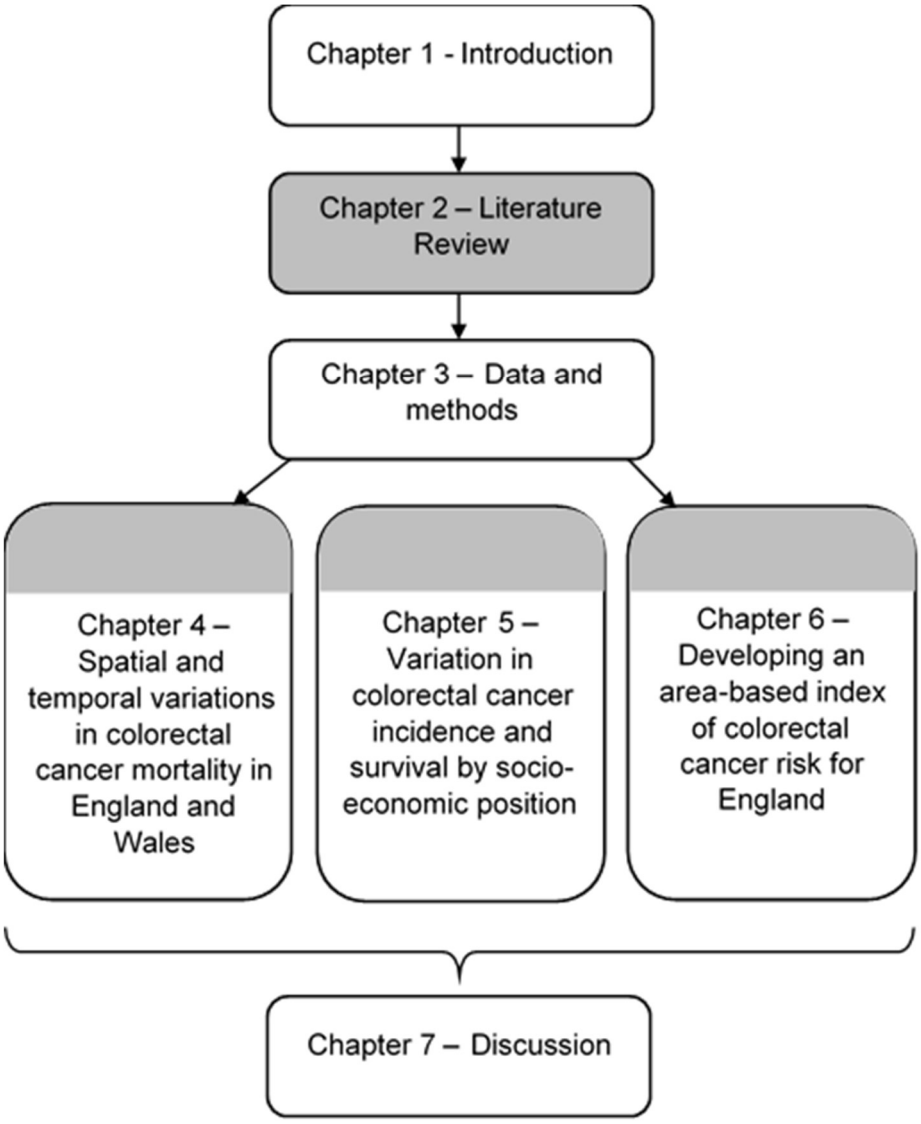


Figure 1.3 Thesis structure

Chapter 2
Literature Review

2.1 Introduction

This chapter will set the thesis in the context of the existing literature. The framework introduced in Chapter 1.4 will be used to structure the review. Spatial and social variations in health will be explored and how they interact with one another and health behaviours to influence health outcomes. For each of the elements (spatial, social, lifestyle), firstly variations in general health and mortality will be considered, then variation in all cancers, and finally variations in colorectal cancer specifically. The literature relevant to the analysis chapters will be discussed in more detail at the start of each of these chapters.

2.2 Methods

A narrative literature review was undertaken to analyse and critique literature within the topic area of interest. An initial literature search was carried out, where Medline and Web of Science databases were searched using the search terms listed in Table 2.1. These search terms were combined using Boolean operators. Papers not written in the English language were excluded as were those published before the year 2000, to focus on recent research findings. Titles and abstracts of search results were reviewed and the relevant papers were imported into Mendeley.

Following the initial literature search a number of key papers and authors were identified. A snowball strategy was adopted, using forward and backward citation searching, to find additional relevant papers. Here, some older studies were included (prior to 2000) to understand how trends have changed over time. Search alerts were set up to keep informed of new publications in the topic area. The review focuses on research in the UK which is most relevant to the setting of this thesis, but international studies in other developed countries are also considered.

Search component	Search terms
General health	health* OR mortality
Cancer	cancer* OR carcinoma* colo?rect* OR colon OR rectal OR rectum OR bowel*
Social	socio?economic* OR socio?demograph* OR geo?demograph* OR social* depriv* OR social* inequalit* OR material* depriv*
Spatial	spatial* analy* OR spatial variation* OR area-level
Outcome	incidence OR mortality OR survival

Table 2.1 Literature review search terms

2.3 Results

There are variations in health by geographic area and area type. Such area characteristics may relate to the composition of an area, for example a deprivation index or geodemographic classification, or the context of the area, such as whether it is in an urban or rural setting (Macintyre et al., 2002). These disparities in health are observed for general health and mortality, and for most cancers. Variation in colorectal cancer by geography and area type is less clear and will be explored throughout this thesis.

2.3.1 Spatial variation in general health

Regional variations in health have been observed in the UK. Geographical inequality in health between the North and South of England has persisted for several decades: overall health is better in the South than the North. This regional health divided has been witnessed across a range of health outcomes. For example, life expectancy has been consistently lower in the North of England than the South (Marmot et al., 2020). There is a gap of eight years in life expectancy for males born in 2016-2018 between those born in Blackpool (North) and Richmond-upon-Thames (South). Healthy life expectancy, the number of years that people can expect to live in good health, is 53.3 years for men born in Blackpool, compared to 71.9 years for those born in Richmond-upon-Thames, a gap of 18.6 years. In an analysis of 18 health indicators across 9 Government Office Regions in England, Ellis and Fry (2010) found that Northern regions generally do worse than the Midlands and London, particularly on indicators such as lung cancer incidence and deaths from all cancer, respiratory diseases and circulatory diseases. They did, however, note some exceptions to this trend such as high levels of childhood obesity in London, a high proportion of drug use in the South East and high levels of breast cancer in the South West. This suggests that geographical patterns in health vary by health indicator. A limitation of the study was that a lot of the indicators were based on sample surveys with small regional samples which highlights the difficulty of analysing health outcomes by geographical area as small numbers can be an issue. More recently, regional inequalities in the health impacts of Covid-19 have been reported, with higher rates of mortality attributable to Covid-19 in the North of England, compared to the rest of the country (Munford et al., 2021).

Inequalities in health have been observed at different spatial scales. In an analysis of premature mortality in Great Britain, Leyland (2004) found persistent differences in mortality between regions but also disparities between districts within regions and that these vary from one region to another. Similarly, in an analysis of mortality by cause of death in countries and cities of the UK, Walsh et al. (2020) identified particular issues in the city-level analysis, such as high mortality rates in Glasgow for all-cause mortality and the majority of causes of death, especially alcohol-related causes. This suggests that regional analysis could mask patterns at sub-regional level and highlights the need for more detailed spatial analysis. The authors noted that trends at city-level can be difficult to interpret given fluctuations in mortality rates, which is an important consideration when conducting spatial analysis at smaller geographical units.

Variations in health have also been observed between rural and urban areas. Based on self-reported health, a positive urban-rural health gradient has been found, with individuals in urban areas reporting the worst health and those in rural areas the best (Allan et al., 2017). A protective “capital city” effect was also noted, whereby residents in London were found to have better than anticipated health. This suggests that there are other factors that impact the health of individuals residing in some cities.

Where we live (residential context) can also influence health-related behaviours and in turn health status. Aspects of the local environment can be health promoting or health damaging. Studies have investigated the influence of features of both the natural and built environment on health. For example, research has shown a positive relationship between the number of fast food outlets in an area and rates of childhood obesity (Fraser and Edwards, 2010). Other studies have shown the density of alcohol-related outlets to be associated with alcohol-related harms (Sherk et al., 2018) and tobacco outlets with risk of smoking (Shortt et al., 2016). On the other hand, accessibility to green space has been shown to be positively associated with physical and mental wellbeing (Mitchell and Popham, 2008). This will be discussed in more detail in Chapter 6, which investigates the association between residential context and colorectal cancer risk.

2.3.2 Spatial variation in cancer

Variations in cancer incidence and survival have been reported at regional level in England. Arik et al. (2020) estimated cancer incidence rates by age, gender, year and region for all-cancer and four common cancer types between 1981 and 2016 using Bayesian modelling. The highest all-cancer incidence rates occurred in the North East for both genders while the lowest rates were mostly observed in London for females and the east for males. Regional patterns in incidence vary by cancer type. Based on data for 2016, Arik at al. (2020) found higher lung cancer incidence in northern regions of England, compared to southern regions. Bowel cancer incidence rates for males were highest in the north east and north west. In contrast, the highest incidence rates for females were in the south of England. The lowest bowel cancer incidence for males and females occurred in London but there was less regional inequality in bowel cancer incidence compared to all-cancer and lung cancer. Breast and prostate cancer rates were highest in the south of England. Less regional variation was noted for breast and prostate cancer incidence, compared to all-cancer, lung cancer and bowel cancer. Regional variation in cancer incidence appears to be increasing over time. Arik et al. (2020) reported a significant increase in regional variation for most types of cancer, although bowel cancer was a notable exception for which regional differences had narrowed over time. Patterns in cancer incidence at sub-regional level were not examined in this study. This is a limitation because there may be considerable variation between areas within regions which is obscured by regional-level patterns.

There is also evidence of variation in cancer survival by geographic region. As described in the context of general health above, a North-South divide in cancer survival has been observed in England, whereby survival is generally lower in northern England, compared to the south. Walters et al. (2011) reported wide geographical variation in cancer survival by cancer network in England. Cancer networks were an NHS regional geography consisting of 28 geographic areas (they were replaced by Cancer Alliances in 2015). Survival estimates were consistently low in northern England and the east Midlands, compared to England overall. While they found that the north-south divide became less pronounced over time, geographical inequalities in survival persisted over time. Breast cancer was a notable exception, for which there was a marked reduction

in geographical inequality in survival. In a more recent analysis using smaller geographic units, there was less geographic inequality in cancer survival by Clinical Commissioning Group (CCG) in England in 2017, compared to 2002 (Public Health England, 2019). There was also less variation in CCG-level survival estimates for breast cancer than in colorectal and lung cancer survival.

Regional differences in cancer survival vary by cancer type. In a recent report, the ONS reported geographical patterns in cancer survival for 3 levels of NHS geographies in England: 7 NHS regions, 21 Cancer Alliances (CAs) and 37 Sustainability and Transformation Partnerships (STPs) (Public Health England, 2020). Across the 21 CAs, the variation in 1-year survival estimates was most narrow for breast cancer, with a range of 1.5 percentage points between the lowest and highest, and widest for brain cancer, for which the range was 13.1 percentage points. Comparing the STPs, the difference between the lowest and highest 1-year survival estimates varied from 2.5 percentage points for breast cancer and melanoma to 17.9 percentage points for ovarian cancer. There was less variation in survival estimates by NHS region compared to CAs and STPs, due to larger, more heterogeneous populations. For both 1-year and 5-year survival, the North East and Yorkshire, North West, Midlands and East of England tended to have lower estimates and conversely, the South East, South West and London tended to have higher estimates.

Spatial variations in cancer risk and outcomes can also be considered in terms of differences between rural and urban areas. Cancer incidence rates and mortality rates, adjusted for population age distribution, vary between rural and urban areas. Rural areas have lower incidence rates of lung cancer and higher incidence of breast, prostate and female colorectal cancers (National Cancer Intelligence Network, 2011). The report found that the variation in cancer incidence and mortality rates by rurality was partly due to the variation in socio-economic status but even when this is taken into account some significant differences remained. The analysis was based on the ONS rural-urban categories and the Index of Multiple Deprivation income domain which were assigned at small-area (LSOA) level. A different study, using similar data, concluded that differences in all-cancer mortality between rural and urban areas were largely accounted for by deprivation (Gartner et al., 2011).

There are also differences in cancer survival between rural and urban areas. Whether a patient lives in a rural or urban area can influence their access to health care services. Distance to health care services can impact on a patients cancer stage at diagnosis, treatment type, survival and quality of life (Ambroggi et al., 2015). Survival for patients living in rural areas in Scotland was poorer for lung cancer and colorectal cancer, compared to those living in towns and cities with a cancer centre (Campbell et al., 2000). In a subsequent study, the authors found that patients in rural areas had more advanced disease at diagnosis (Campbell et al., 2001), which suggests advanced stage at diagnosis may explain the poorer survival among patients in these areas. What is meant by an urban or rural area and how rurality is measured may vary in different countries. For example, Scotland is much less densely populated than England so findings may not be generalizable to other areas. In a study in northern England, cancer stage at diagnosis was found to be associated with increased travel time to general practitioners for colorectal and breast cancers, but travel time to hospital and measures of access to public transport were not associated with stage at diagnosis or survival (Jones et al., 2008b). The impact of rurality on quality of life following a cancer diagnosis has also been investigated. In a study of men 18-42 months post diagnosis of prostate cancer in the UK, the influence of rurality and area deprivation on self-assessed health related quality of life were not greater than what would be expected in the general population (Smith et al., 2020).

2.3.3 Spatial variation in colorectal cancer

Geographical variation in colorectal cancer is less well researched. Age-standardised incidence rates of colorectal cancer are highest in the North East of England and lowest in London (Office for National Statistics, 2019). As noted above, when comparing incidence and survival by cancer type, there is less regional inequality in colorectal cancer incidence, compared to all-cancer and lung cancer but more than breast and prostate cancer. Regional differences in colorectal cancer incidence rates have narrowed over time, in contrast to other cancer types for which regional variation has increased (Arik et al. 2020). Despite a narrowing of geographical inequality in colorectal cancer survival by CCG over time between 2002 and 2017, variations in survival by area persist.

There is still a 17% range between CCGs with the lowest and highest 1-year survival (Public Health England, 2019).

Rurality and access to health care services have been found to influence stage at diagnosis, treatment and survival for colorectal cancer patients. Distance to cancer centre is associated with more advanced disease at diagnosis amongst colorectal cancer patients in Scotland (Campbell et al., 2001). Similarly, advanced colon cancer stage at diagnosis was associated with patient travel distance to hospital in a US study (Massarweh et al., 2014). Furthermore, increasing distance to primary care provider (Parsons and Askland, 2007) and increasing travel time to GP (Jones et al., 2008b) is associated with late stage at diagnosis for colorectal cancer patients.

The likelihood of receiving radiotherapy was reduced for patients with breast, colon, rectal, lung, ovary and prostate cancer with increasing travel time to the nearest radiotherapy centre in northern England (Jones et al., 2008a). In addition, rectal patients were less likely to receive chemotherapy if they lived furthest from these hospitals (Jones et al., 2008a). An increase in driving distance or drive time was associated with a greater time-to-treatment after diagnosis with colorectal cancer in the US (Scoggins et al., 2012). Distance to treatment centre was associated with poorer colon cancer survival in France after adjustment for stage at diagnosis (Dejardin et al., 2008). Among rectal cancer patients in Queensland, those living further from a radiotherapy facility were more likely to die from rectal cancer than those living within 50km of such a facility (Baade et al., 2011). This evidence suggests access to health care services influences receipt of timely treatment across different health care settings.

2.3.4 Social variations in general health

As well as geographical variations in health, inequalities are also observed between social groups. Health status is closely related to people's socio-economic circumstances. These circumstances have been measured in a number of different ways. A common summary measure of socio-economic circumstances across the population is socio-economic deprivation. One such measure is the Index of Multiple Deprivation (IMD), which summarises the degree of deprivation within an area based on a range of factors that includes

levels of income, employment, education and crime. Census-based measures of deprivation are also widely used, for example the Townsend Index of Deprivation (Townsend et al., 1988) and the Carstairs index (Carstairs and Morris, 1991). These combine selected census variables to derive a summary measure of deprivation for each area. These measures reflect the fact that the socio-economic environment is multidimensional, incorporating factors such as income, employment, social class and living conditions. Moreover, each dimension measures a different aspect of deprivation so taken separately they are unlikely to capture the full range of socio-economic disadvantage. Furthermore, life course measures of socio-economic deprivation are increasingly being used to monitor long-term trends in socio-economic position and outcomes such as health (Jivraj et al., 2020).

There is a gradient in health across socio-economic groups within every region in England: poorer health with higher deprivation. This pattern is evident across a range of health outcomes. For example, the difference in life expectancy at birth between the least and most deprived areas in England is 9.4 years for males and 7.6 years for females (Office for National Statistics, 2021b). Moreover, the gradient in healthy life expectancy is steeper than that for life expectancy. People in the least deprived areas can expect to live roughly 19 more years in good health than those in the most deprived areas (Office for National Statistics, 2021b). So not only do people in the most deprived areas have the shortest life expectancy overall, they are also spending more of their lives in ill health than those in less deprived areas. A similar gradient can also be seen for other measures of deprivation, such as level of income and education (The King's Fund, 2015).

The social gradient in health is evident for specific morbidities, such as cardiovascular disease (Mackenbach et al., 2000) and diabetes (Espelt et al., 2008), in self-reported health (Kunst et al., 2005) and in all-cause mortality and specific causes of death (Mackenbach et al., 2008). A socio-economic gradient in Covid-19 mortality has also emerged. The age-standardised mortality rate for deaths involving Covid-19 in the most deprived areas in England between March to June 2020 was more than double the rate in the least deprived areas (Office for National Statistics, 2020). A report by the Northern Health Science Alliance (NHSA) found that around half of the higher Covid-19 mortality

observed in North of England, compared to the rest of the country, and two-thirds of the higher all-cause mortality were explained by greater levels of deprivation and worse pre-pandemic health in the North (Munford et al., 2021).

This social gradient in health can explain some of the geographical variation in health described above. The north of England has a higher concentration of deprived neighbourhoods than the south of England and therefore, a greater proportion of areas where health outcomes are likely to be worse. The relationship between social and spatial inequalities in health is, however, more complex. Poor areas in the north of England tend to have worse health than places with similar levels of deprivation in the rest of England. Furthermore, the social gradient in health is steeper in the north than the rest of England i.e. the gap in health between the most and least deprived groups is greater. For example, life expectancy is nearly five years less for people living in deprived areas in the North East compared to those living in similarly deprived areas in London and the gap in life expectancy between the least and most deprived areas in the North East is greater than in London (Marmot et al., 2020). This highlights the importance of place in health inequalities: place has an impact on health over and above that explained by general relationships with deprivation.

Despite policy that aims to reduce social and geographical inequalities in health (Department of Health, 2009), there has been a widening of the gap in health between the most and least deprived areas in recent years. There has been an increase in mortality among the most deprived and widening inequalities in mortality by level of deprivation (Walsh et al., 2020). The difference in life expectancy at birth increased by 0.4 years for males and 0.9 years for females between 2010-2012 and 2016-2018 (Marmot et al., 2020). Growth in life expectancy have stalled for men and women in England since 2010 and this slowdown has been greatest in the more deprived areas (Marmot et al., 2020). Life expectancy declined for females living in the most deprived 10% of areas between 2010-12 and 2016-2018.

These patterns also play out differently by region. Life expectancy for women in the most deprived 10 per cent of small areas decreased in every region except London, the West Midlands and the North West. Life expectancy for men in the most deprived 10 per cent of small areas decreased in the North East, Yorkshire and the Humber and East of England. Life expectancy for men and

women living in the least deprived 10 per cent of neighbourhoods has increased in every region (Marmot et al., 2020).

Different diseases have contributed to social and spatial inequalities in health. The rate of improvement in Years of Life Lost (YLL) and life expectancy slowed more substantially for cardiovascular disease and (to a lesser extent) breast, colorectal and lung cancers (Steel et al., 2018). This highlights the importance of looking at patterns by cause of death and the role of cancer, and specifically, colorectal cancer in health inequalities.

While there have been reductions in the rate of growth in life expectancy in other countries, they have been largest in the UK (Leon et al., 2019). There is evidence to suggest that the recent stalling in life expectancy in the UK is likely to be associated with the implementation of government policy from 2010 which included cuts to public services and social security (Loopstra et al., 2016; Hiam et al., 2018) .The recent rise in infant mortality has been attributed in part to rising child poverty (Taylor-Robinson et al., 2019). The cuts in Local Authority funding were not uniform across the country, with the most disadvantaged areas disproportionately affected and the impact felt more heavily in the north of England than in the south (Hiam et al., 2018). This highlights how the broader socio-political environment can contribute to widening inequalities in health.

Individual socio-economic status

Social inequalities in health are also measured in term of individual socio-economic characteristics, such as educational attainment, employment, social class and housing tenure. Such information is usually collected in surveys or longitudinal studies, as data for the whole population is not usually disseminated at this level due to issues of confidentiality. Individual-level indicators have been associated with a range of health outcomes. For example, diabetes prevalence is higher among people with a lower educational level (Espelt et al., 2008) and mortality from cardiovascular disease is higher among people with lower occupational class and lower educational level (Mackenbach et al., 2000)

Often area-level indicators are used as a proxy for individual-level information, where this data has not been collected. Area-level indicators can, however, impact health regardless of individual characteristics. Furthermore, individual

and area-level deprivation may interact such that area deprivation has a disproportionate effect on those in low socio-economic groups (Stafford and Marmot, 2003; Luben et al., 2019).

2.3.5 Social variations in cancer

There are variations in cancer incidence and mortality by measures of socio-economic circumstances. Differences in cancer incidence rates have been observed between high and low deprivation groups. Based on the IMD, the highest all-cancer incidence rates were in the most deprived areas and the lowest rates in the least deprived areas (Arik et al., 2021). This pattern was evident in all regions in England, however, as found for general health indicators, the social gradient (the difference in incidence rates between the most and least deprived deciles) was steepest in the North East.

The association between cancer incidence and socio-economic circumstances varies by cancer type. Lung cancer incidence has the strongest association with socio-economic circumstances. Males and females in more deprived areas have higher rates of lung cancer compared to those in less deprived areas (Arik et al., 2020). Lung cancer is strongly linked to tobacco smoking and rates of smoking are significantly higher in more deprived areas (Office for National Statistics, 2017) which is likely to explain the association between socio-economic deprivation and lung cancer incidence. There is evidence that inequalities in lung cancer incidence have widened among women in recent years (Tweed et al., 2018). Lung cancer incidence rates have increased for females, particularly in more deprived areas, and this gap has widened over time (Arik et al., 2020). Again, this pattern was most evident in the north of England. Breast and prostate cancer incidence, on the other hand, show an inverse relationship with socio-economic circumstances: incidence is lower among people living in more deprived areas (Tweed et al., 2018). A possible explanation for high breast cancer incidence in the least deprived areas is that women in these areas are more likely to attend breast cancer screening (Maheswaran et al., 2006). Furthermore, women in the least deprived areas are more likely to be better educated which often leads to them having children later in life and having less children, both of which are associated with an increased risk of breast cancer due to hormonal differences (Al-Ajmi et al., 2018). Similarly, rates of prostate-specific antigen testing in the UK are higher in the

least deprived areas, leading to greater detection of prostate cancers among men in these areas (Williams et al., 2011). Using an alternate area-based measure of socio-economic deprivation, the Townsend index, (Pollock and Vickers, 1997) found a strong positive correlation between deprivation and incidence rates of lung cancers in the South Thames region but no association between deprivation and incidence of breast and colorectal cancers.

The evidence summarised above has come from studies using area-based measures of deprivation. Most recent cancer research in England and Wales uses data from cancer registries which do not record individual socio-economic indicators, hence area-based measures tend to be used as a proxy for these. Few studies have investigated the association between individual measures of socio-economic status and cancer risk. Sharpe et al. (2014) linked census and mortality data from the Scottish Longitudinal Study (SLS) to data from the Scottish Cancer Registry to investigate the effect of country of birth, marital status, area deprivation and individual socio-economic variables (economic activity, education, occupational social class, car ownership, housing tenure) on risk associated with lung cancer, upper aero-digestive tract (UADT) cancers and all cancers combined. They found that different and independent socio-economic variables were inversely associated with different cancer risks in both sexes and therefore no one socio-economic variable can capture all aspects of socio-economic circumstances. This highlights the importance of using multiple measures of socio-economic status in epidemiological studies.

There is strong evidence for socio-economic differences in cancer survival and mortality for many cancers and in many populations (Kogevinas and Porta, 1997). Large inequalities in cancer survival by socio-economic deprivation have been consistently identified in England and Wales (Coleman et al., 2004; Woods et al., 2006). Among adults living in the most deprived areas who were diagnosed between 1996-1999, 5-year survival was lower than for those in the least deprived areas for 28 of the 33 cancer-sex combinations analysed (Coleman et al., 2004).

Pollock and Vickers (1997) found higher mortality rates among patients living in the most deprived areas, based on the Townsend deprivation index, for all three tumour sites analysed (breast, lung and colorectal) and significantly lower survival rates among breast and colorectal patients living in the most deprived

areas. Poor survival in these groups was not explained by differences in incidence, unlike lung cancer for which incidence and mortality were positively correlated with deprivation, but no socio-economic gradient was found for survival. Therefore there must be other factors contributing to socio-economic disparities in survival among breast and colorectal cancer patients.

The deprivation gap in cancer survival has widened. In a population-based study of 2.2 million patients in England diagnosed with one of the 20 most common cancers, survival improved for most cancers in both sexes during the 1990s. For many cancers, however, survival improved more among those living in the least deprived areas than those living in the most deprived areas and as a result the existing deprivation gap in survival widened (Coleman et al., 2004). In a study of recent trends in cancer mortality, the lowest all-cancer mortality rates and steepest improvements in mortality since 2001, for both males and females, were in the more affluent groups (Arik et al., 2021). Furthermore, while there were differences in all-cancer mortality between the most and least deprived areas in all regions, the biggest gap was again in the North East of England.

Some studies have attempted to quantify the public health impact of inequalities in cancer survival. For patients diagnosed during 2004-2006, Ellis et al. (2012) estimate that 7,122 of the 64,940 excess deaths a year (11%) would have been avoided if 3-year survival for all patients had been as high as in the most affluent group (based on ecological measure of deprivation). While there was a reduction in the number of avoidable deaths over time (i.e. an improvement in cancer survival) this did not reflect a narrowing in the deprivation gap. Therefore, eliminating that social gradient in cancer survival should be a key policy aim.

A small number of studies have investigated the association between cancer survival or mortality and individual-level measures of socio-economic status. Sloggett et al. (2007) investigated socio-economic differences in cancer survival in England and Wales using four individual indicators of socio-economic status: Registrar General's Social Class, housing tenure, household access to a car and the Carstairs indicator of relative area deprivation and using relative survival methods. For all cancers combined, the four indicators of socio-economic status showed similar associations, but for individual cancers there were different associations between socio-economic indicators. For example,

there were survival differences by the car access indicator for five cancer types (lung, colorectal, bladder, cervix/uterus and ovarian), whereas the housing tenure showed survival differences for six cancers (lung, colorectal, breast, bladder, leukaemia, ovarian). Where there was an association, all indicators showed poorer survival with lower socio-economic status. This again highlights the importance of using different measures of socio-economic circumstances. In a large population-based study in Belgium, Hagedoorn et al. (2018) reported an association between individual socio-economic position and cancer mortality. Compared to individuals of high socio-economic position, all-cancer mortality was significantly higher among individuals who were primary educated, non-working and those living in low quality housing. Furthermore, individuals living in the most deprived neighbourhoods had significantly higher mortality for all-cancer, lung cancer and pancreatic cancer among men and women and for female colorectal cancer compared to individuals living in the least deprived neighbourhoods after controlling for individual socio-economic position.

2.3.6 Social variations in colorectal cancer

Incidence

The relationship between socio-economic deprivation and colorectal cancer is less clear than that for all-cancers combined and other common cancer types. Findings from studies investigating the association between socio-economic deprivation and colorectal cancer incidence have been inconsistent. In the 1980s, risk of colorectal cancer was higher among more affluent populations in Europe (Faggiano et al., 1997; Kogevinas and Porta, 1997), however, by the end of the decade no association between deprivation and colorectal cancer incidence was evident in England, based on area-based measures (Pollock and Vickers, 1997). In the early 1990s, a small excess of rectal cancers in men only was reported in more deprived groups in England and Wales (Quinn et al., 2001) and since then an association between increasing levels of area deprivation and colorectal cancer incidence in men has become clearer in England and Wales (National Cancer Intelligence Network, 2014) and Scotland (Oliphant et al., 2011). No similar association was reported among women. Recently, an emerging socio-economic gradient in colorectal cancer among women has been reported in Scotland (Tweed et al., 2018). In light of the apparent shift in the relationship between socio-economic deprivation and

colorectal cancer incidence, continued monitoring is needed to see how the pattern is changing. These studies will be explored in more detail in Chapter 5 in relation to the analysis of colorectal cancer incidence in the ONS Longitudinal Study.

Mortality

There is evidence of an association between colorectal cancer mortality in England, whereby mortality rates are higher among people living in the most deprived areas. Using the census-based Townsend deprivation score, Pollock and Vickers (1997) found higher mortality among the most deprived groups in the south of England. Based on another area-based measure of deprivation (IMD), the National Cancer Intelligence Network (2014) also reported higher mortality rates among people living in the most deprived areas compared with the least deprived areas, although the association was small for females. Higher mortality from colorectal cancer among lower socio-economic groups (both among men and women) has also been reported in European populations (measured using educational level) (Menvielle et al., 2008) and in the US (Steenland et al., 2004; Singh et al., 2011). The relationship between deprivation and colorectal cancer in England will be explored further in Chapter 4.

There is a strong body of evidence for an association between socio-economic deprivation and colorectal cancer survival. Significantly lower survival rates have been reported among colorectal cancer patients living in the most deprived areas compared to the least deprived areas in England (Pollock and Vickers, 1997). In a large, cancer registry based study of patients in England and Wales, the deprivation gap in 5-year survival for patients diagnosed during 1996-1999 was 6% and 7% for male and female colon cancer respectively and 9% and 8% for male and female rectal cancer respectively (Mitry et al., 2008a; Mitry et al., 2008b). Survival from colon and rectal cancer increased during the 1990s but improvements in survival were notably faster among patients in the least deprived areas and as a result the deprivation gap in 5-year survival became significantly steeper (Coleman et al., 2004). More recent studies of colorectal cancer patients in England have reported similar differences in survival between the least and most deprived groups determined using the IMD (Møller et al., 2012; Syriopoulou et al., 2019). Interestingly, Moller et al. (2012)

found that the excess mortality in the socio-economic deprived groups was a short-term phenomenon, most evident in the first month of follow-up and after that largely within the first year after diagnosis for colon cancer and two years for rectal cancer. Higher survival differences in the early period of follow-up have been reported in other studies (Ellis et al., 2012; Syriopoulou et al., 2019; Belot et al., 2019). Socio-economic disparities in survival have been found to vary by stage, however, with differences remaining over time in stages 2 and 3 for colon cancer and stage 2 for rectal cancer (Kajiwara Saito et al., 2021). These studies will be discussed in more detail in relation to the survival analysis conducted in Chapter 5.

Removing cancer-related inequalities in survival would result in a substantial gain in life years (Syriopoulou et al., 2019) and an estimated annual reduction of nearly 700 deaths in England (Møller et al., 2012). This highlights the importance of efforts to eliminate these differences.

Similar socio-economic patterns in colorectal cancer survival have been found in other developed countries. In Australia, people from remote and more disadvantaged areas had lower survival than those living in major cities and the least disadvantaged areas after adjustment for cancer stage and individual-level characteristics (Baade et al., 2013). Recent research investigating variations in survival from colorectal cancer in New Zealand found differences in risk of death by ethnicity and deprivation but no difference in risk of death by rurality (Sharples et al., 2018).

Worse survival among colorectal cancer patients in the least deprived groups may reflect socio-economic differences in stage at diagnosis, co-morbidities, lifestyle, treatment and participation in screening.

Stage at diagnosis

The stage of the cancer at the time of diagnosis is the prognostic variable most commonly thought to explain differences in cancer survival between subgroups of the population. However, evidence from studies investigating the contribution of stage at diagnosis to socio-economic variation in survival has been inconsistent. Some studies have reported a major role of stage (Ionescu et al., 1998; Lejeune et al., 2010), while others have determined that stage has little or no influence (Schrijvers et al., 1995; Hole, D J; McArdle, 2002). In a review,

Woods et al. (2006) concluded that while differences between socio-economic groups in the stage of disease at diagnosis explain at least part of the differentials in survival, the deprivation gap in survival may also be associated with factors such as access to optimal treatment, patient lifestyles and health-seeking behaviours and provision of other health care services. Thus, the current emphasis on stage at diagnosis is likely to be too simplistic to explain these inequalities.

Co-morbidity

The influence of health status on survival has been less widely studied. There is evidence that variation in co-morbidity could contribute to differences in mortality between socio-economic groups (Aarts et al., 2010). In a study of colorectal cancer survival in the South West of England, patients living in more deprived areas had worse health on presentation and this partly explained the deprivation gradient in all-cause survival (non-colorectal cancer deaths), but was not associated with cause-specific survival (colorectal cancer deaths) (Wrigley, 2003). The recording of comorbidity and stage information is often incomplete, particularly in patients with a short survival time, which should be taken into consideration in statistical analysis.

Health care factors

Factors related to the health care system may contribute to differences in survival, whereby patients in different socio-economic groups receive different treatments or have different treatment preferences (Aarts et al., 2010). In a large study of patients in South East England, residents of deprived areas with colorectal, lung or breast cancer were more likely to be admitted to hospital as emergencies than those from more affluent areas (Pollock and Vickers, 1998). Patients from deprived areas with lung or breast cancers were also less likely to receive surgical treatment (Pollock and Vickers, 1998). In an analysis of colorectal cancer patient records from three cancer registries in the UK, deprived patients (based on area-level socio-economic information) had poorer survival, were less likely to receive any treatment within 6 months and, if treated, were more likely to receive late treatment, compared to more affluent patients. Notably, the socio-economic gradient was substantially reduced among patients who received early treatment, even after taking account of differences in age at diagnosis and tumour stage (Lejeune et al., 2010).

Variation in the type of treatment received for rectal cancers by socio-economic group has also been reported (Morris et al., 2008). It is difficult to disentangle cause and effect when investigating the association between treatment and cancer survival because stage and comorbidity both strongly influence clinical decisions regarding treatment options.

Screening

Participation in national screening programmes may also contribute to survival differentials among socio-economic groups. In a cross-sectional analysis of colorectal cancer screening uptake data for the first 28 months of the national Bowel Cancer Screening Programme (BCSP), a strong gradient by deprivation was observed, ranging from 35% in the most deprived areas to 61% in the least deprived areas (von Wagner et al., 2011). This is in line with earlier findings from a randomised control trial in which people in more deprived areas were less likely to accept an invitation to be screened (Whynes et al., 2003). A more recent study of participation in the English Bowel Scope Screening programme also found a clear socio-economic gradient in uptake (McGregor et al., 2016).

Individual and area-level socio-economic status has also been associated with uptake of colorectal cancer screening in international studies. Low household income (Bernardo et al., 2018) and lack of health insurance (Davis et al., 2017; Bernardo et al., 2018) were associated with lower odds of colorectal cancer testing, whereas people living in an urban area had higher odds of colorectal cancer screening (Davis et al., 2017). Increasing area-level deprivation was associated with lower levels of screening (Pornet et al., 2010; Davis et al., 2017) whilst Buron et al. (2017) found uptake was lower in the most and least deprived quintiles and highest in the intermediate quintiles in a study in Barcelona.

Patient characteristics such as lifestyle and health-related behaviours also influence cancer risk and outcomes, directly and indirectly by interacting with treatment decisions. These factors will be discussed in more detail in the next section.

2.3.7 Lifestyle and health behaviours

General health

Non-communicable diseases such as cardiovascular disease, diabetes, respiratory diseases and cancer, represent major causes of disability and death globally (GBD 2019 Diseases and Injuries Collaborators, 2020). Several known risk factors for these chronic diseases are modifiable including smoking behaviour, diet, alcohol consumption, body weight and physical activity (GBD 2010 Diseases and Injuries Collaborators, 2012). These risk factors are leading contributors to ill health and premature death worldwide (Ezzati and Riboli, 2013).

The prevalence of these risk factors varies both socially and spatially. Lifestyle risk factors are unequally distributed in the population and associated with income, social class, education and measures of deprivation (NHS Digital, 2019). Behaviours that adversely affect health and risk of disease are more common in the most disadvantaged groups. Furthermore, these groups are more likely to engage in more than one risky behaviour (Khaw et al., 2008). Research using a large European cohort study has shown that having multiple risk factors contributes to greater ill health (Fransen et al., 2014) and likelihood of premature death (Khaw et al., 2008).

The distribution of lifestyle risk factors in the population of England has changed over time. A study by the King's Fund (Buck and Frosini, 2012) found that while there was a significant reduction in the overall proportion of the population that engages in multiple unhealthy behaviours, these reductions were mainly among people in higher socio-economic and educational groups. Therefore, the poorest and those with lowest levels of education will benefit the least from resultant improvement in health, leading to widening inequalities.

Spatial variations in lifestyle risk factors can manifest as a result of compositional factors (more deprived socio-economic groups in some areas) and contextual factors, such as aspects of the physical and built environment that may influence health behaviours. These contextual factors will be explored in more detail in Chapter 6.

Cancer

It is estimated that 40% of cancer cases in the UK are a result of exposure to modifiable risk factors (Brown et al., 2018). Smoking is the leading risk factor for cancer in the UK, followed by overweight and obesity. Lung cancer, colorectal cancer, melanoma skin cancer and breast cancer account for nearly two-thirds of all preventable cancer cases in the UK (Brown et al., 2018).

A number of risk factors are associated with colorectal cancer. The International Agency for Research on Cancer and World Cancer Research Fund classify the following factors as having ‘sufficient’ or ‘convincing’ evidence of a causal association with colorectal cancer risk: cigarette smoking, alcoholic drinks, body fatness/BMI, processed meat, insufficient dietary fibre, and insufficient physical activity (Lauby-Secretan et al., 2016; World Cancer Research Fund, 2018; International Agency for Research on Cancer, 2021).

It is estimated that over 50% of colorectal cancer cases in the UK are linked to preventable risk factors (Brown et al., 2018; Goon et al., 2021). The latest research estimates that preventable cases could be as high as 67% of cases in men and 60% of cases in women (Goon et al., 2021). Again, there are spatial and social variations. Differences in the proportion of all cancer cases attributed to modifiable risk factors have been observed between countries in the UK (Brown et al., Goon et al.). This is likely to be due to demographic differences, such as levels of socio-economic deprivation, which are associated with differences in health behaviours, such as rates of tobacco smoking. Furthermore, many lifestyle ‘choices’ are driven by environmental and social factors, such as cost, access to healthy food, and nutritional knowledge. Exposure to these factors varies by socio-economic position. A healthy diet is generally more expensive than one high in energy dense, nutrient poor foods (Morris et al., 2014). Furthermore, there may be a lack of healthy food options in disadvantaged areas. For example, the availability of fast food outlets is greater in more deprived areas (Fraser et al., 2010). Environmental factors will be explored in more detail in Chapter 6.

Other individual factors

There are other individual demographic factors that influence colorectal cancer risk. Old age is the main risk factor for cancer, therefore the age structure of the population should be taken into account when comparing rates between different areas. Rates of colorectal cancer incidence and mortality are higher in males compared to females. Colorectal cancer is more common in people of White ethnicity than in those of Asian or Black ethnicity (National Cancer Intelligence Network, 2009).

While not a focus of this research, it should be noted that family history and certain hereditary conditions increase an individual's risk of developing colorectal cancer (Fearnhead et al., 2002).

2.3.8 Summary and context

There are spatial and social gradients in health. These gradients are found in general health and specific morbidities such as cardiovascular disease, respiratory disease, diabetes and cancer - both overall and for many cancer types. The pattern for colorectal cancer is, however, less clear. There is strong evidence of an association between socio-economic deprivation and colorectal cancer survival, whereby patients in the most deprived areas have poorer survival than those in the least deprived areas. Similarly, colorectal cancer mortality rates are higher among those in the most deprived areas, compared to the least deprived areas.

The mechanisms to explain the relationship between socio-economic deprivation and cancer survival are complex and not well understood. While there is some evidence that stage at diagnosis and access to treatment play a part, other factors such as lifestyle, health-behaviours, co-morbidity and other health-care factors may also influence the deprivation gap in survival. The evidence for an association between colorectal cancer incidence and socio-economic deprivation is less clear. There is emerging evidence of a social gradient in colorectal cancer incidence in both men and women, but previous findings have been inconsistent. Continued monitoring of the relationship between socio-economic deprivation and colorectal cancer incidence is needed to understand whether these patterns are changing. Chapter 5 of this thesis examines the association between socioeconomic deprivation and colorectal

cancer incidence and survival using both individual-level and area-based indicators for a more up-to-date cohort of the population.

Geographic variations in colorectal cancer incidence and survival have also been reported at regional level, although they are not as marked as those for all-cancers combined and lung cancer. More detailed analysis using smaller geographic units is required to better understand these variations. Colorectal cancer mortality rates are examined at Local Authority level in Chapter 4 of this thesis. The role of the local environment in influencing health outcomes is increasingly being recognised. Features of the local environment are considered in the context of colorectal cancer outcomes in Chapter 6 of this thesis.

It is clear from the review of evidence that place matters for health. Health varies both between different areas and between social groups within areas. Furthermore, the steepness of the social gradient varies by area so it is important to consider social variations within a geographic context. Geographic scale is also important to understand the level at which inequalities operate. While much of the previous research has been at regional level this may mask variations at smaller geographic units which are important for targeting public health interventions. A range of different measures of socio-economic status have been used, both individual and area-based, to study the association with health outcomes. This is important to understand the different aspects of socio-economic deprivation that may influence health. Finally, monitoring of trends over time is required to understand how patterns are changing. There is evidence that inequalities in health are widening so it remains a vital area of research. An updated picture of trends in colorectal cancer is needed to see how patterns compare to those for other cancers and general health and mortality.

This thesis will examine the spatial and social variations in colorectal cancer incidence, survival and mortality in England and Wales by both individual and area-based measures of socio-economic status and area type over time according the objectives outlined in Chapter 1 section 1.6. The next chapter will outline the available data sources and how they can be used to address these aims.

Chapter 3
Data and Methods

3.1 Overview

The aims and objectives of the thesis were outlined in Chapter 1 section 1.6. The existing literature in this area was reviewed in Chapter 2 and some gaps in knowledge were identified. This chapter will appraise the existing data sources available for conducting research to address these aims and objectives. Appropriate methods to analyse these types of data will then be considered.

3.2 Data

No one single data set can provide all the information required to meet the aims and objectives of this thesis, therefore a range of data sources were considered. The types of data sources included traditional cancer registries and death registries, cohort studies and novel data repositories.

An initial data audit was carried out to assess the suitability of different data sources to meet the aims and objectives of the thesis and to assess what information they could provide and gaps in knowledge they could fill (Table 3.1).

Data set	Coverage	Geographic Scale	Time period	Sample size	Data Type	Outcome	Individual SES	Area Types	Lifestyle Info.	Treatment Info.
Mortality	England & Wales	LA	1990-2012	341,234 CRC deaths	Cross sectional	Mortality	✗	✓	✗	✗
ONS LS	England & Wales	Region	1971-2011	14,400 CRC cases (as of April 2015)	Longitudinal	Incidence Survival	✓	✓	✗	✗
CORECT-R	England	LSOA	1997-2018	600,000 CRC cases	Longitudinal	Incidence Survival	✗	✓	✗	✓
UK Women's Cohort Study	Great Britain	Region	1995-Present	491 CRC cases (as of Dec. 2013)	Longitudinal	Incidence Survival	✓	✓	✓	✗
UK Biobank	Great Britain	LSOA?	2010-Present	2,284 prevalent cases 2,275 incident cases 1,176 CRC deaths (as of Sept. 2016)	Longitudinal	Incidence Survival	✓	✓	✓	✗

Table 3.1 Data audit summary

Three different data sets were selected and used in the subsequent analysis chapters (Chapter 4-6): ONS mortality data, the ONS Longitudinal Study and the COloRECTal cancer data Repository (CORECT-R), respectively. Access to other data sets, including the UK Women's Cohort Study and UK Biobank, were explored but these were not pursued due to limitations with the sample sizes and geographic coverage, data access processes and costs. Potential future avenues of research, linking data sets used in this thesis to new data sources, such as consumer data and data from the Bowel Cancer Screening Programme, will be considered in later discussions.

The following sections outline the data sources used in the thesis in terms of the geographic coverage of the data, the geographic detail they provide, the time period they cover, the variables they contain and the sample sizes available. The pros and cons of each data set for this type of research will be considered. The application process to obtain data from each source is also outlined. The diagram in Figure 3.1 illustrates the different dimensions each data set contributes to.

A theme running throughout the thesis will be the use of area types to investigate variation in colorectal cancer. Different types of area-based classifications are also described in this section along with the definitions of spatial measures used in the analysis.

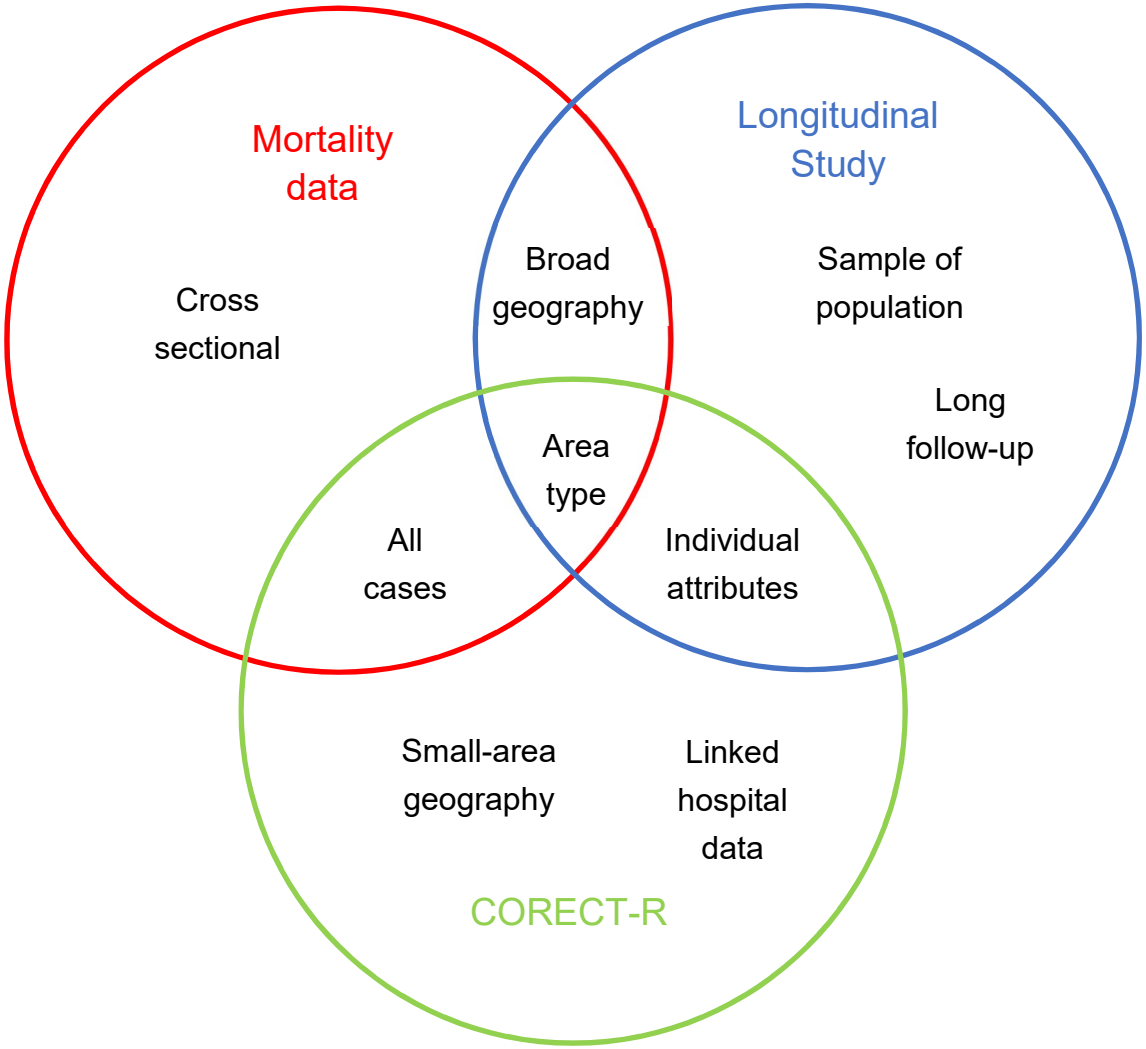


Figure 3.1 Summary of thesis research data sets

3.2.1 ONS Mortality statistics

Mortality statistics for England and Wales are compiled by the ONS from information supplied when deaths are certified and registered as part of the civil registration process, which is a legal requirement. The figures represent the number of deaths registered in a calendar year (rather than the number of deaths that actually occurred in that period). The majority of deaths are registered within 5 days of the date of death, but there are some situations that result in the registration of the death being delayed, such as deaths that are reported to a coroner. Therefore some deaths will not be registered in the same calendar year in which they occurred. Furthermore, where there was an inquest or post mortem, the cause of death may be amended. Deaths due to cancer rarely require an inquest, so there is little delay in death registration. Therefore death statistics based on registration year provide an accurate picture of cancer mortality trends (Office for National Statistics, 2018).

The dataset contains mortality statistics for England and Wales broken down by calendar year of registration, age, sex, underlying cause of death and area of usual residence of the deceased. Cause of death information is coded using the World Health Organisation (WHO) International Classification of Diseases and Related Health Problems (ICD) coding system, Version 9 and 10.

Annual mortality statistics for England and Wales were obtained via the UK Data Service (UKDS) for the time period 1990 to 2012 at Local Authority (LA) level. Rules for accessing death registration data changed post 2012 and a new application to UKDS is required for secure access to individual death registration records for the most recent years of data. Mortality statistics for the period 2013-2020 are published on NomisWeb (Nomis, 2021a), however, these data contain supressed counts so are not usable for analysis of deaths by specific causes for which the pre-aggregated counts of deaths are small when split by age group and geographic area. An application has been made to UKDS to access individual-level death registration data up to 2018 (see Appendix B).

The advantages of this data source for analysis of colorectal cancer mortality trends is that it includes all deaths occurring in England and Wales over a long time period. The data is of high quality as it is a legal requirement to register a

death so there is very little missing data. Area types can be appended to the data, but the geographic level at which the linkages can be performed depends on the granularity at which the data is released. A limitation of the data set is the geographic detail, which for colorectal cancer deaths is currently Local Authority level due to small number suppression in the publicly-available data for more detailed geographic units. Furthermore, the timeliness of the data is restricted due to issues in obtaining access to individual-level death registration records (which is discussed further in Chapter 7).

Colorectal cancer mortality trends in comparison to mortality by all cancers are examined in Chapter 4. The analysis will provide an understanding of all deaths in England and Wales, but is limited to a LA level.

3.2.2 ONS Longitudinal Study

The ONS Longitudinal Study (LS) was set up in 1974 with the initial sample drawn from the resident population enumerated at the 1971 Census (Hattersley et al., 1995). Selection into the LS is by birth date. Four undisclosed dates of birth were used to select a random, 1 per cent sample of the population of England and Wales. The study is a continuous, multi-cohort study with subsequent samples drawn at each census, based on the same selection criteria, and linked into the study. Between censuses, new members enter the study via birth on an LS birth date or by immigration (if born on an LS birth date) and existing members exit the study through death or emigration. The original sample size of the LS was over 500,000 and has remained consistently high over time. Data for approximately 1.1 million individuals has been collected over the 40 years of the LS (Shelton et al., 2018).

The LS contains census records of sample members and life events data, linked between five successive censuses (1971, 1981, 1991, 2001 and 2011). Census information is also available for people living in the same household as an LS member. Administrative data routinely collected by ONS is linked to LS members via the National Health Service Central Register (NHSCR). Events linked to LS member's include enlistments into the Armed Forces, embarkations, re-entries to the NHS from enlistment or embarkation, immigration, new births, live and still births to sample mothers, widowhoods, deaths and cancer registrations.

The LS is maintained and updated by the ONS. The LS data are available to anyone in the UK who has obtained ONS Accredited Researcher status via the ONS Approved Researcher Scheme (Office for National Statistics, 2021a). The application process involves submitting a research project application which is reviewed by the Data Owner, LS Research Board and Research Accreditation Panel. The data can be accessed through the Secure Research Service (SRS) safe setting rooms at ONS offices in London, Newport and Titchfield. Outputs are released to researchers through an output clearance process. Information and support for UK-based researchers is provide by the Centre for Longitudinal Study Information and User Support (CeLSIUS) (University College London, 2021a).

I successfully completed Safe User of Research data Environments (SURE) training and obtained ONS Accredited Researcher status in April 2018. An application to use the ONS LS to investigate variations in colorectal cancer registrations and survival by socio-economic group was approved in September 2018 (see Appendix C).

An advantage of the LS is that it contains a large, representative sample of the population of England and Wales. Individual-level indicators of socio-economic status from the census are available for LS members, which are not collected by cancer registries. Additional area-based attributes can be appended to the LS data using the LS members' postcode, although an additional x-file request is required for the ONS to perform these linkages. The area groupings are returned but the postcode data is not released. As the LS is only a 1% sample of the population of England and Wales, disclosure controls, to ensure individuals within the data set are not identifiable, may limit the number and level of geographic identifiers which are appended to the data. In this case, regional level geographies only were appended to the LS data.

In summary, the LS provides a rich, longitudinal sample for a subset of the population of England and Wales with more granular socio-economic information and detailed area type groupings. The longitudinal nature of the data allows for time-to-event analysis to be performed which is not possible using cross-sectional data. These data and methods are used to analyse colorectal cancer diagnoses and deaths to LS members in Chapter 5.

3.2.3 CORECT-R

The COloRECTal cancer data Repository (CORECT-R) is a single research repository for colorectal cancer data (Downing et al., 2021). The resource contains multiple linked data sets which provide information on many aspects of the patient pathway from initial cancer diagnosis, treatments received and subsequent outcomes. Data sources within CORECT-R include routine administrative datasets such as national cancer registry data curated by the National Cancer Registration and Analysis Service (NCRAS) in England, inpatient and outpatient activity data provided by Hospital Episode Statistics (HES) data and mortality data from the ONS. CORECT-R also facilities linkage to colorectal cancer research studies, such as clinical trials, cohort studies and surveys, and biobanks containing biological samples. The information within the repository will be refreshed annually.

At present, the database contains information on all individuals diagnosed with colorectal and anal cancer in England between January 1997 and December 2018. This includes information on more than 600,000 cases of the disease. CORECT-R is a new data source, established in 2018, and expansion of the resource to cover the whole of the UK is currently under way.

CORECT-R has been developed by the UK Colorectal Cancer Intelligence Hub, a Cancer Research UK funded collaboration between academics, clinical experts, patients and the public (UK Colorectal Cancer Intelligence Hub, 2021). Access to the CORECT-R resource is managed by the UK Colorectal Cancer Intelligence Hub team. The application process involves the development of a project protocol that describes the information required and sets out the study objectives. Access to the data with CORECT-R is free to academic researchers with the necessary regulatory approvals. The data is held in a secure Trusted Research Environment (TRE), an analytical area that, following both project and user approval, is accessed via two-factor authentication and a virtual desktop. An application to use data held with CORECT-R was approved in April 2021 (see Appendix D).

The strengths of the CORECT-R data set are the timeliness of the data and large sample size, containing all colorectal cancer diagnoses in England. Additionally, area-level information can be linked (at source) using the patient's postcode of residence at diagnosis, including administrative geographies and

area types. Small-area geographic units (Lower Layer Super Output Areas) are available, which is more detail than provided by the other data sets. A limitation of the CORECT-R data is that the coverage is currently only England and it does not contain any individual-level socio-economic information about cancer patients.

In summary, CORECT-R is a new resource for colorectal cancer research which contains a wider variety of datasets, linked for all patients in England. It is used to analyse colorectal cancer incidence by local area in Chapter 7.

3.2.4 Area data

The data sources noted above can be used in conjunction with area data and over a variety of time frames. First below, I specify what geographical scales will be used (alongside notes on geographies from the literature which are not used). Then, I note the sources of area attribute data used before describing the types of derived measures incorporated which use area variables as inputs. Finally, I provide some detail on the spatial and statistical measures used for analyses. The information below will be expanded on in the relevant empirical chapter and in later discussions.

3.2.4.1 Geographies

A range of administrative and health geographies are referenced within this thesis (Office for National Statistics, 2021e).

Local Authorities

Local Authorities (LA) are administrative geographies made up of unitary authorities, metropolitan and non-metropolitan districts and London boroughs in England and unitary authorities in Wales. There are a total of 348 LAs in England and Wales.

LAs were used in the geographic analysis of colorectal cancer mortality in Chapter 4. While we might expect to see more variation between smaller geographic units (as there is less heterogeneity in the demographics of the population within each small area, but more heterogeneity between areas), LAs are the geographic units at which Public Health is administered in England. Furthermore, it not always possible to get data for smaller geographic units due

to the need to protect confidentiality. The balance between geographic detail and confidentiality is discussed further in Section 7.4.1.

Lower Layer Super Output Areas

LSOAs are small areas designed to be of a similar population size, with an average of approximately 1,500 residents or 650 households. They are produced by the ONS for the reporting of small area statistics. There are 32,844 LSOAs in England.

LSOAs are used in Chapter 7 to append additional information to the cancer registry data. These were the smallest geographic units that could be linked to the data. Small geographic units (LSOAs) were used to represent a patients local area, however, the use of administrative boundaries, while often used due to availability of data does have limitations in terms of their relevance to the definitions of neighbourhoods. The uncertainty around geographic context for research studies is discussed further in Section 7.4.4.

Health geographies

Currently, health geography in England comprises 7 NHS England Regions, within which there are 42 Sustainability and Transformation Partnerships and 21 Cancer Alliances. At the primary care level, there are 135 Clinical Commissioning Groups. However, over time, the health geographies defined by the NHS have changed frequently with different political regimes (Smith et al., 2001) and it would be distracting at this point to include comprehensive detail. These changes make direct comparison of results in the literature difficult.

3.2.4.2 Sources

Census data

A census of all people and households in England and Wales takes place every 10 years. The most recent census was on 21 March 2021. A large amount of demographic and socio-economic data is gathered at each census including information about education, employment, marital status, housing tenure and household composition. Questions asked on the census have changed over time. For example, questions about ethnicity and self-reported limiting long-term illness were introduced in the 1991 Census. Aggregate results of the census are

published for a range of census geographies. Area census data can be linked to the individual records of LS members.

Colorectal cancer registrations linked to LS members and survival by individual indicators of socio-economic status from the census are analysed in Chapter 5. Census variables are also used as inputs to the Townsend Deprivation Index described below and used in Chapter 4 and 5.

Population estimates

Annual mid-year population estimates are available from the ONS for a range of geographies. The data contains population estimates by year, age and sex for each area. Mid-year population estimates relate to the usually resident population on 30 June of each year. Annual population estimates for LAs are used in Chapter 4 and for LSOAs in Chapter 7 when calculating colorectal cancer mortality and incidence rates, respectively.

Access to Healthy Assets and Hazards (AHAH) index

The Access to Healthy Assets and Hazards (AHAH) index is an open-access multi-dimensional index of the accessibility to health-related features of the environment for small areas (LSOAs) across Great Britain (Green et al., 2018). The overall index is made up of four domains of accessibility: retail outlets, health services, air quality and natural environment domain. Each domain contains a series of measures of health-related features of neighbourhoods.

As reported in Green et al. (2018), the index utilises a range of nationally extensive data sources to create the health-related indicators. Data on retail outlets (gambling outlets, fast food outlets, pubs/bars/nightclubs, off licences and tobacconists) and leisure facilities were acquired from the Local Data Company (LDC). The location of health services (GP surgeries, A&E hospitals, dentists, pharmacies) were obtained from NHS Digital. For all postcodes in Great Britain, the road network distance to the nearest of each service was calculated. Open data from Ordnance Survey was used create two variables related to the distance to the nearest green space and the total green space areas available to each postcode within a 900-meter buffer range. A key part of this definition was that green space needed to be available to the public (i.e. not private land). Blue space locations such as beaches were acquired from OpenStreetMap and mainland water bodies (e.g. lakes and rivers) were

sourced from the European Settlement Map. Measures of air quality were calculated from data provided by DEFRA. Data for all indicators was from 2016. The data was aggregated (by taking the mean) to small geographical areas (2011 LSOAs for England and Wales and 2011 Data Zones for Scotland).

This source is used as an input in this thesis, though it contains derived measures, and as such there is some overlap with the next section. It is included here as it is a data source for the analysis in Chapter 6. A range of the individual indicators of the AHAH index relevant to colorectal cancer were used in Chapter 6 as inputs to derive area-based measures of colorectal cancer risk.

3.2.4.3 Derived area measures

Deprivation Indices

Several methods exist to measure the level of deprivation in an area. The release of small area statistics from the census enabled the development of deprivation indices (composite measures of socio-economic deprivation), which have facilitated research into the relationship between area-level socio-economic deprivation and health outcomes. Early work using this approach by (Townsend et al., 1988) and (Carstairs and Morris, 1991) provided evidence for the link between area deprivation and mortality. The small area approach also allowed for the investigation of inequalities in health beyond measuring differences in mortality and using the traditional measure of social class. For example, using postcode units to link records to an area enabled health information such as that from cancer registries to be interrogated (Townsend et al., 1988).

The Carstairs Index and the Townsend Index are census-based indices derived from a selection of census variables that are combined to create a single measure of the relative deprivation of an area. They can be calculated at different geographic scales, provided that census data is available. A comparable measure can be calculated for other countries if similar census variables are available. The Townsend Index is preferred over the Carstairs Index in this thesis as the input variables are more consistent in definition and in their availability over time.

The Townsend Index is constructed from four census variables: households without a car, overcrowded households, households not owner-occupied and

person's unemployed. Each variable is divided by the appropriate count of persons or households to obtain a percentage. The unemployment and overcrowding variables tend to be positively skewed so the percentages are log-transformed in order to normalise the values. All four input variables are then standardised using a z-score. The Townsend index is then calculated by summing the four z-scores to obtain a single value. Positive values of the index indicate more deprived areas, relative to the reference area, whereas negative values indicate relative affluence. A score of zero represents an area with mean values.

Townsend scores were used in the analysis of colorectal cancer mortality trends in Chapter 4 and appended to LS member records in Chapter 5. The limitations of the Townsend Index are discussed in more detail in section 4.7.3.

The approach to measuring area-level deprivation is the subject of much debate, regarding the appropriate geographical scale, the selection of variables and the methods used to combine them into a single measure. Census variables provide consistent measures that are available for the entire population, however, the use of census variables some constraints on the ability to capture all aspects of disadvantage. More recently, the availability of administrative data from government departments at small-area level has led to the construction of new measures of disadvantage based on different variables. The benefit of using administrative data is that it is available more regularly and recently than census data, collected only once every ten years.

The English Index of Multiple Deprivation (IMD) is a measure of relative deprivation for small-areas (LSOAs) in England in which areas are ranked from most to least deprived (Ministry of Housing Communities and Local Government, 2019). It combines information from seven domains of deprivation into a single measure according to their corresponding weights: Income deprivation (22.5%); Employment deprivation (22.5%); Education, skills and training deprivation (13.5%); Health deprivation and disability (13.5%); Crime (9.3%); Barriers to housing and services (9.3%); Living environment deprivation (9.3%). Each country in the UK produces its own version of the IMD, therefore it is not possible to make direct comparisons between the indices. For this reason the IMD was not an appropriate measure of area deprivation to use in Chapter 4 to analyse mortality trends in England and Wales. In addition, the incorporation

of a health domain in the IMD may over emphasise the relationship between deprivation and health, so the income domain alone is commonly used in health research.

While different approaches have been taken there are strong correlations between different deprivation measures, including between Townsend and IMD (Norman, 2010). Most measures are calculated cross-sectionally, however, recently measures of deprivation over time have been constructed, which allow the impact of changes in area-deprivation on health outcomes to be investigated (Norman and Riva, 2012).

Population density categories

Rurality of areas was defined by population density. A population density variable is derived by ONS using the number of persons from the census and area in hectares from the GIS Shapefile (Nomis, 2013).

The groupings developed by the Organisation for Economic Co-operation and Development (OECD, 1994) were used to assign a rural-urban category to each LSOA based on an approximation of the persons per hectare within each category (Table 3.2).

Population density category	Description	Persons per hectare
1	Most urban	>33
2	Very urban	26-33
3	Urban	13-26
4	Rural	1-13
5	Most rural	<1

Table 3.2 Population density categories

These population density categories are used in Chapter 4 to examine mortality trends by rurality and in Chapter 6 to control for rurality when modelling cancer incidence by area types.

3.3 Methods

Appropriate methods that I considered for analysing the types of data sets outlined above are described here. There is some cross-over of methods between the analysis chapters (Chapters 4, 5 & 6). The general methods are described here but how they are applied to the specific data sets and how they are interpreted is explained in the relevant chapter.

3.3.1 Statistical methods

3.3.1.1 Rates

Crude rates are calculated by dividing the number of cancer cases (incidence) or number of deaths (mortality) by the population at risk. Colorectal cancer incidence and mortality is strongly related to age, therefore comparing crude rates between two or more populations or time periods could be misleading because the populations being compared may have different age structures. Methods that control for differences in the age structure of populations are therefore required in this context.

Standardisation is a set of techniques used to overcome the effects of confounding variables when comparing summary measures across populations (Schoenbach and Rosamond, 2000). There are two methods of standardisation commonly used in epidemiological studies: direct and indirect standardisation. For direct standardisation the observed age-specific rates of each study population are applied to the age distribution of a given standard population. This produces age-standardised rates that the study population would have experienced if they had the same age distribution as the standard population. This method requires that age-specific rates are available for all the populations being studied. In the indirect method, rather than taking a single population structure as standard and applying sets of rates to it, a set of age-specific rates from a standard population is applied to each of the populations being compared to calculate standardised incidence or mortality ratios. This method is commonly used when age-specific rates are unavailable as it only requires the total number of cases or deaths and the age structure of the study population to be known.

Directly standardised rates can be readily compared to each other and to the standard population because they are all based on the same set of weights (the age distribution of the standard population). Comparison of indirectly standardised rates is more difficult as the weights applied to the standard age-specific rates depend on the age structure of the study population. If there are large differences in age-structure between the standard populations the SMRs calculated would be based on different weights and not comparable. Indirect standardisation is, however, preferable when there are small numbers in particular age groups. In this case, age-specific rate estimates from the direct method will be unstable.

The direct method of standardisation was preferred over the indirect method in Chapter 4 to allow for valid comparisons to be made between geographical areas and over time. The mortality and population data was split by age group so it was possible to calculate age-specific mortality rates. Similarly, in Chapter 7, direct standardisation was used to calculate incidence rates to allow for comparison across area types.

Directly age-standardised rates were generated using the 2013 European Standard Population (ESP). This is a hypothetical population and assumes the age structure is the same in both sexes, therefore allowing comparisons to be made between the sexes as well as between geographical areas. The ESP was first introduced in 1976 and has been revised by the statistical office of the European Union (Eurostat) taking into account changes in the age-structure of the population that occurred in the European Union member states since the mid-1970s (Eurostat, 2013).

3.3.1.2 Logistic regression

Binary logistic regression is used when predicting an outcome that has only two possible values. Binary logistic regression was used in Chapter 7 to investigate variation in stage at diagnosis by neighbourhood characteristics. For this analysis, colorectal cancer patients were dichotomised into early stage (Stage 1 and 2) and late stage (Stage 3 and 4) cases.

3.3.1.3 Time-to-event analysis

Time-to-event analysis is used to investigate the time it takes for an event of interest to occur. This approach accounts for individuals who do not experience the event during the study follow-up period and for whom survival times are unknown. This feature is known as censoring and may arise if: an individual has not experienced the event of interest within the study time period; an individual is lost to follow-up during the study period; an individual experiences a different event that makes further follow-up impossible (Clark et al., 2003).

Survival data are described and modelled using two related properties: the survival probability and the hazard probability (Clark et al., 2003). The survival probability, also known as the survivor function S(t), is the probability that an individual survives from the time of origin to a specified future time t. The hazard, denoted by h(t), is the probability that an individual who is under observation at time t has an event at that time.

The survival probability can be estimated from observed survival times using the Kaplan Meier method (Kaplan and Meier, 1958). Kaplan-Meier survival curves, a plot of the Kaplan-Meier survival probability against time, provide a summary of the survival data and can be used to estimate the median survival time. Two or more survival curves can be compared by various tests, the most widely used of which is the log-rank test. The log-rank test compares the number of observed events in each group to the number that would be expected if they all had the same survival curve. Difference between groups is identified by a p-value of less than 0.05.

The Cox proportional hazards regression model compares the hazard rate in different groups. The results are usually expressed in terms of the hazard ratio (i.e. the ratio of the rates). A hazard ratio greater than 1 indicates a factor is positively associated with the event probability and hence negatively associated with the length of survival (a hazard ratio of 1 means no difference). A key assumption of the proportional hazards model is that the hazard of the event in any group is a constant multiple of the hazard in any other. This assumption implies that the hazard curves for different groups should be proportional and should not cross. Log-log plots were produced to assess whether this assumption holds. An advantage of the Cox model over other time-to-event

methods is that it allows for adjustment of other factors that might affect the risk of the event occurring. In addition, it provides the effect size for each factor.

Time-to-event analysis was used in Chapter 5 to examine the association between indicators of socio-economic status and colorectal cancer registration and survival in the ONS LS.

3.3.2 Spatial methods

3.3.2.1 Mapping

Choropleth maps were produced in Chapter 4, for each census year, to visualise the variation in mortality rates at LA level in England and Wales and in Chapter 6 to visualise the distribution of the colorectal cancer risk index.

3.3.2.2 Spatial cluster analysis

Spatial cluster analysis encompasses a range of statistical techniques used to quantify spatial patterns in data that may not be apparent using simple visualisation methods. They identify areas with unusually high or low concentrations of a particular characteristic or event. The Getis-Ord Gi* statistic is used in Chapter 4 to examine the spatial association of mortality rates at Local Authority level. This method was chosen as it is an appropriate technique for analysing data that has been aggregated to geographic units, such as administrative boundaries. The Getis-Ord Gi* statistic is a measure of local spatial autocorrelation, which identifies the local association between a feature and its neighbours within a specified neighbourhood. It indicates the extent to which each feature is surrounded by similarly high or low values. The sum of values for a feature and its neighbours are compared to what would be expected given all the values in the entire study area. When the observed sum is different to the expected sum, and when the difference is too large to be a result of random chance, a statistically significant z-score is returned. Positive z-scores indicate clustering of high values (hot spots) and negative z-scores indicate clustering of low values (cold spots). This measure has been applied previously in epidemiological studies, for example to detect hot spots of colorectal cancer mortality in the US (Siegel et al., 2015).

An alternative measure of spatial autocorrelation, Anselin Local Moran's I (Anselin, 1995) was considered which, in addition to hot and cold spots, also

identifies spatial outliers. Numerous other spatial scan techniques exist, but most methods use point data so they were not appropriate approaches to analyse the aggregate area-level data used in Chapter 4.

The results of spatial cluster analysis are sensitive to the size of the areas being analysed and the definition of the surrounding neighbourhood, which is discussed in section 4.7.3.

3.4 Ethics

Ethical review was required as part of the application made to the UKDS to obtain approval to access individual death registration data for England and Wales (Chapter 4). The data set is categorised by the UKDS as controlled data and must be accessed in a secure environment. There were some changes to the application process in 2020 and applications for ONS Secure Access data now require ethical approval from the applicant's institution. An application was made to the University of Leeds Faculty of Medicine and Health Research Ethics Committee for ethical review (MREC 20-009) and confirmation that the study was considered favourably was received in October 2020 (see Appendix A). An ethics-self assessment form was also required by the UKDS for consideration by the UK Statistics Authority.

The study using data from the LS (Chapter 5) did not require full ethical approval as it involved secondary analysis of established anonymised data. It was approved by the ONS Microdata Release Panel (see Appendix C).

The CORECT-R resource, and analyses based upon the data within it, has received approval from the South West-Central Bristol research ethics committee (18/SW/0134). Therefore separate ethical review for the study in Chapter 7 using CORECT-R data was not required.

3.6 Summary

This chapter has outlined the data sources available and how each data set can contribute to different aspects of the aims and objectives of this thesis. Appropriate methods for analysing these types of data have been described and will be utilised in Chapters 4-6.

The themes raised in this chapter including timeliness of data, data access and confidentiality in research will be returned to in the discussion (Chapter 7), in the context of challenges faced when analysing these data.

Chapter 4

Spatial and temporal variations in colorectal cancer mortality in England and Wales

4.1 Overview

This chapter will explore trends in colorectal cancer mortality by geographic location and area-level deprivation over a 20-year period using death registration data from the ONS. The long time period covered by the data set will make it possible to see how trends in mortality in relation to deprivation and rurality have changed over time. The geographic information will enable spatial analysis of mortality rates at sub-regional level. Mortality trends will be analysed by gender and age, but the data set does not contain individual indicators of socio-economic status.

4.2 Background

Trends in colorectal cancer mortality over time

Colorectal cancer is the second most common cause of cancer death in the UK, accounting for around 16,600 deaths a year (Cancer Research UK, 2018). Colorectal cancer mortality is strongly linked to age, with the highest mortality rates found in the older age groups. In the UK, approximately 58% of deaths each year are in people aged 75 and over (Cancer Research UK, 2018) which reflects higher incidence and lower survival for colorectal cancer in the older age groups. Colorectal cancer mortality rates are significantly higher for males than females.

There has been a considerable decrease in colorectal cancer mortality rates over the past 50 years in the UK (Quinn et al., 2001). The decrease in mortality rates over time could reflect changes in prevalence of cancer risk factors, improvements in diagnosis, including screening in recent years, and improvements in the treatment of colorectal cancer.

Variation in colorectal cancer mortality by deprivation

Despite efforts to reduce geographic and socio-economic inequalities in cancer outcomes, disparities persist. Research has largely focused on quantifying socio-economic differences in cancer outcomes. There is a large body of

evidence relating to variation in colorectal cancer incidence, survival and mortality by socio-economic deprivation (Aarts et al., 2010). In the UK, using the area-based Townsend deprivation score, Pollock and Vickers (1997) found higher colorectal cancer mortality among the most deprived groups and significantly lower survival rates among patients living in deprived areas. In a more recent analysis, based on data from 1997-2011, an association between colorectal cancer mortality and deprivation, measured by the IMD, was reported, although the association was small for females (National Cancer Intelligence Network, 2014). It is estimated that there would be around 860 fewer deaths each year in England if everyone experienced the same mortality rates as the least deprived.

Higher colorectal cancer mortality among low SES groups could reflect higher incidence, however, evidence of an emerging association between deprivation and colorectal cancer incidence has only recently been reported (Oliphant et al., 2011; Tweed et al., 2018). There is, however, strong evidence of association between deprivation and colorectal cancer survival. Survival has been significantly lower among men and women living in the most deprived areas since the 1980s (Mitry et al., 2008a; Mitry et al., 2008b). Poorer survival among the most deprived socio-economic groups may reflect socio-economic differences in stage at diagnosis, co-morbidities, lifestyle (diet and physical activity), receipt of treatment and participation in screening.

Studies using an area-based measure of deprivation at a single point in time, such as census year do not take into account the level of deprivation in area over time. The deprivation trajectory of an area could impact on the health outcomes of people living in those areas. In a recent US study of colorectal cancer incidence, Zhang et al. (2018) stratified areas based on neighbourhood deprivation at two time points 10 years apart. Long-term low neighbourhood socio-economic status and decreasing neighbourhood socio-economic status were associated with a higher risk of colorectal cancer compared with long-term high neighbourhood socio-economic status.

Geographic variation in colorectal cancer mortality

Geographic variations in colorectal cancer morbidity and mortality are less well described. Some regional variation in colorectal cancer incidence and mortality has been reported in England (Arik et al., 2020; Arik et al., 2021). Bayesian

analysis was used to estimate rates by region of England but rates were not estimated at sub-regional level. Less regional variation in colorectal cancer incidence and mortality compared to all-cancer and lung cancer, but more than breast and prostate cancer, was reported. Furthermore, little variation was found in colorectal cancer mortality rates by deprivation within regions of England. Geographic variation in colorectal cancer survival has been investigated by Cancer Network (Walters et al., 2011). The study reported a narrowing in the range of 1-year survival by geographic area over time for men.

Other factors (other than deprivation), such as rurality, may influence geographic variation in colorectal cancer outcomes. There is evidence for an association between accessibility to healthcare services and stage at disease at diagnosis, with both increasing travel time to GP surgery (Jones et al., 2008b) and distance to cancer centre (Campbell et al., 2001) associated with more advanced disease at diagnosis.

4.3 Objectives

- Describe and critique trends in colorectal cancer mortality in England and Wales at Local Authority level and by area-based socio-economic deprivation.
- Compare colorectal cancer mortality trends to all cancers combined.
- Identify whether there are statistically significant hot spots of colorectal cancer and all-cancer mortality rates by geographic area and over time.

4.4 Data

4.4.1 Death registration data

Annual death registration data was obtained from the UK Data Service from 1990 to 2012 at Local Authority (LA) level for England and Wales, with colorectal cancer as the underlying cause of death. Colorectal cancer was defined according to the World Health Organisation (WHO) International Classification of Diseases and health related problems (ICD), the international standard for mortality and morbidity statistics, versions 9 and 10 (World Health Organisation, 2019). For deaths prior to 2001 colorectal cancer was classified using ICD-9 codes 153 and 154. From 1 January 2001, ICD-10 was used to code cause of cause of death, replacing ICD-9. The ICD-10 codes C18-21 were used to define colorectal cancer.

LA was chosen as the geographic scale for the analysis as this is the lowest level of geography at which the mortality data for this time period is openly available. There are a total of 348 LAs in England and Wales.

The data set contained the annual number of deaths in each LA split by sex and five year age-group. There were two LAs, City of London and Isles of Scilly, for which there was no corresponding mortality data. These areas were also removed from the analysis which left a total of 346 LAs.

Annual mortality data for the same time period was obtained for all neoplasms at LA level, split by sex and five year age-group. This was to enable comparisons to be made between colorectal cancer and all-cancer mortality trends.

Rules for accessing death registration data changed post 2012 and a new application to access disclosive data is required to obtain an update. Published data for the period 2013-2018 contains supressed counts which led to problems in calculating directly-standardised mortality rates for colorectal cancer due to low counts when split by age group and geographic area. An application has been made to UKDS to access individual-level death registration data up to 2018 (see Appendix B).

4.4.2 Population estimates

Annual mid-year population estimates for the period 1990 to 2012 were sourced from the Office for National Statistics (ONS). The data contained the population by sex and five year age-group for each LA. Mid-year population estimates relate to the usually resident population on 30 June of each year.

4.4.3 Townsend Deprivation Index

Townsend Deprivation Index scores for LAs in England and Wales were used as an area-based measure of material deprivation. The Townsend Index was chosen as it can be calculated for all LAs in England and Wales, unlike other measures of area deprivation which are country-specific.

The Townsend Index is expressed as a single figure index which places each area's level of deprivation relative to a large reference area. In this case, the study area is England and Wales and the scores were calculated for each LA in 1991, 2001 and 2011.

Positive values of the index indicate more deprived areas, relative to the reference area, whereas negative values indicate relative affluence. A score of zero represents an area with mean values. To enable comparison of mortality trends by area type, the index scores were categorised into quintiles. Population weighted quintiles were generated by sorting the LAs, and corresponding mid-year populations, according to the index scores from negative to positive. The LAs were then grouped into quintiles (with 1 being the least deprived LAs and 5 being the most deprived LAs) so that the sum of the populations of each quintile is as equal as possible. Three sets of quintiles were assigned to each LA, based on the Townsend index scores calculated at each census year (1991, 2001 and 2011). These represent the start-, mid- and end-point of deprivation within the study period.

In order to measure changing deprivation over time, comparable Townsend scores over time were provided. These scores were calculated by apportioning the raw data for each census within the study time period (1991, 2001, 2011) from the original geographies at which they were released to be geographically consistent with the 2011 LA definitions for England and Wales (Norman and Darlington-Pollock, 2017). The data were apportioned using address count weighted postcode distributions (as a proxy for population distributions),

whereby the apportionment weight was determined by the proportion of postcodes in the overlapping area divided by all postcodes in the original geographic area. The input variables were standardised using z scores calculated as the observation for any one area in a year expressed relative to the mean and standard deviation for all areas and time points. Quintiles were calculated to have equal populations across all years. An increase or decrease in deprivation scores or quintiles thereby represents worsening or improving deprivation, respectively (Norman, 2017).

4.4.4 Population density

Population density was used as a proxy urban-rural measure in this analysis, as there are no consistently adopted UK-wide urban-rural measures.

Population density, measured as the number of persons per hectare, is derived by ONS from the number of persons at the census and area in hectares from the GIS Shapefile. Population density data was sourced for LAs in England and Wales at the 1991, 2001 and 2011 census. Each area was assigned a population density category based on the classification outlined in Chapter 3 Table 3.2. The classification comprises five categories, with 1 being the most urban and 5 being the most rural.

4.4.5 Boundary data

Digital boundary data in ESRI Shapefile format was sourced for Local Authorities and Regions in England and Wales from the ONS (Office for National Statistics, 2021d).

4.5 Methods

4.5.1 Descriptive statistics

Summary statistics showing the number and percentage of colorectal cancer (and all cancer) deaths by sex, age group, Townsend Deprivation Index score and population density category were produced.

4.5.2 Mortality rates

A time series of mortality rates by area type was constructed by first aggregating the mortality and population data across Townsend deprivation quintiles. Age-standardised mortality rates were calculated annually from 1990 to 2012 and for a 3-year rolling average. The rolling average was calculated by pooling 3 years of (mortality and population) data around each year within the study period, excluding 1990 and 2012 as data was not obtained outside of these bounds. A rolling average was taken in order to remove any yearly fluctuations in the mortality rates.

The direct method of standardisation was preferred over the indirect method in this study to allow for valid comparisons to be made between geographical areas and over time. The mortality and population data was split by age group so it was possible to calculate age-specific mortality rates.

The calculation of the age-standardised rate is below. Deaths are expressed per 100,000 persons:

$$\text{Age-standardised rate} = \frac{\sum(P_k m_k)}{\sum P_k}$$

Where:

P_k = Standard population in sex/age group k

m_k = Observed mortality rate (deaths per 100,000) in sex/age group k

k = age/sex group 0-4, 5-9, … , 80-84, 85 years and over

Directly age-standardised mortality rates were calculated using the 2013 European Standard Population (ESP) (Eurostat, 2013). Age-standardised

mortality rates were calculated separately for males and females and for premature mortality, defined (according to the definition used by Public Health England) as deaths before the age of 75.

To investigate mortality trends by rurality, the data were also aggregated by population density categories and a time-series of directly age-standardised rates calculated.

Mortality rates were also calculated at LA level for each year and using 3-years of data pooled around each census year.

4.5.3 Mapping

Choropleth maps were produced, for each census year, to visualise the variation in mortality rates at LA level in England and Wales.

4.5.4 Spatial cluster analysis

Hot spot analysis was employed in this analysis to identify areas with an excess of colorectal cancer deaths. The input data was the directly-standardised mortality rates by LA.

The local G-statistic was calculated for each LA within England and Wales to identify the locations of hot and cold spots of colorectal and all-cancer mortality rates. Three years' worth of mortality and population data was pooled around each census year. The neighbourhood around each LA was defined based on a 50 kilometre radius. This method is recommended when there is a large variation in polygon size (as is the case for LAs) and a consistent scale of analysis is desired. The resultant z-scores were mapped to visualise the location of hot and cold spots of high or low colorectal and all-cancer mortality.

Statistical significance in this study was considered at the confidence level of 95% or higher, associated with a p-value <0.05 and z-score > 1.96. The analysis was carried out in ArcGIS (Esri, 2015). The results of the spatial cluster analysis were adjusted for multiple hypothesis testing using the false discovery rate (FDR) method. The FDR method reduces the p-value threshold which determines the statistical significance, to account for multiple testing. Multiple testing is an issue in local spatial statistics as a test is performed for every feature (in this case LA) in the dataset, which may increase the number of false positives i.e. spatial clusters identified which are in fact random.

4.6 Results

4.6.1 Descriptive statistics

There were 341,234 colorectal deaths in England and Wales between 1990 and 2012. Over the same time period, there were 2.8 million deaths from all cancers (other than colorectal cancer) combined Table 4.1.

The proportion of colorectal cancer deaths was slightly higher in males (52%), compared to females. Just under half (47%) of colorectal cancer deaths were at premature age (before age 75). There were 70,408 deaths (21%) in the least deprived quintile compared to 55,716 deaths (16%) in the most deprived quintile based on Townsend Deprivation Index scores at the 2001 census. The highest proportion of colorectal cancer deaths were in areas categorised as “Rural” (51%).

Colorectal cancer deaths accounted for approximately 11% of all-cancer deaths. Compared to deaths from all-cancers combined (excluding colorectal cancer deaths), a higher proportion of colorectal cancer deaths were in the older age groups. Fifteen per cent of all-cancer deaths were among those aged 85 and over, compared to 20% of colorectal cancer deaths. A higher proportion of all-cancer deaths were in the most deprived areas (18%), compared to colorectal cancer deaths (16%). The distribution of colorectal cancer and all-cancer deaths was similar by population density category, but there with a slightly higher proportion of all-cancer deaths in the most urban areas, and a slightly higher proportion of colorectal cancer deaths in the most rural areas.

The overall directly standardised colorectal cancer mortality rate for England and Wales was 41.9 deaths per 100,000 (95% CI 41.2-42.5) at the start of the study period in 1990 and 28.6 deaths per 100,000 (95% CI 28.1-29.0) at the end of the study period in 2012. For males the corresponding rates were 52.9 per 100,000 (95% CI 51.8-54.1) in 1990 and 36.0 per 100,000 (95% CI 35.2-36.8) in 2012 and for females 34.9 per 100,000 (95% CI 34.2-35.7) and 22.8 per 100,000 (95% CI 22.3-23.4).

Characteristics		CRC deaths		All cancer deaths (excluding CRC)	
		n	%	n	%
Total		341,234		2,820,194	
Sex	Male	177,890	52.1	1,473,322	52.2
	Female	163,344	47.9	1,346,872	47.8
Age group	0-50	11,467	3.4	154,079	5.5
	50-55	10,783	3.2	109,905	3.9
	55-60	18,430	5.4	170,973	6.1
	60-65	28,060	8.2	255,415	9.1
	65-70	39,758	11.7	350,303	12.4
	70-75	51,263	15.0	439,132	15.6
	75-80	58,012	17.0	476,330	16.9
	80-85	56,895	16.7	429,216	15.2
	85+	66,566	19.5	434,841	15.4
Townsend deprivation score	1 - Least deprived	70,408	20.6	547,369	19.4
	2	74,090	21.7	591,295	21.0
	3	72,499	21.2	588,001	20.8
	4	68,521	20.1	586,441	20.8
	5 - Most deprived	55,716	16.3	507,088	18.0
Population density	Most urban	59,368	17.4	529,078	18.8
	Very urban	17,887	5.2	145,263	5.2
	Urban	61,740	18.1	525,404	18.6
	Rural	173,237	50.8	1,398,279	49.6
	Most rural	29,002	8.5	222,170	7.9

Table 4.1 Colorectal cancer and all-cancer (excluding colorectal) deaths by demographic characteristics

4.6.2 Geographic variation in cancer mortality

The mortality rates by LA presented below are based on 3-years of data pooled around 2011 (the most recent Census year). Rates for 1991 and 2001 have also been mapped, using comparable quintiles, to show the trends in mortality rates over time.

The spatial pattern of colorectal cancer mortality in England and Wales is displayed in Figure 4.1. Mortality rates ranged from 14 deaths per 100,000 people (95% CI 9.5-18.3) in Rugby to 52 deaths per 100,000 people (95% CI 41.3-62.3) in Eden. Higher mortality rates were generally found in coastal and peripheral areas, but there was not a clear spatial pattern.

Figure 4.2 shows the colorectal cancer mortality rates on a comparable scale for 1991, 2001 and 2011. The mortality rates have decreased in most areas over the time period, but there are some areas with consistently higher mortality rates.

There was a clearer spatial pattern in all-cancer mortality (Figure 4.3), compared to colorectal cancer mortality. High rates of all-cancer mortality were found in the north of England and south Wales. All-cancer mortality rates were generally lower in the south and east of England, except for some higher cancer rates in some coastal areas in the south.

When comparing all-cancer mortality rates over time (Figure 4.4), despite generally decreasing mortality rates there were some areas in the north of England with consistently higher mortality rates.

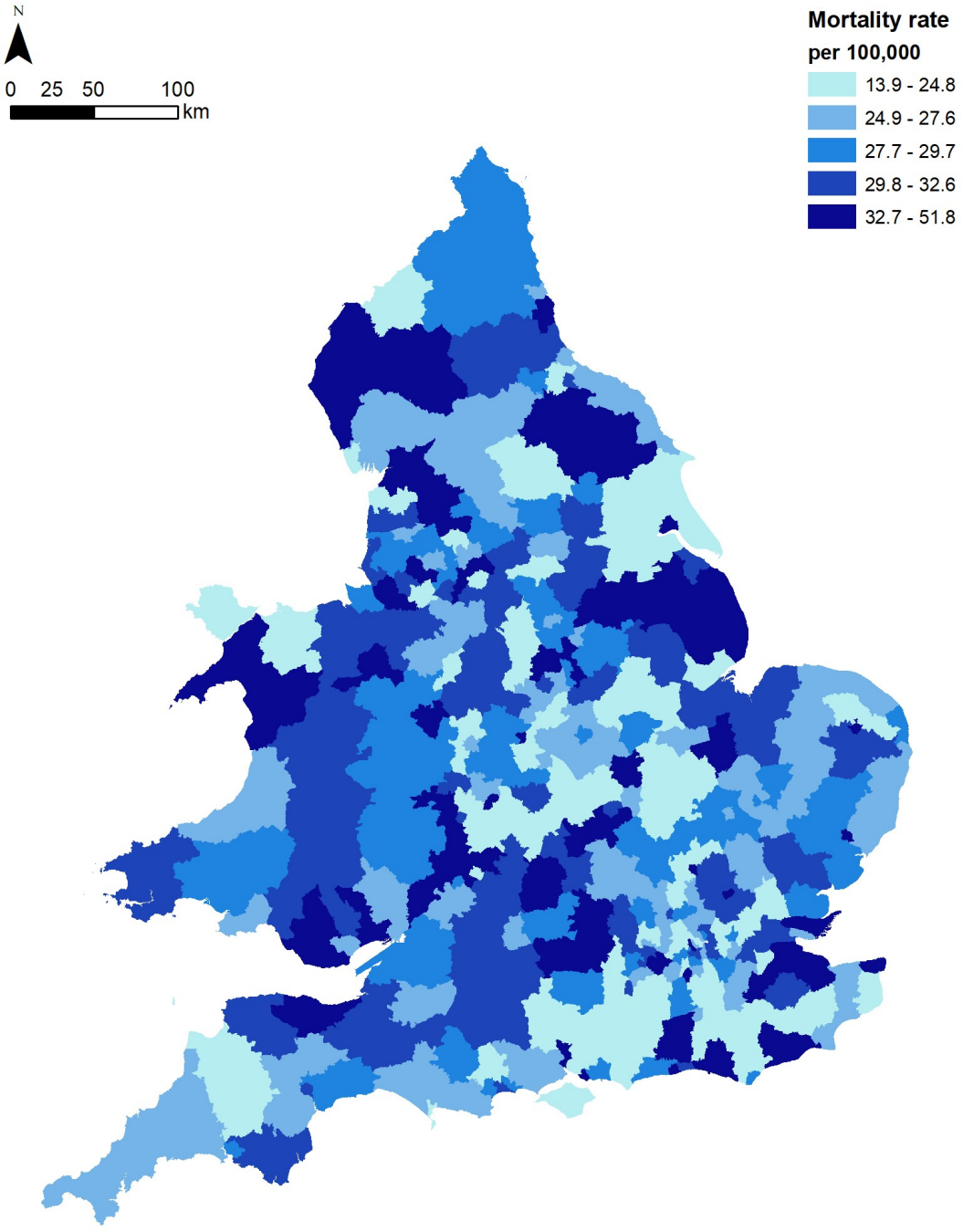


Figure 4.1 Age-standardised colorectal cancer mortality rates, by Local Authority, England and Wales, 2011

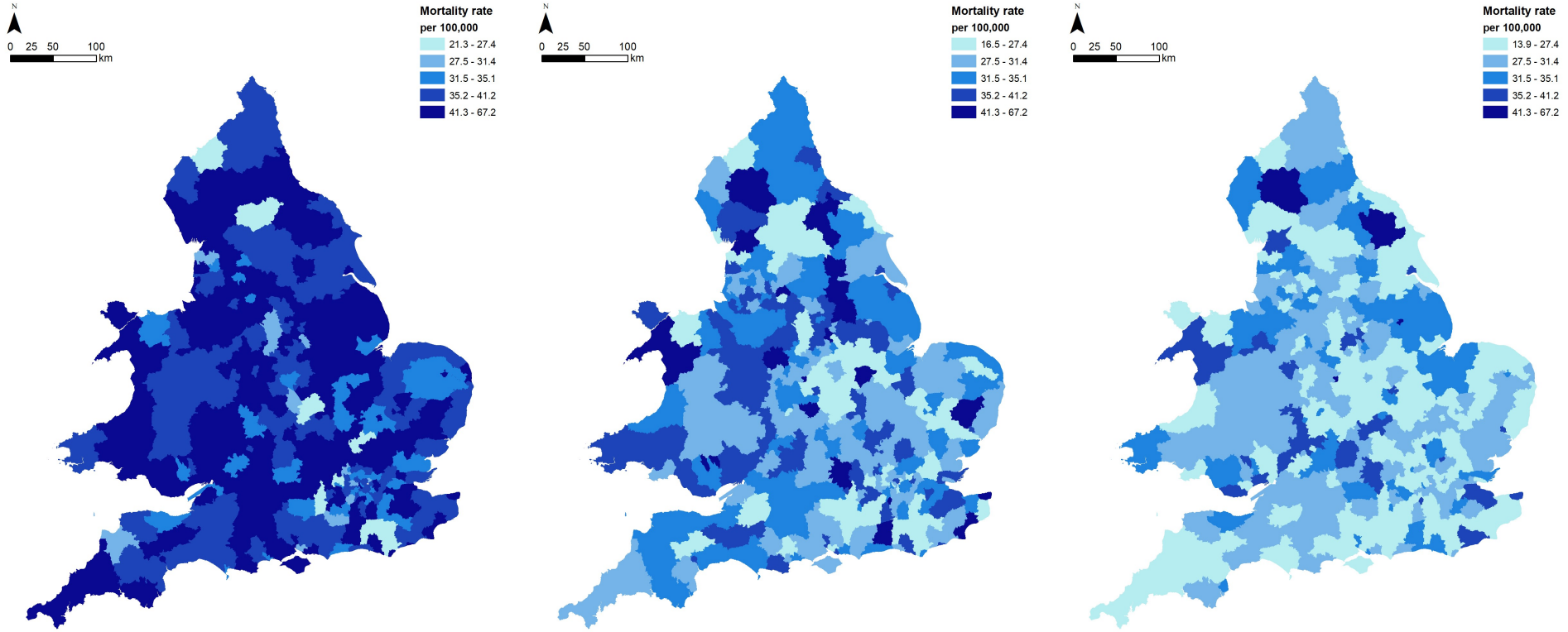


Figure 4.2 Age-standardised colorectal cancer mortality rates, by Local Authority, England and Wales: 1991, 2001, 2011

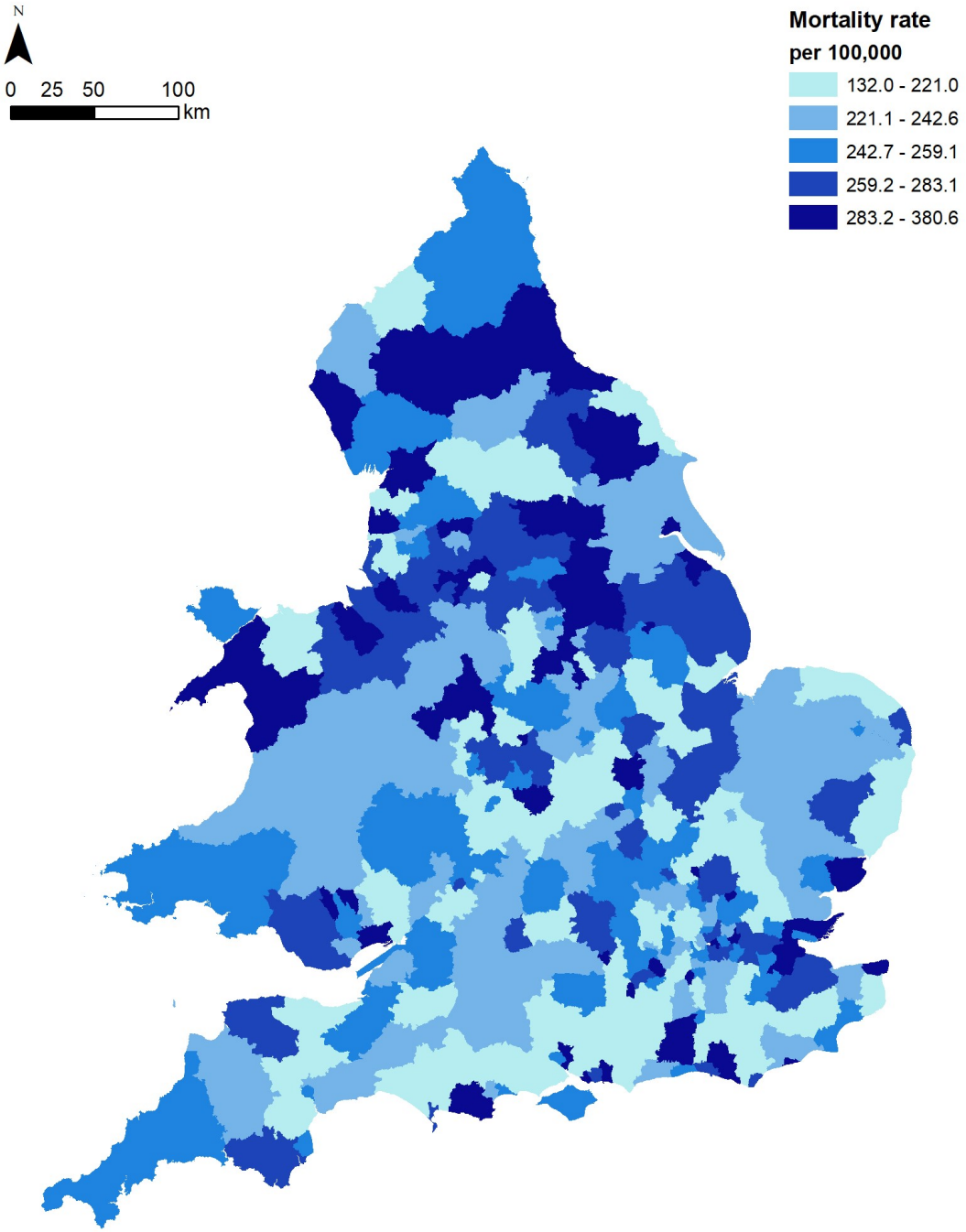


Figure 4.3 Age-standardised all-cancer (excluding colorectal) mortality rates, by Local Authority, England and Wales, 2011

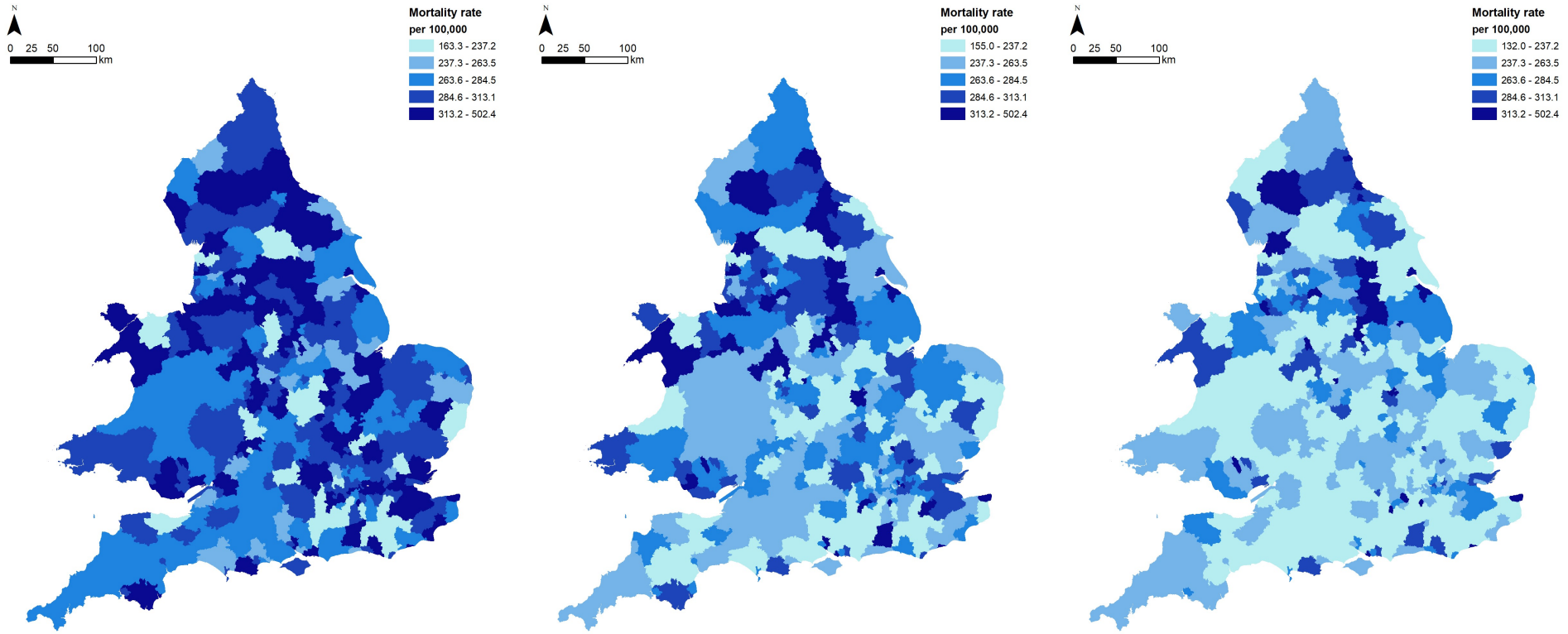


Figure 4.4 Age-standardised all-cancer (excluding colorectal) mortality rates, by Local Authority, England and Wales: 1991, 2001, 2011

4.6.3 Spatial cluster analysis

The spatial variability in colorectal cancer and all-cancer mortality was quantified using the Getis-Ord Gi* statistic. The mortality rates mapped in Figure 4.5 and Figure 4.6 are based on 3-years of mortality data pooled around each census year within the study period (1991, 2001 and 2011).

Statistically significant hot spots of high colorectal cancer mortality in 2011 were identified in south Wales, Lincolnshire and Cumbria. Conversely, statistically significant cold spots of low colorectal cancer mortality rates were found in the south east of England, centred around Surrey and Local Authority areas in London, Hertfordshire and Buckinghamshire. There was also a statistically significant cold spot of colorectal cancer mortality in the East Midlands, around the Local Authority of Corby.

A similar spatial pattern was observed at each of the census time points. The cold spot of low colorectal cancer mortality rates in south-east England was more pronounced in 1991 and 2001 and in the Midlands in 2001. Hot spots of high colorectal cancer mortality rates were identified in Lincolnshire, Cheshire and the north east of England, particularly in 2001. The hot spot in south Wales was not significant in 1991.

After adjustment for multiple hypothesis testing the clusters identified in 2011 were no longer statistically significant. The cold spots of low colorectal cancer mortality rates around London in 1991 and 2001 remained significant, as did the hot spot of high colorectal cancer mortality in Lincolnshire (in 1991 and 2001) and south Wales (in 2001).

Statistically significant hotspots of all-cancer mortality (Figure 4.6) were identified in the North West and north east of England, Lincolnshire and south Yorkshire. Cold spots of low all-cancer mortality were identified in the south east and south west of England. There was a similar pattern at the two previous census time points, although the cold spot in the south east of England was more pronounced in 2011. After adjustment for multiple hypothesis testing the hot spots in the North West and north east of England and the cold spot in the south east remained statistically significant (although less pronounced).

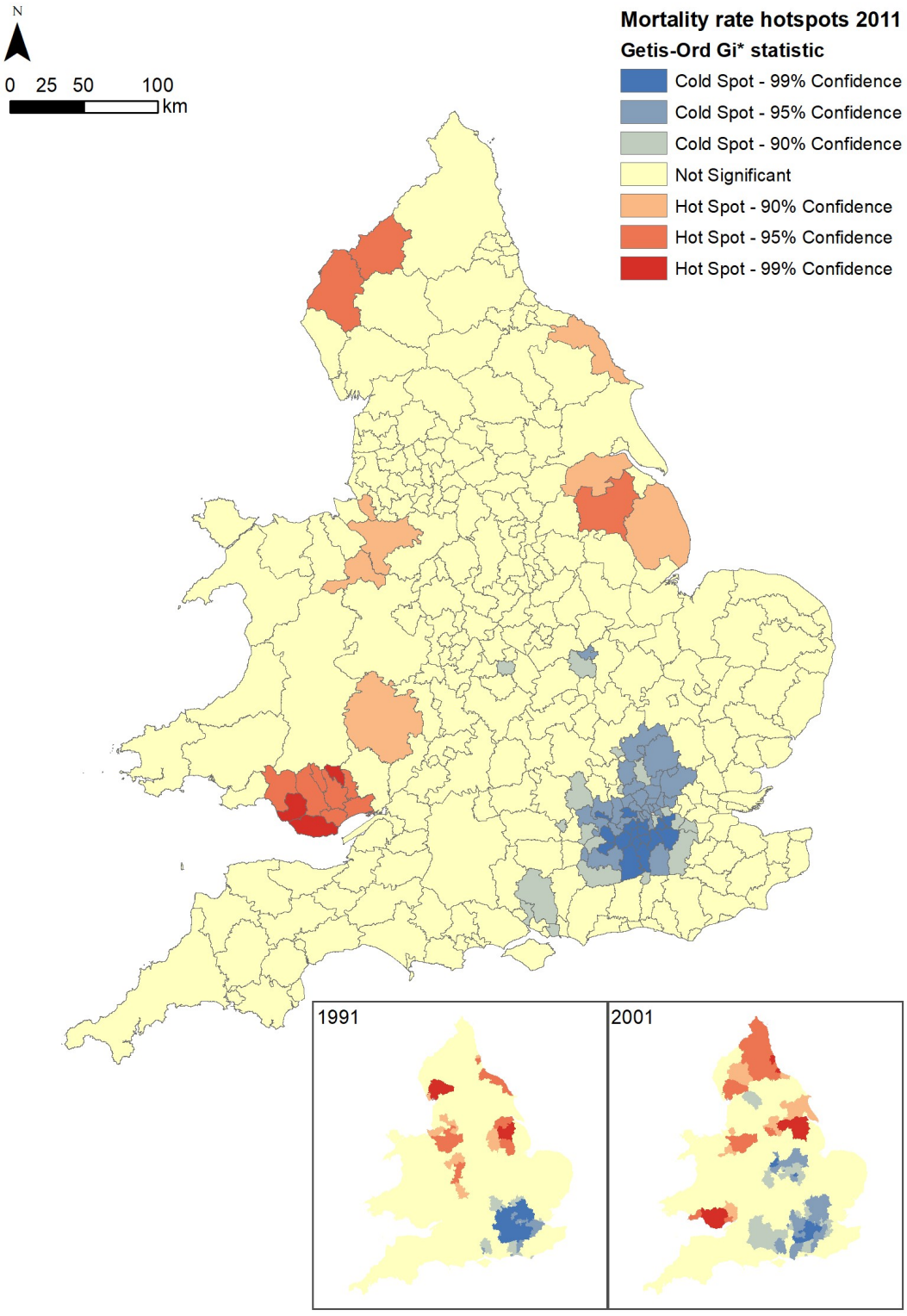


Figure 4.5 Hot spot analysis (Getis-Ord Gi*), 2011 colorectal cancer mortality, England and Wales

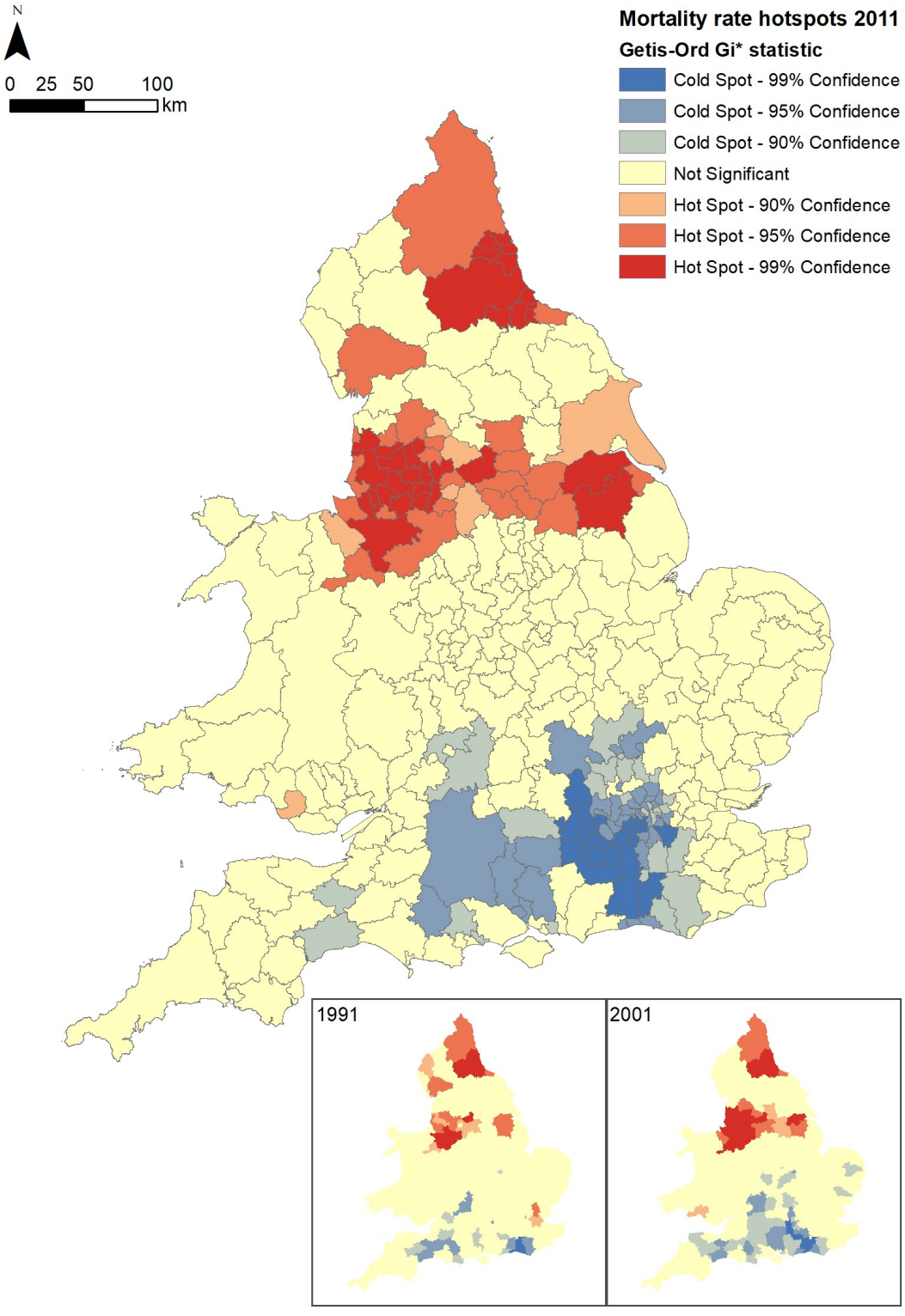


Figure 4.6 Hot spot analysis (Getis-Ord Gi*), 2011 all-cancer (excluding colorectal) mortality, England and Wales

4.6.4 Cancer mortality by population density

The spatial distribution of colorectal cancer mortality rates showed some higher rates in rural areas. To investigate this further, the rates have been stratified by population density, used as a proxy for rurality. The lowest colorectal cancer mortality rates were consistently found in the most densely populated areas, but there was not a clear gradient by population density (Figure 4.7 and Table 4.2).

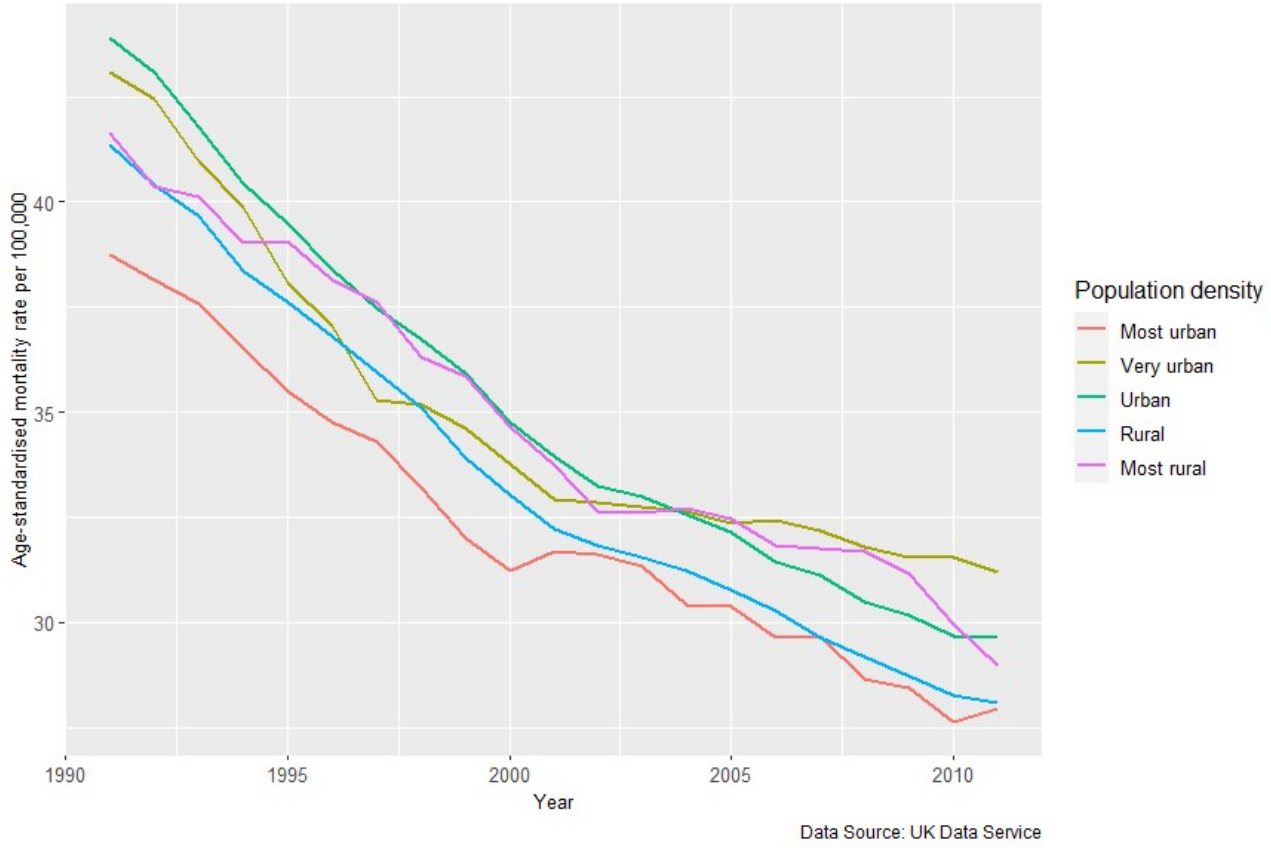


Figure 4.7 Age-standardised colorectal cancer mortality rates, by population density, England and Wales, 1990-2012

	1991 ASMR (95% CI)	2001	2011
Most urban	38.8 (38.0-39.5)	31.7 (31.0-32.4)	27.9 (27.3-28.6)
Very urban	43.1 (41.5-44.7)	32.9 (31.5-34.3)	31.2 (29.9-32.5)
Urban	43.9 (43.0-44.8)	33.9 (33.2-34.7)	29.7 (29.0-30.3)
Rural	41.4 (40.8-41.9)	32.2 (31.8-32.7)	28.1 (27.7-28.4)
Most rural	41.6 (40.3-42.9)	33.7 (32.6-34.8)	29.0 (28.1-29.9)

Table 4.2 Age-standardised colorectal cancer mortality rates, by population density and census year, England and Wales

All-cancer mortality rates were more clearly stratified by categories of population density (Figure 4.8). Mortality rates were consistently higher in the urban categories (Urban, Very urban and Most urban) compared to the rural categories (Most rural and Rural), although there is not a gradient across all five categories. Mortality rates in areas categorised as “Very urban” have not decreased at the same rate as other areas, which is also apparent for colorectal cancer mortality rates in these areas.

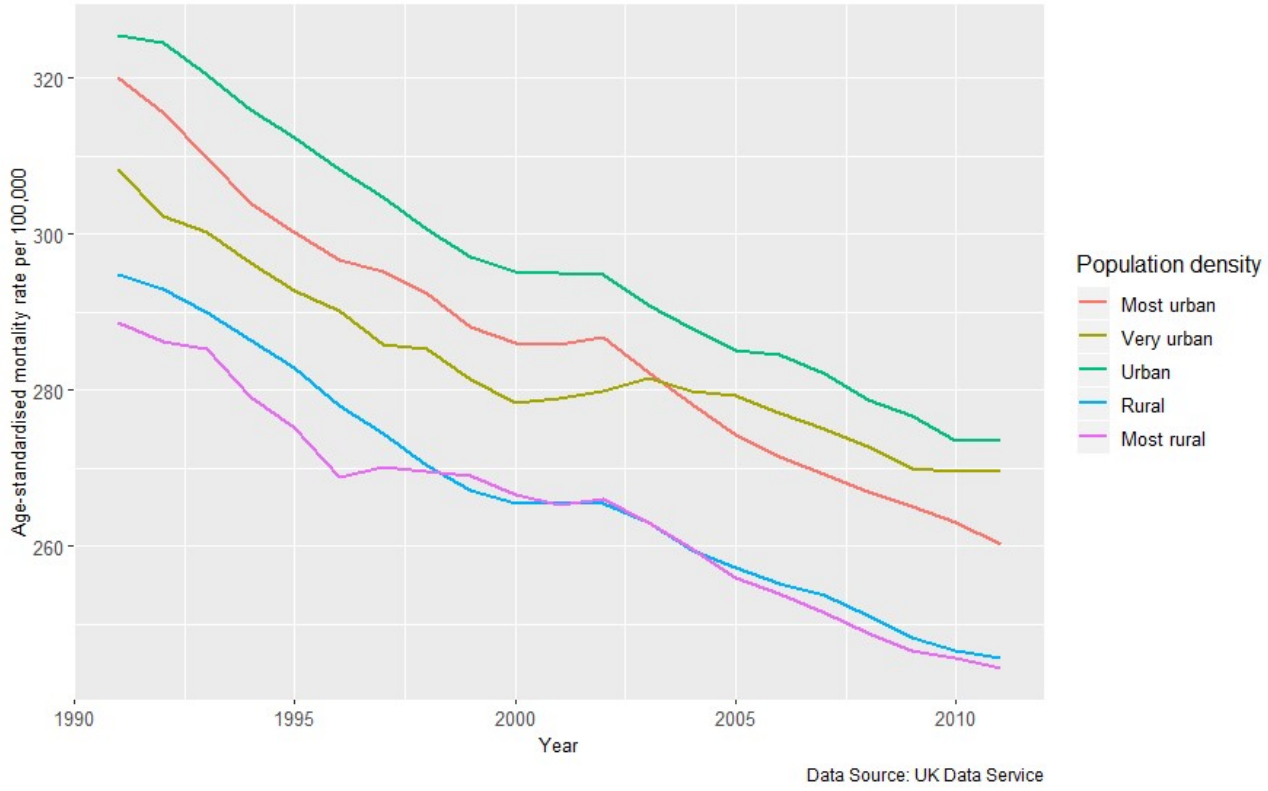


Figure 4.8 Age-standardised all-cancer (excluding colorectal) mortality rates, by population density, England and Wales, 1990-2012

4.6.5 Cancer mortality by area deprivation

This section examines variation in colorectal and all-cancer mortality by the Townsend Deprivation Index. The trends presented below are based on the mid-point of deprivation (2001 Townsend quintiles) as this was judged to best represent the level of deprivation in LAs across the study period. Results for the 3-year rolling average of mortality rates are presented here as there were some fluctuations in the annual mortality rates, due to small numbers of deaths when split by year and geographic area.

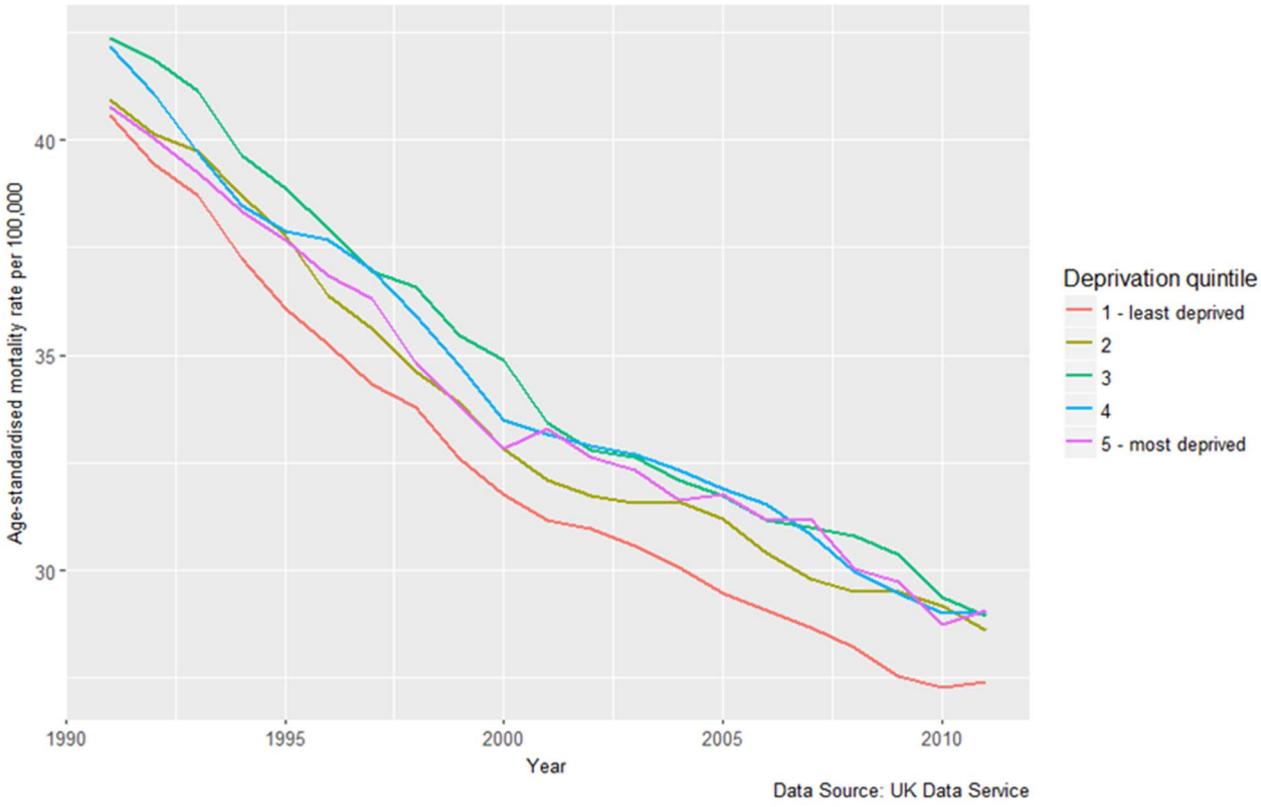


Figure 4.9 Age-standardised colorectal cancer mortality rates, by deprivation quintile, England and Wales, 1990-2012

	1991 ASMR (95% CI)	2001	2011
Quintile 1 – least deprived	40.6 (39.8-41.4)	31.2 (30.5-31.8)	27.4 (26.8-28.0)
Quintile 2	40.9 (40.2-41.7)	32.1 (31.4-32.8)	28.6 (28.0-29.2)
Quintile 3	42.4 (41.6-43.2)	33.4 (32.8-34.1)	29.0 (28.4-29.6)
Quintile 4	42.2 (41.4-43.0)	33.2 (32.4-33.9)	29.0 (28.4-29.7)
Quintile 5 – most deprived	40.8 (39.9-41.6)	33.3 (32.5-34.1)	29.1 (28.4-29.8)

Table 4.3 Age-standardised colorectal cancer mortality rates, by deprivation quintile and census year, England and Wales

There was a general pattern of improving colorectal cancer mortality between 1990 and 2012. Mortality rates fell rapidly over the time period across all deprivation quintiles (Figure 4.9 and Table 4.3). The lowest mortality rates were observed among the least deprived group throughout the study period. There was no consistent difference in the mortality rates of the other four groups. By the end of the time period the mortality rates among these four groups had converged and were significantly higher than that in the least deprived group (29.1 deaths per 100,000 people in the most deprived group compared to 27.4 deaths per 100,000 people in the least deprived group).

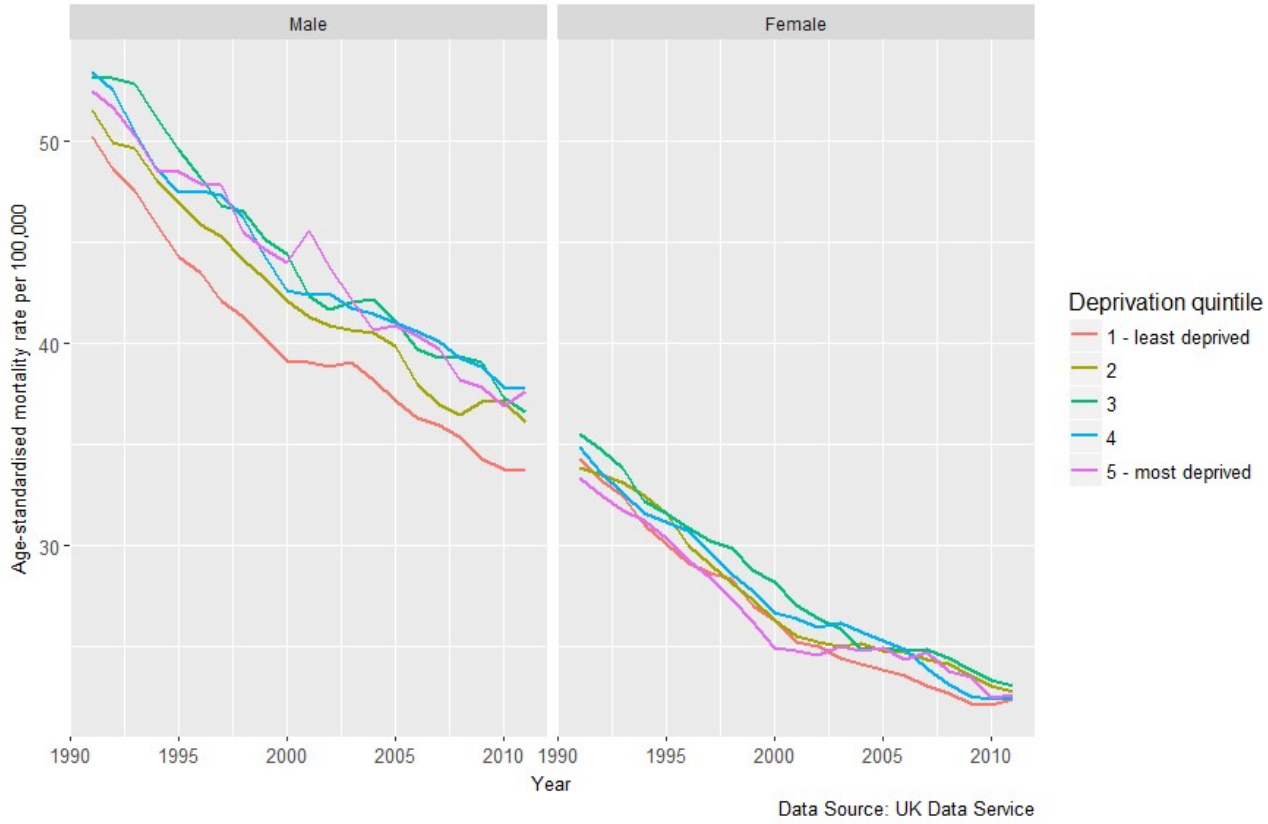


Figure 4.10 Age-standardised colorectal cancer mortality rates, by sex and deprivation quintile, England and Wales, 1990-2012

Both males and females experienced a considerable decline in colorectal cancer mortality rates over the study period, however, much higher mortality rates were consistently observed among males compared to females. The lowest mortality rates were found in the least deprived quintile, followed by the

second least deprived quintile in men, however there was very little variation in female mortality rates by deprivation quintile (Figure 4.10).

Figure 4.11 shows rates of premature mortality by deprivation quintile. Premature mortality is defined as deaths before age 75. The lowest rates of premature mortality were observed in the least deprived quintile throughout the study period, followed by the next least deprived quintile. There was little variation in mortality rates among the remaining three deprivation quintiles.

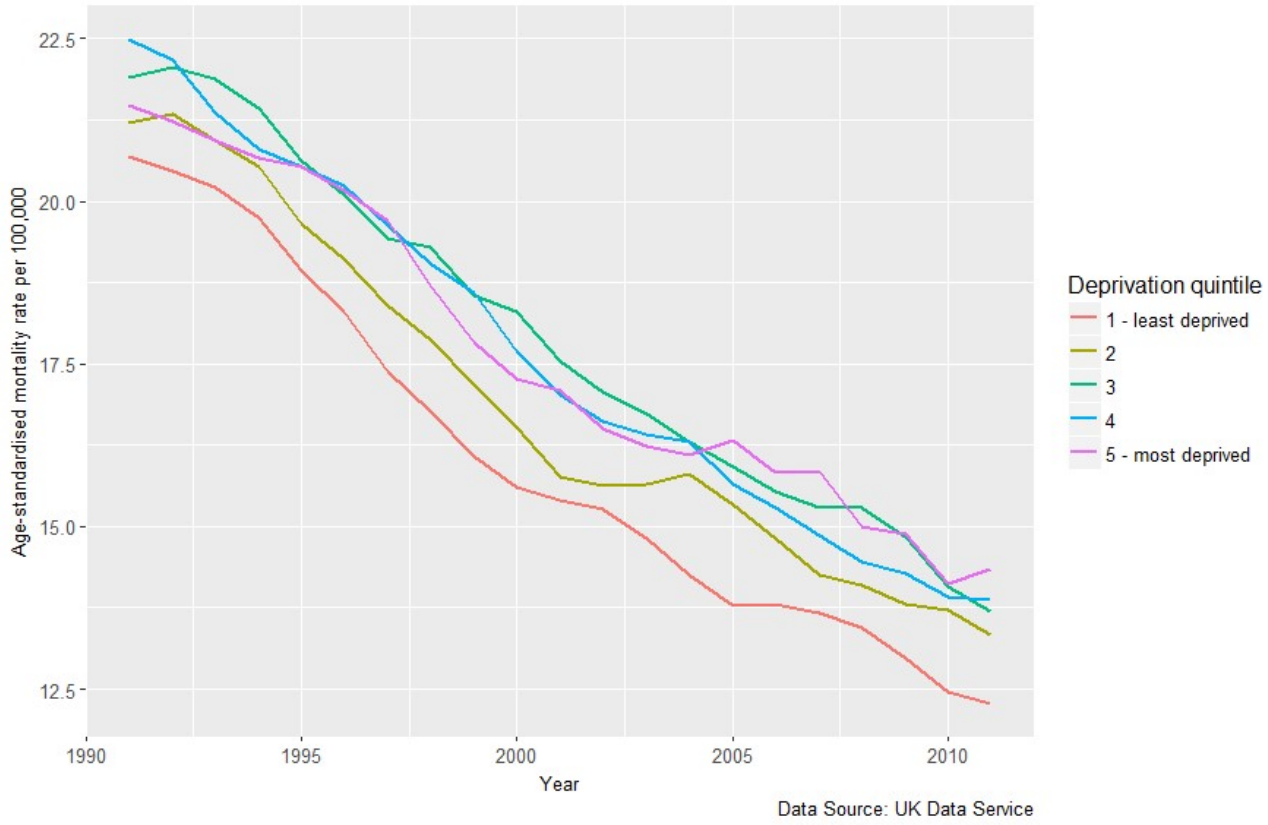


Figure 4.11 Age-standardised premature colorectal cancer mortality rates, by deprivation quintile, England and Wales, 1990-2012

When split by gender (Figure 4.12) the lower rates of premature mortality found in the two least deprived quintiles was observed in males but not in females. There was very little variation in female premature mortality rates by deprivation quintile, whereas in males rates were lower in the two least deprived quintiles but there was little variation in the remaining three quintiles.

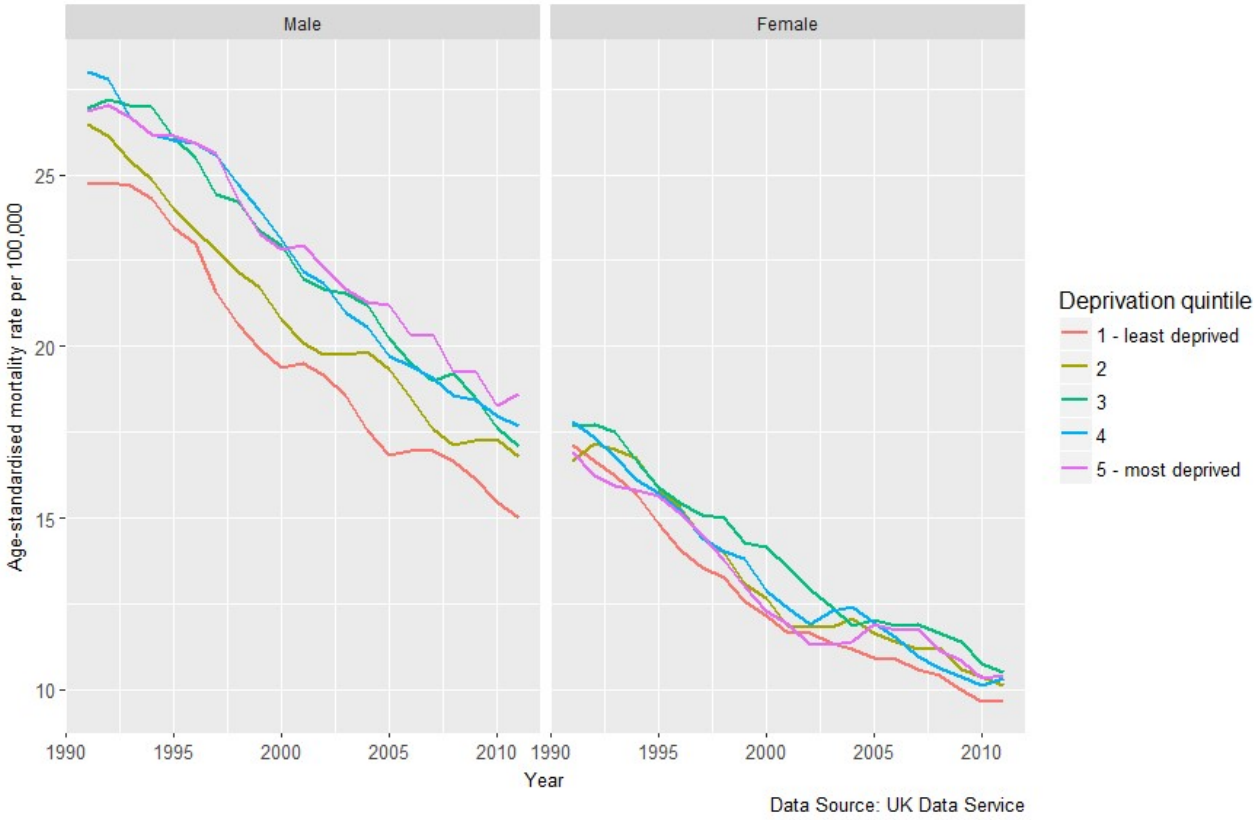


Figure 4.12 Age-standardised premature colorectal cancer mortality rates, by sex and deprivation quintile, England and Wales, 1990-2012

Mortality trends for all cancers (Figure 4.13) across the same time period are presented below. The figures are based on the number of cancer deaths minus colorectal cancer deaths. As for colorectal cancer, there was a sharp decline in all-cancer mortality rates across all deprivation quintiles during the study period, however the rates appeared to start levelling off after 2010. There was a more pronounced deprivation gradient in all-cancer mortality rates compared to those for colorectal cancer, with the lowest rates observed among the least deprived quintile and the highest rates found among the most deprived quintile (Figure 4.13). In comparison, there was little variation in colorectal cancer mortality rates by deprivation quintile, apart from lower rates among the most deprived group. The slight increase in all-cancer mortality rates in the early 2000s could be due to the change in coding from ICD-9 to ICD-10 in 2001.

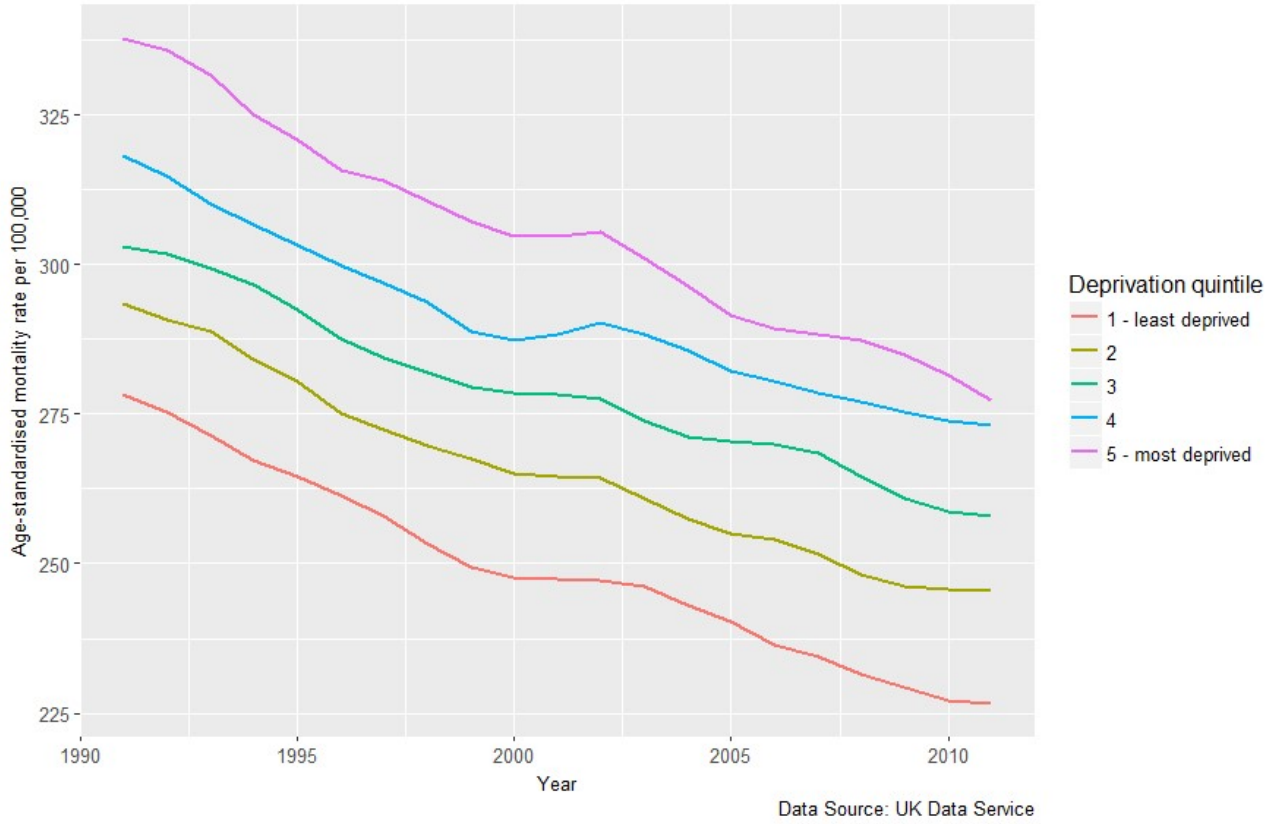


Figure 4.13 Age-standardised all-cancer (excluding colorectal) mortality rates, by deprivation quintile, England and Wales, 1990-2012

4.7 Discussion

4.7.1 Summary

There was some variation in colorectal cancer mortality rates by Local Authority area, but no clear spatial pattern was observed. This was in contrast to all-cancers combined for which mortality rates were generally higher in the north of England and south Wales and lower in the south of England, except for some higher rates in coastal areas. These findings concur with regional level analysis which found less geographic variation in colorectal cancer compared to all-cancer combined (Arik et al., 2021). There was no clear gradient in colorectal cancer mortality rates by population density, despite evidence that colorectal cancer patients in rural areas are less likely to receive timely treatment (Jones et al., 2008a).

The apparent north-south divide in all-cancer mortality compared to colorectal cancer mortality could be explained by a stronger relationship with area deprivation.

There has been a downward trend in colorectal cancer mortality in England and Wales across all deprivation groups between 1990 and 2012. The pattern of colorectal cancer mortality rates by deprivation differs to that for all cancers. There is a clear deprivation gradient in all-cancer mortality rates, whereas colorectal cancer mortality rates are lowest among the least deprived group but there is little variation among the other groups.

Possible explanations for lower colorectal cancer mortality rates in the least deprived areas could include differences in stage at diagnosis, co-morbidities and access to diagnostic and treatment services between areas. There is some evidence that patients in deprived areas are more likely to present at a late stage (Lejeune et al., 2010) but findings have been inconsistent (Schrijvers et al., 1995). Differential uptake of screening by demographic group has also been reported (McGregor et al., 2016), with lower rates of uptake in the most deprived area. It may be too early to detect the impact of the Bowel Cancer Screening Programme on colorectal cancer mortality in this data set as the programme was only introduced in England in 2006. It would be interesting to investigate this with more up-to-date data.

Colorectal cancer mortality rates were consistently higher in males, compared to females over the study period. This is in line with reported national trends in colorectal cancer mortality rates, which are significantly higher for males than for females (Cancer Research UK, 2021). A review of sex-related differences in colorectal cancer by White et al. (2018) suggested the higher colorectal cancer mortality rate in males is primarily due to the higher incidence rate in males, compared to females. The review found that there were relatively small differences by sex in route to diagnosis, cancer stage at diagnosis and 5-year survival, but sex and gender differences in biological and behavioural factors were more apparent in the pathway up to the point of colorectal cancer diagnosis. For example, men are more likely to have a diet high in red and processed meat (Public Health England, 2016), are more likely to be heavy consumers of alcohol (Schutze et al., 2011) and are more likely to smoke (Chang et al., 2014), all of which are associated with an increased risk of colorectal cancer. Furthermore, bowel cancer screening uptake is lower in men than women (von Wagner et al., 2011), although women are more likely to be diagnosed via an emergency presentation than men (White et al., 2018).

In this study, trends in colorectal cancer mortality by deprivation varied by gender. Male colorectal cancer mortality rates were lower in the two least deprived Townsend quintiles, but there was little variation in female mortality rates by deprivation. This could reflect the lower incidence rates among males living in the least deprived areas compared to the most deprived areas (National Cancer Intelligence Network, 2014), while for females the incidence rates are similar for those living in the least and most deprived areas. The sex differences in incidence could be due to the increased likelihood of men in deprived areas to have a lifestyle associated with colorectal cancer risk factors (Office for National Statistics, 2017).

4.7.2 Strengths

A strength of this study is the comprehensive death registration data, which includes all deaths in England and Wales between 1990 and 2012. The 20-year time period allows for mortality trends over time to be examined.

While access to more detailed geographic information was not possible within the time scales of this thesis, the available data did allow for sub-regional level

analysis which extends previous research at regional level. Advanced spatial analysis techniques were applied to identify whether there was evidence of spatial clustering of colorectal cancer mortality rates in England and Wales.

4.7.3 Limitations

A challenge of analysing mortality trends for specific-causes of death at sub-regional level is that there are often small numbers of deaths when the data is split by area type and five-year age band. The analysis was repeated using ten-year age bands which generated very similar patterns by deprivation. The analysis by LA is sensitive to low numbers of deaths when split by LA and age band. To address this, the mortality rates in this analysis were calculated based on data pooled across 3-years.

This study used an area-based measure of deprivation (Townsend deprivation score). Not all individuals within each LA will experience the same level of deprivation, therefore the association between mortality rates and deprivation at the individual level cannot be determined from associations at LA level (ecological fallacy).

There are limitations to the Townsend Index as a measure of deprivation. The inclusion of non-car ownership in the index has been critiqued as in heavily urbanised areas with good provision of public transport, particularly in London, many people choose not to own a car irrespective of economic background. While it is a strength of the Townsend Index that it can be calculated in the same way at different time points, societal changes may mean some of the input measure are not comparable over time, for example due to changes in the nature of home ownership.

A considerable amount of information is lost when calculating summary measures such as standardised mortality rates which may conceal heterogeneity. For example, mortality differences between populations may be much greater in older ages. The choice of standard population can also affect the relative magnitude of standardised rates depending on which age groups are weighted most heavily. Similarly, the results of spatial clustering analysis are heavily dependent on the conceptualisation of the spatial relationship between the features (i.e. the neighbourhood definition). Further sensitivity analysis could have been carried out to investigate the impact of using different

neighbourhood definitions on the results. Results could also be compared using different spatial clustering techniques.

The change in ICD codes from version 9 to version 10 in 2001 may have affected how some of the mortality cases were coded, however, results from the bridge coding study showed the change only had a small impact for coding of colorectal tumours (Cook et al., 2002).

A major limitation of this study was the lack of mortality data for the recent time period. An application has been made to the UKDS (see Appendix B) to obtain access to individual mortality records for the period up to 2018. Individual-level data are required in order to append area-types to the mortality records at smaller geographic units prior to aggregation, which is not possibly using the openly published data.

4.7.4 Further work

A challenge in this type of analysis is the use of decennial measures of deprivation for an annual time series of outcomes. The Townsend deprivation index is based on census variables and therefore can only be calculated at each census time point, whereas the mortality figures are reported for each year. A mid-point of deprivation was used in this study (2001 Townsend deprivation quintiles), however this assumes that the level of deprivation in an area stayed the same over the time period. In reality, areas may become more or less deprived over time. Some studies have sought to overcome this by incorporating area trajectories into the methodology design. For example, by comparing the relative deprivation of an area at the start and end of the time period and grouping areas based on the deprivation trajectory (Zhang et al., 2018). Whilst this captures changing deprivation across the whole time period it does not account for the intervening period.

Work is planned to extend the analysis presented here to examine the deprivation trajectories of LAs, based on the Townsend deprivation scores at each census time point within the study period. Time comparable measures of the Townsend Index have been calculated, using raw census data for each census from 1981 to 2011, whereby each variable is expressed as a z-score relative to the mean and standard deviation across all areas and over time. Accordingly, a reduction in scores represents an improvement in relative

deprivation, whereas an increase in score represents worsening levels of deprivation in an area. The time-comparable scores were then been classified into groups, using a k-means cluster analysis method, to capture similarities in deprivation trajectories for LAs in England and Wales. Cluster solutions with different numbers of clusters were tested to find the optimum number of clusters with distinct deprivation trajectories. A time-series of mortality rates by deprivation clusters will then be calculated (as in this study for the cross-sectional measures of deprivation).

Additionally, the LAs will be stratified by their mortality trajectories to identify areas with persistently high or low mortality that could not be explained by factors already investigated (i.e. level of deprivation or rurality). A k-means cluster analysis will be performed again, this time using annual mortality rates as the input variables. Census data will then be used to describe the characteristics of these areas and determine whether there are distinctive characteristics of areas with high (or low) colorectal cancer mortality rates. A further question will be to determine if London is different to other areas. Areas in London are often found to have better health than expected given their level of deprivation. This relationship could impact on the observed pattern of colorectal cancer by deprivation. It may be necessary to separate areas in London from the rest of the country.

This work is awaiting access to the more recent and granular data. The more recent data would also be used to update the trends presented here and append the Townsend scores at smaller geographic units.

In addition to census data, other sources of area-based data could be used to explore the other factors which may influence colorectal cancer mortality. For example, access to healthcare services and the location of specialist cancer care centres and the prevalence of risk factors associated with colorectal cancer such as obesity rates, levels of physical activity and diet and food consumption patterns. This is considered further in later discussions.

4.8 Conclusion

Key messages:

- There was some geographic variation in colorectal cancer mortality rates, but no clear spatial pattern was observed at Local Authority level. A clearer spatial pattern was seen for all-cancers combined.
- There was no gradient in colorectal cancer mortality rates by area deprivation
- There was little evidence of spatial clustering of colorectal cancer mortality rates

Chapter 5
Variation in colorectal cancer incidence and survival by socio-economic position

5.1 Overview

This chapter will investigate variations in colorectal cancer (CRC) incidence and survival by indicators of socio-economic status in the ONS Longitudinal Study. While colorectal cancer mortality trends were previously examined by area-based measures of deprivation (Chapter 4), the LS enables analysis by individual characteristics such as educational attainment, social class and housing tenure along with linkages to small area deprivation. Furthermore, the longitudinal nature of the LS, which links individual records across multiple censuses, facilitates survival analysis over a long follow-up period. This is not possible with the time-series of cross-sectional mortality data as used in the previous chapter.

5.2 Background

5.2.1 Trends in colorectal cancer incidence by deprivation

Findings from previous studies investigating the relationship between socio-economic deprivation and colorectal cancer incidence have been inconsistent. There is evidence that the association between socio-economic deprivation and colorectal cancer incidence has changed direction over recent decades. In the 1980s, affluence was associated with an increased risk of colon and rectal cancer in Europe (Faggiano et al., 1997). In an analysis of colorectal cancer cases diagnosed between 1987 and 1992 in the South Thames region of England, however, no association was found between deprivation and incidence of colorectal cancer (Pollock and Vickers, 1997). The study used an area-based proxy indicator of deprivation status, the Townsend deprivation index, derived from 1991 Census variables and assigned these based on each patient's home postcode and corresponding enumeration district. Standardised incidence ratios were calculated for deciles of the Townsend index. The disadvantage of proxy measures is that some patients will be misclassified.

In a study of socio-economic differences in colorectal cancer incidence using individual indicators of socio-economic status from the ONS Longitudinal Study, trends in colorectal cancer incidence in England and Wales were not consistent by indicator (Brown et al., 1998). Using a measure of social class from the 1971 Census, women in more advantaged social groups experienced higher incidence while there was no significant association for men. When a later measure of social class was used (from the 1981 Census), however, the direction of the association appeared to have changed among women, with higher incidence in the manual social classes. Among men, incidence was higher in both social classes IIIN (skilled non-manual) and IIIM (skilled manual) when the later measure of social class was used.

In the early 1990s, a clear (but not large) deprivation gradient in rectal cancer for males was reported in England and Wales, with incidence rates around 25% higher in the more deprived groups than in the affluent groups, based on the Carstairs index of deprivation (Quinn et al., 2001). There was, however, no variation in rectal cancer incidence by deprivation in females or colon cancer incidence by deprivation group for either males or females. Since then, a clear association between colorectal cancer incidence and deprivation in men has been found in England and Wales, based on incidence data between 1996 and 2010 and using the English IMD (National Cancer Intelligence Network, 2014) and in Scotland, from 2005 onwards using the Scottish IMD (Oliphant et al., 2011). No association was reported among women. More recently, however, a deprivation gradient in colorectal cancer in women has become apparent in Scotland, again based on the Scottish IMD (Tweed et al., 2018).

Given the apparent shift in the relationship between socio-economic deprivation and colorectal cancer incidence, continued monitoring is needed with more recent data to see if the patterns are changing. Furthermore, analysis by different measures of deprivation is needed, both individual-level indicators, which may capture variations in (dis-)advantage (e.g. educational attainment), and area-based measures.

5.2.2 Trends in colorectal cancer survival by deprivation

Large inequalities in colorectal cancer survival by socio-economic deprivation in England have been consistently reported (Coleman et al., 2004). Lower survival for colorectal cancer patients living in deprived areas was first reported in the UK in a study of patients in the South Thames region between 1987 and 1992 (Pollock and Vickers, 1997). Significantly lower five-year relative survival rates were found among colorectal cancer patients living in the most deprived areas compared with the least deprived, using the area-based Townsend deprivation index. The differences in survival by deprivation status were not explained by differences in the incidence of colorectal cancer in the study population.

Survival from cancers of the colon and rectum increased during the 1990s, which may be attributable to both earlier diagnosis and improved treatment. Improvements in survival were notably faster for the least deprived groups, however, and as a result the deprivation gap in 5-year survival became significantly steeper (Coleman et al., 2004). This suggests that more affluent patients have benefited preferentially from progress in early diagnostic procedures and in access to optimal treatment over this period. In a large, cancer-registry-based studies of colon and rectal cancer patients in England and Wales, the deprivation gap in 5-year survival for patients diagnosed during 1996-1999, using the area-based IMD measure of deprivation, was 6% and 7% for males and female colon cancer respectively and 9% and 8% for male and female rectal cancer respectively (Mitry et al., 2008a; Mitry et al., 2008b).

Recent studies of colorectal cancer patients in England have found similar survival differences by socio-economic deprivation, using area-based measures of socio-economic status (Møller et al., 2012; Syriopoulou et al., 2019). A common finding among recent studies is that survival differences by deprivation are higher in the early period following diagnosis. Moller et al. (2012) found that the excess death rate in the most deprived groups was most evident in the first month of follow-up and after that was largely restricted to the first year after diagnosis for colon cancer and two years for rectal cancer. Belot et al. (2019) also found that most of the difference in the excess mortality between deprivation groups happened during the beginning of follow-up and as follow-up increased the difference between the most deprived group and other deprivation groups narrowed.

Few studies using cancer registry data contain individual-level indicators of socio-economic status. Sloggett et al. (2007) investigated socio-economic differences in cancer survival in England and Wales by four individual indicators of socio-economic status: Registrar General's Social Class, housing tenure, household access to a car and the Carstairs indicator of area deprivation using data from the ONS Longitudinal study for participants diagnosed between 1981 and 1997. The car access and tenure variables showed a statistically significant association with colorectal cancer survival (poorer survival with lower socio-economic status) but no association was apparent by the Carstairs measure or by social class. The authors concluded that socio-economic differentials in survival from colorectal cancer may vary by indicator used. More up to date research is needed using a range of individual-level indicators of socio-economic status and over a long follow-up period.

5.3 Objectives

- Describe the demographic and socio-economic attributes of LS members with a diagnoses of colorectal cancer.
- Investigate the association between colorectal cancer incidence and individual and area-based indicators of socio-economic status using time-to-event analysis.
- Investigate the association between all-cause and cause-specific survival and individual and area-based indicators of socio-economic status over a 15-year follow-up period among people with a colorectal cancer diagnosis using time-to-event analysis.

5.4 Data

5.4.1 The Office for National Statistics Longitudinal Study

5.4.1.1 Overview

The Office for National Statistics (ONS) Longitudinal Study (the LS) is a cohort study containing linked census and life events data for a 1 per cent sample of the population of England and Wales. It was set up in 1974 with the initial sample drawn from the resident population enumerated at the 1971 Census (Hattersley et al., 1995). Selection into the LS is by birth date. Four undisclosed dates of birth were used to select a random, 1 per cent sample of the population. The study was designed as a continuous, multi-cohort study with subsequent samples drawn at each census, based on the same selection criteria, and linked into the study. Between censuses, new members enter the study via birth on an LS birth date or by immigration (if born on an LS birth date) and existing members exit the study through death or emigration.

The original sample size of the LS was over 500,000 and has remained consistently high over time. Data for approximately 1.1 million individuals has been collected over the 40 years of the LS (Shelton et al., 2018).

Data about LS members is obtained from answers to census questions and from linking administrative data routinely collected by ONS (formerly known as OPCS) to the sample. These data are linked to LS members by the National Health Service Central Register (NHSCR). Data from five censuses (1971, 1981, 1991, 2001 and 2011) are currently available for both LS members and their households (University College London, 2021b). Events linked to LS member's include enlistments into the Armed Forces, embarkations, re-entries to the NHS from enlistment or embarkation, immigration, new births, live and still births to sample mothers, widowhoods, deaths and cancer registrations.

The LS is maintained and updated by the ONS. The LS data are available to anyone in the UK who has obtained ONS Accredited Researcher status via the ONS Approved Researcher Scheme (Office for National Statistics, 2021a). The application process involves submitting a research project application which is reviewed by the Data Owner, LS Research Board and Research Accreditation Panel. The data can be accessed through the Secure Research Service (SRS)

safe setting rooms at ONS offices in London, Newport and Titchfield. Outputs are released to researchers through an output clearance process. Information and support for UK-based researchers is provide by the Centre for Longitudinal Study Information and User Support (CeLSIUS) (University College London, 2021a).

5.4.1.2 Census data

The census is a survey of all people and households in England and Wales that takes place every 10 years. A large amount of sociodemographic data is gathered at each census including information about employment, education, marital status, housing tenure and household composition. The data consist of the responses to the census questions and some additional variables (e.g. social class) derived from relevant census variables. Questions asked on the census have changed over time (e.g. questions about ethnicity and self-reported limiting long-term illness were introduced in the 1991 Census).

Census data is linked to LS members and their households at each census time point. The matching process involves taking a new census record for a person born on an LS birth date and identifying a corresponding record in the LS. For those who are matched, the new census data is added to their existing LS record. Unmatched individuals are added as new members to the LS. Linkage rates between censuses are generally very high at around 90 per cent (Shelton et al., 2018).

5.4.2 Cancer registration data

Cancer registrations are collected by the National Cancer Registration and Analysis Service in England and the Welsh Cancer Intelligence and Surveillance Unit. These data are linked to LS member records through the NHSCR. The NHSCR maintain records of all individuals in England and Wales who are registered with the NHS. To link a cancer registration to a LS record, the LS member must be traced (have an NHS number recorded at NHSCR) at the time of the cancer registration. Event data in the LS are usually added annually, with a two year delay but there may be a three year delay or more for cancer registrations.

Cancer registrations have been linked to the LS from 1971 onwards. The cancer registration data in the LS was last updated with information on cancer

diagnoses up to and including 2015. Since 2016, the cancer registry has been managed by the National Cancer Registration and Analysis Service (NCRAS) with Public Health England (PHE). A data sharing agreement with PHE for the linkage of cancer registration has not yet been put in place, therefore cancer registrations after 2015 have not been linked to the LS.

The World Health Organisation International Classification of Diseases and Related Health Problems (ICD) coding system is used to record the type and site of cancer registrations. The ICD code current at the time of the cancer registration is used. Since 1st January 1995 the ICD-10 coding system has been used to code cancer registrations (World Health Organisation, 2019), prior to this the ICD-8 (1971-1978) and ICD-9 (1979-1994) classifications were used.

5.4.3 Death registration data

Deaths of LS members are identified through two sources: routine notifications to the NHSCR and searches of the ONS deaths database. It is a statutory obligation to register a death, therefore the linkage rate of deaths of LS members is very high.

Cause of death information is coded using the WHO ICD coding system. Three versions of the ICD are used to code deaths in the LS: ICD8 (1971-81), ICD9 (1981-2000) and ICD10 (2001 onwards). One underlying cause of death is identified and up to eight contributory causes of death may also be recorded for each death.

The LS database is updated with death registrations annually by year of registration with a two-year delay. There may be some deaths that occurred in one year but are not registered until the following year. Furthermore, where there was an inquest or post mortem, the cause of death may be amended. The death data in the LS was last updated with information on deaths up to and including 2016.

5.4.4 Embarkation data

LS members can exit the LS by emigration out of England and Wales – known as embarkation. An LS member can have more than one embarkation event: if they embark, re-enter the country and then re-embark. The data available on embarkations is limited as most people do not give prior notice of an

embarkation (either by returning their NHS card or informing their GP). Therefore, it is estimated that only around 50 per cent of embarkations are reported to NHSCR.

5.5 Methods

5.5.1 Data cleaning and recoding

LS members were classified for the work reported here by three established indicators of socio-economic position: educational attainment, occupational social class and housing tenure. A fourth, area-based measure of deprivation (Townsend Deprivation Index) was appended to LS member records based on each member's area of residence.

These four measures were chosen to investigate different elements of socio-economic position which may influence health outcomes. Educational attainment could be related to awareness of disease symptoms and likelihood to seek medical advice. Occupational social class is a proxy for income and is also related to occupational hazards. Housing tenure is a measure of material wealth and could also be related to living conditions. The Townsend Deprivation Index is a measure of area deprivation. It was chosen over the Index of Multiple Deprivation (IMD) as the LS includes people in both England and Wales and there is no UK-wide measure of multiple deprivation.

All variables were recorded at the 2001 Census. Educational qualifications were categorised into two groups based on higher education attainment: degree and no degree. People aged 15 and under and 75 and over were not required to answer this question in the 2001 Census. Information regarding educational attainment was infilled from the previous census in 1991 for study members for which this question was not applicable in 2001 (i.e. those aged under 15 and those aged 75 and over at the 2001 Census). This is an advantage of the LS as members data is linked between censuses. Only those LS members aged 50+ in 2001 were included in this study (outlined in section 1.5.2), therefore it is unlikely that someone's level of educational attainment would have changed since the previous census so this was considered an appropriate method to increase the proportion of members allocated to a category.

The Registrar General Social Class categories were used: Social Class I (professional), II (intermediate), IIIN (skilled non-manual), IIIM (skilled manual), IV (semi-skilled manual) and V (unskilled manual). This classification was used as it is comparable over time. The categories were further aggregated into two groups: Non-manual (I, II, IIIN) and Manual (IIIM, IV, V). As with educational

attainment, questions regarding occupation were not asked of people aged 15 and under and 75 and over at the 2001 Census. This information was infilled from the 1991 Census for study members aged 75 and over at the 2001 Census. If an individual did not have a social class category recorded at the previous Census in 1991 (for example, if they were not in employment), this information was recorded as missing.

Housing tenure was categorised as: owner-occupied (owned outright, owned with a mortgage or loan, shared ownership), privately rented (private landlord or letting agency, employer of a household member, relative or friend of household member, other, lives rent free) or social rented (rented from council, other social rented). People living in communal establishments or missing housing tenure information were not included in the analysis.

Marital status was grouped into three categories: married (including first marriage and re-marriage); separated, divorced or widowed; and single (never married).

The 16 ethnic groups in the 2001 classification were combined into four categories (White, Black, Indian/Pakistani/Bangladeshi, Chinese/Mixed/Other) due to low numbers of cancer registrations in some ethnic groups.

The Townsend Deprivation Index scores were categorised into quintiles (1= least deprived, 5 =most deprived). A Townsend quintile was appended to each LS member's record based on their small-area of residence, Lower Layer Super Output Area (LSOA), at each census time point. This was to enable the potential for investigating the influence of living in differently deprived areas over time (though see later discussions). The linkages were carried out by ONS using a lookup file provided as part of this work using a method developed by Norman (Norman and Riva, 2012).

5.5.2 Study cohort

The study cohort comprised all LS members aged 50 years and over who were recorded at the 2001 Census. This age group was chosen as incidence of colorectal cancer among people aged under 50 is very low and hence there were only a small number of colorectal cancer registrations among LS members in this age group. Furthermore, there is some evidence that disease aetiology is different among younger adults (under 50) than older adults (Araghi et al.,

2019). The cohort were followed up and colorectal cancer registrations recorded until December 2015.

Only LS members traced to the NHSCR were included in the study cohort as cancer registrations cannot be linked to untraced members. This avoids denominator bias. Members traced at the 1991 Census or earlier were included to enable cancer registrations in the period preceding the start of following up (April 2001) to be identified. LS members with a colorectal cancer diagnosis prior to the start of the follow-up period were excluded from the analysis.

Colorectal cancer cases were identified in the linked cancer registration event data based on ICD10 codes C18-21 and ICD9 codes 153 and 154. Only information on primary cancers is kept in the LS. Where an LS member had more than one primary case of colorectal cancer, the date of the first colorectal cancer registration was taken. The date of cancer registrations was given in month and year format. Access to information on the specific day of cancer registration was not permitted to protect confidentiality of LS members.

In the analysis of colorectal cancer deaths, the study cohort comprised LS members who had a colorectal cancer diagnosis aged 50 and over between April 2001 and December 2015 and were followed up until December 2016. Deaths of LS members were identified from the linked death registration event data. Deaths for which colorectal cancer (ICD10 C18-21) was recorded as the underlying cause of death were flagged to enable cause-specific survival analysis.

5.5.3 Directed acyclic graphs

Directed acyclic graphs (DAGs) were used to inform the statistical modelling (Figure 5.1). The diagrams clarify how the variables relate to one another. These relationships were based on evidence in the literature.

DAGs were used to help identify appropriate confounders to adjust for in statistical models modelling the association between indicators of socio-economic status and colorectal cancer outcomes. Separate diagrams were produced for each exposure of interest: education, social class, housing tenure and area of residence. A set of diagrams were created for colorectal cancer registration as the outcome (Figures 1-4) and for cancer survival as the outcome (Figures 5-8). All possible variables were included in the diagrams, as is convention, even though some of them are not available in this data set. The direction of association is shown with an arrow in the diagram. This was based on the temporal order of the variables and the hypothesised pathways between the variables, for example, educational attainment is likely to influence occupational social class which will in turn influence housing tenure. Confounders were identified as ancestors of the exposure and the outcome.

DAGitty software was used to create and analyse the DAGs (Textor et al., 2016).

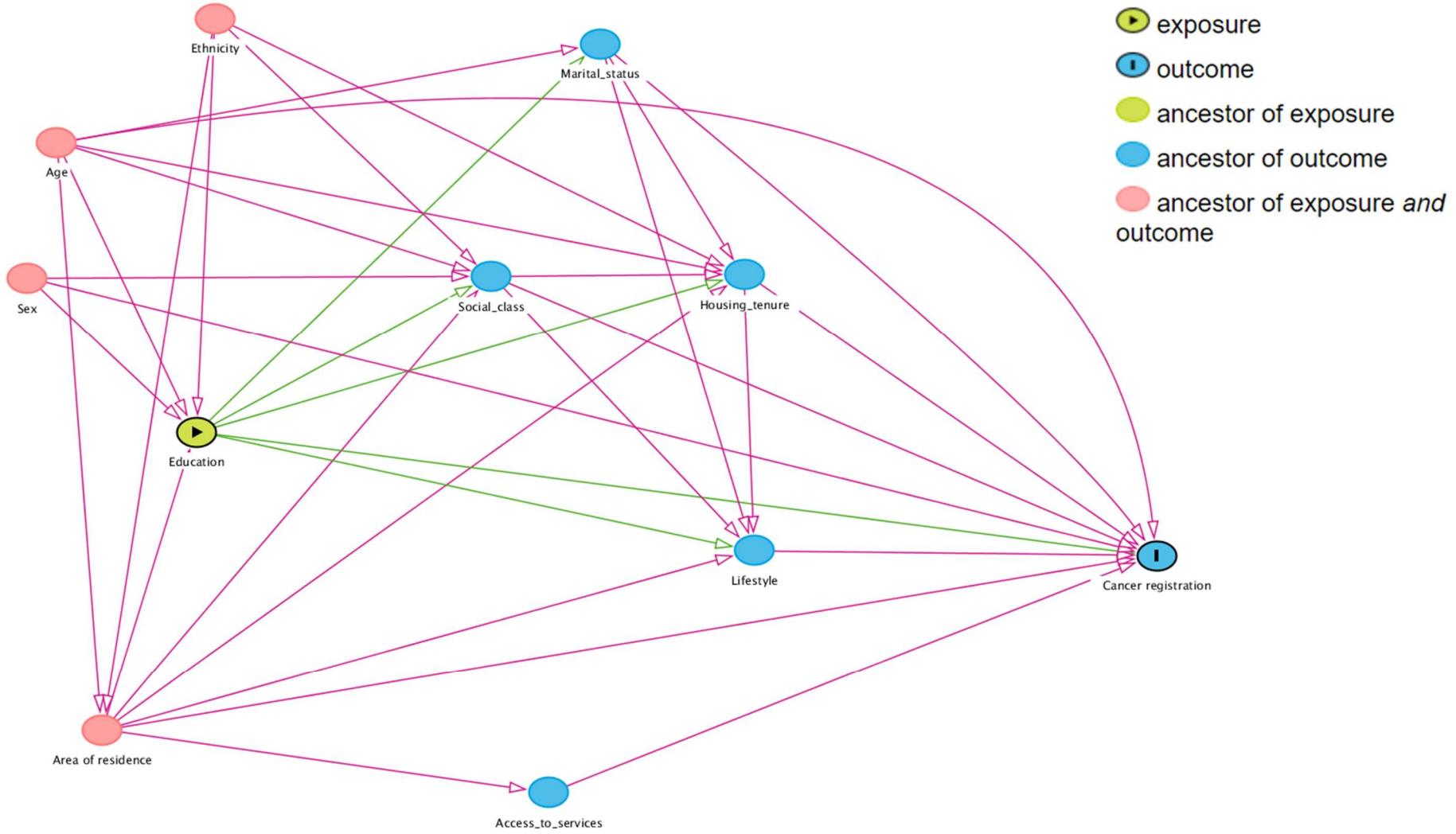


Figure 5.1 Directed acyclic graph showing relationship between educational attainment status and CRC registration

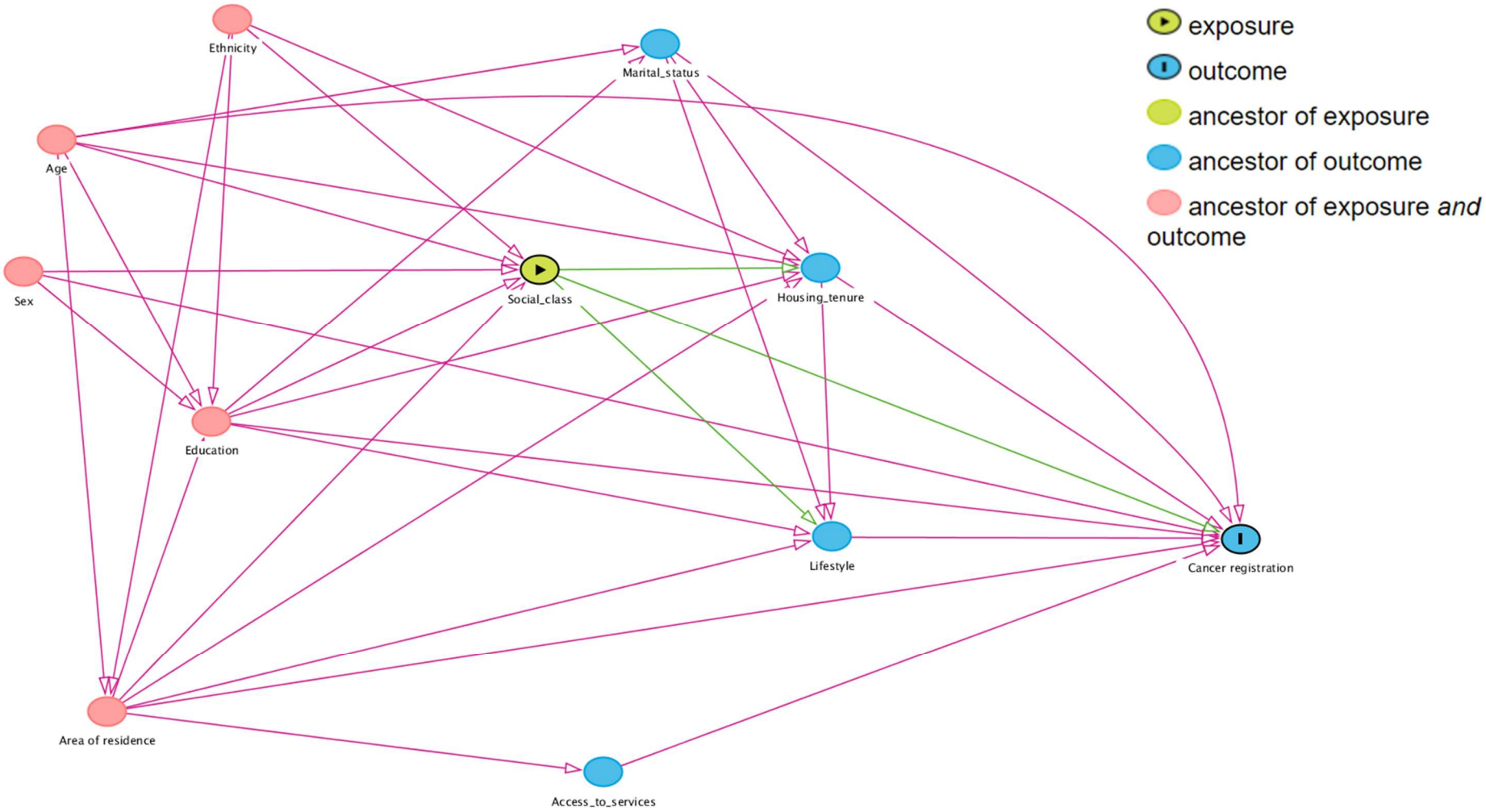


Figure 5.2 Directed acyclic graph showing relationship between social class and CRC registration

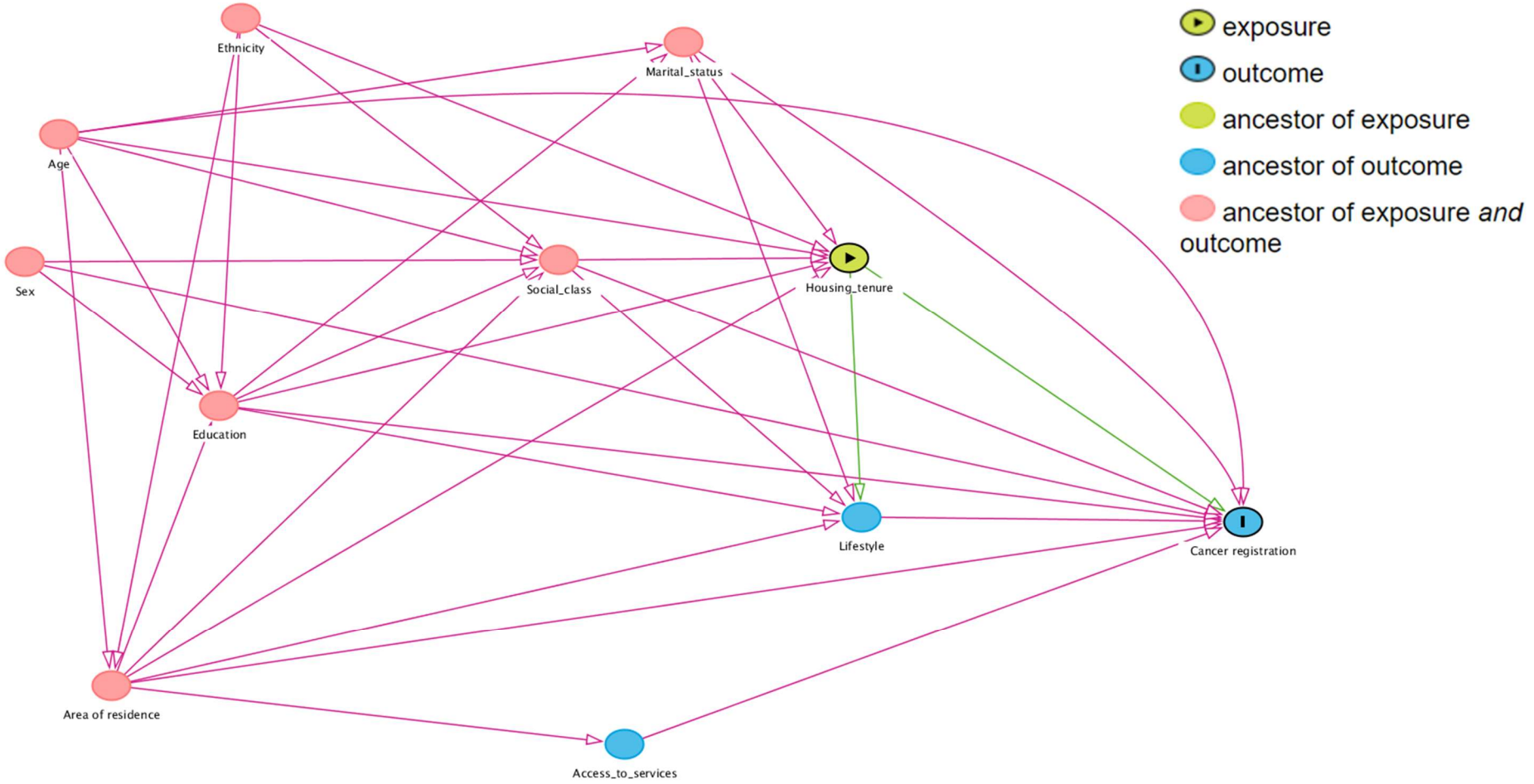


Figure 5.3 Directed acyclic graph showing relationship between housing tenure and CRC registration

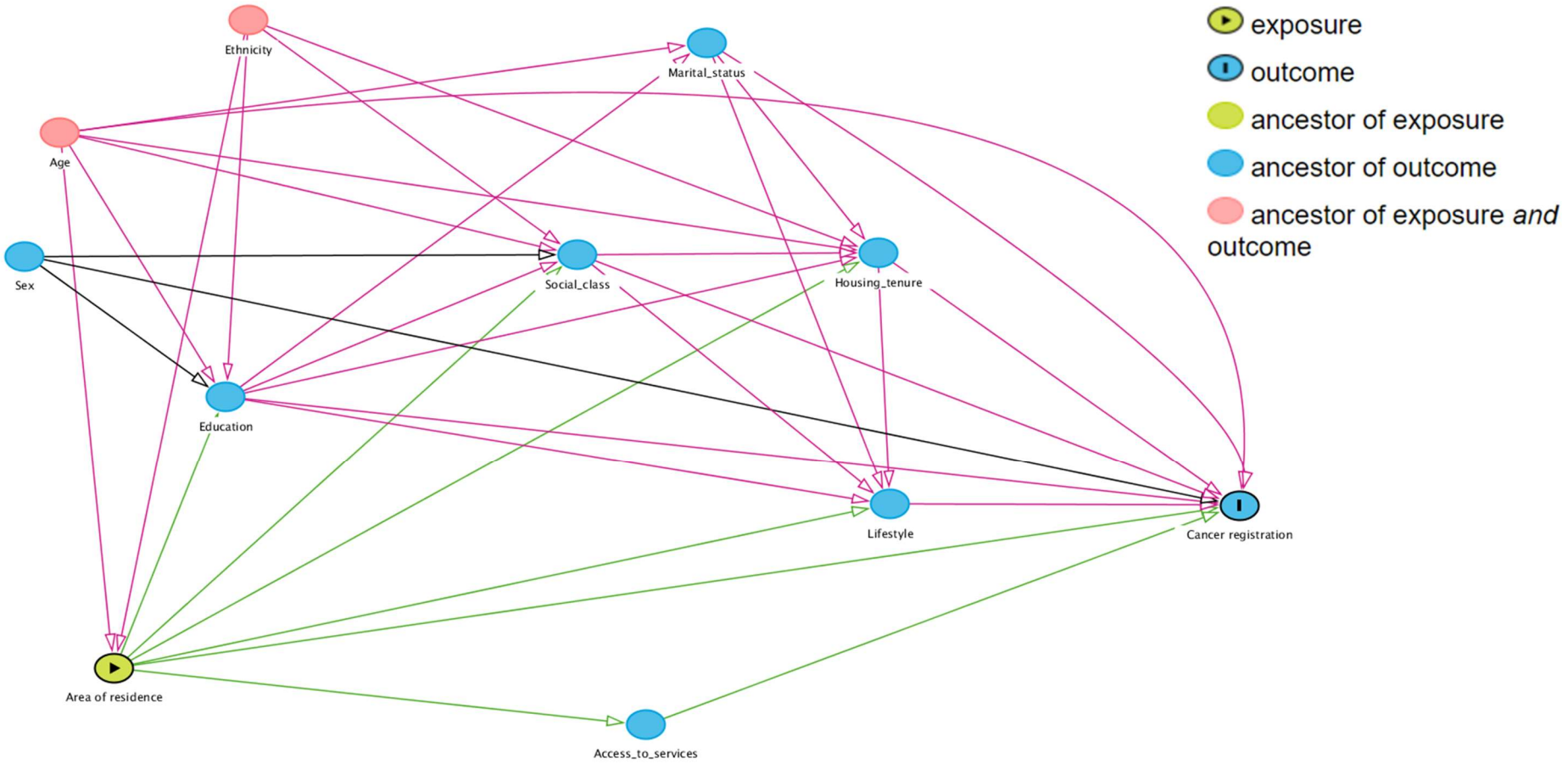


Figure 5.4 Directed acyclic graph showing relationship between area of residence and CRC registration

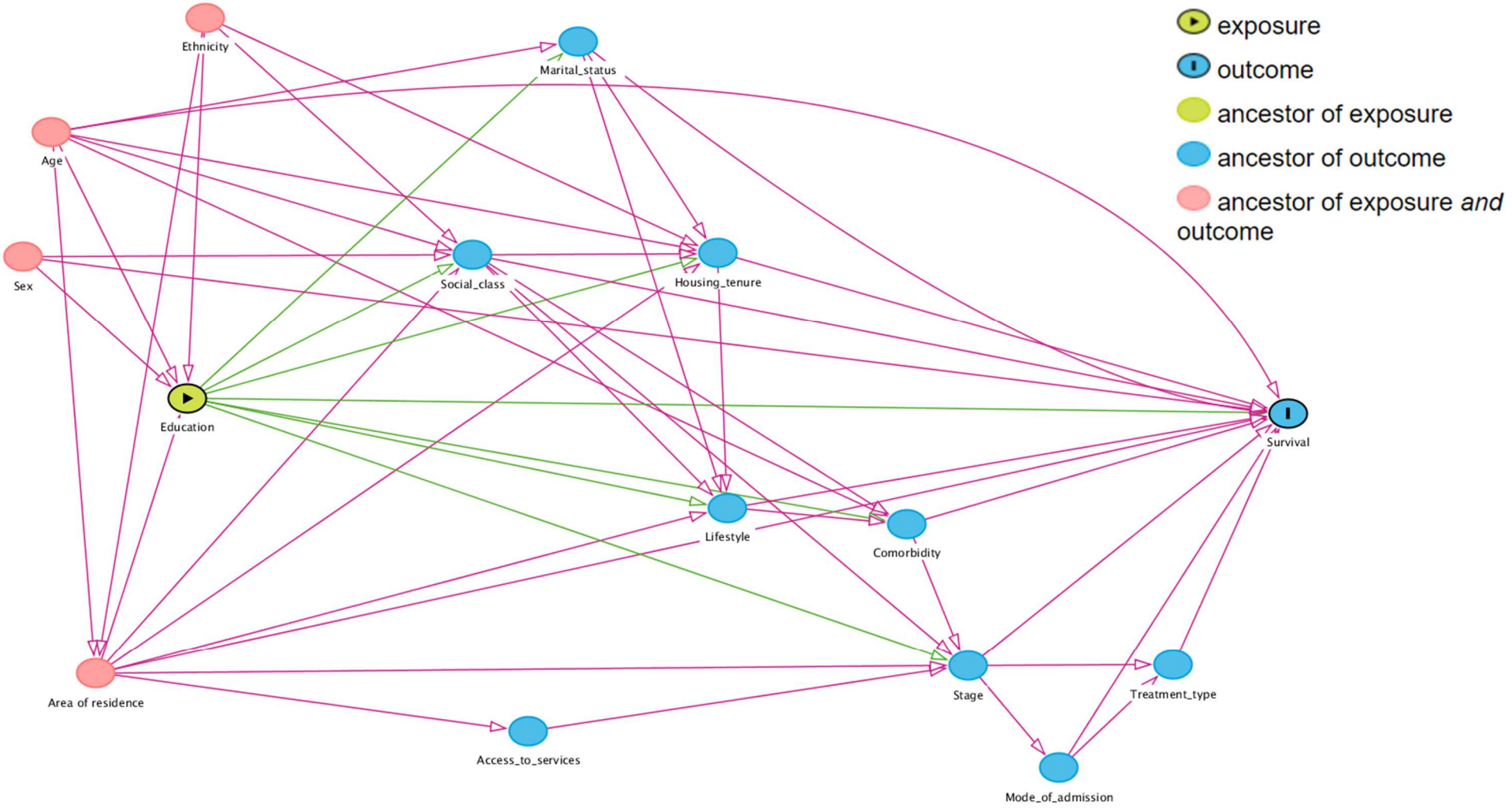


Figure 5.5 Directed acyclic graph showing relationship between educational attainment and CRC survival

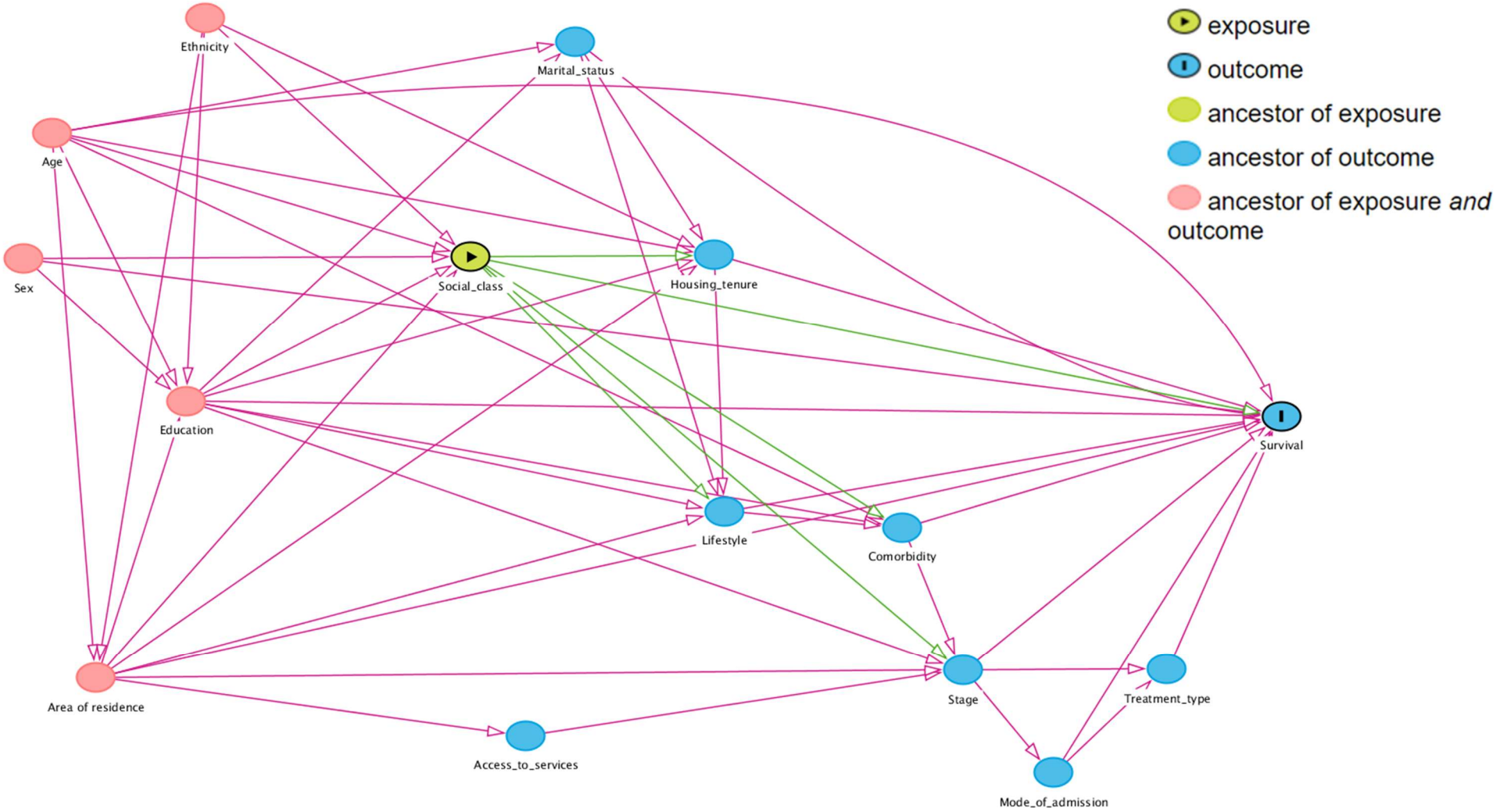


Figure 5.6 Directed acyclic graph showing relationship between social class and CRC survival

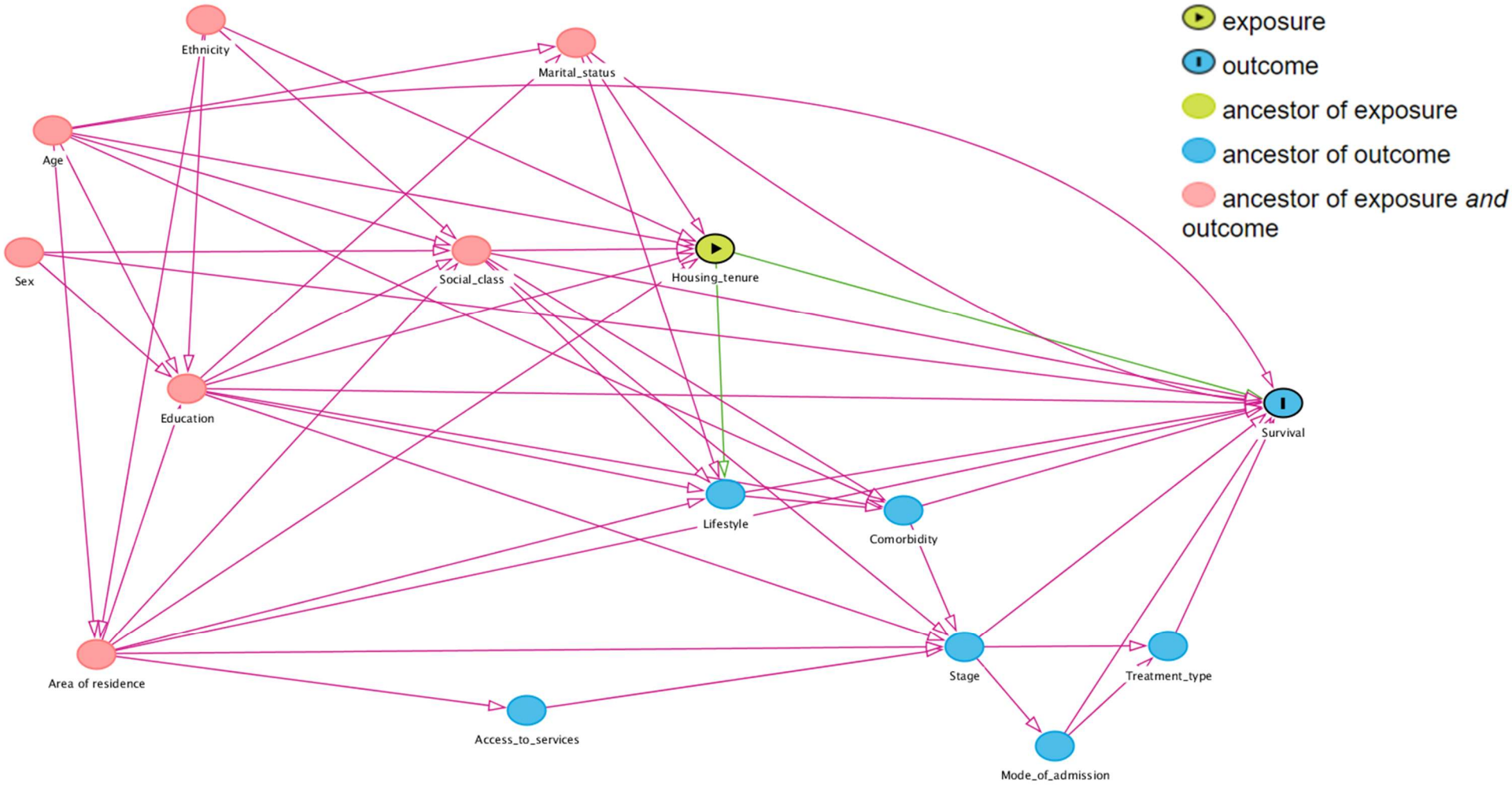


Figure 5.7 Directed acyclic graph showing relationship between housing tenure and CRC survival

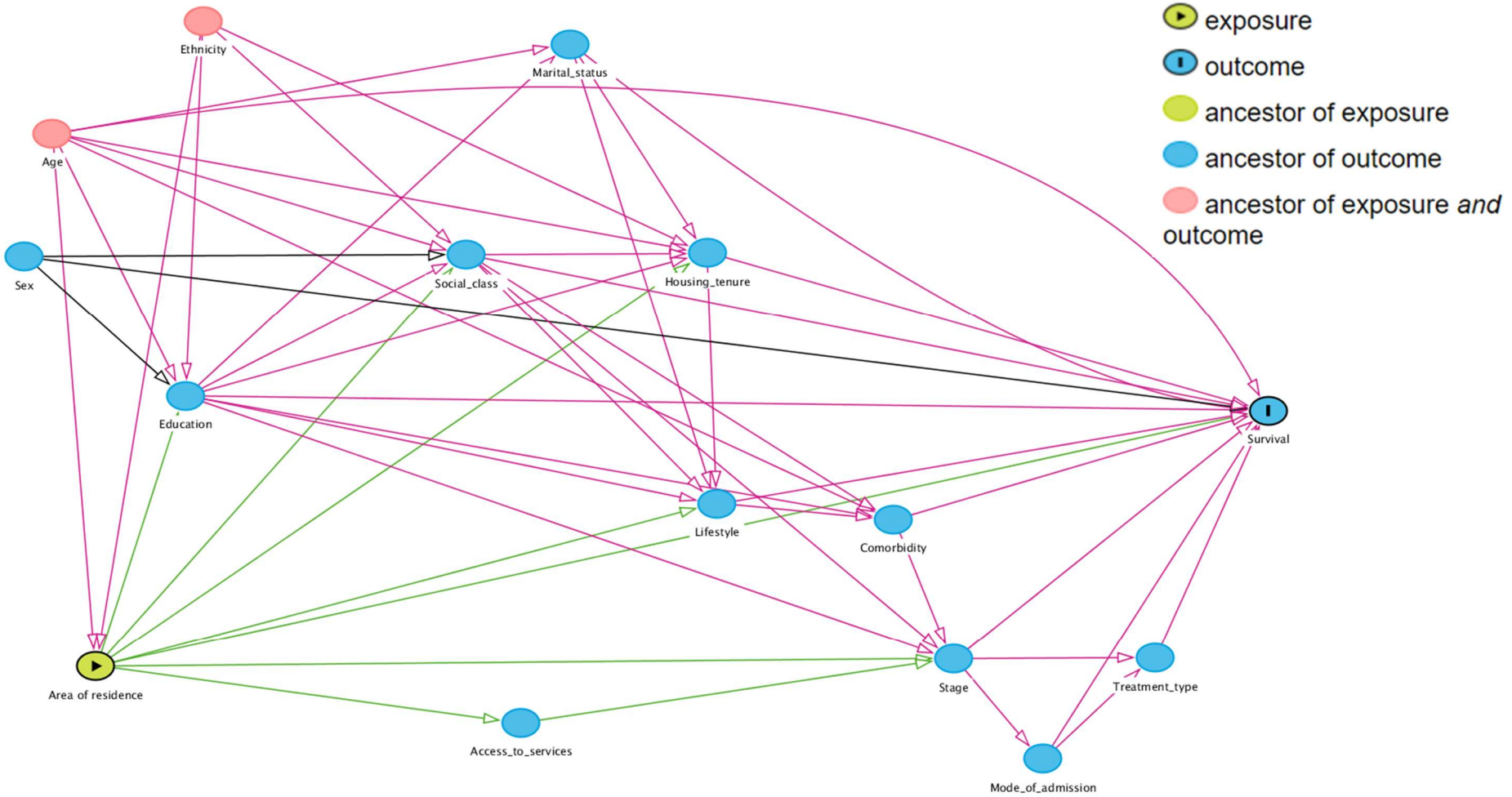


Figure 5.8 Directed acyclic graph showing relationship between area of residence and CRC survival

5.5.4 Descriptive statistics

Descriptive statistics are presented for the study cohort in two cohort subsets. The first cohort was split according to whether or not they had a colorectal cancer registration during follow-up. The second cohort was split according to whether or not they died during follow-up and whether or not the deaths were from all-causes or from colorectal cancer as the underlying cause of death. Chi-squared tests were used to detect differences between groups.

5.5.5 Time-to-event analysis

Time-to-event analysis was used in this study to examine the association between indicators of socio-economic status and three outcomes: colorectal cancer registration, survival from all-causes and cause-specific survival.

5.5.5.1 Incidence

In the analysis of colorectal cancer registrations the ‘event’ was a diagnosis of colorectal cancer (recorded by the cancer registry). Time-to-event (in months) was calculated for each study member from the start of the follow-up period (April 2001) to the earliest date of either diagnosis of colorectal cancer, exit from the study or the end of follow up (December 2015). As only access to month and year of diagnosis was given, 0.5 of a month was added to all the times to avoid a time-to-event of zero (which would be deleted from the time-to-event analysis).

Kaplan-Meier cumulative incidence curves were produced to show the probability of a colorectal cancer registration. Cumulative incidence, or cumulative failure probability, is calculated as 1-S(t). Observed versus expected colorectal cancer cases by socio-economic group were compared using the log rank test for difference. Cox Proportional Hazards regression was performed to estimate the hazard ratios (HR) and confidence intervals (CI) of colorectal cancer registration for each indicator of socio-economic status: educational attainment, occupational social class, housing tenure and area deprivation. Two models were run for each exposure variable. Model 1 was unadjusted and Model 2 was adjusted for all variables (available in the data set) identified in the corresponding causal diagram. In Model 2, when educational attainment was the exposure the model was adjusted for age, sex, ethnicity and area

deprivation; when social class was the exposure the model was adjusted for age, sex, ethnicity, education and area of residence; when housing tenure was the exposure the model was adjusted for age, sex, education, social class, marital status and area of residence; and when area of residence was the exposure the model was adjusted for age and ethnicity. The proportional hazards assumption was assessed graphically using complimentary log-log plots. If the proportional hazards assumption holds (that the effect of the factors being investigated on the hazard is the same over the follow-up time) the lines should be parallel i.e. at a fixed height above each other.

5.5.5.2 Survival

In the analysis of survival after a diagnosis of colorectal cancer, two events of interest were examined: death from all-causes and death from colorectal cancer.

For the analysis of all-cause deaths, survival time (in months) was calculated from diagnosis of colorectal cancer to death or exit from the study or the end of follow up (December 2016) for all LS members with a colorectal cancer diagnosis. If a study member exited the study more than once, the first (earliest) embarkation date was used to exclude individuals at the point of exit, as beyond this they are no longer ‘at risk’ of having a (linkable) cancer registration.

For the analysis of cause-specific deaths, survival time (in months) was calculated from diagnosis of colorectal cancer to death from colorectal cancer or exit from the study or the end of follow-up (December 2016) for all LS members with a colorectal cancer diagnosis. If a study member died of other causes during follow-up the survival time was censored at the date of death. Embarkations were treated as above.

Kaplan-Meier survival curves were produced to display the survival probability by socio-economic group for both all-cause and cause-specific survival. Observed versus expected deaths (all-cause and cause specific) by socio-economic group were compared using the log rank test for difference. Cox Proportional Hazards regression was performed to estimate the hazard ratios (HR) and confidence intervals (CI) of death from all-causes and death from colorectal cancer. Two models were run for each of the exposure variables, with

adjustments detailed above. Complimentary log-log plots were used to assess the proportional hazards assumption.

Crude probabilities of death due to colorectal cancer and due to other causes were calculated for each socio-economic group to allow for an assessment of the probability of dying from colorectal cancer in the presence of competing risks.

5.6 Results

The results are presented in three sections, according to the outcome of interest. Firstly, results for the analysis examining associations between indicators of socio-economic status and colorectal cancer incidence are presented, this is followed by the results for all-cause survival and finally, cause-specific survival.

5.6.1 Colorectal cancer incidence

Figure 5.9 details the number of LS members enumerated at the 2001 Census and how the study-specific sample size is reduced by the exclusions imposed. Of the 186,687 LS members aged 50+ present at the 2001 Census, 7,029 had not been traced at the 1991 Census (or earlier). A further 1,542 had a primary colorectal cancer diagnosis before April 2001. This gave a final study sample of 178,116 individuals.

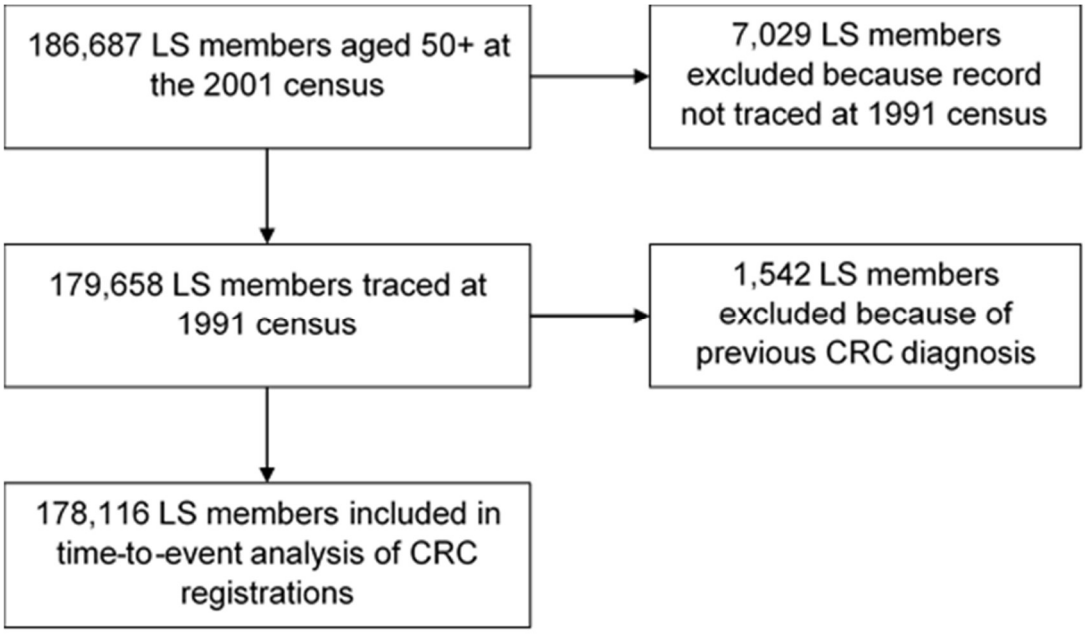


Figure 5.9 Flow diagram of study sample for analysis of CRC registrations

The socio-demographic characteristics of the study members are reported in Table 5.1. There were slightly more males than females (54% and 46% respectively). The vast majority of study participants were of White ethnicity (96%). Nearly two-thirds of the study sample were married (64%), 29% were separated, divorced or widowed and just 6.5% were single. The majority of the

study sample did not have degree level qualifications (86.5% compared to 13% with a degree). A higher proportion of study participants were employed in non-manual occupations (44%), compared to manual occupations (36%). Despite infilling educational attainment and social class information from the previous census in 1991, nearly a fifth of records are not assigned a social class as this information can only be recorded if they have had an occupation and there will be some people, particularly women of this age group, who have never worked. There were fewer people living in more deprived areas compared to the least deprived areas. This is a function of using the comparable deprivation quintiles over time. On average, area deprivation eases over time (using the Townsend Deprivation Index input variables) and as people age into mid-life, on average there is a move towards less deprived areas.

Variable	Category	n	%
Total		178,116	
Sex	Male	81,802	45.9
	Female	96,314	54.1
Age group	50-59	67,234	37.7
	60-69	50,317	28.2
	70-79	39,256	22.0
	80+	21,308	12.0
	Missing	1	0.0
Ethnic group	White	170,367	95.6
	Black	1,686	0.9
	Indian/Pakistani/Bangladeshi	4,887	2.7
	Chinese/Mixed/Other	1,128	0.6
	Missing	48	0.0
Marital status	Married	114,434	64.2
	Separated/Divorced/Widowed	52,024	29.2
	Single	11,657	6.5
	Missing	1	0.0
Educational attainment	No degree	154,152	86.5
	Degree	22,555	12.7
	Missing	1,409	0.8
Occupational social class	Non manual	78,804	44.2
	Manual	64,881	36.4
	Missing	34,431	19.3
Housing tenure	Owner occupied	136,239	76.5
	Privately rented	9,503	5.3
	Social rented	28,103	15.8
	Missing (Communal establishment)	4,271	2.4
Area deprivation	1- Least deprived	78,444	44.0
	2	42,669	24.0
	3	27,624	15.5
	4	20,022	11.2
	5 - Most deprived	9,325	5.2
	Missing	32	0.0

Table 5.1 Characteristics of study sample for analysis of CRC registrations (Data source: ONS LS)

There were 4,421 cases of colorectal cancer during a mean follow up time of 11.8 years. Cases prevalent at baseline were excluded. There were 3 cases diagnosed after embarkation (exit from the study) which were excluded from the analysis. This left the sample with 4,418 incident colorectal cancers for the time-to-event analysis.

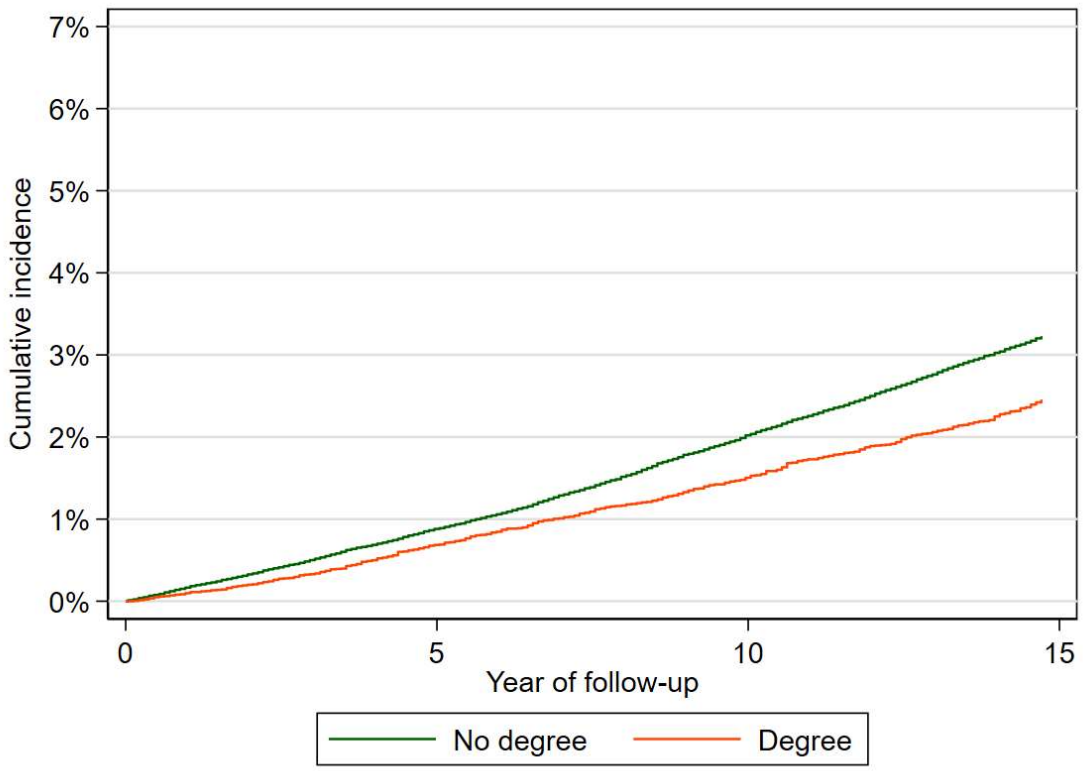
Colorectal cancer cases and non-cases by socio-demographic characteristics are shown in Table 5.2. The percentage of colorectal cancer cases was significantly different by sex (higher in men), age group (higher in those aged 60-79), ethnic group (higher in the White ethnic group), marital status (higher among those married), educational attainment (higher among those without a degree) and social class (higher in manual social classes). There was no significant difference by housing tenure or Townsend deprivation quintile.

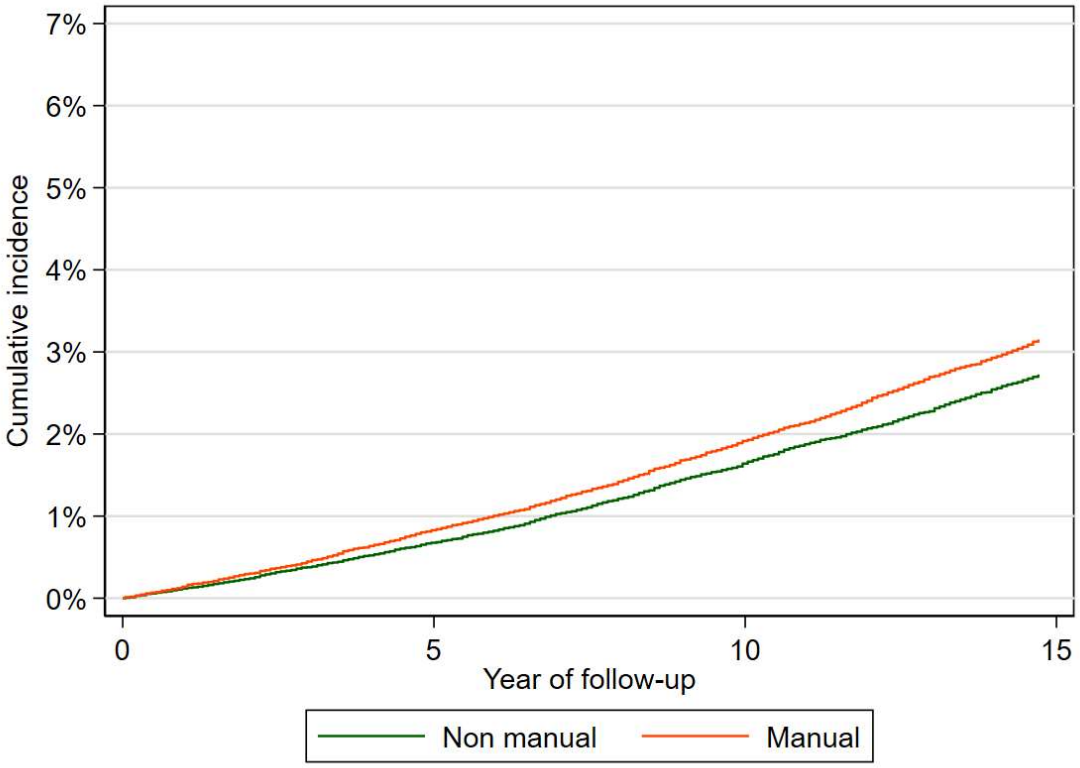
		Total	Case		Non-case		Chi-squared
Variable	Category		n	%	n	%	
Total		178,116	4,418	2.5	173,698	97.5	
Sex	Male	81,802	2,378	2.9	79,424	97.1	p < 0.001
	Female	96,314	2,040	2.1	94,274	97.9	
Age group	50-59	67,234	1,039	1.5	66,195	98.5	p < 0.001
	60-69	50,317	1,508	3.0	48,809	97.0	
	70-79	39,256	1,374	3.5	37,882	96.5	
	80+	21,308	497	2.3	20,811	97.7	
Ethnic group	White	170,367	4,313	2.5	166,054	97.5	p < 0.001
	Black	1,686	35	2.1	1,651	97.9	
	Indian/Pakistani/Bangladeshi	4,887	51	1.0	4,836	99.0	
	Chinese/Mixed/Other	1,128	19	1.7	1,109	98.3	
Marital status	Married	114,434	2,942	2.6	111,492	97.4	p = 0.002
	Separated/Divorced/Widowed	52,024	1,226	2.4	50,798	97.6	
	Single	11,657	250	2.1	11,407	97.9	
Educational attainment	No degree	154,152	3,891	2.5	150,261	97.5	p < 0.001
	Degree	22,555	497	2.2	22,058	97.8	
Occupational social class	Non manual	78,804	1,884	2.4	76,920	97.6	p < 0.001
	Manual	64,881	1,694	2.6	63,187	97.4	
Housing tenure	Owner occupied	136,239	3,437	2.5	132,802	97.5	p = 0.776
	Privately rented	9,503	229	2.4	9,274	97.6	
	Social rented	28,103	713	2.5	27,390	97.5	
Area deprivation	1- Least deprived	78,444	2,021	2.6	76,423	97.4	p = 0.063
	2	42,669	1,068	2.5	41,601	97.5	
	3	27,624	655	2.4	26,969	97.6	
	4	20,022	452	2.3	19,570	97.7	
	5 - Most deprived	9,325	222	2.4	9,103	97.6	

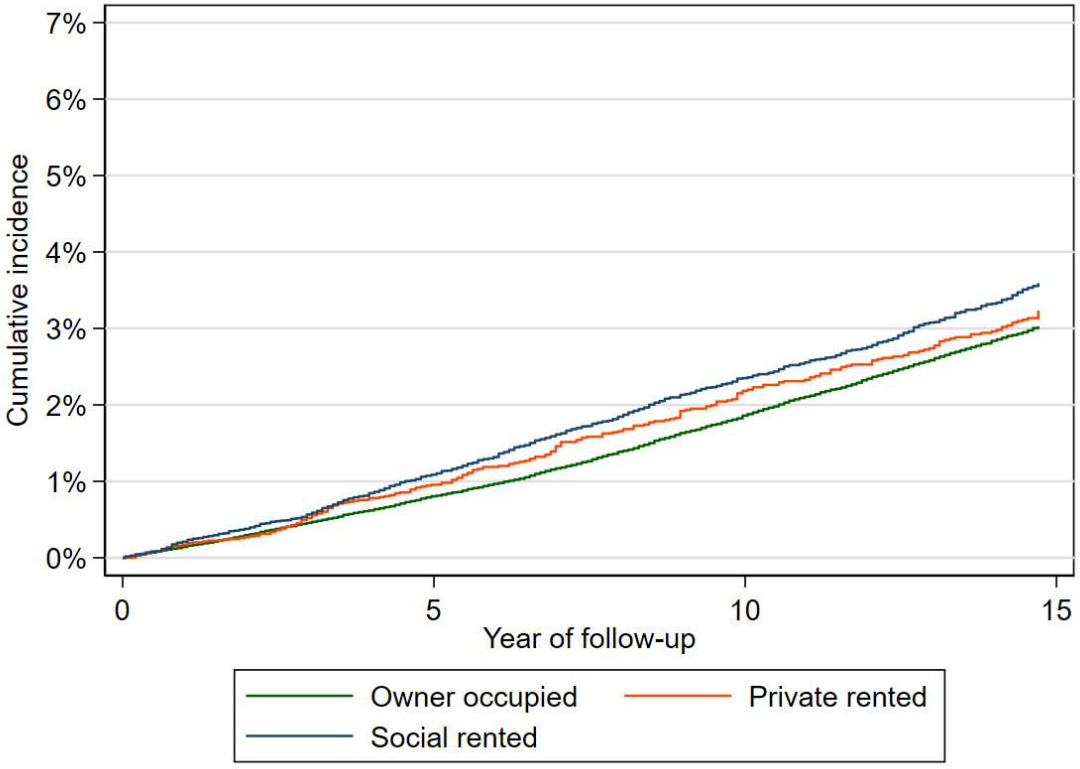
Table 5.2 Study sample by demographic characteristics and colorectal cancer cases and non-cases (Data source: ONS LS)

5.6.1.1 Time-to-event analysis

Kaplan-Meier cumulative incidence curves by indicator of socio-economic status are displayed in Figure 5.10. The cumulative incidence trends are summarised in Table 5.3. Cumulative incidence of colorectal cancer registrations at 5, 10 and 14 years of follow-up are shown (follow-up was from April 2001 to December 2015 for cancer registrations so there was not a full fifteenth year of follow-up). At each time point, cumulative colorectal cancer incidence was higher among study members with no degree, compared to those with a degree; among those in manual occupations, compared to those in non-manual occupations; and among those living in rented housing, compared to those living in owner-occupied housing. There was no gradient in cumulative colorectal cancer incidence by quintile of area deprivation.




7%
6%
5%
4%
3%
2%
1%
0%
Cumulative incidence
0
5
10
15
Year of follow-up
Non manual
Manual


7%
6%
5%
4%
3%
2%
1%
0%
Cumulative incidence
0
5
10
15
Year of follow-up
Owner occupied
Private rented
Social rented

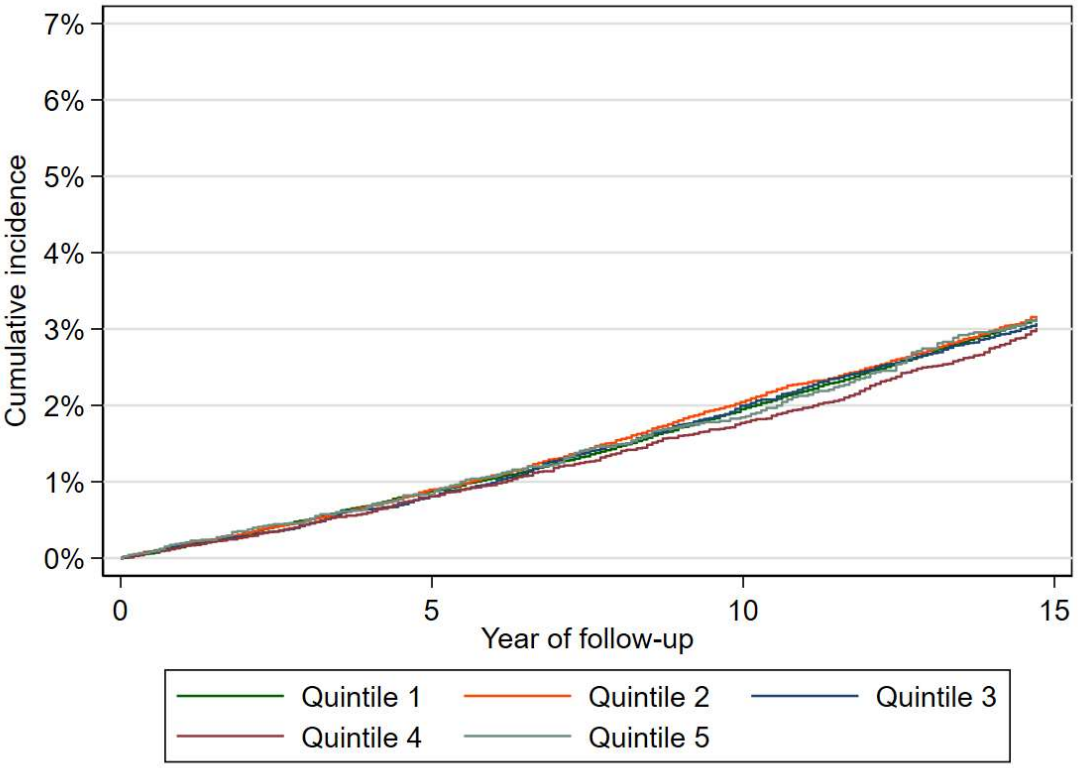


Figure 5.10 Kaplan-Meier cumulative incidence curves for probability of colorectal cancer registration by: a) educational attainment; b) social class; c) housing tenure; d) area deprivation (Data source: ONS LS)

Variable	Category	Cumulative incidence per cent [95% CI]		
		5-year	10-year	14-year
Educational attainment	No degree	0.9 [0.8-0.9]	2.0 [2.0-2.1]	3.0 [2.9-3.1]
	Degree	0.7 [0.6-0.8]	1.5 [1.4-1.7]	2.3 [2.1-2.5]
Social class	Non manual	0.7 [0.6-0.7]	1.6 [1.6-1.7]	2.5 [2.4-2.7]
	Manual	0.8 [0.8-0.9]	1.9 [1.8-2.0]	2.9 [2.8-3.1]
Housing tenure	Owner occupied	0.8 [0.8-0.9]	1.9 [1.8-1.9]	2.8 [2.7-2.9]
	Private rented	1.0 [0.8-1.2]	2.2 [1.9-2.5]	3.0 [2.6-3.4]
	Social rented	1.1 [1.0-1.2]	2.4 [2.2-2.6]	3.3 [3.1-3.6]
Area deprivation	1 - Least deprived	0.9 [0.8-0.9]	2.0 [1.9-2.1]	2.9 [2.8-3.1]
	2	0.9 [0.8-1.0]	2.0 [1.9-2.2]	3.0 [2.8-3.2]
	3	0.8 [0.7-0.9]	2.0 [1.8-2.2]	2.9 [2.7-3.1]
	4	0.8 [0.7-0.9]	1.8 [1.6-2.0]	2.8 [2.5-3.0]
	5 - Most deprived	0.8 [0.7-1.1]	1.8 [1.6-2.2]	3.0 [2.6-3.4]

Table 5.3 Cumulative incidence of colorectal cancer registrations over time by indicator of socio-economic status (Data source: ONS LS)

Log-rank tests for difference (Table 5.4) show statistically significant differences in the observed versus expected cases by educational attainment, social class and tenure. There were more observed cases than expected among study members without a degree and among those in manual occupations (compared to fewer observed than expected cases among those with a degree and non-manual occupations). For tenure, there were more observed than expected cases in the private rented and social rented categories, compared to fewer observed than expected cases in the owner occupied category. No significant differences in colorectal cancer risk were observed by area deprivation (Townsend deprivation quintile).

Variable	Category	Observed cases	Expected cases	Log-rank test
Educational attainment	No degree	3,891	3754	p < 0.001
	Degree	497	634	
Occupational social class	Non manual	1,884	2016	p < 0.001
	Manual	1,694	1562	
Housing tenure	Owner occupied	3,437	3553	p < 0.001
	Privately rented	229	218	
	Social rented	713	608	
Area deprivation	1- Least deprived	2,021	2013	p = 0.820
	2	1,068	1047	
	3	655	663	
	4	452	474	
	5 - Most deprived	222	221	

Table 5.4 Log-rank test results for indicators of socio-economic status showing observed versus expected CRC cases and test for difference (Data source: ONS LS)

In Cox Proportional Hazards regression (Table 5.5), a significantly reduced risk of colorectal cancer registration was observed among study members with a degree, compared to those without a degree (unadjusted Model 1). This pattern remained in Model 2 (adjusted for age, sex, ethnicity and area deprivation) (HR 0.89 [0.81-0.98]). Conversely, a significantly increased risk was observed among study members employed in manual occupations, compared to those in non-manual occupations, however this association was not significant when adjusted for confounders. Significant variation in colorectal cancer risk by housing tenure was observed, but only for those in social rented housing, who were at an increased risk of colorectal cancer compared to owner-occupiers. This pattern remained in Model 2 (HR 1.11 [1.00-1.24]). There was no significant difference in colorectal cancer risk among study members in private

rented accommodation compared to those in owner-occupied housing. No significant variation in colorectal cancer risk was found by area deprivation (Townsend quintile).

The complimentary log-log plots for the analysis of colorectal cancer registrations (see Appendix E) suggest that the proportional hazard assumption holds for the educational attainment and social class indicators. On the log-log plot for the housing tenure indicator, there is some crossing of the line for private rented with the line for owner-occupation, however, the lines for social rented and owner-occupation and pretty parallel throughout the study period. This suggests the hazard ratio for private rented housing tenure, in comparison to owner-occupation and social rented housing tenure may have changed over time. The bumpy plot for private rented may also be due to a smaller sample size in this category. It was difficult to assess the proportional hazards assumption for the indicator of area deprivation as there is very little difference in the hazard ratios by deprivation quintile, however, there does appear to be some crossing of the lines which might suggest the proportional hazards assumption does not hold for this indicator.

Variable	Category	Model 1		Model 2	
		HR [95% CI]	p-value	HR [95% CI]	p-value
Educational attainment	No degree	1		1	
	Degree	**0.76 [0.69-0.83]**	**<0.001**	**0.89 [0.81-0.98]**	**0.016**
Social class	Non manual	1		1	
	Manual	**1.16 [1.09-1.24]**	**<0.001**	1.04 [0.97-1.12]	0.270
Housing tenure	Owner occupied	1		1	
	Private rented	1.09 [0.95-1.24]	0.227	1.05 [0.90-1.23]	0.530
	Social rented	**1.21 [1.12-1.31]**	**<0.001**	**1.11 [1.00-1.24]**	**0.046**
Area deprivation	1 - Least deprived	1		1	
	2	1.02 [0.94-1.09]	0.677	0.99 [0.92-1.07]	0.806
	3	0.98 [0.90-1.07]	0.722	0.97 [0.89-1.06]	0.456
	4	0.95 [0.86-1.05]	0.331	0.96 [0.87-1.06]	0.427
	5 - Most deprived	1.00 [0.87-1.15]	0.966	1.08 [0.94-1.24]	0.293

Model 1 is unadjusted. Model 2 is adjusted for: age, sex, ethnicity and area deprivation (Educational attainment model); age, sex, ethnicity, educational attainment, area deprivation (Social class model); age sex, education, social class, marital status and area deprivation (Housing tenure model); age and ethnicity (Area deprivation model)

Table 5.5 Cox proportional hazards regression model results for risk of colorectal cancer registration April 2001-December 2015 (Data source: ONS LS)

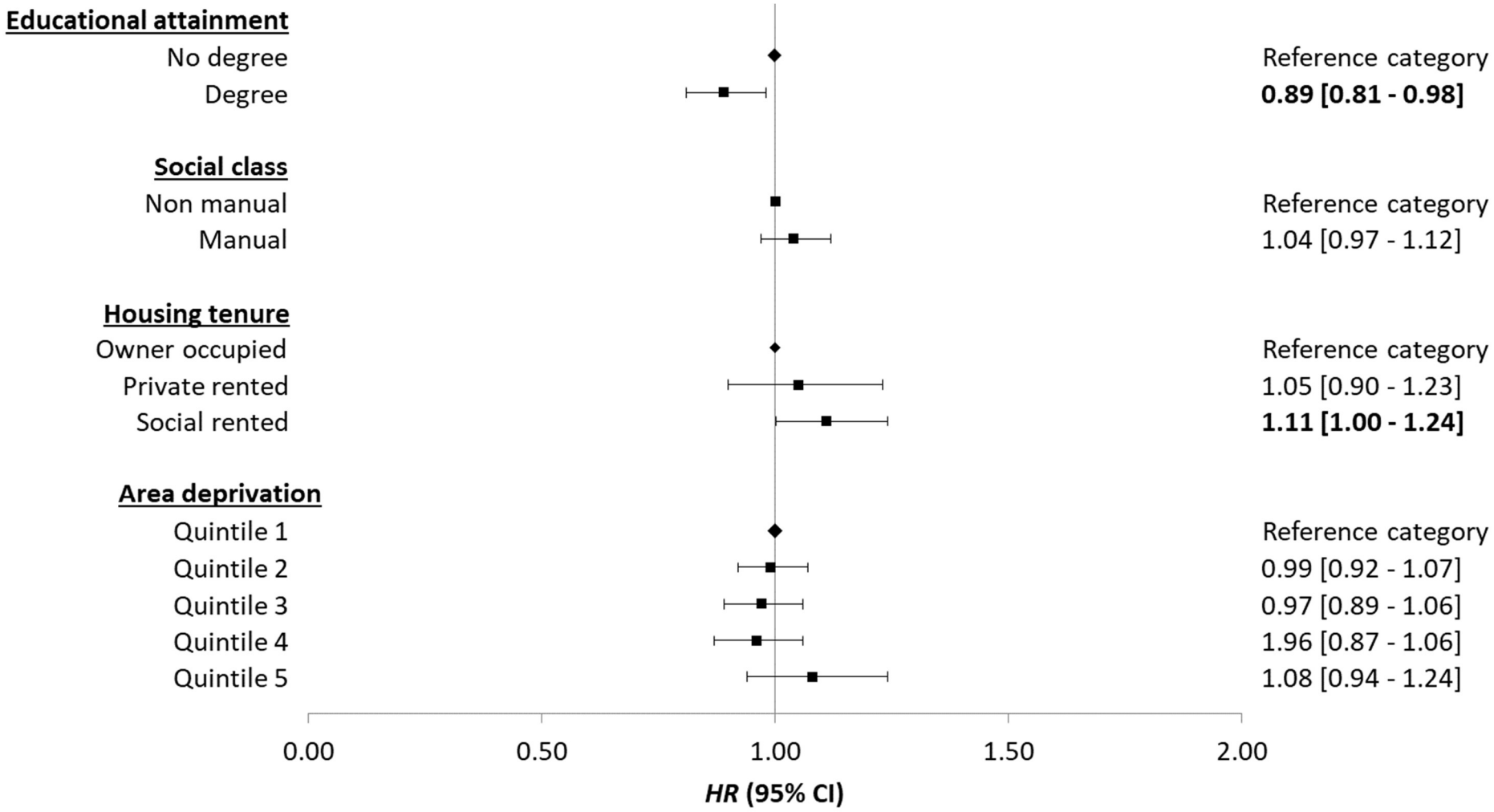


Figure 5.11 Forest plot of adjusted hazards ratios of colorectal cancer registration by indicator of socio-economic status (Data source: ONS LS)

5.6.2 Colorectal cancer survival (all-cause)

This section presents results of survival analysis for patients diagnosed with colorectal cancer at age 50 or over between April 2001 and December 2015 and followed up until December 2016. The event in this analysis is death from all causes. The study sample and exclusions are outlined in Figure 5.12. The number of colorectal cancer registrations is greater than that among the previous cohort (section 1.6.1) as it includes all diagnoses between April 2001 and December 2015 to LS members aged 50+ at diagnosis (i.e. it includes some people aged under 50 in 2001 but who were 50+ at diagnosis, whereas the analysis of colorectal cancer registrations was restricted to only those aged 50+ at the April 2001 Census). Descriptive statistics of the survival times are presented, followed by results of the log-rank test for difference and Cox proportional hazards regression models.

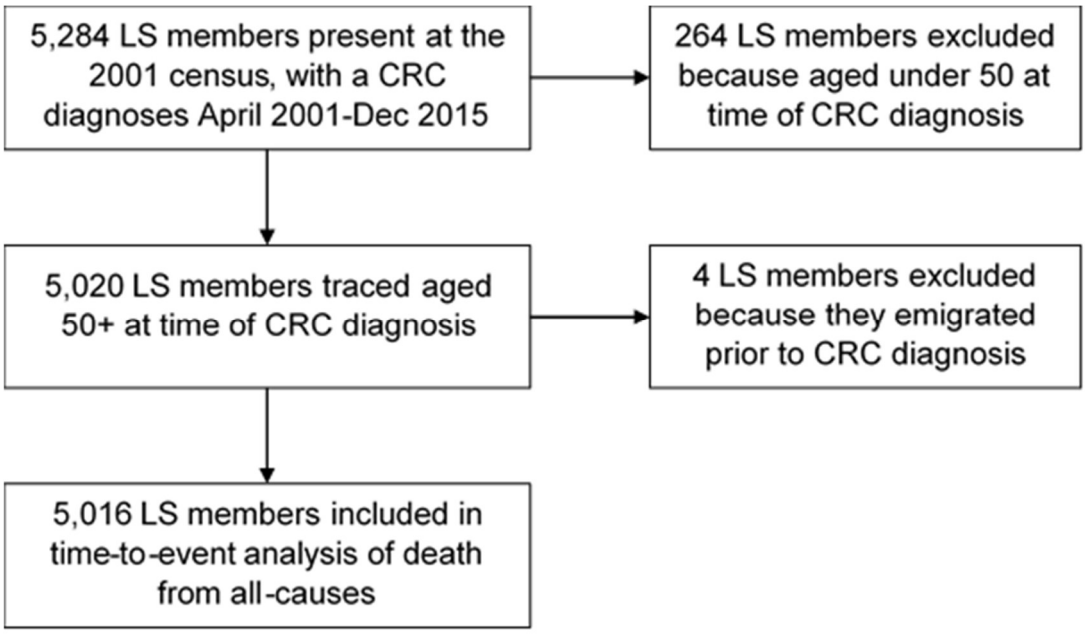


Figure 5.12 Flow diagram of study sample for analysis of CRC survival from all-causes of death

The socio-demographic characteristics of the study members are reported in Table 5.6. There were slightly more males than females (54% and 46% respectively). The 70-79 age group had the largest proportion (33%) of colorectal cancer diagnoses. The vast majority of the cohort were in the White ethnic group (97%). The majority of the study cohort did not have a degree

(87% compared to 12% with a degree). Forty-four percent of the cohort were employed in non-manual occupations, compared to 38% in manual occupations. Again, as in the previous analysis, social class information could not be recorded for a large number of LS members. The majority of the cohort lived in owner-occupied housing (77%), followed by social rented housing (16.3%) and private rented housing (5.6%). A higher proportion of sample members lived in the least deprived areas compared to the most deprived areas.

Variable	Category	n	%
Total		5,016	
Sex	Male	2,696	53.7
	Female	2,320	46.3
Age group	50-59	590	11.8
	60-69	1,244	24.8
	70-79	1,654	33.0
	80+	1,528	30.5
Ethnic group	White	4,862	96.9
	Black	46	0.9
	Indian/Pakistani/Bangladeshi	78	1.6
	Chinese/Mixed/Other	29	0.6
	Missing	1	0.0
Marital status	Married	3,337	66.5
	Separated/Divorced/Widowed	1,352	27.0
	Single	327	6.5
Educational attainment	No degree	4,347	86.7
	Degree	608	12.1
	Missing	61	1.2
Occupational social class	Non manual	2,183	43.5
	Manual	1,913	38.1
	Missing	920	18.3
Housing tenure	Owner occupied	3,875	77.3
	Privately rented	279	5.6
	Social rented	818	16.3
	Missing (Communal establishment)	44	0.9
Area deprivation	1- Least deprived	2,262	45.1
	2	1,203	24.0
	3	759	15.1
	4	527	10.5
	5 - Most deprived	264	5.3
	Missing	1	0.0

Table 5.6 Characteristics of study sample for analysis of CRC survival from all-causes of death (Data source: ONS LS)

There were 3,048 deaths from all causes among the study population during a mean follow up time of 3.8 years. Three deaths of study members who died after embarkation (exit) were excluded from the analysis which left a sample of 3,045 deaths for the time-to-event analysis.

The proportion of deaths among study members was significantly different by age group, educational attainment, social class, housing tenure and area deprivation (Table 5.7). There was no significant difference by sex.

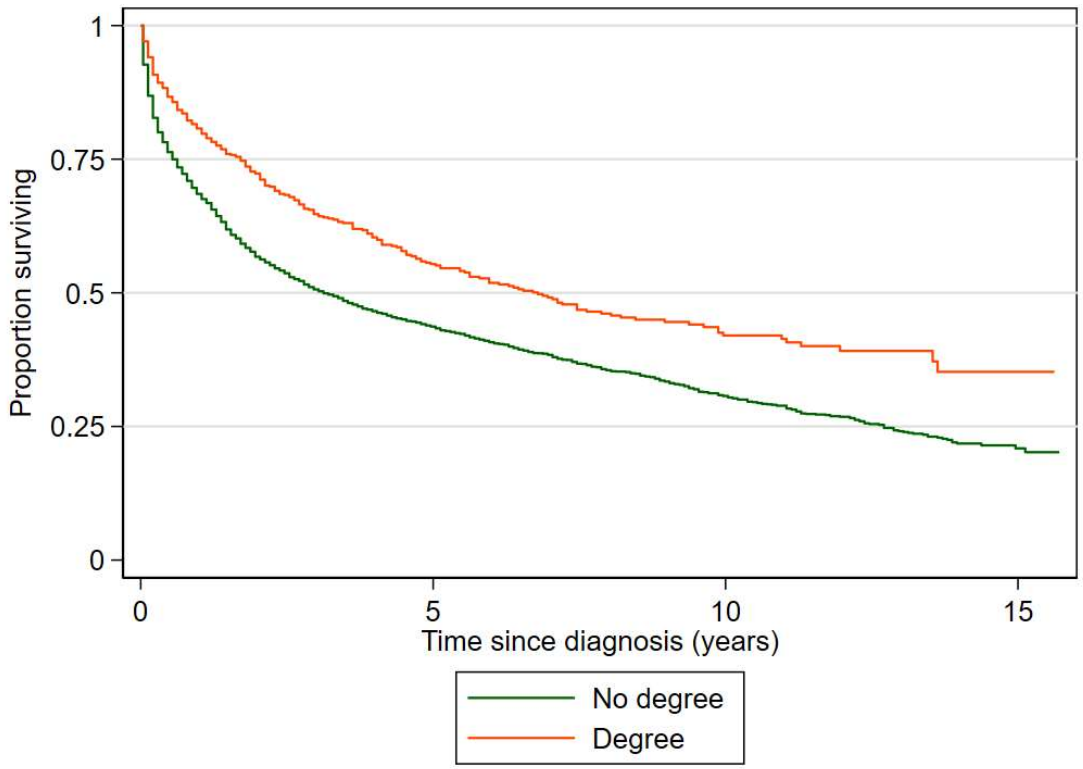
The proportion of members who died during follow up increased with age. For individual-level indicators of socio-economic status, the proportion of deaths was highest among those with no degree, those employed in manual occupations and those in social rented housing. By area deprivation, the proportion of deaths was highest in the most deprived Townsend quintile and lowest in the least deprived quintile, however there was not a clear gradient in the intermediate groups.

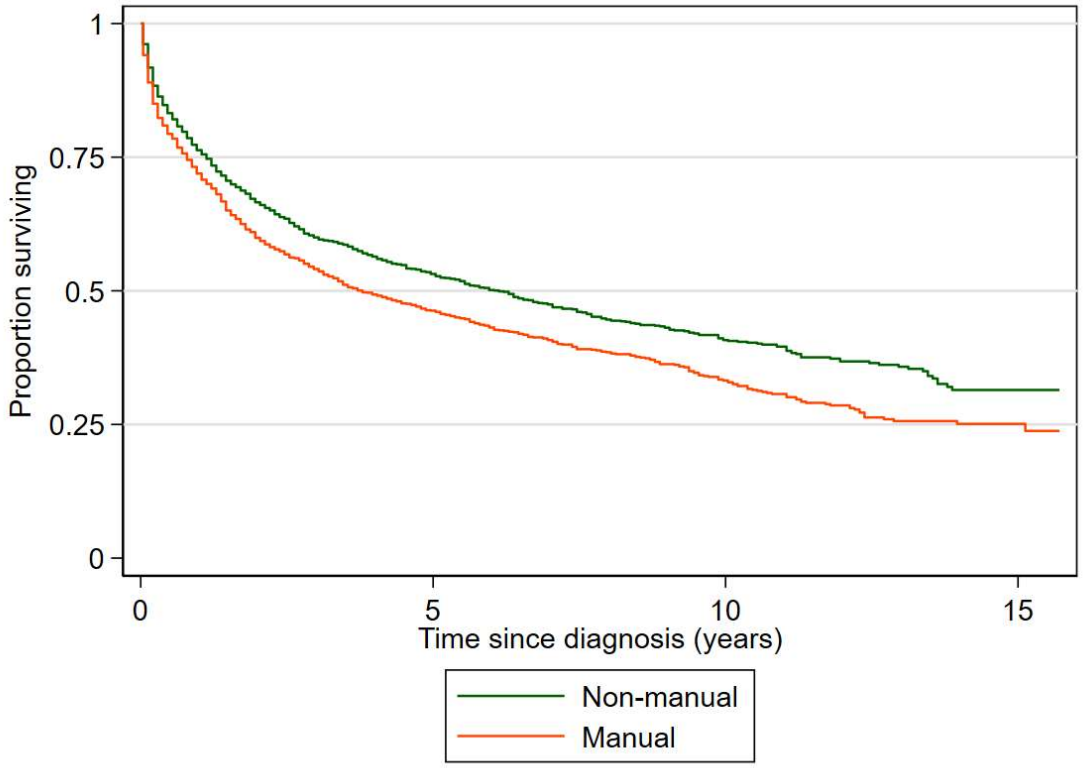
		Total	Dead		Alive		Chi-squared
Variable	Category		n	%	n	%	
Total		5,016	3,045	60.7	1,971	39.3	
Sex	Male	2,696	1,615	59.9	1,081	40.1	p = 0.210
	Female	2,320	1,430	61.6	890	38.4	
Age group	50-59	590	237	40.2	353	59.8	p < 0.001
	60-69	1,244	560	45.0	684	55.0	
	70-79	1,654	978	59.1	676	40.9	
	80+	1,528	1,270	83.1	258	16.9	
Ethnic group	White	4,862	2,960	60.9	1,902	39.1	p = 0.232
	Black	46	29	63.0	17	37.0	
	Indian/Pakistani/Bangladeshi	78	39	50.0	39	50.0	
	Chinese/Mixed/Other	29	16	55.2	13	44.8	
Marital status	Married	3,337	1,871	56.1	1,466	43.9	p < 0.001
	Separated/Divorced/Widowed	1,352	965	71.4	387	28.6	
	Single	327	209	63.9	118	36.1	
Educational attainment	No degree	4,347	2,696	62.0	1,651	38.0	p < 0.001
	Degree	608	291	47.9	317	52.1	
Occupational social class	Non manual	2,183	1,120	51.3	1,063	48.7	p < 0.001
	Manual	1,913	1,127	58.9	786	41.1	
Housing tenure	Owner occupied	3,875	2,221	57.3	1,654	42.7	p < 0.001
	Privately rented	279	182	65.2	97	34.8	
	Social rented	818	600	73.3	218	26.7	
Area deprivation	1- Least deprived	2,262	1,305	57.7	957	42.3	p < 0.001
	2	1,203	755	62.8	448	37.2	
	3	759	479	63.1	280	36.9	
	4	527	321	60.9	206	39.1	
	5 - Most deprived	264	184	69.7	80	30.3	

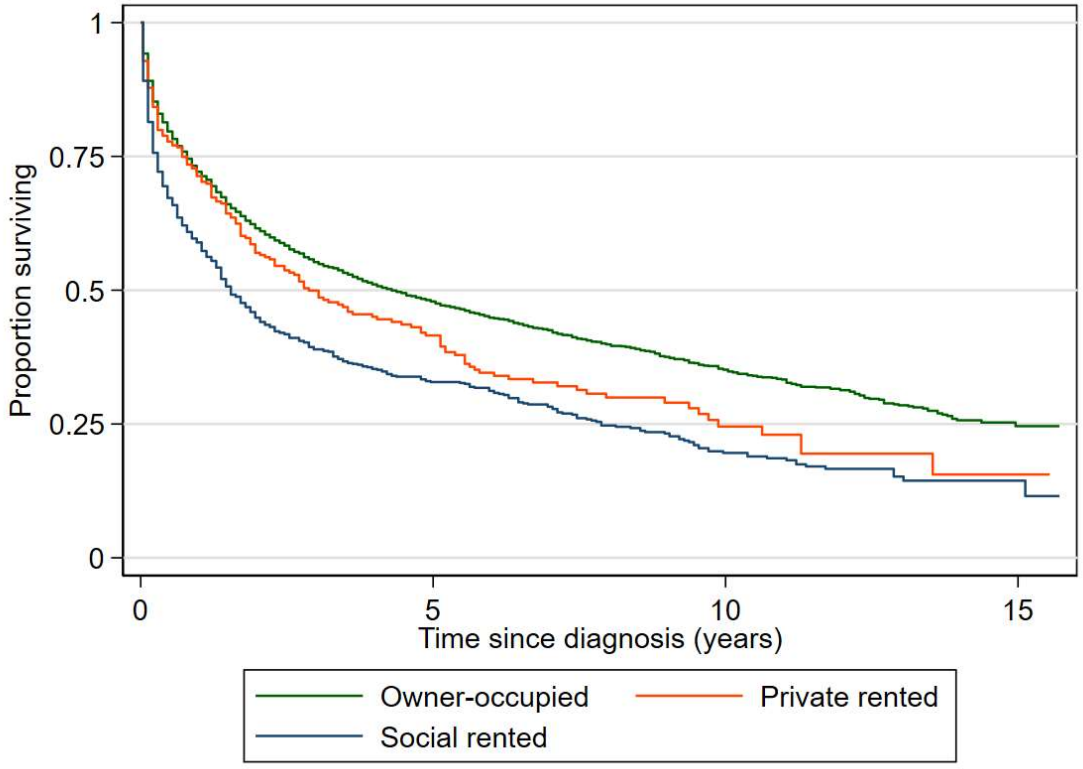
Table 5.7 Study sample by demographic characteristics and death from all-causes (Data source: ONS LS)

5.6.2.1 Time-to-event analysis

Kaplan-Meier survival curves are shown in Figure 5.13 showing the proportion of the study sample surviving over the follow-up time by indicator of socio-economic status. The survival trends are summarised in Table 5.8. Study members with a degree had better all-cause survival than those without a degree (5-year survival of 55%, compared to 44%). Those employed in non-manual occupations had better survival than those in manual occupations (5-year survival of 53%, compared to 46%). By housing tenure, 48% of those living in owner occupied housing survived for at least 5-years, compared to 42% and 33% of those in private rented and social rented housing respectively. Study members living in the most deprived areas had poorer survival than those living in the least deprived areas (5-year survival of 38% and 48% respectively). There was little variation in survival between the intermediary quintiles of area deprivation.




Proportion surviving
1
0.75
0.5
0.25
0
0
5
10
15
Time since diagnosis (years)
Non-manual
Manual


Proportion surviving
1
0.75
0.5
0.25
0
0
5
10
15
Time since diagnosis (years)
Owner-occupied
Private rented
Social rented

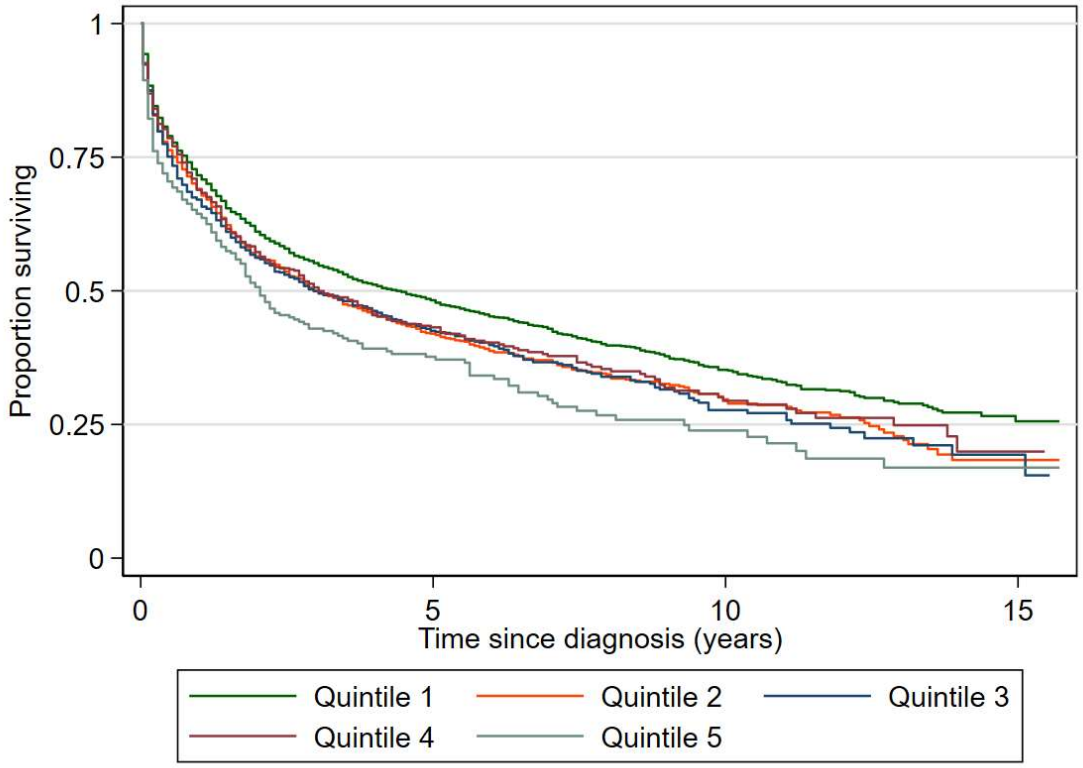


Figure 5.13 Kaplan-Meier survival curves for probability of death from all-causes by: a) educational attainment; b) social class; c) housing tenure; d) area deprivation (Data source: ONS LS)

		Proportion surviving over time [95% CI]		
Variable	Category	5-year	10-year	15-year
Educational attainment	No degree	0.44 [0.42-0.45]	0.31 [0.29-0.32]	0.21 [0.18-0.23]
	Degree	0.55 [0.51-0.60]	0.42 [0.37-0.47]	0.35 [0.28-0.42]
Social class	Non manual	0.53 [0.51-0.55]	0.41 [0.38-0.43]	0.31 [0.28-0.35]
	Manual	0.46 [0.44-0.49]	0.33 [0.31-0.36]	0.25 [0.22-0.28]
Housing tenure	Owner occupied	0.48 [0.46-0.50]	0.35 [0.33-0.37]	0.25 [0.22-0.28]
	Private rented	0.42 [0.35-0.48]	0.25 [0.18-0.31]	0.16 [0.08-0.26]
	Social rented	0.33 [0.29-0.36]	0.20 [0.16-0.23]	0.14 [0.11-0.18]
Area deprivation	1 - Least deprived	0.48 [0.46-0.50]	0.35 [0.33-0.38]	0.26 [0.22-0.29]
	2	0.42 [0.39-0.45]	0.30 [0.27-0.33]	0.18 [0.14-0.23]
	3	0.43 [0.39-0.46]	0.28 [0.24-0.32]	0.19 [0.14-0.26]
	4	0.43 [0.39-0.48]	0.29 [0.24-0.35]	0.20 [0.13-0.28]
	5 - Most deprived	0.38 [0.32-0.44]	0.24 [0.18-0.30]	0.17 [0.11-0.24]

Table 5.8 Summary of all-cause survival trends by indicator of socio-economic status (Data source: ONS LS)

There was a statistically significant difference in all-cause survival by educational attainment, social class, housing tenure and area deprivation. There were more deaths observed than expected among study members in manual occupations, private and social rented housing and in the two most deprived Townsend quintiles. Conversely, there were fewer deaths observed than expected among members with a degree, in non-manual occupations, owner-occupied housing and in the least deprived Townsend quintile (Table 5.9).

Variable	Category	Observed cases	Expected cases	Log-rank test
Educational attainment	No degree	2,696	2,574	p < 0.001
	Degree	291	413	
Occupational social class	Non manual	1,120	1,235	p < 0.001
	Manual	1,127	1,012	
Housing tenure	Owner occupied	2,221	2,420	p < 0.001
	Privately rented	182	163	
	Social rented	600	421	
Area deprivation	1- Least deprived	1,305	1,433	p < 0.001
	2	755	712	
	3	479	444	
	4	321	310	
	5 - Most deprived	184	145	

Table 5.9 Log-rank test results for indicators of socio-economic status showing observed versus expected number of deaths from all causes and test for difference (Data source: ONS LS)

In Cox Proportional Hazards regression (Table 5.10), study members with a degree had a statistically significantly decreased risk of death compared to those without a degree (unadjusted Model 1). This pattern remained in Model 2 (adjusted for age, sex, ethnic group and area deprivation) (HR 0.82 [0.72-0.92]). Conversely, study members employed in manual occupations had a statistically significantly increased risk of death compared those employed in non- manual occupations, which remained in Model 2 (adjusted for age, sex, ethnic group, education and area deprivation) (HR 1.17 [1.07-1.28]). Study members living in private rented and social rented housing had a statistically significantly increased risk of death compared to those living in owner occupied housing. The increased risk for those in private rented housing was no longer significant in Model 2 (adjusted for age, sex, educational attainment, social class, marital status and area deprivation) but remained for those in social rented housing

(HR 1.31 [1.16-1.49]). There was a statistically significant increased risk of death among those living in the most deprived Townsend quintile compared to the least deprived quintile which remained in Model 2 (adjusted for age and ethnic group) (HR 1.36 [1.16-1.59]). Quintiles 2, 3, and 4 also had a statistically significantly increased risk of death compared to Quintile 1 (least deprived) in Model 1 and 2 but there was not a clear gradient among these groups.

Complimentary log-log plots (see Appendix E) suggest the proportional hazards assumptions holds for the educational attainment and social class indicators. There was some divergence of the lines on the plot for housing tenure, although the lines for owning occupied and social rented housing remained pretty parallel. On the plot for area deprivation, there was some crossing of the lines for the intermediate Townsend quintiles, but the lines for Quintile 1 (least deprived) and Quintile 5 (most deprived) remained generally parallel throughout.

		Model 1		Model 2	
Variable	Category	HR [95% CI]	p-value	HR [95% CI]	p-value
Educational attainment	No degree	1		1	
	Degree	**0.67 [0.60-0.76]**	**<0.001**	**0.82 [0.72-0.92]**	**0.001**
Social class	Non-manual	1		1	
	Manual	**1.23 [1.13-1.33]**	**<0.001**	**1.17 [1.07-1.28]**	**0.001**
Tenure	Owner occupied	1		1	
	Private rented	**1.22 [1.05-1.42]**	**0.010**	1.16 [0.96-1.40]	0.125
	Social rented	**1.55 [1.42-1.70]**	**<0.001**	**1.31 [1.16-1.49]**	**<0.001**
Area deprivation	1 - Least deprived	1		1	
	2	**1.16 [1.06-1.27]**	**0.001**	**1.16 [1.06-1.27]**	**0.001**
	3	**1.18 [1.07-1.31]**	**0.002**	**1.18 [1.06-1.31]**	**0.002**
	4	**1.14 [1.01-1.28]**	**0.040**	**1.17 [1.04-1.33]**	**0.010**
	5 - Most deprived	**1.39 [1.19-1.62]**	**<0.001**	**1.36 [1.16-1.59]**	**<0.001**

Model 1 is unadjusted. Model 2 is adjusted for: age, sex, ethnicity and area deprivation (Educational attainment model); age, sex, ethnicity, educational attainment, area deprivation (Social class model); age sex, education, social class, marital status and area deprivation (Housing tenure model); age and ethnicity (Area deprivation model)

Table 5.10 Cox proportional hazards regression model results for risk of death from all-causes April 2001-December 2016 (Data source: ONS LS)

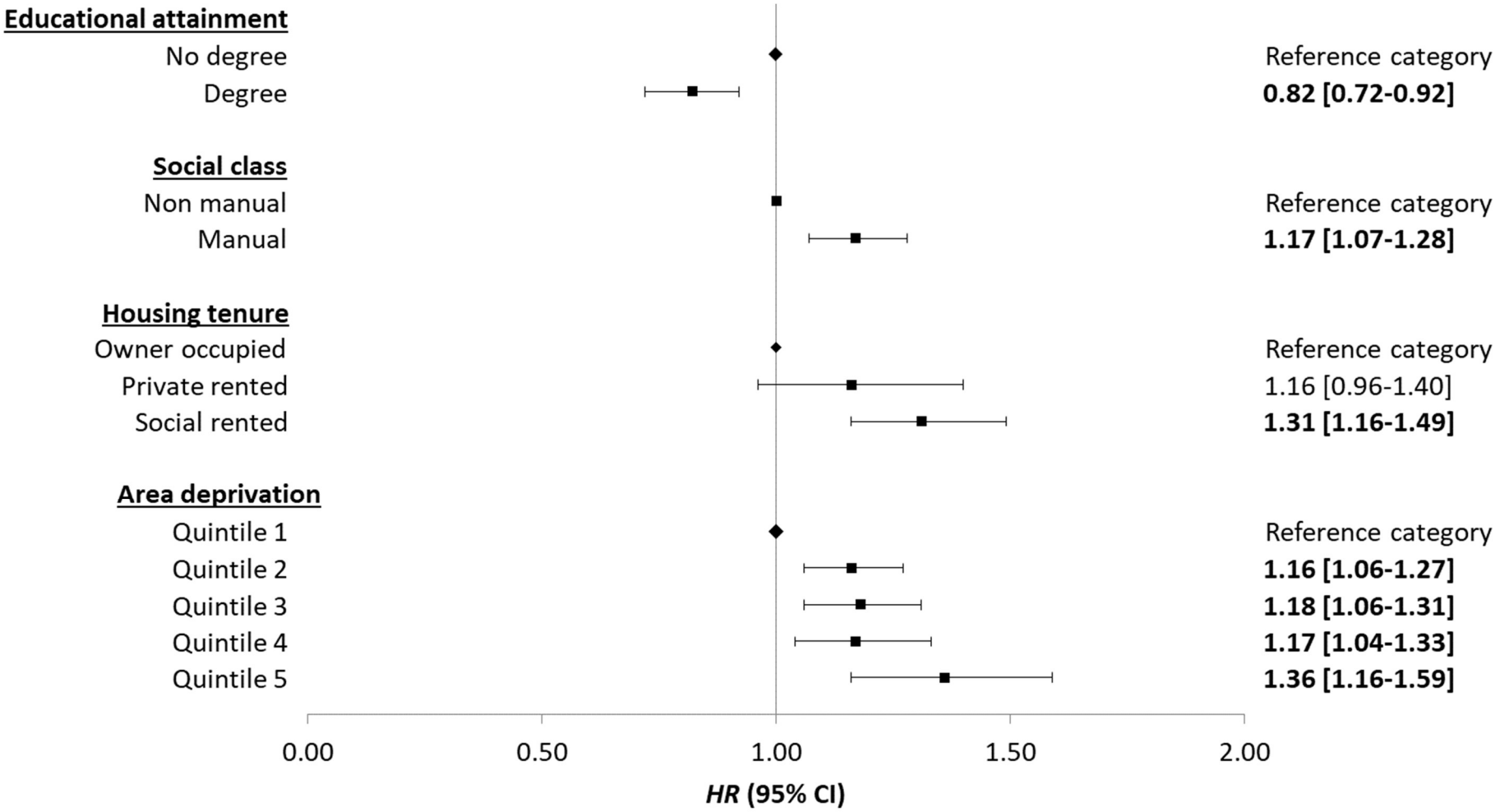


Figure 5.14 Forest plot of adjusted hazards ratios of death from all causes by indicator of socio-economic status (Data source: ONS LS)

5.6.3 Colorectal cancer survival (cause-specific death)

This section presents results of survival analysis for the same cohort as section 1.6.2 (patients diagnosed with colorectal cancer at age 50 or over between April 2001 and December 2015 and followed up until December 2016) but the event of interest in this analysis is death from colorectal cancer. Of the 3,045 deaths of study members during the follow-up period, there were 1,825 deaths (60%) from colorectal cancer and 1,220 (40%) from other causes. LS members who died from other causes were censored at the date of death in this analysis.

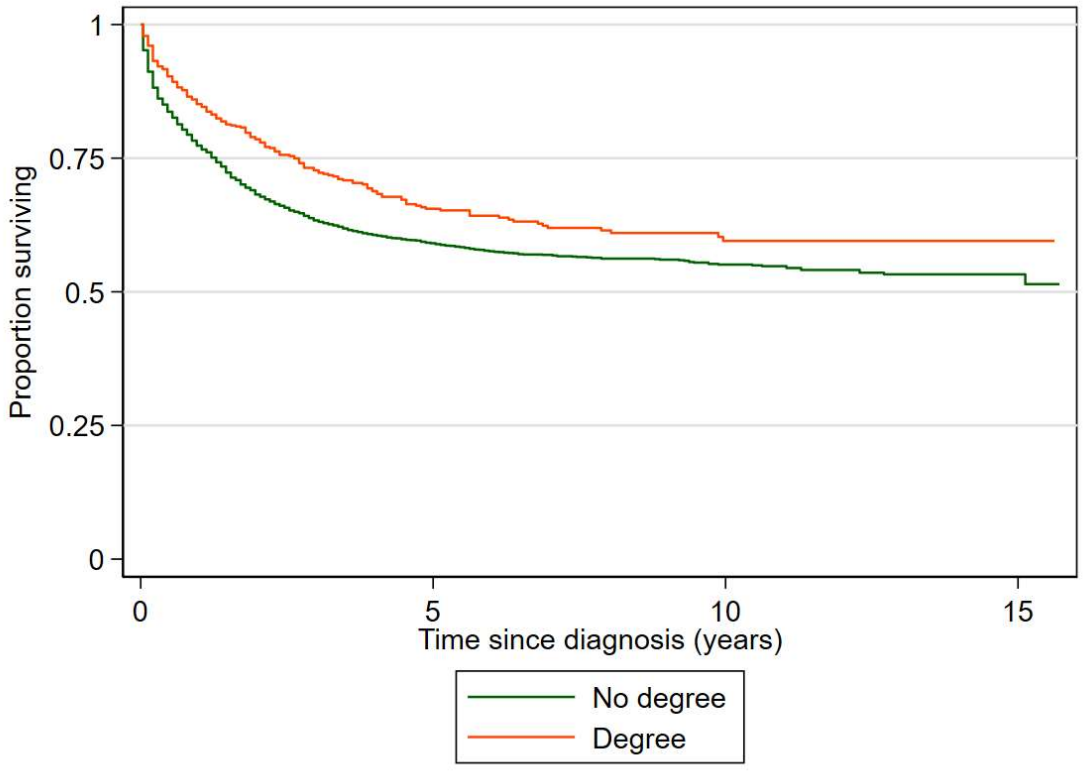
The proportion of the study cohort who died of colorectal cancer was significantly different by sex (higher in females), age group (higher in older age groups), educational attainment (higher among those without a degree), social class (higher among those in manual occupations), housing tenure (higher among those living in social rented housing) and area deprivation (higher in the most deprived quintile compared to the least deprived quintile) (Table 5.11).

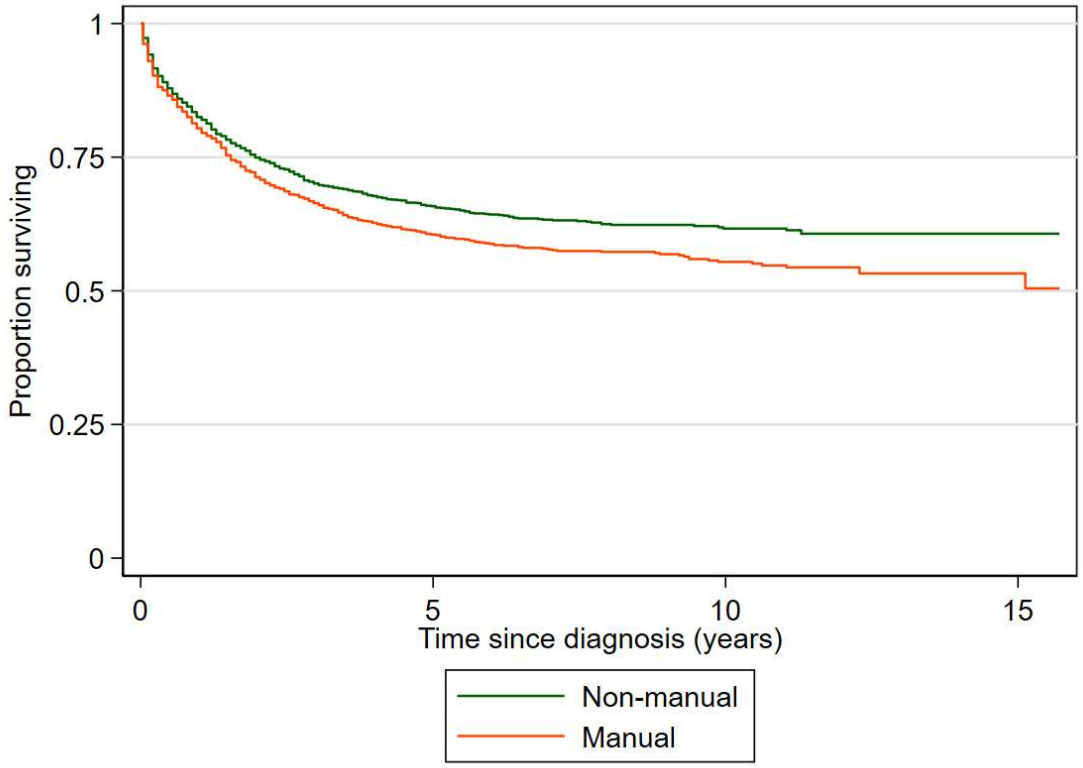
		Total	Dead from CRC		Not dead from CRC		Chi-squared
Variable	Category		n	%	n	%	
Total		5,016	1,825	36.4	3,191	63.6	
Sex	Male	2,696	928	34.4	1,768	65.6	p = 0.002
	Female	2,320	897	38.7	1,423	61.3	
Age group	50-59	590	168	28.5	422	71.5	p < 0.001
	60-69	1,244	359	28.9	885	71.1	
	70-79	1,654	562	34.0	1,092	66.0	
	80+	1,528	736	48.2	792	51.8	
Ethnic group	White	4,862	1,772	36.4	3,090	63.6	p = 0.157
	Black	46	20	43.5	26	56.5	
	Indian/Pakistani/Bangladeshi	78	20	25.6	58	74.4	
	Chinese/Mixed/Other	29	12	41.4	17	58.6	
Marital status	Married	3,337	1,120	33.6	2,217	66.4	p < 0.001
	Separated/Divorced/Widowed	1,352	578	42.8	774	57.2	
	Single	327	127	38.8	200	61.2	
Educational attainment	No degree	4,347	1,602	36.9	2,745	63.1	p = 0.011
	Degree	608	192	31.6	416	68.4	
Occupational social class	Non manual	2,183	690	31.6	1,493	68.4	p = 0.009
	Manual	1,913	679	35.5	1,234	64.5	
Housing tenure	Owner occupied	3,875	1,331	34.3	2,544	65.7	p < 0.001
	Privately rented	279	103	36.9	176	63.1	
	Social rented	818	364	44.5	454	55.5	
Area deprivation	1- Least deprived	2,262	775	34.3	1,487	65.7	p = 0.039
	2	1,203	457	38.0	746	62.0	
	3	759	281	37.0	478	63.0	
	4	527	200	38.0	327	62.0	
	5 - Most deprived	264	111	42.0	153	58.0	

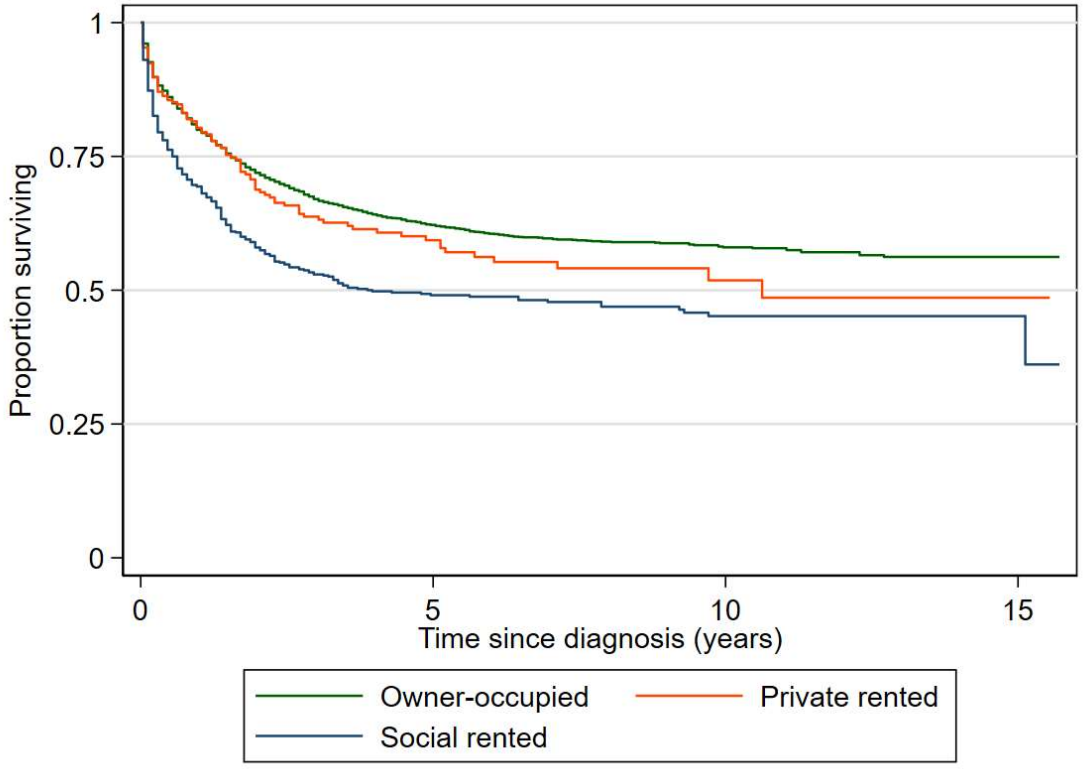
Table 5.11 Study sample by demographic characteristics and cause-specific death (Data source: ONS LS)

Kaplan-Meier survival curves show cause-specific survival over the follow-up period by indicator of socio-economic status (Figure 5.15). The survival trends are summarised in Table 5.12.

Study members with a degree had better cause-specific survival than those without a degree (5-year survival of 66% compared to 59%). Those employed in mon-manual occupations had better cause-specific survival than those in manual occupations (5-year survival of 66% compared to 61%). Conversely, members living in social rented housing (5-year survival of 49%) had worse survival than those in owner occupied (62%) and private rented housing (59%). People living in the least deprived areas had the best cause-specific survival and those in the most deprived areas had the worst survival (5-year survival of 63% and 52% respectively).




1
0.75
0.5
0.25
0
Proportion surviving
0
5
10
15
Time since diagnosis (years)
Non-manual
Manual


1
0.75
0.5
0.25
0
Proportion surviving
0
5
10
15
Time since diagnosis (years)
Owner-occupied
Private rented
Social rented

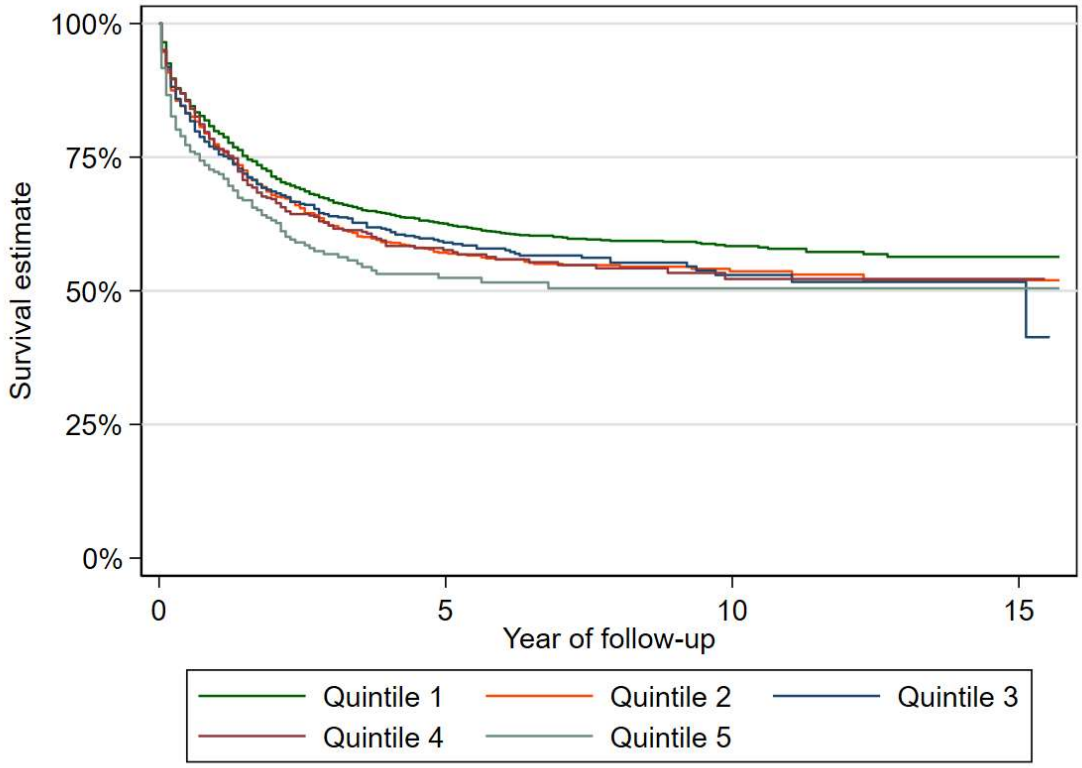


Figure 5.15 Kaplan-Meier survival curves for probability of colorectal cancer death by: a) educational attainment; b) social class; c) housing tenure; d) area deprivation (Data source: ONS LS)

		Proportion surviving over time [95% CI]		
Variable	Category	5-year	10-year	15-year
Educational attainment	No degree	0.59 [0.57-0.61]	0.55 [0.53-0.57]	0.53 [0.51-0.55]
	Degree	0.66 [0.61-0.70]	0.60 [0.54-0.64]	0.60 [0.54-0.64]
Social class	Non manual	0.66 [0.64-0.68]	0.62 [0.59-0.64]	0.61 [0.58-0.63]
	Manual	0.61 [0.58-0.63]	0.55 [0.52-0.58]	0.53 [0.50-0.57]
Housing tenure	Owner occupied	0.62 [0.60-0.64]	0.58 [0.56-0.60]	0.56 [0.54-0.58]
	Private rented	0.59 [0.52-0.66]	0.52 [0.43-0.60]	0.49 [0.38-0.58]
	Social rented	0.49 [0.45-0.53]	0.45 [0.41-0.50]	0.45 [0.41-0.50]
Area deprivation	1 - Least deprived	0.63 [0.60-0.65]	0.58 [0.56-0.61]	0.56 [0.53-0.59]
	2	0.57 [0.54-0.60]	0.54 [0.50-0.57]	0.52 [0.48-0.56]
	3	0.59 [0.55-0.63]	0.53 [0.48-0.58]	0.52 [0.46-0.57]
	4	0.58 [0.53-0.62]	0.52 [0.46-0.58]	0.52 [0.46-0.58]
	5 - Most deprived	0.52 [0.45-0.59]	0.50 [0.43-0.57]	0.50 [0.43-0.57]

Table 5.12 Summary of cause-specific survival trends by indicator of socio-economic status (Data source: ONS LS)

The difference in cause-specific survival by educational attainment, social class, housing tenure and area deprivation was statistically significant (Table 5.13). There were more observed than expected colorectal cancer deaths among those in manual social class, social rented housing and in the most deprived areas. Conversely, there were fewer colorectal cancer deaths observed than expected among those with a degree, in owner-occupied housing and in the least deprived areas.

Variable	Category	Observed cases	Expected cases	Log-rank test
Educational attainment	No degree	1,602	1549	p < 0.001
	Degree	192	245	
Occupational social class	Non manual	690	750	p = 0.001
	Manual	679	619	
Housing tenure	Owner occupied	1,331	1444	p < 0.001
	Privately rented	103	99	
	Social rented	364	255	
Area deprivation	1- Least deprived	775	849	p = 0.004
	2	457	430	
	3	281	269	
	4	200	188	
	5 - Most deprived	111	88	

Table 5.13 Log-rank test results for indicators of socio-economic status showing observed versus expected number of colorectal cancer deaths and test for difference (Data source: ONS LS)

In Cox Proportional Hazards regression models (Table 5.14), study members employed in manual occupations had an increased risk of colorectal cancer death compared to those in non-manual occupations (unadjusted Model 1), which remained statistically significant in Model 2 (HR 1.19 [1.06-1.34]).

Study members living in social rented housing had an increased risk (HR 1.31 [1.12-1.53]) of colorectal cancer death compared to those living in owner-occupied housing, however, there was not a statistically significant difference in risk among those living in private rented housing compared to owner-occupied housing.

Study members living in the most deprived areas had an increased risk of colorectal cancer death compared to those living in the least deprived areas (HR 1.35 [1.10-1.65]). While there was a small but statistically significant

increase in risk of death in Townsend quintiles 2 and 4, compared to quintile 1 (least deprived), there was not a clear gradient by area deprivation.

People with a degree had a decreased risk of colorectal cancer death, compared to those without a degree but this was not significant in the adjusted model.

The complimentary log-log plots (see Appendix E) suggest the proportional hazards assumption holds for the educational attainment and social class variables. Again, there is some divergence in the lines for private rented housing and owner-occupied housing in the plot for housing tenure which suggest the relationship may have changed over time and there was some crossing of the lines for the intermediate quintiles of area deprivation.

		Model 1		Model 2	
Variable	Category	HR [95% CI]	p-value	Hazard ratio	p-value
Educational attainment	No degree	1		1	
	Degree	**0.76 [0.65-0.88]**	**<0.001**	0.89 [0.77-1.04]	0.140
Social class	Non-manual	1		1	
	Manual	**1.19 [1.07-1.33]**	**0.001**	**1.19 [1.06-1.34]**	**0.003**
Tenure	Owner occupied	1		1	
	Private rented	1.13 [0.92-1.38]	0.247	1.06 [0.83-1.35]	0.666
	Social rented	**1.55 [1.38-1.74]**	**<0.001**	**1.31 [1.12-1.53]**	**0.001**
Area deprivation	1 - Least deprived	1		1	
	2	**1.16 [1.04-1.31]**	**0.010**	**1.16 [1.03-1.30]**	**0.011**
	3	1.14 [1.00-1.31]	0.056	1.14 [0.99-1.31]	0.063
	4	1.17 [1.00-1.36]	0.052	**1.20 [1.03-1.40]**	**0.021**
	5 - Most deprived	**1.38 [1.13-1.68]**	**0.001**	**1.35 [1.10-1.65]**	**0.004**

Model 1 is unadjusted. Model 2 is adjusted for: age, sex, ethnicity and area deprivation (Educational attainment model); age, sex, ethnicity, educational attainment, area deprivation (Social class model); age sex, education, social class, marital status and area deprivation (Housing tenure model); age and ethnicity (Area deprivation model)

Table 5.14 Cox proportional hazards regression model results for risk of colorectal cancer death April 2001 – December 2016 (Data source: ONS LS)

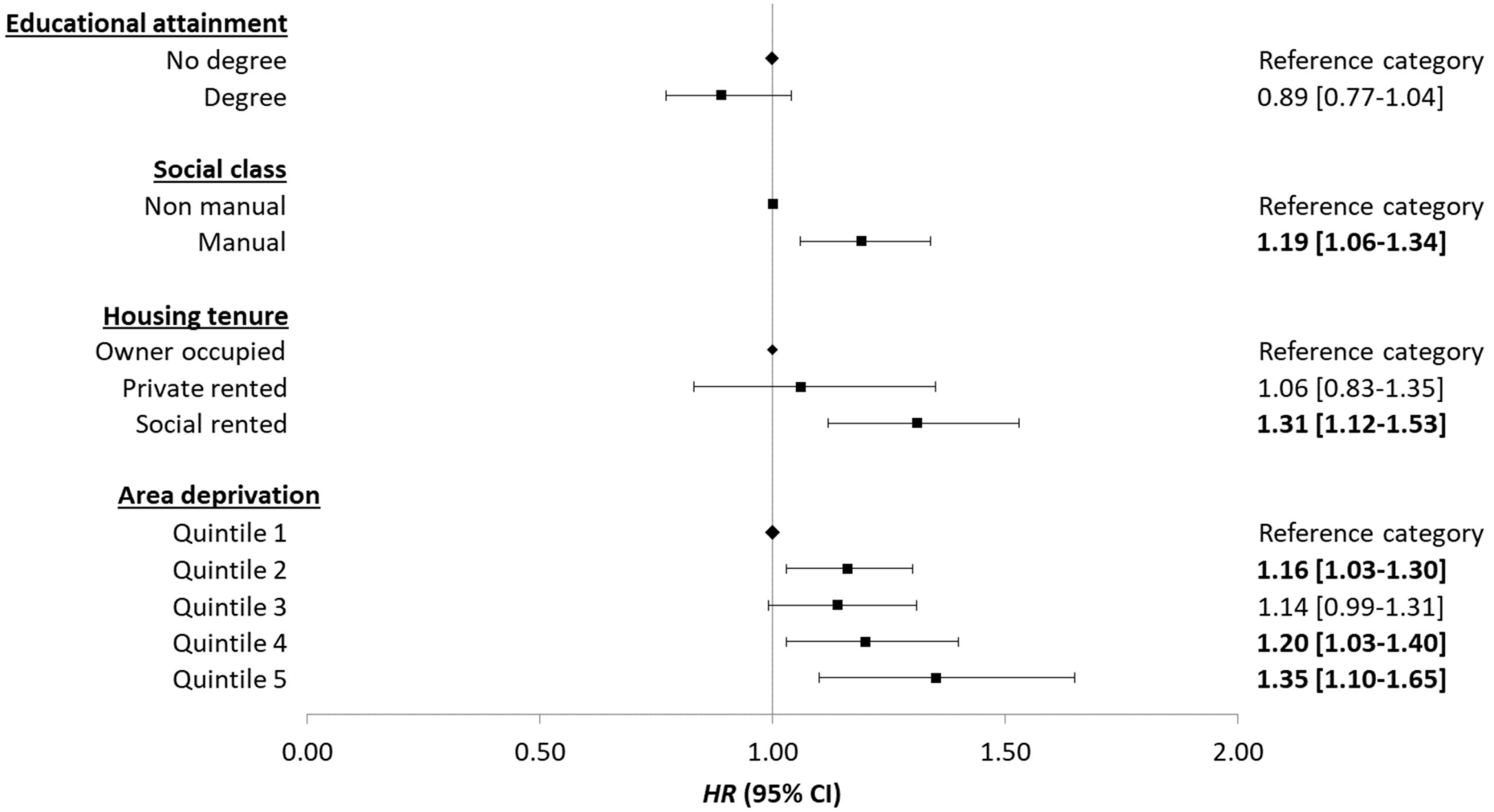


Figure 5.16 Forest plot of adjusted hazards ratios of death from colorectal cancer by indicator of socio-economic status (Data source: ONS LS)

The crude probability of death graph (Figure 5.17) shows the probability of death due to colorectal cancer and due to other causes over the following up period. There was a higher proportion of colorectal cancer deaths in the first 3-years of follow-up. This then levels off at around 5-years and the proportion of deaths from other causes continues to increase, making up a larger proportion of the overall deaths. A similar pattern was observed for all indicators of socio-economic status.

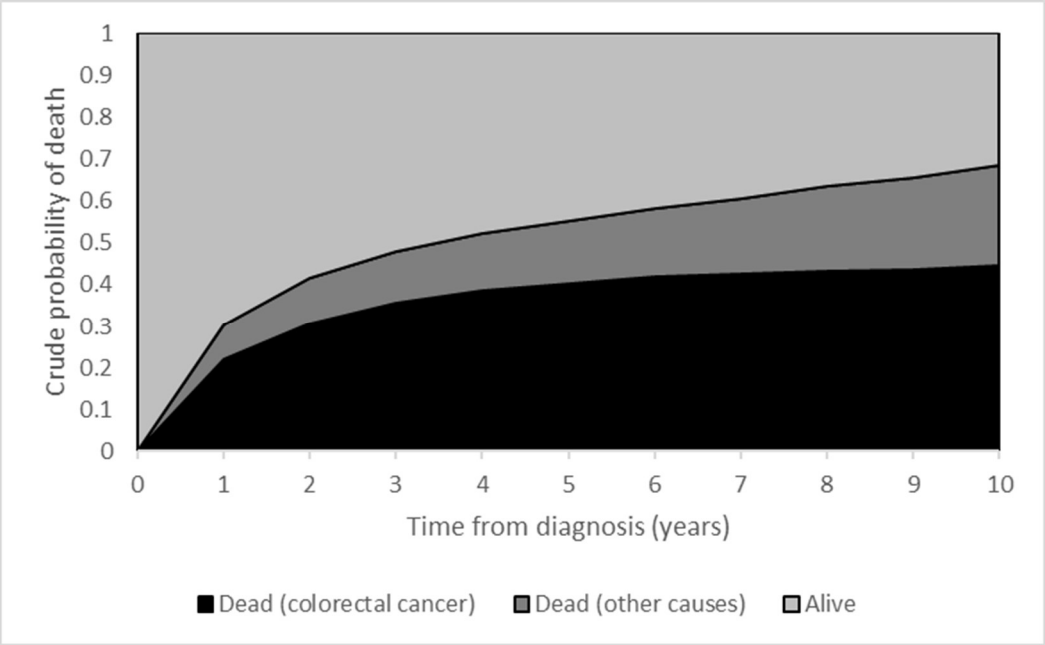


Figure 5.17 Crude probability of death from colorectal cancer and other causes (Data source: ONS LS)

5.7 Discussion

5.7.1 Summary

Socio-economic status and colorectal cancer incidence

This study found a small but statistically significant decreased risk of colorectal cancer among people with a degree, compared to those without a degree. Conversely, a small but significant increase in colorectal cancer risk was observed among people living in social rented housing, compared to owner occupiers. There was an increased risk of colorectal cancer among people in manual occupations compared to non-manual occupations, but this was not statistically significant when adjusted for confounders.

No association was found between CRC risk and area deprivation in this study, which is contrary to recent research which has reported an emerging association between colorectal cancer incidence and deprivation, measured at the area-level. Different measures of area deprivation were used which makes comparisons more difficult. This study used the Townsend Deprivation Index whereas previous studies have used Scottish IMD (Oliphant et al., 2011; Tweed et al., 2018) and English IMD (National Cancer Intelligence Network, 2014) which use different variables to generate the indices. Previous research using the Townsend Deprivation Index in the South Thames region of England found also found no association between colorectal cancer incidence and area deprivation (Pollock and Vickers, 1997), however the study was cross-sectional, whereas the longitudinal nature of the LS data enabled time-to-event analysis to be employed. It is unlikely to be the specifics of the deprivation measures which lead to variations in findings since there are strong correlations between the Townsend index and the IMDs (Norman, 2010).

The disadvantage of using proxy measures of deprivation is that some individuals will be misclassified. Not all individuals living in deprived areas will experience the same level of deprivation. Furthermore, recent research suggests area-level deprivation is not necessarily a good indicator of individual socio-economic circumstances (Ingleby et al., 2020). This could explain the lack of area affects observed in this study despite associations between individual indicators of socio-economic position and CRC incidence.

Most previous studies have used area-level measures of deprivation (due to the lack of individual-level socio-economic data collected by cancer registries), however, one previous study using the LS found the trend in colorectal cancer incidence was inconsistent over time by measurement of social position (Brown et al., 1998). They found incidence was higher in women in the non-manual classes when measured at the 1971 Census but when measured at the 1981 Census incidence was higher among women in the manual classes. In this study, using social class measured at the 2001 Census, an increased risk of CRC incidence was also observed among people in manual occupations (compared to those in non-manual occupations) but the association was not statistically significant when adjusted for confounders.

One explanation for an association between colorectal cancer incidence and socio-economic status could be differential exposure to modifiable risk factors (Brown et al., 2018). There is strong evidence to link socio-economic disadvantage with exposures known to increase risk of colorectal cancer such as unhealthy diet and smoking (Office for National Statistics, 2017). Furthermore, there have been changes over time in the distribution of risk factors, which is likely to have an influence on colorectal cancer trends (Brown et al., 2018).

Socio-economic status and colorectal cancer survival

This study found a statistically significant association between all-cause survival and all four measures of socio-economic status among LS members with a diagnosis of colorectal cancer. An increased risk of death was observed among those without a degree, compared to those with a degree; those employed in manual occupations, compared to those in non-manual occupations; those in social rented housing, compared to owner-occupiers and those in the most deprived areas, compared to the least deprived.

A statistically significant relationship was also found between cause-specific survival and occupational social class, housing tenure and area deprivation. A significant association was not observed with educational attainment. Previous research using the LS found an association between colorectal cancer survival and socio-economic status based on indicators of housing tenure and car access, but not by social class and Carstairs area deprivation indicators (Sloggett et al., 2007). While recent studies have found that variation in relative

(colorectal cancer) survival by socio-economic group is mainly in the short-term after diagnosis (Møller et al., 2012), survival differences in this study persisted throughout the follow-up period.

One explanation for variations in survival observed in this study could be differences in stage at diagnosis by socio-economic status. Evidence for an association between socio-economic status and survival is, however, inconsistent and the relationship between deprivation and survival from colorectal cancer has been found to remain significant after controlling for stage (Dejardin et al., 2014). This suggests there are other factors to explain this relationship. This could include lifestyle factors, co-morbidities, access to and uptake of treatment and participation in screening.

5.7.2 Strengths

A strength of this study is the large, representative sample in the Longitudinal Study population. This improves both the reliability and generalisability of the results. The tracing rate of the LS sample to the NHSCR is high and there is minimal loss to follow-up, due to the quality of the data sources (census data, vital statistics and cancer registration data) and methods of linkage. The proportion of missing values within the LS is also low.

A key strength of the LS data is the ability to investigate cancer outcomes by a range of individual-level indicators of socio-economic status. Most cancer registry-based studies only include area-based measures of deprivation as proxies for individual socio-economic status. Furthermore, the LS has a long follow-up period for analysing cancer outcomes which is an advantage over other cohort studies.

Robust, well established (time-to-event) methods were used to investigate the association between indicators of socio-economic status and colorectal cancer registrations and outcomes. A novel aspect of the methodological approach was the use of DAGs to inform the statistical models.

5.7.3 Limitations

Defining and measuring deprivation is difficult and can make it hard to compare results from different studies. A weakness of the Townsend Index of Deprivation is the applicability of the measures of socio-economic status. For example, the

proportion of people with a degree has increased over time. There has been a recent increase in the proportion of people renting their homes versus home ownership. The distinction between manual and non-manual jobs is not as relevant a measure of wealth as previously (ONS now uses NS-SEC). This highlights the importance of using different measures to investigate the relationship between deprivation and cancer outcomes to validate results.

For the educational attainment and social class variables, where this information was not collected at the 2001 census, it was infilled from the previous census. Only those LS members aged 50+ in 2001 were included in this study, therefore it is unlikely that someone's level of educational attainment or occupational social class would have changed since the previous census so this was considered an appropriate method to increase the proportion of members allocated to a category. Despite infilling, however, there was still a high proportion of study members missing social class information. This could potentially bias the results as women of this age group are more likely to have never worked (and therefore not have been assigned as social class category) than men. Future work could employ alternative strategies to deal with missing data, such as multiple imputation.

A weakness of the LS is that it does not include lifestyle information as this type of information is not collected by the census. Lifestyle is one of the mechanisms by which socio-economic status may influence colorectal cancer risk but it was not possible to investigate this using this data set.

A limitation of the time-to-event analysis was that only month and year of diagnosis and death were used to calculate the survival time which may affect the accuracy of the results (Woods et al., 2012). Day of diagnosis and death could not be used due to access restrictions necessary to protect the confidentiality of LS members. A weakness of using cause of death information to identify colorectal cancer death is the possible misclassification of cause of death on death certification. Other studies have preferred a relative survival approach over analysis of cause specific deaths (Sloggett et al., 2007).

5.7.4 Further work

Further work could investigate whether the association between socio-economic status and colorectal cancer incidence and survival has changed over time in the LS (by repeating the analysis for different cohorts).

An interesting next step would be to investigate if and how change in individual socio-economic status and area deprivation over the life course influences the risk of colorectal cancer registration and/or survival after a diagnosis of colorectal cancer. Although there are associations of individual attributes and area deprivation with longer survival times, there are no quality of life indicators in the LS. It could be that people are living longer but with debilitating conditions. Stratifying the LS study sample by those persons with and without limiting long-term illness, as self-reported in the censuses since 1991, could be an informative line of enquiry.

5.8 Conclusion

Key messages:

- This study used a large, representative sample of the population of England and Wales to examine colorectal cancer incidence and survival by individual and area-based indicators of socio-economic position.
- Individual measures of socio-economic position were associated with risk of colorectal cancer registration, but there was no association with the area-based measure.
- Socio-economic variation in colorectal cancer survival persists. All indicators were associated with risk of death from all-causes and colorectal cancer death.
- The proportion of deaths from colorectal cancer, compared to all-causes, was highest in the first 3 years following diagnosis and levelled off after this period.

Chapter 6

Developing an area-based index of colorectal cancer risk for England to investigate the association between the local environment and colorectal cancer incidence

6.1 Overview

This chapter will explore variation in colorectal cancer incidence by features of the local environment. It outlines the development of a bespoke area-based risk index that stratifies areas based on expected colorectal cancer risk, according to local environment characteristics. Data from CORECT-R enables linkage to detailed geographic information at small-area (LSOA) level, which was not possible in the geographic analysis of mortality trends in Chapter 4. Furthermore, the CORECT-R data contains more detailed information about cancer registrations, such as stage at diagnosis.

6.2 Background

Over 50% of colorectal cancer cases in the UK are estimated to be preventable (Brown et al., 2018) based on population exposure to modifiable risk factors. It is estimated that preventable cases could be as high as 67% of cases in men and 60% of cases in women, according to the most recent research (Goon et al., 2021).

A number of risk factors are associated with colorectal cancer. There is ‘sufficient’ or ‘convincing’ evidence of a causal association between the following factors and the risk of developing colorectal cancer, according to the International Agency for Research on Cancer (2021) and the World Cancer Research Fund (2018). The proportion of colorectal cancer cases attributable to these risk factors have since been estimated (Brown et al., 2018). Insufficient dietary fibre contributed the largest proportion of preventable colorectal cancer cases (28%), followed by processed meat (13%), overweight and obesity (11%), tobacco smoking (7%), alcohol drinking (6%), insufficient physical activity (5%) and ionising radiation (2%).

The proportion of preventable cancer cases also varied by geographic area, which is likely to reflect demographic and socio-economic differences that influence exposure to modifiable risk factors (Brown et al., 2018). For example,

areas with higher levels of socio-economic deprivation have higher rates of tobacco smoking (Office for National Statistics, 2017). Furthermore, many health behaviours are driven by social and economic factors, such as cost and access, which may disproportionately affect those in deprived areas. For example, access to healthy food options.

While previous research has focused on individual-level exposure to modifiable cancer risk factors, this study will investigate how the local 'neighbourhood' environment is associated with exposure to these risk factors. The broader geographical determinants of health have been considered within the health geography literature. The location of different types of retail outlets is one area of interest. For example, studies have shown a positive relationship between the number of fast food outlets in an area and the rate of childhood obesity (Fraser and Edwards, 2010). The density of alcohol-related outlets has been found to be associated with alcohol-related harms (Sherk et al., 2018) and the density of tobacco outlets associated with the risk of smoking (Shortt et al., 2016). Studies have also investigated the impact of the natural environment on health. There is evidence that accessibility to green space may have benefits for physical health (Maas et al., 2006) and mental wellbeing (Astell-Burt et al., 2014). Furthermore, income-related inequalities in health have been found to be less marked in populations with greater exposure to green space (Mitchell and Popham, 2008). Accessibility to health and support services may also influence health outcomes. Previous research investigating access to health services and cancer risk has found increasing travel time to GP surgery was associated with late stage diagnosis among colorectal cancer patients (Jones et al., 2008b). Furthermore, distance to health care services has been found to be associated with treatment type, quality of life and survival among cancer patients (Ambroggi et al., 2015).

Research on the role of the neighbourhood environment and colorectal cancer risk is limited. Studies examining neighbourhood-level factors have tended to include only socio-economic status. A recent cohort study by Canchola et al. (2017) in the US investigated the association between a set of neighbourhood obesogenic attributes (socio-economic status, population density, restaurant and retail food environments, numbers of recreational facilities and businesses, commute patterns, traffic density and street connectivity) and risk of colorectal

cancer. A modest association between higher traffic density and increased colorectal cancer risk was found, possibly indicating a less walkable environment which could have a negative impact on physical activity, but no other associations were observed. A similar study based on the same cohort examined the impact of changes over time in participants' neighbourhood environment on colorectal cancer risk. A relative increase in population density was associated with higher colorectal cancer risk among male and female non-movers. At the same time, a relative decrease in population density was also associated with a higher risk of colorectal cancer among female non-movers. An increase in the number of recreational facilities was associated with lower colorectal cancer risk among female non-movers (Shvetsov et al., 2020). A recent ecological analysis by Gibson et al. (2020) examined the relationship between the local food environment and colorectal cancer incidence in Texas. Colorectal cancer incidence was not strongly associated with census tracts with higher unhealthy food availability, but the study did not examine other aspects of the built environment.

This study will investigate the relationship between the neighbourhood environment and colorectal cancer incidence in a UK context through the development of an area-based index of colorectal cancer risk.

Area-level multivariate indices are widely used to measure multiple socio-economic deprivation in the UK (for example the Townsend Index and Carstairs index), however, similar composite measures for physical environmental deprivation have only recently been constructed for the UK. The first example of an environmental multivariate risk index for the UK was the Multiple Environmental Deprivation Index (MEDIx) (Richardson et al., 2010). The index combined area-level data across five dimensions of the physical environment with evidence of being either detrimental or beneficial to population health: air pollution, climate, proximity to industrial facilities, exposure to UV radiation and quantity of green space. The MEDIx has been used to demonstrate that exposure to multiple physical environmental deprivation is related to socio-economic deprivation. Air pollution and proximity to industrial facilities (considered detrimental for health) tended to increase as income deprivation increased, whereas the proportion of greenspace and the UVB index (for which higher values are considered beneficial for health) reduced as income

deprivation increased (Pearce et al., 2010). Multiple physical environmental deprivation, measured using the MEDIx, has also been shown to be related to risk of mortality and morbidity. After adjustment for age and sex, all-cause mortality was lowest in areas with the least environmental deprivation and there was a downward gradient in health as environmental deprivation increased. Furthermore, the relationship between MEDIx and heath remained after adjustment for the level of socio-economic deprivation in an area, which suggests that in the UK environmental deprivation is related to health, independently of the age, sex and socio-economic make-up of an area (Pearce et al., 2010). The gradient in health by environmental deprivation was observed across all income-deprivation quintiles, suggesting multiple environmental influences health regardless of the level of income deprivation (Pearce et al., 2010).

A more recent example of a multivariate environment index is the AHAH index (described in more detail in Chapter 3), which combines indicators across four domains of accessibility: retail environment, health services, physical environment and air quality. The AHAH has been used to study the association between multiple features of the neighbourhood environment (hypothesised as being good or bad for health) and physical and mental health outcomes. A significant association between the worst performing areas for health, as identified by the index, and a decrease in mental wellbeing was reported (Green et al., 2018).

It is important to understand the role of exposure to multiple dimensions of environmental deprivation on health outcomes to identify the population most ‘at risk’ and to inform policies which can prioritise and target interventions to ultimately reduce cancer incidence.

6.3 Objectives

Objectives:

- Describe individuals with a colorectal cancer diagnosis by individual attributes (age, sex, ethnicity) and characteristics of the local area in which they reside (area deprivation, population density).
- Develop an area-based index of the types of places with higher risk of colorectal cancer incidence, encompassing three domains: the retail environment, health services, and the natural environment.
- Model the relationship between colorectal cancer incidence and the colorectal cancer risk index, adjusting for demographic and socio-economic characteristics of areas.
- Estimate the odds of late stage diagnosis by the colorectal cancer risk index.

6.4 Data

6.4.1 CORECT-R

Information on colorectal cancer diagnoses in England was sourced from the COloRECTal cancer data Repository (CORECT-R). CORECT-R is a secure data repository containing multiple linked routine data sets from across the cancer pathway of diagnosis, treatment and outcome (Downing et al., 2021). The CORECT-R database contains information for all colorectal (ICD-10 C18-C20) and anal (ICD-10 C21) tumours diagnosed between January 1997 and December 2018 in England.

An application was made to access data held within CORECT-R by submitting a project protocol to the UK Colorectal Cancer Intelligence Hub (UK Colorectal Cancer Intelligence Hub, 2021) (see Appendix D). Once approval was granted, the data was accessed via a secure TRE.

The CORECT-R data extract for this study contained all diagnoses of colorectal cancer (ICD10 C18-C20) in England between 1st January 2014 and 31st December 2018. The CORECT-R definition of colorectal cancer does not include anal tumours (ICD-10 C21). This is not consistent with the definition of colorectal cancer in previous chapters, which include anal tumours, but the number of anal cancers is very small so it is unlikely to affect the results. This time period was chosen as it contains the most recent data available within CORECT-R and spans the year in which data for the area-based index (described below) was sourced. The CORECT-R data is currently only available for England so the coverage of the study was England only.

Selected variables from the CORECT-R data included patient characteristics (sex, age at diagnosis, ethnic group, year of diagnosis), tumour information (stage at disease at diagnosis, site of tumour), geographic information (Lower Layer Super Output Areas) and an indicator of area-based socio-economic deprivation (Index of Multiple Deprivation).

Lower Layer Super Output Areas (LSOA) are small areas designed to be of a similar population size, with an average of approximately 1,500 residents or 650 households. They are produced by the ONS for the reporting of small area statistics. LSOAs were linked to the CORECT-R data at source, using the patients' postcode of residence, and the LSOA level data was made available

for this study. Additional area-level information was then appended to the patient data based on the LSOA in which they reside.

6.4.2 Access to Healthy Assets and Hazards

Data from the Access to Healthy Assets and Hazards (AHAH) resource were used to construct an area-based index of colorectal cancer risk. The AHAH resource (described in more detail in Chapter 3) contains a set of open-source measures of accessibility to health-related features of the environment at LSOA level for Great Britain (Daras et al., 2019).

Features of the environment identified as relevant to colorectal cancer risk within the literature were selected for inclusion in this study. Eight indicators were selected from across three of the AHAH domains: retail outlets, health services and natural environment. Table 6.1 summarises the selected AHAH indicators and their relevance to the colorectal cancer risk factors described above.

AHAH Domain	AHAH Indicator	Relevance to CRC risk
Retail	Fast food outlet accessibility	Processed meat consumption
	Off-licence accessibility	Alcoholic drink consumption
	Tobacconist accessibility	Cigarette smoking
Health services	GP practice accessibility	CRC (early) diagnosis
	A&E hospital accessibility	CRC diagnosis
	Leisure service accessibility	Physical activity
Natural environment	Green spaces (active) accessibility	Physical activity
	Green spaces (passive) accessibility	Physical activity

Table 6.1 Health indicators relevant to colorectal cancer risk

The source data for the AHAH indicators were the mean average postcode distance (in km) to the nearest service for each LSOA. The active green space

indicator is the average distance to the nearest greenspace (in km), whereas the passive green space indicator is the average amount of green space (in km^2) within a 900-metre buffer of postcodes within each LSOA.

6.4.3 Population estimates

Mid-year population estimates by sex and five-year age group for LSOAs in England for the years 2014 to 2018 were sourced from the ONS via NomisWeb (Nomis, 2021b).

6.4.4 Census data

Count and percentages of the population by ethnic group for LSOAs in England at the 2011 census were downloaded from NomisWeb. An indicator of population density, population per hectare, for LSOAs in England at the 2011 census was also downloaded.

6.4.5 Index of Multiple Deprivation

The Index of Multiple Deprivation (IMD) is a multi-dimensional index of socio-economic deprivation for small areas in England. It is a measure of relative levels of deprivation in 32,844 LSOAs in England. The LSOA with a rank of 1 is the most deprived and the LSOA with a rank of 32,844 is the least deprived.

The IMD is an overall measure of deprivation constructed by combining seven domains of deprivation according to their respective weights: Income deprivation (22.5%); Employment deprivation (22.5%); Education, skills and training deprivation (13.5%); Health deprivation and disability (13.5%); Crime (9.3%); Barriers to housing and services (9.3%); Living environment deprivation (9.3%).

The IMD is an appropriate measure of deprivation to use in this study as it is analysing areas in England only (whereas previous chapters have included both England and Wales which cannot be compared using the IMD as it is country-specific measure (Norman, 2010). The 2015 version of the IMD (income domain only) is appended to the patient data within CORECT-R. Quintiles of the IMD scores have also been created within CORECT-R with the rank values flipped so that Quintile 1 represents the least deprived fifth of areas and Quintile 5 represents the most deprived fifth of areas.

For adjustment in the area-based analysis of colorectal cancer incidence rates, the overall IMD index was used (because the IMD income domain information within CORECT-R was only available for LSOAs for which there were patients). The overall IMD quintiles were appended to the patient data so it was consistent. The general pattern of IMD by colorectal cancer index quintile is similar by overall IMD and the income domain only.

6.4.6 Boundary data

ONS digital boundary data were download in ESRI Shapefile format for 2011 LSOAs in England (Office for National Statistics, 2011).

6.5 Methods

6.5.1 Bespoke area-based index

An area-based index of colorectal cancer risk for LSOAs in England was developed using the selected set of open-access health indicators created as individual inputs for the AHAH index outlined in Table 6.2. The colorectal cancer risk index was created following the methodology used by Green et al. (2018) to develop the AHAH index.

Small areas (LSOAs) in England only were included in the colorectal cancer risk index, as the study uses data from the CORECT-R repository, which is currently only available for patients in England. There are a total of 32,844 LSOAs in England.

The colorectal cancer risk index is made up of eight indicators within three domains. Table 6.2 details the indicators and the hypothesised direction of association with colorectal cancer risk. Each of the indicators in the retail domain were hypothesised as increasing colorectal cancer risk. A low value on these indicators (i.e. shorter distances) were assumed to increase risk of engaging in behaviours associated with colorectal cancer risk (e.g. processed meat consumption, alcohol use and smoking), and conversely a high values (i.e. longer distances) were assumed to reduce risk of engaging in these behaviours. Each of the indicators in the health domain were hypothesised as reducing risk of colorectal cancer. Low values on this domain (i.e. short distances) were assumed to increase engagement with health services and leisure facilities and vice versa. A lack of green space was presumed to increase behaviours associated with risk of colorectal cancer (e.g. lack of physical activity), therefore high values of the active green space indicator (i.e. longer distances) were hypothesised as increasing colorectal cancer risk, whereas high values of the passive green space indicator (i.e. larger amount of green space) were hypothesised as reducing colorectal cancer risk.

Domain of CRC risk index	Indicator	Weight within domain	Hypothesised association to CRC*	
			Low value	High value
Retail outlets	Accessibility to fast food outlets	1/3	+	-
	Accessibility to off-licences	1/3	+	-
	Accessibility to tobacconists	1/3	+	-
Health services	Accessibility to GP practices	1/3	-	+
	Accessibility to A&E hospitals	1/3	-	+
	Accessibility to leisure centres	1/3	-	+
Green space	Accessibility to green spaces (active)	1/2	-	+
	Accessibility to green spaces (passive)	1/2	+	-

*A positive value means that a value is positively associated with risk of colorectal cancer (and vice versa).

Table 6.2 Colorectal cancer risk index indicators with weights and direction of association

Each indicator was standardised, due to differences in their distributions and units, by ranking the LSOAs from lowest to highest risk, based on the hypothesised directions of association with colorectal cancer risk described above. The values of the indicators in the retail domain and the passive green space indicator were flipped so the values for all indicators were ranked in the same direction (i.e. from lowest to highest risk). Each standardised variable was transformed and the indicators within each domain were then combined together to create a domain score. Indicators were equally weighted within each domain (see Table 2). The domain scores were then standardised to give them a similar distribution by ranking them from lowest to highest colorectal cancer risk. An exponential transformation was then applied to the ranked domain scores and they were combined, with equal weights, to create an overall colorectal cancer risk score. Quintiles of the overall and domain scores were calculated by splitting the rank values.

Larger scores of the index are interpreted as areas with hypothesised higher risk of colorectal cancer and lower scores as areas with hypothesised lower risk. The interpretation is similar for each of the domain scores. For the retail domain, larger values represent areas closer on average to the retail outlets. Larger values of the health services domain represent areas further away from these services. Larger scores in the green space domain represent areas with a lack of green space. In each case, large scores have been hypothesised to be associated with increased colorectal cancer risk.

6.5.2 Data preparation

The CORECT-R data extract was de-duplicated to create one record (colorectal cancer case) per patient. Where patients have multiple tumours recorded, the rule applied was to take the first tumour (date wise), and if two or more tumours were diagnosed on the same day, the tumour of most advanced stage was retained.

An age-group variable was derived by grouping the patient age at diagnosis variable into 18 five-year age categories: 0-4, 5-9, 10-14…85+. The 18 minor ethnic group categories were further aggregated into five categories based on the ONS recommended groupings for ethnicity (Race Disparity Unit, 2021): White, Mixed, Asian, Black, Other.

A binary variable was created to indicate if the tumour was diagnosed at a late stage. Late stage at diagnosis was defined as stage 3 or stage 4 tumours.

Additional ‘unknown’ categories were created for the ethnicity and stage at diagnosis variables as they contain a large proportion of missing data. Approximately 10% of records in the cancer registry are missing information on stage at diagnosis. Ethnicity data derived from linked HES data is also incomplete (Smith et al., 2017).

Population density data (person per hectare) was sourced for LSOAs in England and Wales at the 2011 census. A population density category was assigned to each LSOA using on the classification outlined in Chapter 3 Table 3.2. The classification comprises five categories, with 1 being the most urban and 5 being the most rural.

6.5.4 Statistical methods

Descriptive statistics are reported showing the number and proportion of people with a colorectal cancer diagnosis in the England by patient characteristics (age, sex, ethnic group) and area-based attributes (rurality, socio-economic deprivation). The local area attributes were categorised based on quintiles of their distribution across England.

A series of choropleth maps were produced to visualise the distribution of the colorectal cancer risk index and sub-domains across England.

Crude incidence rates were calculated by risk index and sub-domain quintiles by dividing the number of cases by the person-years at risk. When calculating rates for the entire population, population at risk is generally approximated by the mid-year population which is based on the previous census and updated annually to account for demographic change (Office for National Statistics, 2021c). Cases for several years of observation (2014-2018) were used (to ensure robustness of incident rate estimates) and the average annual incidence rate calculated. The denominator is estimated as person-years, by summing up the mid-year population estimates for each of the years under investigation (Boyle and Parkin, 1991), in this case mid-year population estimates for 2014-2018. As is conventional in cancer studies, the incidence rates are expressed as cases per 100,000 person-years. It should be noted that when population estimates are used to approximate person-years at risk, the denominator of the rates will include a few person who are not truly at risk, however, for the study of incidence rates of particular cancer, this makes little difference, since the number of individuals in the population who are alive and already have a cancer of a specific site is relatively small (Boyle and Parkin, 1991).

Age-specific rates by risk index and sub-domain quintiles were estimated by dividing the number of cases in each age-group by the person-years at risk in the corresponding age group in the study time period. Age-standardised rates were estimated by applying the age-specific rates to the European Standard Population. This method is described in more detail in Chapter 4 Section 4.5.2. Age-standardised rates account for differences in the age-structure of the population in areas within each index quintile.

Poisson regression models were constructed to estimate Incidence Rate Ratios (IRRs) and 95% confidence intervals for colorectal cancer by quintiles of the overall colorectal cancer risk index and sub-domains, adjusted for socio-economic characteristics of the areas. Counts of colorectal cancer cases and population by index quintile, sex, age group, IMD quintile and population density category were compiled to construct the Poisson regression models. An IRR of one denotes no differences in rates, an IRR above one denotes higher incidence and an IRR below one lower incidence.

A binary logistic regression model was constructed to estimate the odds of late stage diagnosis by quintiles of the overall index and sub-domains. A multivariate binary logistic regression model was constructed to estimate the odds of late stage diagnosis, adjusting for patient socio-demographic characteristics (sex, age, ethnicity, IMD and population density). 95% confidence intervals were reported alongside the odds ratios.

6.6 Results

6.6.1 Spatial analysis

Figure 6.1 shows the spatial distribution of the overall colorectal cancer risk index score in England. There is a clear urban-rural divide, with urban areas generally having lower hypothesised colorectal cancer risk than rural areas. This is driven by the relative remoteness of many rural areas from health services. Despite the apparent rural-urban divide, urban areas did not always have the lowest risk. Areas in inner London, central Liverpool and other towns and cities including Bolton, Sheffield and Nottingham were among the top ten per cent of LSOAs with the highest hypothesised colorectal cancer risk. This was largely due to the density of retail outlets in these areas, which were assumed to be bad for health. The areas with the lowest hypothesised colorectal cancer risk appear to be located in suburban areas surrounding cities. This is due to a combination of good access to green space in these areas, being located relatively nearer to health services and relatively further from retail outlets. Figure 6.2, Figure 6.3 and Figure 6.4 provide further insight into these patterns by presenting the spatial distribution of the colorectal cancer risk index domains. Rural areas have lower hypothesised colorectal cancer risk than urban areas based on the retail domain scores due to their location relatively further from retail outlets (which were assumed to be bad for health). On the other hand, urban areas have lower hypothesised colorectal cancer risk based on the health domain scores due to their relative proximity to health services, compared to rural areas. The spatial distribution of the green space domain scores are slightly more dispersed, although there is still a notable rural-urban divide, with urban areas generally having lower hypothesised colorectal cancer risk than rural areas. This is slightly surprising but may relate to the definition of green space used for this health indicator which only includes green space accessible to the public i.e. not private land.

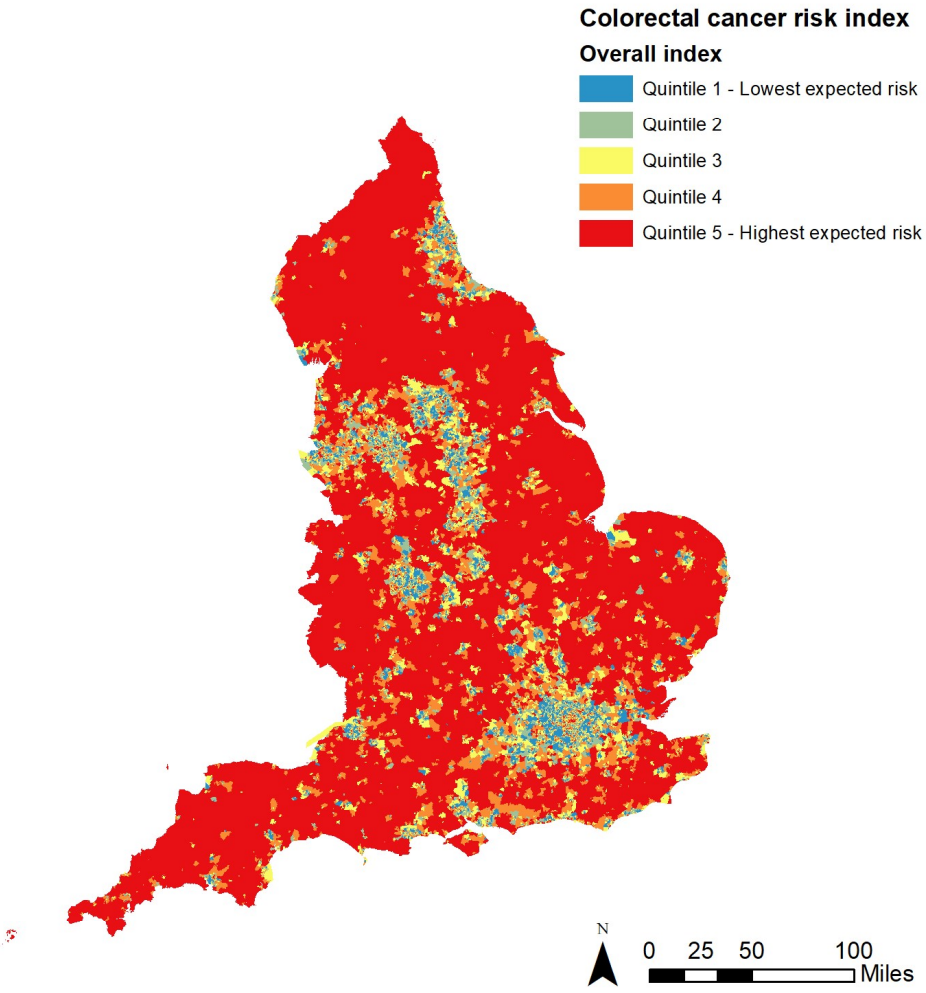


Figure 6.1 Spatial distribution of colorectal cancer risk index

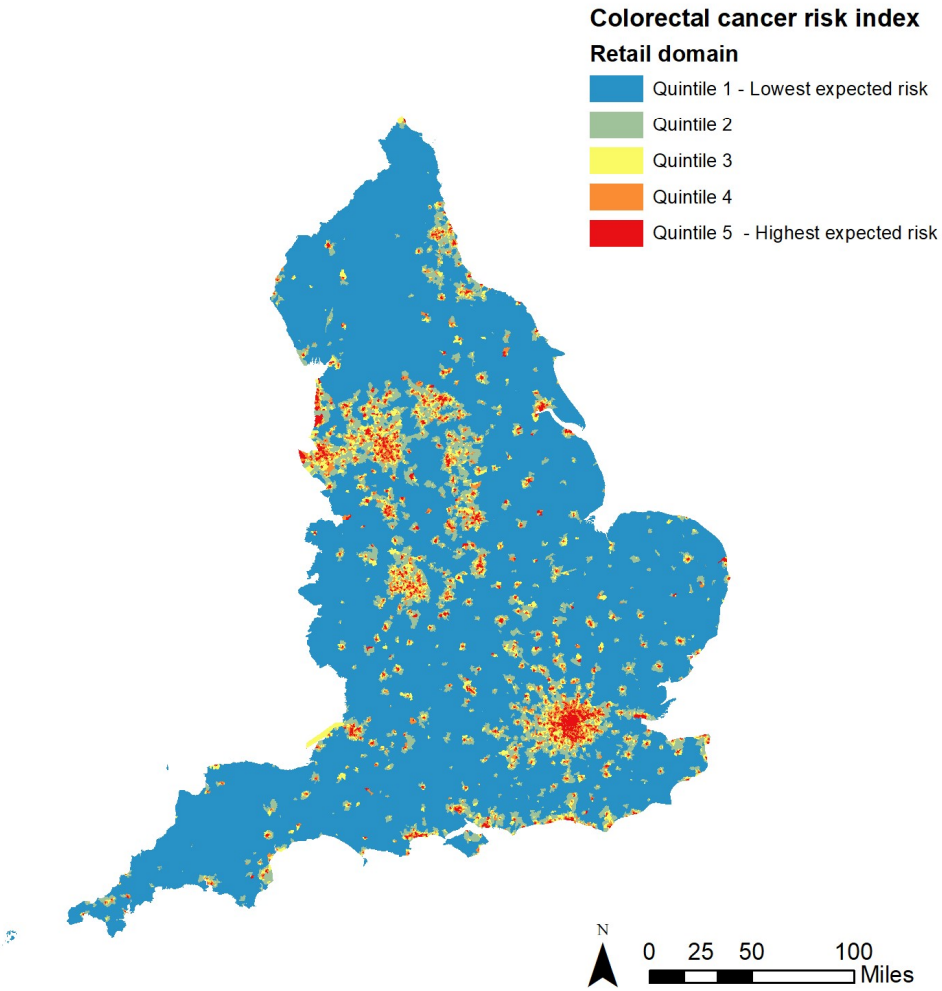


Figure 6.2 Spatial distribution of retail domain of colorectal cancer risk index

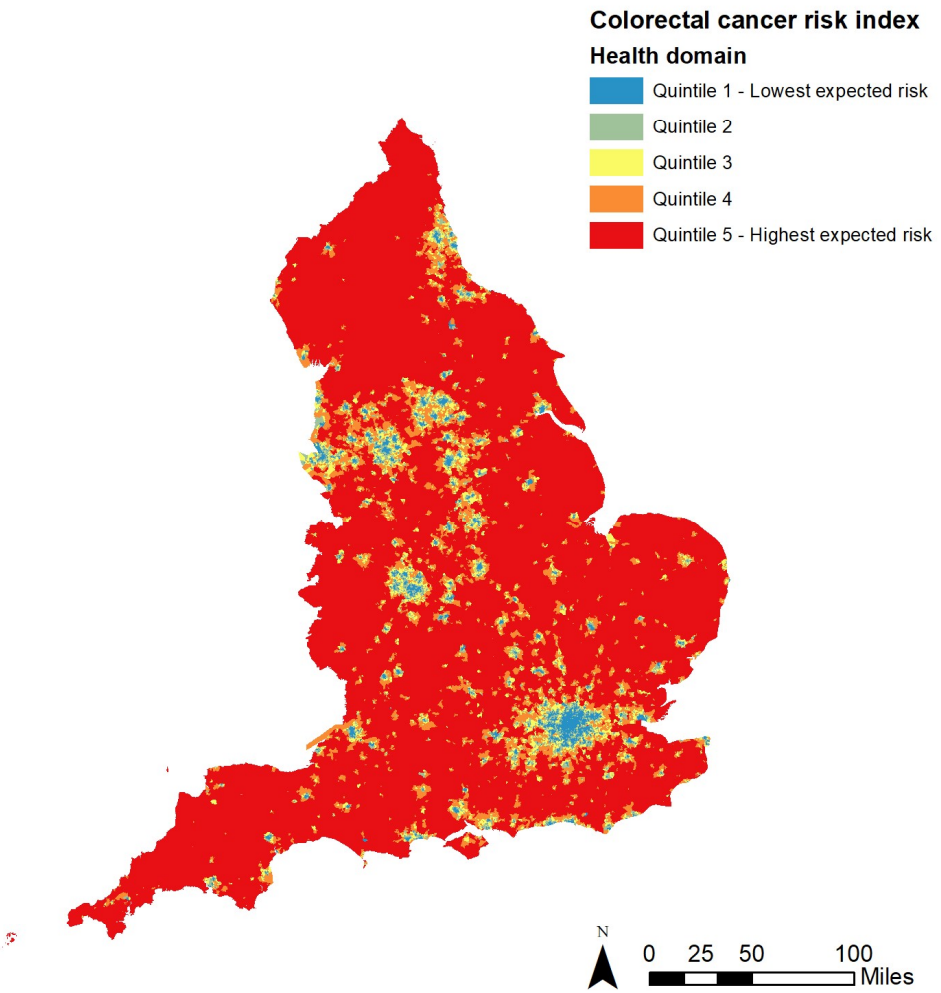


Figure 6.3 Spatial distribution of health domain of colorectal cancer risk index

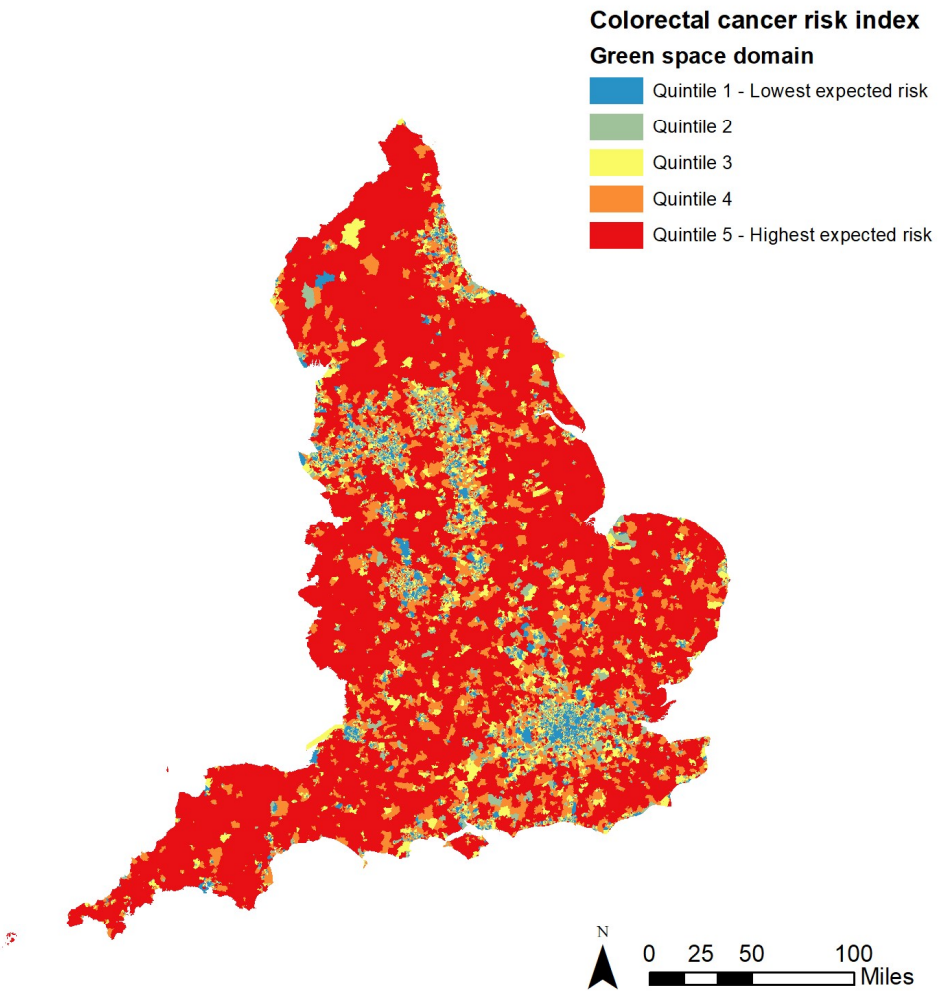


Figure 6.4 Spatial distribution of green space domain of colorectal cancer risk index

6.6.2 Study population

Table 6.3 shows the characteristics of the study population by individual attributes and area-based characteristics of the small-areas in which they live. There were 170,387 patients diagnosed with a primary colorectal tumour between January 2014 and December 2018. Colorectal cancer diagnoses were higher in males (55%) compared to females (45%). The majority of cases were diagnosed in the older age groups (59% of cases were in those aged over 70). Ninety per cent of the study population were of White ethnicity, 2.2% of Asian ethnicity, 1.5% of Black ethnicity, 1% Other ethnicity and 0.3% of Mixed ethnicity. Around 5% of the study population were missing ethnic group information.

A higher proportion of the study population lived in the least deprived areas based on the IMD (income domain), compared to the most deprived areas (22% and 16% respectively). The largest proportion of the study population lived in the most urban areas (44%) and the smallest proportion in the most rural (9%).

Characteristics		n	%
	Total	170,387	
Sex	Male	94,480	55.5
	Female	75,907	44.5
Age group	0-50	10,618	6.2
	50-54	7,701	4.5
	55-59	12,021	7.1
	60-64	17,269	10.1
	65-69	22,039	12.9
	70-74	25,964	15.2
	75-79	25,128	14.7
	80-84	24,543	14.4
	85+	25,104	14.7
Ethnic group	White	153,161	89.9
	Mixed	542	0.3
	Asian	3,682	2.2
	Black	2,475	1.5
	Other	1,783	1.0
	Unknown	8,744	5.1
IMD (income domain)	1 – Least deprived	36,928	21.7
	2	38,664	22.7
	3	35,750	21.0
	4	31,126	18.3
	5 – Most deprived	27,919	16.4
Population density	1 – Most urban	74,658	43.8
	2	14,565	8.5
	3	26,700	15.7
	4	38,837	22.8
	5 – Most rural	15,627	9.2

Table 6.3 Characteristics of the study population

Table 6.4 shows the distribution of the study population by quintiles of the overall colorectal cancer risk index (quintile 1 being areas with the lowest hypothesised risk of colorectal cancer, and quintile 5 being areas with the highest hypothesised risk). Areas with the highest risk have the largest proportion of patients of White ethnicity (92%) and the smallest proportions of the patients of Mixed (0.2%), Asian (0.9%) or Black (0.5%) ethnicity. Conversely, areas with the lowest risk had a relatively smaller proportion of patients of White ethnicity (88%) and the highest percentages of patients of Mixed (0.4%), Asian (3.3%) and Black (2.1%) ethnicity. There was a slightly higher proportion of patients of unknown ethnicity in the higher risk areas.

Areas with highest colorectal cancer risk had the largest proportion of patients in the least deprived IMD quintile (27%), compared to areas with the lowest risk (17%). Conversely, areas with the lowest risk had the largest proportion of patients in the most deprived quintile (25%), compared to 7% in the highest risk areas.

The majority (59%) of patients in areas with the lowest colorectal cancer risk lived in the most urban areas. Conversely, areas with the highest colorectal cancer risk had a higher proportion of patients living in the most rural (35%) and rural (32%) areas.

		Index quintiles				
Characteristics		1 - Lowest risk	2	3	4	5 - Highest risk
Sex	Male	55.2	55.3	54.9	55.8	55.9
	Female	44.8	44.7	45.1	44.2	44.1
Age group	0-50	6.9	6.6	6.8	6.0	5.1
	50-54	5.0	4.7	4.6	4.3	4.1
	55-59	7.4	7.5	7.1	6.8	6.6
	60-64	10.4	10.1	10.1	9.9	10.2
	65-69	12.5	12.5	12.7	13.3	13.6
	70-74	14.5	14.7	15.1	15.3	16.4
	75-79	14.5	14.6	14.4	14.8	15.3
	80-84	14.0	14.3	14.4	14.7	14.5
	85+	15.0	14.9	14.8	14.9	14.3
Ethnic group	White	88.1	88.7	89.8	90.5	91.9
	Mixed	0.4	0.4	0.3	0.3	0.2
	Asian	3.3	2.9	2.3	1.8	0.9
	Black	2.1	1.9	1.6	1.4	0.5
	Other	1.4	1.3	1.1	1.0	0.6
	Unknown	4.8	4.8	4.9	5.1	5.8
IMD (income domain)	1 – Least deprived	16.5	19.6	21.1	23.1	26.6
	2	17.1	18.7	21.1	23.8	30.7
	3	19.2	20.0	22.0	20.9	22.4
	4	22.3	20.7	19.4	17.4	13.0
	5 – Most deprived	25.0	21.0	16.5	14.8	7.2
Population density	1 – Most urban	58.9	53.9	49.8	42.6	19.9
	2	10.9	11.1	9.5	8.6	3.8
	3	18.4	18.3	17.9	16.5	8.7
	4	11.8	16.4	21.8	28.5	32.4
	5 – Most rural	0.1	0.2	1.0	3.7	35.2

Table 6.4 Patient characteristics by colorectal cancer risk index quintile

		All			Male			Female		
Index quintiles		Cases	Pop.	Rate	Cases	Pop.	Rate	Cases	Pop.	Rate
Overall index	1 - Lowest risk	30,927	54,614,560	56.6	17,080	26,777,937	63.8	13,847	27,836,623	49.7
	2	32,198	55,025,393	58.5	17,799	27,107,242	65.7	14,399	27,918,151	51.6
	3	32,942	55,139,608	59.7	18,092	27,208,970	66.5	14,850	27,930,638	53.2
	4	34,743	55,542,623	62.6	19,386	27,495,761	70.5	15,357	28,046,862	54.8
	5 - Highest risk	39,577	55,645,436	71.1	22,123	27,662,487	80.0	17,454	27,982,949	62.4
Retail domain	1 - Lowest risk	42,342	54,779,718	77.3	23,764	26,977,461	88.1	18,578	27,802,257	66.8
	2	35,788	53,335,824	67.1	19,739	26,049,775	75.8	16,049	27,286,049	58.8
	3	34,020	53,985,281	63.0	18,848	26,430,299	71.3	15,172	27,554,982	55.1
	4	31,309	55,408,947	56.5	17,268	27,346,553	63.1	14,041	28,062,394	50.0
	5 - Highest risk	26,928	58,457,850	46.1	14,861	29,448,309	50.5	12,067	29,009,541	41.6
Health domain	1 - Lowest risk	25,152	59,203,499	42.5	13,837	29,813,617	46.4	11,315	29,389,882	38.5
	2	31,519	55,062,467	57.2	17,370	27,161,847	63.9	14,149	27,900,620	50.7
	3	34,379	53,177,982	64.6	18,914	25,992,977	72.8	15,465	27,185,005	56.9
	4	36,294	53,343,765	68.0	20,213	26,105,283	77.4	16,081	27,238,482	59.0
	5 - Highest risk	43,043	55,179,907	78.0	24,146	27,178,673	88.8	18,897	28,001,234	67.5
Green space domain	1 - Lowest risk	28,684	55,974,766	51.2	15,834	27,694,858	57.2	12,850	28,279,908	45.4
	2	31,329	55,862,191	56.1	17,405	27,630,610	63.0	13,924	28,231,581	49.3
	3	33,777	55,266,480	61.1	18,553	27,261,530	68.1	15,224	28,004,950	54.4
	4	35,987	54,445,876	66.1	19,967	26,786,575	74.5	16,020	27,659,301	57.9
	5 - Highest risk	40,610	54,418,307	74.6	22,721	26,878,824	84.5	17,889	27,539,483	65.0

Table 6.5 Crude colorectal cancer incidence rates by risk index quintiles

6.6.3 Colorectal cancer incidence

Table 6.5 shows the estimated crude colorectal cancer incidence rates by quintiles of the risk index. Crude rates were lowest in the lowest risk quintile of the index, and highest in the highest risk quintile. There was a gradient: incidence rates increased as index scores increased (larger values represent areas with higher expected colorectal cancer risk). Incidence rates were much higher in areas with the highest expected colorectal cancer risk (quintile 5). The pattern by index quintile was similar for males and females but incidence rates were higher in men compared to women in all quintiles.

Disaggregating the index into domain scores reveals different patterns by sub-domain. Crude rates were highest in the lowest risk quintile of the retail domain and lowest in the highest risk quintile (i.e. areas with less access to health-damaging retail outlets had higher colorectal cancer rates). This could reflect the interaction between the retail environment and population density and the different age structure of the population in rural and urban areas (and the associated age-related risk of colorectal cancer). Conversely, crude incidence rates were lowest in the lowest risk quintile of the health and green space domains and highest in the highest risk quintile (i.e. areas with greater access to health services and green space had lower colorectal cancer rates). In all of the domains, there was a gradient across the quintiles of index scores. The pattern was similar for males and females in all domains, but rates were higher among men compared to women which reflects higher colorectal cancer incidence observed among men.

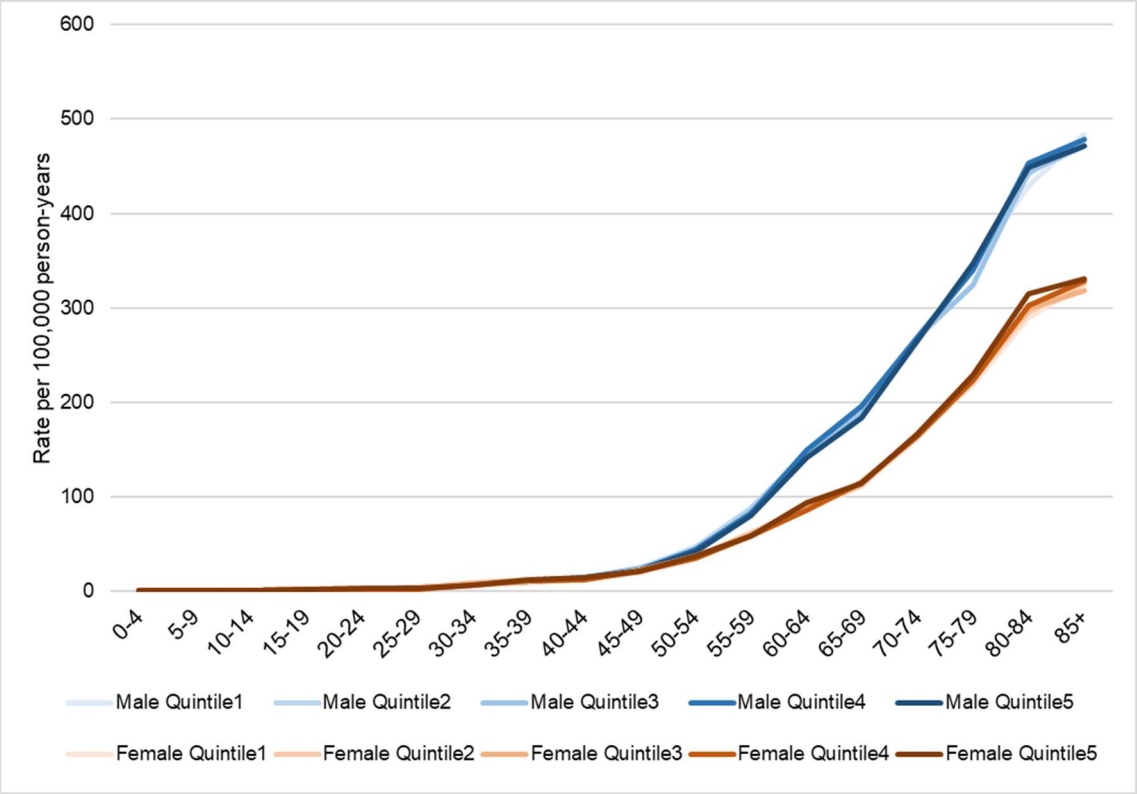


Figure 6.5 Age-specific colorectal cancer incidence rates by quintiles of the overall colorectal cancer index

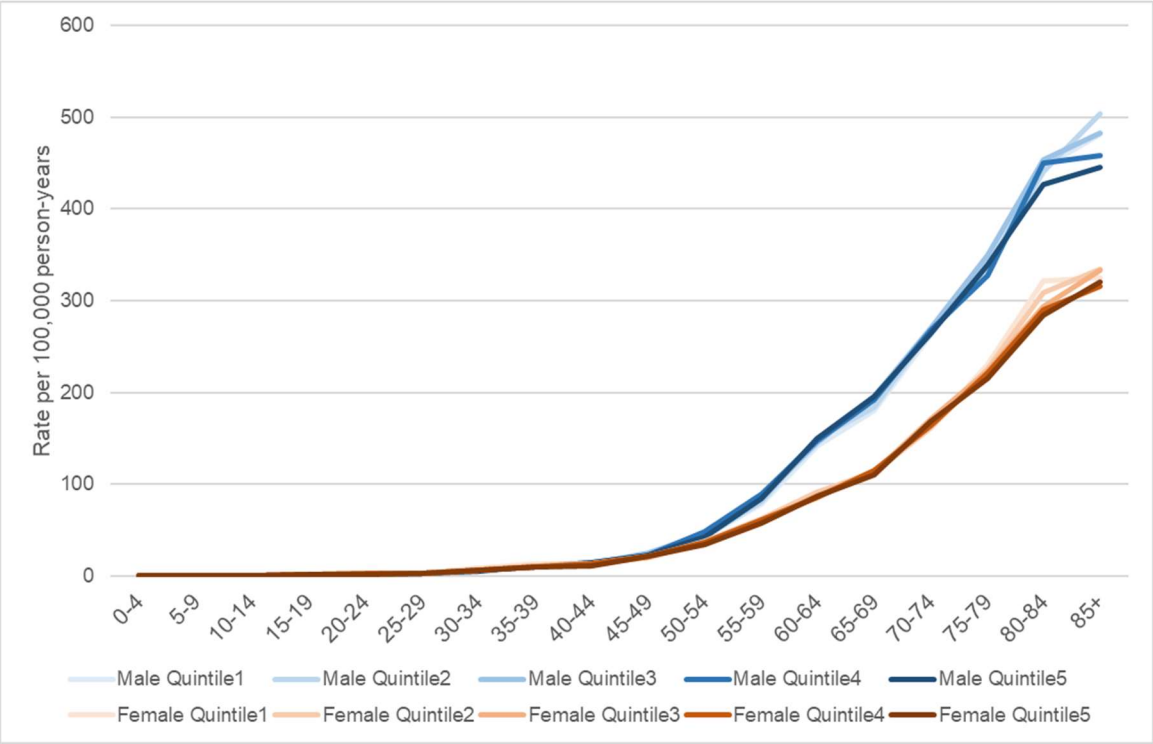


Figure 6.6 Age-specific colorectal cancer incidence rates by quintiles of the colorectal cancer index retail domain

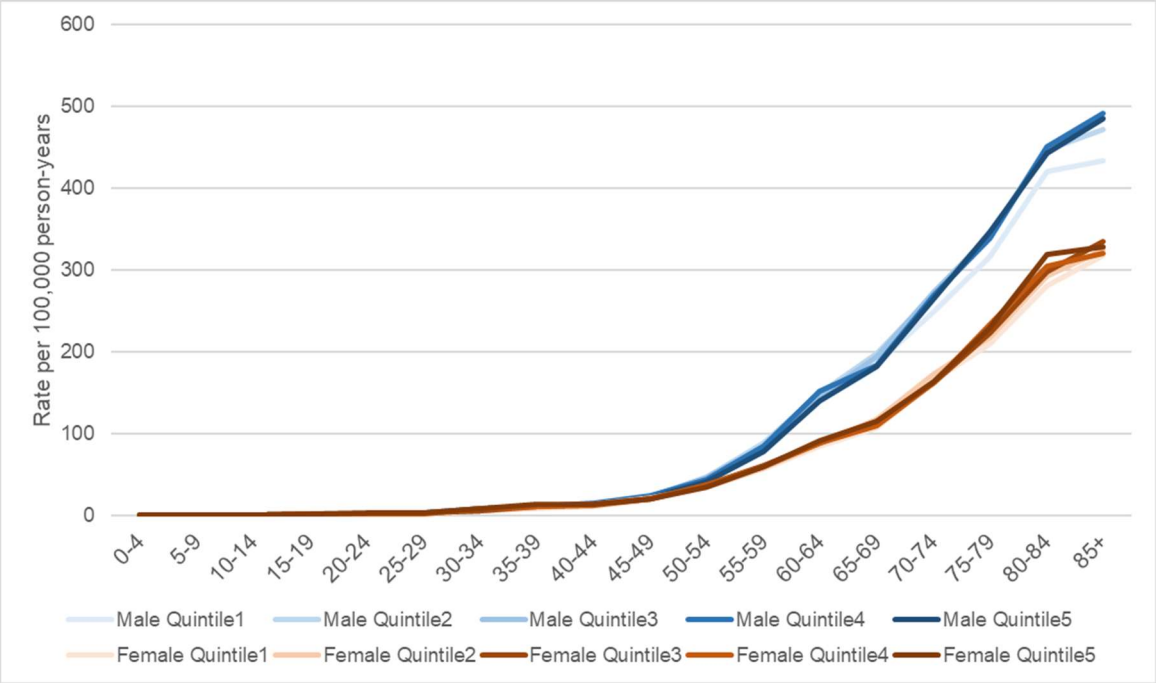


Figure 6.7 Age-specific colorectal cancer incidence rates by quintiles of the colorectal cancer index health domain

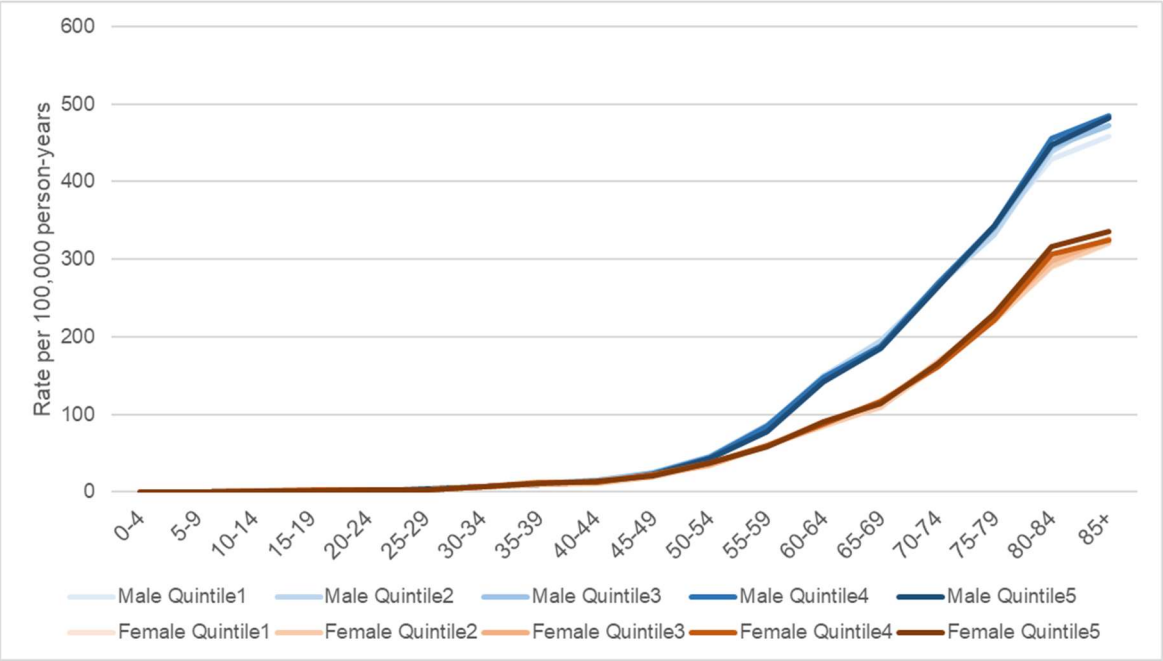


Figure 6.8 Age-specific colorectal cancer incidence rates by quintiles of the colorectal cancer index green space domain

Age-specific colorectal cancer incidence rates were similar in each quintile of the overall risk index scores (Figure 6.5). Disaggregating by index domain scores, there were some differences in the age-specific rates among the older age groups. Age-specific incidence rates among the oldest age-groups were higher in the lowest risk quintile of the retail domain (Figure 6.6) than the highest risk quintile. Conversely, age-specific incidence rates among the oldest age-groups in the health domain (Figure 6.7) and green space domain (Figure 6.8) were higher in the highest risk quintile than lowest risk quintile, particularly among males.

Index quintiles		All		Male		Female	
		Cases	Rate [95% CI]	Cases	Rate [95% CI]	Cases	Rate [95% CI]
Overall index	1 - Lowest risk	30,927	66.5 [65.8-67.2]	17,080	81.9 [80.7-83.2]	13,847	53.9 [53.0-54.8]
	2	32,198	66.9 [66.2-67.6]	17,799	82.1 [80.9-83.3]	14,399	54.3 [53.5-55.2]
	3	32,942	66.6 [65.9-67.3]	18,092	81.1 [79.9-82.3]	14,850	54.7 [53.8-55.6]
	4	34,743	67.1 [66.4-67.8]	19,386	82.4 [81.2-83.5]	15,357	54.3 [53.4-55.1]
	5 - Highest risk	39,577	67.3 [66.7-68.0]	22,123	80.5 [79.5-81.6]	17,454	55.8 [55.0-56.6]
Retail domain	1 - Lowest risk	42,342	67.3 [66.7-67.9]	23,764	80.6 [79.5-81.6]	18,578	55.7 [54.9-56.5]
	2	35,788	67.5 [66.8-68.2]	19,739	82.0 [80.8-83.1]	16,049	55.4 [54.6-56.3]
	3	34,020	67.6 [66.9-68.4]	18,848	83.1 [82.0-84.3]	15,172	54.8 [53.9-55.7]
	4	31,309	66.4 [65.7-67.2]	17,268	81.7 [80.5-82.9]	14,041	53.9 [53.0-54.8]
	5 - Highest risk	26,928	65.6 [64.8-66.4]	14,861	80.7 [79.4-82.0]	12,067	53.1 [52.1-54.0]
Health domain	1 - Lowest risk	25,152	63.9 [63.1-64.7]	13,837	77.9 [76.6-79.2]	11,315	52.2 [51.2-53.1]
	2	31,519	67.5 [66.8-68.3]	17,370	83.2 [82.0-84.4]	14,149	54.7 [53.8-55.6]
	3	34,379	67.7 [67.0-68.4]	18,914	83.0 [81.8-84.1]	15,465	55.2 [54.3-56.0]
	4	36,294	67.4 [66.7-68.1]	20,213	82.5 [81.4-83.7]	16,081	54.8 [54.0-55.7]
	5 - Highest risk	43,043	67.5 [66.8-68.1]	24,146	80.7 [79.7-81.7]	18,897	55.9 [55.1-56.7]
Green space domain	1 - Lowest risk	28,684	66.1 [65.4-66.9]	15,834	81.2 [79.9-82.5]	12,850	53.7 [52.7-54.6]
	2	31,329	66.3 [65.6-67.1]	17,405	81.9 [80.7-83.1]	13,924	53.5 [52.6-54.4]
	3	33,777	67.1 [66.4-67.8]	18,553	81.7 [80.5-82.9]	15,224	55.0 [54.2-55.9]
	4	35,987	67.4 [66.7-68.1]	19,967	82.4 [81.2-83.5]	16,020	54.9 [54.1-55.8]
	5 - Highest risk	40,610	67.4 [66.8-68.1]	22,721	80.7 [79.7-81.8]	17,889	55.8 [55.0-56.6]

Table 6.6 Age-standardised colorectal cancer incidence rates by quintiles of risk index

Age-standardised colorectal cancer incidence rates were slightly higher in areas with the highest hypothesised colorectal cancer risk, compared to areas with the lowest hypothesised risk (Table 6.6), but when split by gender this pattern was only seen for females.

Disaggregating the index into quintiles of domain scores reveals some statistically significant differences in age-standardised incidence rates. Significantly lower colorectal cancer incidence rates were observed among females in the two highest risk quintiles of the retail domain compared the lowest risk quintile. In contrast, there were significantly lower colorectal cancer incidence rates among both males and females in the lowest risk quintile of the health domain compared to the other quintiles (i.e. areas with relatively good access to health services had lower colorectal cancer rates). Colorectal cancer incidence rates among females in the highest risk quintile of the green space domain were significantly higher than those in the lowest risk quintile (i.e. females in areas with poor access to green space had higher incidence of colorectal cancer).

Index quintiles		All IRR [95% CI]	p-value	Male IRR [95% CI]	p-value	Female IRR [95% CI]	p-value
Overall index	1 - Lowest risk	1		1		1	
	2	1.01 [0.99-1.02]	0.37	1.01 [0.98-1.03]	0.60	1.01 [0.99-1.03]	0.43
	3	1.01 [0.99-1.02]	0.50	1.00 [0.98-1.02]	0.82	1.02 [0.99-1.04]	0.20
	4	1.01 [1.00-1.03]	0.14	1.01 [0.99-1.04]	0.17	1.01 [0.98-1.03]	0.49
	5 - Highest risk	1.01 [0.99-1.03]	0.22	1.00 [0.97-1.02]	0.71	**1.03 [1.00-1.06]**	**0.02**
Retail domain	1 - Lowest risk	1		1		1	
	2	1.00 [0.99-1.02]	0.73	1.01 [0.98-1.03]	0.64	1.00 [0.98-1.02]	1.00
	3	1.00 [0.98-1.02]	0.93	1.01 [0.99-1.04]	0.30	0.99 [0.96-1.01]	0.30
	4	**0.98 [0.96-1.00]**	**0.03**	0.99 [0.96-1.01]	0.39	**0.97 [0.94-0.99]**	**0.02**
	5 - Highest risk	**0.96 [0.94-0.98]**	**<0.01**	**0.96 [0.94-0.99]**	**0.01**	**0.95 [0.92-0.98]**	**<0.01**
Health domain	1 - Lowest risk	1		1		1	
	2	**1.07 [1.05-1.09]**	**<0.01**	**1.08 [1.06-1.11]**	**<0.01**	**1.05 [1.03-1.08]**	**<0.01**
	3	**1.08 [1.06-1.10]**	**<0.01**	**1.09 [1.06-1.11]**	**<0.01**	**1.06 [1.04-1.09]**	**<0.01**
	4	**1.08 [1.06-1.10]**	**<0.01**	**1.10 [1.07-1.12]**	**<0.01**	**1.06 [1.03-1.09]**	**<0.01**
	5 - Highest risk	**1.08 [1.06-1.11]**	**<0.01**	**1.09 [1.06-1.12]**	**<0.01**	**1.08 [1.05-1.11]**	**<0.01**
Green space domain	1 - Lowest risk	1		1		1	
	2	1.01 [0.99-1.02]	0.31	1.02 [0.99-1.04]	0.15	1.00 [0.98-1.02]	0.92
	3	**1.02 [1.01-1.04]**	0.01	1.02 [1.00-1.04]	0.08	**1.03 [1.00-1.05]**	**0.02**
	4	**1.03 [1.01-1.05]**	**<0.01**	**1.03 [1.01-1.06]**	**<0.01**	**1.03 [1.00-1.05]**	**0.03**
	5 - Highest risk	**1.03 [1.01-1.05]**	**<0.01**	**1.02 [1.00-1.05]**	**0.06**	**1.04 [1.01-1.07]**	**<0.01**

Table 6.7 Colorectal cancer incidence rate ratios by quintiles of risk index

There were no statistically significant differences in the IRRs by quintile of the overall index scores, when adjusted for sex, age, socio-economic deprivation and rurality of areas (Table 6.7). IRRs were lower in the two highest risk quintiles of the retail domain (small but statistically significant difference), compared to the reference category (areas with lowest expected colorectal cancer risk). Conversely, IRRs were higher in the highest risk quintiles of the health and green space domain scores compared to the reference category (areas with lowest expected colorectal cancer risk). The difference was small but statistically significant. The different directions of the relationship by sub-domain is likely to have been cancelled out in the overall index.

6.6.4 Late stage diagnosis

There was little variation in the proportion of late stage colorectal cancer diagnoses by index and domain scores (Table 6.8). The proportion of late stage cases increased slightly by hypothesised increased risk of colorectal cancer in the retail domain. In contrast, there was a slightly higher proportion of late stage cases in areas with the lowest expected colorectal cancer risk in the health domain. Similarly, the proportion of late stage cases in the green space domain was higher in the lowest risk areas, compared to the highest risk areas.

These patterns could reflect the interaction between the risk index domain scores and area deprivation. Areas with good access to health services (low risk on the health domain) and located relatively near to retail outlets (high risk on the retail domain) are likely to be urban areas with a higher proportion of deprived areas.

		Early stage		Late stage	
Index quintiles		n	%	n	%
Overall index	1 - Lowest risk	15,154	49.0	15,773	51.0
	2	15,735	48.9	16,463	51.1
	3	16,175	49.1	16,767	50.9
	4	17,130	49.3	17,613	50.7
	5 - Highest risk	19,625	49.6	19,952	50.4
Retail domain	1 - Lowest risk	21,179	50.0	21,163	50.0
	2	17,573	49.1	18,215	50.9
	3	16,685	49.0	17,335	51.0
	4	15,289	48.8	16,020	51.2
	5 - Highest risk	13,093	48.6	13,835	51.4
Health domain	1 - Lowest risk	12,155	48.3	12,997	51.7
	2	15,430	49.0	16,089	51.0
	3	16,941	49.3	17,438	50.7
	4	17,758	48.9	18,536	51.1
	5 - Highest risk	21,535	50.0	21,508	50.0
Green space domain	1 - Lowest risk	14,031	48.9	14,653	51.1
	2	15,412	49.2	15,917	50.8
	3	16,621	49.2	17,156	50.8
	4	17,587	48.9	18,400	51.1
	5 - Highest risk	20,168	49.7	20,442	50.3

Table 6.8 Proportion of late stage colorectal cancer diagnoses by index quintiles

Table 6.9 shows the odds of late stage colorectal cancer diagnosis by quintiles of the risk index. In the univariate analysis (Model 1), there were very slightly higher odds of late stage diagnosis in the areas with higher expected risk compared to reference category (areas with the lowest expected risk). This was no longer statistically significant in the adjusted model (Model 2). There were very slightly lower odds of late stage diagnosis in the highest risk quintile of the health domain (i.e. furthest from health services) compared to reference category (i.e. areas located relatively near to health services), but this was not statistically significant in the adjusted model. There were slightly higher odds of late stage diagnosis in the second higher risk quintile of the environment domain in the adjusted model (Model 2), but there was no gradient by quintile of the environment domain. No other statistically significant differences in the odds of late stage diagnosis by quintiles of the overall index or sub-domains were observed in the multivariate analysis.

Index quintiles		Model 1 Odds ratio [95% CI]	p-value	Model 2 Odds ratio [95% CI]	p-value
Overall index	1 - Lowest risk	1		1	
	2	1.01 [0.97-1.04]	0.74	1.01 [0.98-1.04]	0.47
	3	1.00 [0.97-1.03]	0.80	1.00 [0.98-1.04]	0.67
	4	0.99 [0.96-1.02]	0.43	1.01 [0.98-1.04]	0.66
	5 - Highest risk	0.98 [0.95-1.01]	0.12	1.02 [0.98-1.05]	0.36
Retail domain	1 - Lowest risk	1		1	
	2	**1.04 [1.01-1.07]**	**0.01**	1.02 [0.99-1.05]	0.29
	3	**1.04 [1.01-1.07]**	**0.01**	1.01 [0.98-1.05]	0.51
	4	**1.05 [1.02-1.08]**	**<0.01**	1.01 [0.97-1.05]	0.69
	5 - Highest risk	**1.06 [1.03-1.09]**	**<0.01**	1.00 [0.96-1.04]	0.86
Health domain	1 - Lowest risk	1		1	
	2	0.98 [0.94-1.01]	0.14	1.00 [0.97-1.04]	0.87
	3	0.96 [0.93-0.99]	0.02	1.00 [0.97-1.04]	0.78
	4	0.98 [0.95-1.01]	0.14	1.03 [0.99-1.07]	0.10
	5 - Highest risk	**0.93 [0.91-0.96]**	**<0.01**	1.00 [0.96-1.04]	0.89
Green space domain	1 - Lowest risk	1		1	
	2	0.99 [0.96-1.02]	0.50	1.00 [0.97-1.04]	0.84
	3	0.99 [0.96-1.02]	0.47	1.01 [0.98-1.04]	0.49
	4	1.00 [0.97-1.03]	0.91	**1.04 [1.01-1.07]**	**0.02**
	5 - Highest risk	0.97 [0.94-1.00]	0.05	1.03 [0.99-1.06]	0.11

Model 1 unadjusted; Model 2 adjusted for sex, age at diagnosis, ethnic group, area deprivation and population density

Table 6.9 Odds of late stage diagnosis by index quintiles

6.7 Discussion

6.7.1 Summary

There was no consistent association between the overall colorectal cancer risk index and colorectal cancer incidence. When disaggregated by the index domain scores, some small but statistically significant differences were observed. Incidence rate ratios (IRRs) of colorectal cancer were higher in areas with higher hypothesised risk of colorectal cancer on the health and green space domains of the index (i.e. in areas located relatively further from health services and green space). Conversely, the IRRs of colorectal cancer were lower in the highest risk quintile of the retail domain (i.e. areas with better access to retail outlets). This is contrary to what was expected. Reasons for this could be that the accessibility measures used are not adequate proxies for the retail environment and the availability of unhealthy foods may be influenced by other factors such as transportation and retail preferences. It is perhaps not a surprise that a stronger association was found with the domains rather than the overall index as they each measure specific aspects of the local environment which may be masked when they are combined together into a single index. In the analysis of late stage diagnosis, no association was found between the overall risk index or sub-domains and the likelihood of late stage diagnosis. It would be interesting to examine this outcome by population density as previous studies have found an association between rurality and late stage diagnosis which was difficult to infer from these results given the interaction between population density and area deprivation.

Previous studies have investigated the association between individual risk factors and colorectal cancer incidence. This study builds on previous research by examining the potential influence of features of the local environment on colorectal cancer incidence. Overall, the results are consistent with limited previous research investigating the relationship between the food environment and colorectal cancer incidence. In a prospective cohort study of 80,000 participants living in California, Canchola et al. (2017) did not find an association between the food environment within a 1-mile network buffer of an individual’s residence and being diagnosed with colorectal cancer. The study investigated a range of residential neighbourhood obesogenic attributes

including a Retail Food Environment Index, defined as the ratio of the number of convenience stores, liquor stores and fast food restaurants to supermarkets and farmers markets. An ecological study of census tracts in Texas did not demonstrate an association between unhealthy food availability (defined as the number of limited-service restaurants divided by the total population in each census tract) and colorectal cancer incidence (Gibson et al., 2020). The authors concluded that efforts to limit the availability of unhealthy food stores may not be beneficial for reducing colorectal cancer incidence. These studies used different geographic metrics to define the food environment (buffer distances versus administrative boundaries), the choice of which may impact the associations observed (Wilkins et al., 2019).

There is a relationship between the colorectal cancer risk index and rurality. The relationship differs by domain of the index. For the retail domain, the lowest risk was in the rural areas and highest risk in urban areas, whereas for the health domain the relationship was reversed (highest risk in rural areas and lowest risk in urban areas). The relationship between the green space domain and colorectal cancer risk is less pronounced, but the highest risk tended to be in rural areas and the lowest risk in urban areas. By adjusting for population density in the statistical models it was possible to investigate whether the risk index was related to colorectal cancer incidence independently of population density. However, some of the indicators in the risk index may be on one of the numerous pathways by which rurality influences health. For example, rurality may be detrimental for health due to relative lack of access to health services in rural areas and conversely urbanity may be detrimental for health due to the relative increased opportunities for unhealthy food choices. Therefore, a more appropriate approach may have been to adjust for area deprivation only (and not rurality) in the statistical models to examine whether differences remain by risk index score.

6.7.2 Strengths

A strength of this study is that it uses a population-based cancer registry to capture cases of colorectal cancer for a large population. Linkage of the cancer registrations to small areas (LSOAs) in England enabled the relationship between a range of exposures and colorectal cancer incidence to be examined. Developing an index at small-area level using national level data could provide

a useful policy tool to identify areas to target interventions and to help identify pathways that inventions could address.

Previous research has focused on the association between individual risk factors and cancer incidence, however, such studies do not take into account the reality that individuals may interact with numerous features of the environment simultaneously (Pearce et al., 2010). Furthermore, if particular features of the local environment occur together, it may have a greater impact on health outcomes than individual characteristics. By incorporating a range of health-related features of the local environment into the colorectal cancer risk index, areas that display numerous characteristics relevant to risk of colorectal cancer can be identified. This study builds on previous research investigating geographical determinants of health in the context of colorectal cancer.

6.7.3 Limitations

The study has several weaknesses. Firstly, results from an ecological study cannot be used to infer individual risk of developing colorectal cancer (the ecological fallacy). Secondly, the definitions of accessibility used in this study may not represent an individual's behaviour patterns. For example, they may not use their nearest retail outlet or health service and may travel further (by other means of transportation) to use services or access green space.

There is a temporal mismatch between the health indicators and cancer incidence statistics. Ideally, colorectal cancer cases should have been taken in the period after the health indicators were collected, however cancer registry data was only available up to 2018 which would have resulted in too few cases to produce robust results. The small-area level data on health indicators was not available for earlier periods. A further consideration is that there is likely to be a considerable lag time between cancer development and diagnosis. Cancer risk is likely to be influenced by longer-term environmental exposures, and therefore may be less affected by current characteristics of the environment alone. Life course measures of health-related features of the environment are needed but may be difficult to construct due to limitations of data accessibility. It is also recognised that there will be other health-related features of the environment that could potentially influence colorectal cancer risk that have not

been included in the index. The construction of the index allows for additional variables to be included in the future.

Combining the indicators based on equal weights does not take into account that some factors may have a greater influence on colorectal cancer risk than others. By weighting the indicators equally, the contribution of some indicators to the overall risk index score may be greater than their relative influence on colorectal cancer risk.

It is also recognised that categorising continuous variables, as was done in this study by grouping the index score into quintiles, may result in a loss of information, statistical power and precision. Furthermore the choice of cut points is likely to influence the results (Busch, 2021).

6.7.4 Further work

Further work related to the food environment and colorectal cancer risk could investigate expenditure patterns to get a clearer picture of where people are buying certain goods. Food, drink and tobacco expenditure estimates are available at Local Authority District level for Great Britain (James et al., 2019). This could be extended further by linking data from supermarket loyalty card transactions on food and drink. The potential of data sources such as these is considered further in the Chapter 7.

Future research investigating the environment and cancer outcomes could further develop and validate measures of the local environment and examine associations using different neighbourhood boundaries, buffer distances and accessibility measures. Further consideration is also needed as to how best to combine the separate indicators into a single index. Applying different weights to the indictors based on estimates of their contribution to colorectal cancer risk factors (for example, using population attributable fractions as calculated by Brown et al. (2018) is a logical next step. .

6.8 Conclusion

Key messages:

- No association was observed between the overall colorectal cancer risk index and incidence of colorectal cancer or likelihood of late stage diagnosis.
- Incidence of colorectal cancer was higher in areas hypothesised as having higher risk of colorectal cancer, according to indicators in the health and green space domains of the index, but incidence was lower in areas hypothesised as having higher risk according to the retail domain.
- Future research should develop and validate geographic measures of the local environment and life course measures of environmental exposures.

Chapter 7
Discussion

7.1 Overview

The previous chapters have explored spatial and social variations in colorectal cancer using a variety of datasets. Each analysis chapter has discussed the results in the context of previous research. This final chapter aims to highlight the key findings of the thesis and the potential impact of these on the broad research area and public health policy. The strengths, limitations and challenges of the research are considered.

7.2 Key findings

The findings of this thesis have shown that:

- Comprehensive, nationwide mortality data for England and Wales showed there is geographical variation in colorectal cancer mortality rates at Local Authority level, although no clear spatial pattern was observed. This is in contrast to all-cancer mortality, for which there are higher rates in Local Authorities in the north of England and south Wales and lower rates in south and east England.
- All-cancer mortality is strongly related to deprivation at Local Authority level, although the pattern of colorectal cancer mortality by deprivation is less clear. Similarly, there does not appear to be a gradient in colorectal cancer mortality rates by rurality. These could explain the lack of geographic variation in colorectal cancer mortality rates.
- Linked census and cancer registration data for a large, representative sample of the population of England and Wales showed that individual indicators of socio-economic position were associated with colorectal cancer diagnosis. Risk of colorectal cancer diagnoses was lower among those with a degree, compared to those with no degree and conversely, the risk of diagnosis was higher among those living in social rented housing compared to those living in owner-occupied housing. No associations with colorectal cancer diagnoses were found for the area-based measure of deprivation.

- In the same dataset, individual indicators of socio-economic position (educational attainment, social class and housing tenure) were strongly associated with colorectal cancer survival time. The survival time for all-causes of death was longer among those with a degree, compared to no degree. Conversely, the survival time was shorter among those employed in manual occupations, compared to non-manual occupations and among those in social rented housing, compared to owner-occupiers. Individual indicators of socio-economic position were also associated with survival time for death from colorectal cancer. There was a negative association between area deprivation and survival time for both all-cause and cause-specific death.
- Using a bespoke risk index applied to population-based cancer registry data at small-area (LSOA) level, higher incidence of colorectal cancer was observed in areas located relatively further from health services and in areas with less access to green space, however colorectal cancer incidence was lower in areas with greater access to (assumed health-damaging) retail outlets.
- No associations were found between health-related features of the local environment and stage at diagnosis, according to the risk index.

7.3 Impact

7.3.1 Importance of research area

Scale of public health problem

Understanding spatial and social variations in colorectal cancer is important as, despite improvements in early diagnosis and treatment, the disease remains a major public health concern in the UK. Over 40,000 people are diagnosed with colorectal cancer in the UK each year and there are over 16,000 deaths. Furthermore, it is estimated that over 50% of colorectal cancer cases are preventable, so identifying and targeting interventions to those most 'at risk' could have a considerable impact on morbidity and mortality.

As well as the impact colorectal cancer has on the health of the population, it also places a huge economic burden on health systems and patients. The economic cost of colorectal cancer in the UK in 2015 was estimated to be £1.7 billion (Henderson et al., 2021). This figure includes direct expenditure on colorectal cancer health care, as well as the indirect economic impact which includes the cost of temporary or permanent absence from work due to disability, premature death and unpaid informal care. Understanding where initiatives to improve access to screening, early diagnosis and treatment would have the greatest impact could reduce the economic burden of the disease on individuals and the NHS.

Links to previous research

The model of health determinants in Figure 7.1 was introduced in Chapter 1 and used as a framework to review existing literature and identify research gaps in Chapter 2. This thesis has contributed to research investigating the broader determinants of health in the context of colorectal cancer.

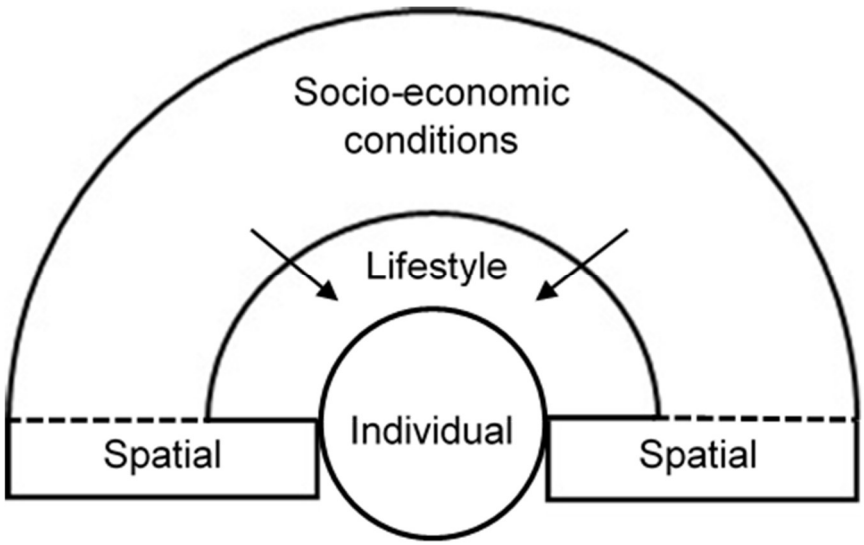


Figure 7.1 Thesis framework

Previous research examining variation in colorectal cancer by geographic area in the UK has been at regional level, based on administrative regions or (old) health geographies. Insight at a more local level is needed as there are considerable demographic and socio-economic differences between local areas within regions which may be masked by regional patterns. This research has tried to address this gap by analysing variation in colorectal cancer mortality at Local Authority (LA) level (Chapter 4). While it is recognised (below) that LAs still contain heterogeneous populations and a range of urban and rural settings, they provide much more detail than regions and are more relevant for Public Health, which operates at LA level. The findings concur with regional level analysis in England which found less geographic and socio-economic variation in colorectal cancer incidence and mortality compared to all-cancer mortality (Arik et al., 2021). Differences by gender in colorectal cancer mortality have also been previously reported. Higher colorectal cancer mortality was found among males, but not females, in more deprived areas, based on the IMD (National Cancer Intelligence Network, 2014). While there was little variation in colorectal cancer mortality rates for females, the results in Chapter 4 showed lower colorectal cancer mortality rates among males in the least deprivation quintile.

Few studies have investigated the association between individual indicators of socio-economic status and colorectal cancer outcomes as these data are not routinely collected by cancer registries. Using a large, representative cohort of

from the ONS Longitudinal Study it was possible to examine colorectal cancer diagnoses and survival by individual census variables (Chapter 5). Previous research has reported inconsistent findings regarding the association between socio-economic status and colorectal cancer incidence (Brown et al., 1998). The results in Chapter 5 showed an association between both educational attainment and housing tenure and colorectal cancer diagnoses. Recent research has reported an emerging socio-economic gradient in colorectal cancer incidence, measured at the area-level (Tweed et al., 2018), but an association was not found between the area-based measure of deprivation and colorectal cancer incidence in the LS using time-to-event-analysis. The findings of the analysis of colorectal cancer survival (Chapter 5) confirmed those of a previous study using the LS for an earlier time period which found poorer colorectal cancer survival among individuals with lower socio-economic status, by indicators of housing tenure and car access (Sloggett et al., 2007), suggesting this issue has not been addressed by policy. It is useful to examine associations by different indicators of socio-economic status as they may reveal patterns that are not apparent when measured by ecological measures. Furthermore, there may be different associations with some socio-economic indicators for different cancers which could aid understanding of how particular aspects of socio-economic status impact on cancer outcomes.

The impact of the local environment on health has been investigated in relation to a range of health outcomes, such as obesity, alcohol-related disease and smoking (Shortt et al., 2016; Burgoine et al., 2018; Sherk et al., 2018). There is less research on the influence of local environment on cancer risk and outcomes (Canchola et al., 2017; Shvetsov et al., 2020; Gibson et al., 2020). Chapter 6 investigated the influence of accessibility to health-related features of the environment, measured at small area level, on colorectal cancer risk. While areas relatively further from health services and with less access to green space had small but significantly higher incidence of colorectal cancer, areas relatively closer to retail outlets (presumed to have negative impact on health) have a small but significantly decreased incidence of colorectal cancer. This unexpected direction in the relationship with the retail environment may be influenced by the level of deprivation and rurality in areas located closer to retail outlets, which warrants further investigation. A number of modifiable risk factors

are associated with colorectal cancer and it is recognised that health behaviours are influenced by broader environmental, societal and economic conditions, therefore this area of research is important for understanding the pathway between socio-economic conditions, lifestyle and cancer outcomes.

7.3.2 Policy and broader relevance

Reducing social and spatial inequalities in cancer outcomes has been a key policy aim under successive governments. Despite this, inequalities have persisted, both within the UK and internationally. Cancer outcomes in the UK have been consistently worse than comparable countries (Coleman et al., 2011). Therefore continued monitoring of colorectal cancer outcomes by geographic area and socio-economic group is vital. While no clear deprivation gradient was found in Chapter 4 for colorectal cancer mortality at area-level, both incidence and survival of colorectal cancer were associated with individual indicators of socio-economic status in Chapter 5.

There is also increasing recognition among policy makers of the wider determinants of health. Examining the influence of the retail environment and physical environment on cancer risk could inform policies with regard to access to healthy food and green space. Linkage of small-area level data about the local environment to cancer registry data (as in Chapter 6) provides a means to examine the influence of such factors on cancer outcomes. It would be interesting to examine these relationship for other cancer types, for which there are known modifiable risk factors.

Understanding colorectal cancer risk at small-area level could inform policies which prioritise and target interventions to higher risk places and, in doing so, hopefully reach more susceptible individuals. This should include screening and education programmes. This would also help to reduce the economic burden of the disease as interventions could be rolled out in the areas where they would have the biggest impact.

This thesis has demonstrated the value of using linked population-based data sets for health research. These types of data are being used in health research internationally, for example the International Cancer Benchmarking Partnership (ICBP) was set up in 2009 to enable comparative research on cancer survival, incidence and mortality trends across high income countries, using a range of

data sources. The need for (near) real-time health data has come to the fore during the COVID-19 pandemic, during which linked population-based data sets have provided insights about the disease and its impact.

During my PhD I have contributed more broadly to advancing population level linked health data for research, building on my previous data science skills (for example in (Sturley et al., 2018)). I went on a research visit to the University of Canterbury, Christchurch, New Zealand which led to me co-authoring a paper reviewing the utility of these types of data, with a particular focus on case studies in the UK and New Zealand (Oldroyd et al., 2021). I undertook a 3-month secondment with Leeds Teaching Hospital NHS Trust (LTHT) at the height of the COVID-19 pandemic to support their Research Data and Informatics team with data access requests for COVID-19 research projects. This included preparing data for the DECOVID project, coordinated by the Alan Turing Institute, which is compiling a research database from routinely collected hospital data, to provide rapid insights into COVID-19.

7.4 Challenges and Learnings

7.4.1 Confidentiality vs. research

Confidentiality of patient data is of utmost importance, but restrictions placed on the data can limit its usability for research. Openly published data have limitations for research due to processes such as suppression and rounding of small numbers. These are necessary to protect the confidentiality of individuals but can cause problems when, for example, calculating mortality rates for diseases in which there are small numbers in particular age groups, such as colorectal cancer, or when split by geographic area. This was the case in Chapter 4, whereby the smallest geographic level at which the data was published was LA and even at this geographic unit the data contained suppression counts. In the UK, confidential data is protected by the Five Safes framework (UK Data Service, 2021).

7.4.2 Data access

Data access issues have been a major barrier to this research. While it is important that rigorous application processes are in place, the administration can be extremely time consuming and slow to process and there have been numerous delays.

Process

The three data sets used in this thesis each had different data providers and different processes for approval to access the data. In order to access individual mortality records held by the Office for National Statistics (ONS), an application to the UK Data Service (UKDS) was required (see Appendix B). The individual-level data is categorised by the UKDS as controlled data, meaning it is potentially disclosive and must be accessed in a secure environment. The application process involved completion of a Digital Economy Act (DEA) Research Project Application form and a Secure Access User Agreement which required a signature from a University solicitor. A further step to the application process was added after the application was originally submitted, requiring ethics approval from the applicant's institution and completion of an ethics self-assessment. In order to obtain access to data from the Longitudinal Study (LS) an application was made to the ONS via the Centre for Longitudinal Study

Information and User Support (CeLSIUS) (see Appendix C). The application was reviewed by the ONS Research Accreditation Panel (RAP). A further application to use restricted data was required to append Townsend quantiles to LS members at small-area level. This is required for any geographic classification with more detail than local authority. An application to use data held in CORECT-R was made to the UK Colorectal Cancer Intelligence Hub (see Appendix D). This comprised a project protocol and data specification which was reviewed by the Hub Access Committee for approval. The committee included input from the Bowel Cancer Intelligence UK Patient-Public Group. A data access agreement was also required and needed to be signed by both the University of Leeds and the University of Oxford. These application processes involved many hours of input by researcher, supervisors and University administrative personnel.

Accreditation and training

A constraint on access to ONS data is that all members of the research team must hold ONS Accredited Researcher status. This applied to both the mortality data and the LS data. Individuals applying for Accredited Researcher status must also complete Safe User of Research data Environments (SURE) training and pass an assessment. In order to gain access to the CORECT-R data, researchers must complete an Information Governance training course.

Remote access

Remote access to research data sets can also be a challenge, which was particularly pertinent during the Covid-19 pandemic (see Covid-19 Impact Statement). Access to the mortality data is via the UKDS Secure Lab, a secure remote service accessible from an approved PC at the University of Leeds. The LS is accessed via the Secure Research Service (SRS) safe setting room at the ONS offices in London. An Assured Organisational Connectivity (AOC) agreement was obtained in January 2021, allowing remote access to the SRS via a safe room at the University of Leeds. Data held within CORECT-R is accessed remotely via a secure TRE and a request must be made for results to be checked and outputted. There was a change to the data hosting service provider for CORECT-R during the period of study which impacted data access and caused delays.

Reflections

It is important to have these processes in place to ensure appropriate use of the data and to protect the confidentiality of individuals whose data is being used. However, the data access process can be extremely slow which can impede progress, particularly when projects have set timescales such is the case with PhD research funding. There were significant delays in obtaining approval to access the mortality data, both from UKDS and in obtaining a University of Leeds signature on the Secure Access User Agreement, which were exacerbated by the Coivd-19 pandemic, as organisations and individuals adjusted to new ways of working. Access to the CORECT-R data was delayed while a Data Sharing Agreement was signed by the Office for Data Release (ODR) within Public Health England (PHE).

The use of three different data sets in this thesis resulted in three times the challenges of data access. A more streamlined process to enable faster access to research datasets for trusted researchers, whilst respecting patient confidentiality, is required. This would facilitate applied research for which multiple data sets are required and improve the timeliness of research outputs.

7.4.3 Geographic scale

Identifying the appropriate scale of geography which would provide adequate detail for the analysis whilst also ensuring the reliability of the results was an important consideration in this research. Despite colorectal cancer being a relatively common disease (~40,000 cases a year in the UK), small number problems were still encountered when attempting to analyse trends by geographic area and demographic group.

In the analysis of mortality trends (Chapter 4), even at Local Authority level there were some small numbers of colorectal cancer deaths when split by age group and geographic area. A solution to overcome this was to pool the mortality data over multiple years. A compromise of conducting the analysis at LA level is the lack of geographic detail. This is a problem because there can be considerable variation in deprivation and rurality within large areas such as LAs. An alternative approach could be to undertake the analysis for smaller areas (such as MSOAs) and calculate indirectly standardised mortality ratios. The LS data is only available for broad geographies due to small numbers of sample

members when split by geographic area and demographic variables. However, area types can be appended at small-area level by ONS and the identifiers removed before the data set is returned. Small-area geographies (LSOAs) are available on the CORECT-R data, however, small number issues are still relevant when analysing these data by geographic area.

A further consideration is the most appropriate geographic units for disseminating this type of research. Public Health initiatives are usually area-based. Local authorities have responsibility for commissioning behavioural and lifestyle campaigns to prevent cancer and are also responsible for addressing local health inequalities. Outcomes for the LA as a whole as well as small-areas with the LA will be of interest for targeting interventions. On the other hand, NHS cancer services are commissioned by Clinical Commissioning Groups (CCGs). Further work could be to look at some of these geographic trends by specific health geographies though a problem will be that health geography areal definitions invariably change more than once per decade making time-series and linkage of individual records challenging.

7.4.4 Individual vs. ecological

This thesis has investigated trends in colorectal cancer risk and outcomes by both individual and area-based indicators of deprivation. Few data sets contain linked individual-level indicators of socio-economic status and health outcome, therefore area-based measures of deprivation are often used as a proxy.

When conducting area-based analysis (as in Chapters 4 and 6) consideration should be given to the ecological fallacy when assessing the generalisability of results. Relationships observed at the area-level or for groups may not be true for individuals within that area or group. A further consideration is the Modifiable Areal Unit Problem (MAUP), whereby the definition of spatial units to which data are aggregated will influence the results of analysis. This is particularly applicable to analysis of trends over time, during which boundaries may have changed. For the analysis in Chapter 4, definitions of geographically consistent boundaries over time were used. Using smaller geographic units (such as LSOAs in Chapter 7) may decrease the effect of MAUP and the ecological fallacy, but they are still relevant even for the smallest spatial units.

Related to this is the uncertain geographic context problem, whereby results about the effects of area-based attributes on individual behaviours or outcomes could be affected by how areas are characterised (Kwan, 2012). Consideration needs to be given to the geographic and temporal configuration of the physical and social factors that influence health outcomes. For example, in the analysis in Chapter 6, there will be areas beyond their local residential neighbourhood that will exert influence on individuals' behaviour, such as their commute, place of work and leisure. The opportunities for physical activity and quality of food may be different near the work place than near home. Different factors operate at different scales and the scale is likely to differ by demographic group, which makes it difficult to define relevant geographic and temporal contexts. A more recent approach has been to define activity spaces to capture contextual exposures (Perchoux et al., 2013).

An additional challenge when dealing with uncertainty in geographical research, is that researchers are often limited in what they can observe by the data and methods available (Franklin, 2022). For example, it may not be possible to link individuals to areas deemed the most relevant spatial unit for the process under investigation due to the data specification, availability and confidentiality constraints. This also applies to ecological analyses, in which ideally analysis might be carried out by smaller spatial units, but this is not possible due to data access challenges and small number issues, which require handling using appropriate methods.

7.4.5 Generalisability of results

The mortality data used in Chapter 4 contained comprehensive death registration records for England and Wales and there is very little missing data. Similarly, CORECT-R contains cancer-registry data for the whole of England and plans are in place to expand this to the rest of the UK. The LS is a large, representative sample of the population of England and Wales, which improves the generalisability of the results generated using these data to the wider population.

The findings of this research are potentially generalisable to other similar countries, although caution should be taken when comparing the results to those in different healthcare settings and geographic contexts, such as

countries in which the definition of urban and rural will include some very remote populations. Furthermore, in England and Wales the NHS is free to use so there may be less disparities in health than in similarly developed countries where health care is not free and requires insurance.

7.4.6 Timeliness

The timeliness of the data available was a limitation to the studies in this thesis. While the most recent (up to 2020) mortality data is openly published at aggregate level, due to delays in accessing the individual-level data described above, it has currently only been possible to analyse the mortality trends up to 2012. Cancer registrations linked to the LS are only available up to December 2016, however, there was a long-follow up period for time-to-event analysis. Cancer registry data up to the end of 2018 is held within CORECT-R, although there are plans to update this. Due to these challenges the time period covered in the analyses is not as up-to-date as hoped.

7.5 Further work

Mortality data

When access to the up-to-date mortality data is approved, colorectal cancer mortality trends for the most recent time period could be examined. It would be interesting to see if colorectal cancer mortality trends reflect those observed for all-cause mortality in recent years, whereby there has been a stalling of improvements in mortality rates in the UK since 2011 (Hiam et al., 2018) and an increase in mortality rates in some demographic groups (Marmot et al., 2020). The impact of Covid-19 on colorectal cancer mortality rates, and variation by area deprivation, could be examined with the most recent years of data. This could reveal where delays to diagnosis and treatment have had the biggest impact on colorectal cancer outcomes. Evidence from population-based NHS data sets has demonstrated that the diagnostic pathway for colorectal cancer was severely disrupted during the COVID-19 pandemic in England, with reductions in urgent 2-week referrals and the use of colonoscopy (Morris et al., 2021). It was estimated that over 3,500 fewer people have been diagnosed and treated for colorectal cancer in England between April and October 2020 than would have been expected.

A further line of enquiry using the mortality data would be to examine trends of colorectal cancer mortality rates by trajectories of deprivation and population density over time. This approach overcomes the issues of using cross-sectional measures of area types for a time-series of outcome data. It was planned to carry out this work, and the methodology has been tested, but it is awaiting more recent and more local-level data.

ONS LS

An advantage of the LS is the long time period over which the cohort has been followed. While this was utilised in Chapter 5 to investigate colorectal cancer outcomes over a long follow-up period, a further step could be to stratify the LS study sample by change in individual socio-economic status or area deprivation over time. Change in area deprivation could be defined both in terms of migration (moving between areas with different levels of deprivation) and changing deprivation of areas over time.

If and when more recent cancer registration data are linked to the LS, it would be interesting to examine socio-economic variation in colorectal cancer diagnoses and outcomes for different cohorts, pre and post the introduction of the Bowel Cancer Screening programme.

CORECT-R

The CORECT-R resource offers huge potential for further research into colorectal cancer using linked data sets. Small-area data linkage to CORECT-R would enable further area-level attributes to be investigated (such as expenditure patterns discussed below). The bespoke index in Chapter 6 could be further developed to encompass such datasets. There are also plans to link cancer screening data to CORECT-R.

The Rapid Registration Data produced by the National Cancer Registration Service (2022) will allow analysis of more recent colorectal cancer data. This data set provides more up-to-date cancer data than the standard cancer registry data as it does not go through the same rigorous data quality and completeness checks that registry data usually would (which is why there is a considerable time lag before it is released).

The Hospital Episode Statistics (HES) and National Radiotherapy data held within CORECT-R could be used to examine outcomes for radiotherapy patients by geographic location. There is evidence from previous research in a single region in England that rectal cancer patients were less likely to receive radiotherapy with increased travel time to the nearest radiotherapy centre so it would be interesting to investigate this using a more recent and nation-wide data set.

Other data sources

Other data sources offer potential avenues for research which would complement the work in this thesis.

Data from the Bowel Cancer Screening Programme (BCSP) could be used to examine the impact of the programme on colorectal cancer incidence and mortality rates. The geographical roll-out of the screening programme was complex, but it would be interesting to see if the timing of the roll-out is reflected in the incidence and mortality trends. Data containing uptake rates of bowel cancer screening at area-level is available via the PHE Fingertips tool (Office for

Health Improvement and Disparities, 2022). It may also be possible to link data from the BCSP to the LS to examine uptake my socio-demographic variables.

Consumer data

Novel data sources, such as consumer data, could also be utilised in epidemiological studies. Supermarket transaction data for loyalty card holders could be used to examine spending patterns on food and drink. These type of data have been used previously to identify and profile dietary patterns of the UK population (Sturley et al., 2018; Clark et al., 2021). These type of data (or profiles derived from them) could potentially be linked to data held with CORECT to examine colorectal cancer incidence by dietary patterns. The Consumer Data Research Centre (CDRC) holds retail and consumer data sets which researchers can apply to use.

7.6 Summary

Reducing inequalities in colorectal cancer is a public health priority. This thesis has used three different datasets with national coverage to investigate spatial and social variations in colorectal cancer in England and Wales. The data sets have complemented one another (
Figure 7.2) to provide a better understanding of colorectal cancer incidence, survival and mortality. The mortality data provided population-level information on death registration over a 20-year period which enabled geographic and socio-economic analysis of colorectal cancer mortality at LA level. The LS includes individual socio-economic attributes, which are not contained in the other data sets but does not include detailed geographic information. The long-follow up period in this data set enables time-to-event analysis which is not possible in cross-section data. CORECT-R contains small-area geographic information which enables contextual analysis, but it does not include individual socio-economic attributes.

While there was some spatial variation in colorectal cancer mortality, there was no clear pattern by geographic area or area deprivation at Local Authority level. More detailed geographic information is needed to investigate these patterns further. A stronger association was found between individual socio-economic attributes and colorectal cancer incidence and survival. Examination of the influence of the local environment is a relatively new area of cancer research. Some associations with colorectal cancer incidence were observed, but not in the expected direction regarding the retail environment.

Access to up-to-date data is needed to monitor current health outcomes. Combining traditional datasets with novel data sources offer the potential to examine these relationships in more detail.

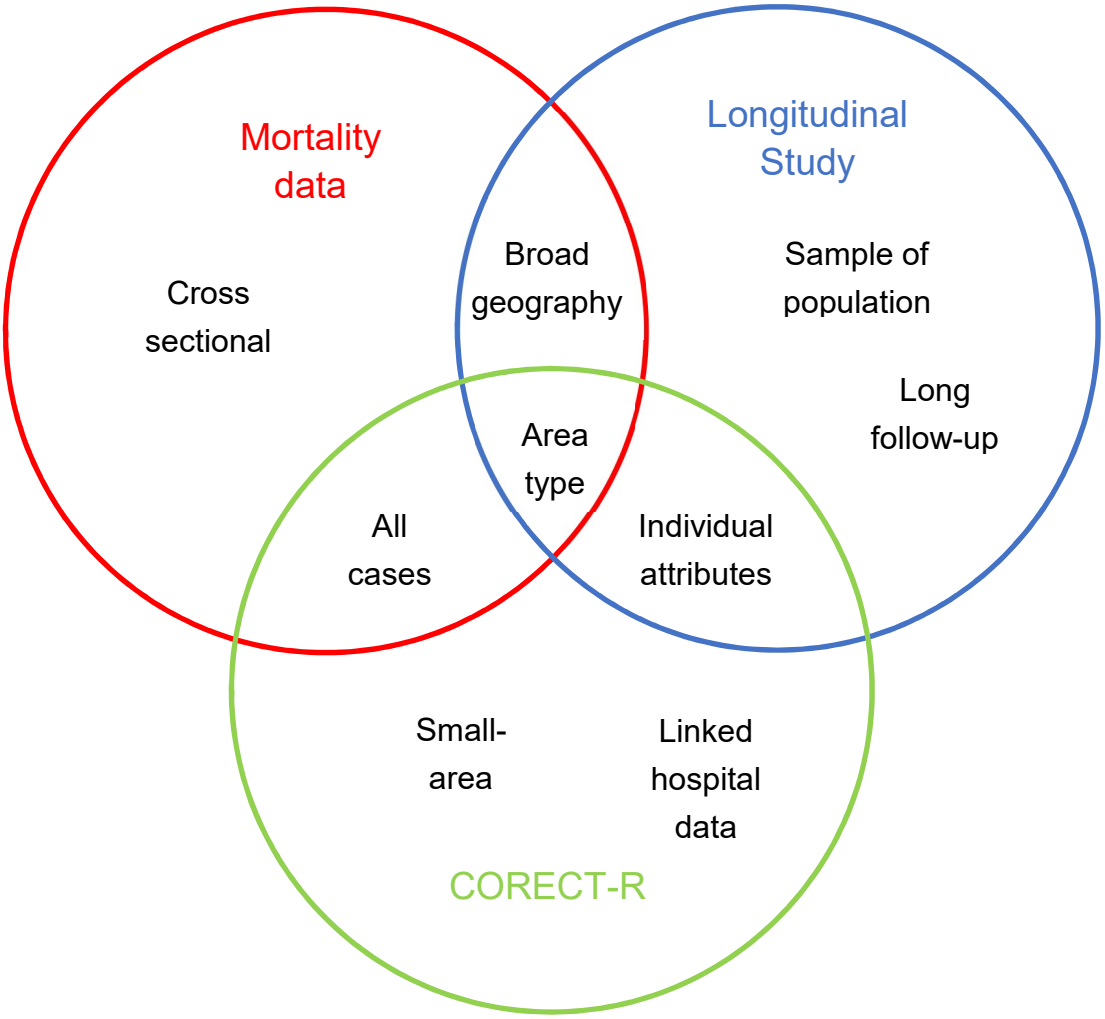


Figure 7.2 Summary of research data sets

Conference presentations

- September 2018 - British Society for Population Studies Annual Conference, Winchester (UK)

  Oral presentation: Spatial variation in colorectal cancer mortality by deprivation: England and Wales

- September 2018 – Leeds Institute for Data Analytics Annual Showcase, Leeds (UK)

  Oral presentation: Spatial variations in colorectal cancer mortality by deprivation: England and Wales

- June 2019 – 18th International Medical Geography Symposium, Queenstown (New Zealand)

  Oral presentation: Variations in colorectal cancer survival by socioeconomic status and area type

- August 2019 – GEOMED, Glasgow (UK)

  Poster presentation: Variations in colorectal cancer survival by socioeconomic status

- April 2021 – GIS Research UK Annual Conference, Cardiff (UK) [Virtual]

  Oral presentation: Spatial variations in colorectal cancer mortality in England and Wales

  Published conference proceedings:

  Sturley, C., Norman, P., Downing, A. and Morris, M. (2021). Spatial variations in colorectal cancer mortality in England and Wales. 29th Annual GIS Research UK Conference (GISRUK), Cardiff, Wales, UK (Online). Available from: https://doi.org/10.5281/zenodo.4669898

References

Aarts, M.J., Lemmens, V.E.P.P., Louwman, M.W.J., Kunst, A.E. and Coebergh, J.W.W. 2010. Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. *European Journal of Cancer*. [Online]. **46**(15), pp.2681–2695. Available from: http://dx.doi.org/10.1016/j.ejca.2010.04.026.

Al-Ajmi, K., Lophatananon, A., Ollier, W. and Muir, K. 2018. Risk of breast cancer in the UK biobank female cohort and its relationship to anthropometric and reproductive factors. *PLoS ONE*. **13**(7), pp.1–18.

Allan, R., Williamson, P. and Kulu, H. 2017. Unravelling urban-rural health disparities in England. *Population Space and Place*. **23**(8).

Ambroggi, M., Biasini, C., Del Giovane, C., Fornari, F. and Cavanna, L. 2015. Distance as a Barrier to Cancer Diagnosis and Treatment: Review of the Literature. *Oncologist*. [Online]. **20**(12), pp.1378–1385. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26512045.

Anselin, L. 1995. Local Indicators of Spatial Association-LISA. *Geographical Analysis*. **27**(2), pp.93–115.

Araghi, M., Soerjomataram, I., Bardot, A., Ferlay, J., Cabasag, C.J., Morrison, D.S., De, P., Tervonen, H., Walsh, P.M., Bucher, O., Engholm, G., Jackson, C., McClure, C., Woods, R.R., Saint-Jacques, N., Morgan, E., Ransom, D., Thursfield, V., Møller, B., Leonfellner, S., Guren, M.G., Bray, F. and Arnold, M. 2019. Changes in colorectal cancer incidence in seven high-income countries: a population-based study. *The Lancet Gastroenterology & Hepatology*. [Online]. **4**(7), pp.511–518. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2468125319301475.

Arik, A., Dodd, E., Cairns, A. and Streftaris, G. 2021. Socioeconomic disparities in cancer incidence and mortality in England and the impact of age-at-diagnosis on cancer mortality. *PLoS ONE*. **16**(7 July).

Arik, A., Dodd, E. and Streftaris, G. 2020. Cancer morbidity trends and regional differences in England-A Bayesian analysis. *PLoS ONE*. **15**(5).

Astell-Burt, T., Mitchell, R. and Hartig, T. 2014. The association between green space and mental health varies across the lifecourse. a longitudinal study. *Journal of Epidemiology and Community Health*. **68**(6), pp.578–583.

Baade, P.D., Dasgupta, P., Aitken, J.F. and Turrell, G. 2011. Distance to the closest radiotherapy facility and survival after a diagnosis of rectal cancer in Queensland. *Medical Journal of Australia*. **195**(6), pp.350–354.

Baade, P.D., Dasgupta, P., Aitken, J.F. and Turrell, G. 2013. Geographic remoteness, area-level socioeconomic disadvantage and inequalities in colorectal cancer survival in Queensland: a multilevel analysis. *BMC Cancer*. **13**.

Belot, A., Ndiaye, A., Luque-Fernandez, M.-A., Kipourou, D.-K., Maringe, C., Rubio, F.J. and Rachet, B. 2019. Summarizing and communicating on survival data according to the audience: a tutorial on different measures illustrated with population-based cancer registry data. *Clinical Epidemiology*. [Online]. **Volume 11**, pp.53–65. Available from:

https://www.dovepress.com/summarising-and-communicating-on-survival-data-according-to-the-audien-peer-reviewed-article-CLEP.

Bernardo, B., Gross, A., Young, G., Baltic, R., Reisinger, S., Blot, W.J. and Paskett, E.D. 2018. Predictors of Colorectal Cancer Screening in two underserved U.S. populations: A parallel analysis. *Frontiers in Oncology*.

Boyle, P. and Parkin, D. 1991. Cancer registration: principles and methods. Statistical methods for registries. *IARC Scientific Publications*. **95**, pp.126–158.

Brown, J., Harding, S., Bethune, A. and Rosato, M. 1998. Longitudinal study of socio-economic differences in the incidence of stomach, colorectal and pancreatic cancers. *Population Trends*. **Winter**(94), pp.35–41.

Brown, K.F., Rumgay, H., Dunlop, C., Ryan, M., Quartly, F., Cox, A., Deas, A., Elliss-Brookes, L., Gavin, A., Hounsome, L., Huws, D., Ormiston-Smith, N., Shelton, J., White, C. and Parkin, D.M. 2018. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. *British Journal of Cancer*. [Online]. **118**(8), pp.1130–1141. Available from: http://dx.doi.org/10.1038/s41416-018-0029-6.

Buck, D. and Frosini, F. 2012. Clustering of unhealthy behaviours over time - Implications for policy and practice. *The Kings Fund*. [Online], pp.1–24. Available from: www.kingsfund.org.uk.

Burgoine, T., Sarkar, C., Webster, C.J. and Monsivais, P. 2018. Examining the interaction of fast-food outlet exposure and income on diet and obesity: Evidence from 51,361 UK Biobank participants. *International Journal of Behavioral Nutrition and Physical Activity*. **15**(1), pp.1–12.

Buron, A., Auge, J.M., Sala, M., Roman, M., Castells, A., Macia, F., Comas, M., Guiriguet, C., Bessa, X., Castells, X. and and the, P. research group 2017. Association between socioeconomic deprivation and colorectal cancer screening outcomes: Low uptake rates among the most and least deprived people. *PLoS One*. [Online]. **12**(6), p.e0179864. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28622365.

Busch, E.L. 2021. Cut points and contexts. *Cancer*. **127**(23), pp.4348–4355.

Campbell, N.C., Elliott, A.M., Sharp, L., Ritchie, L.D., Cassidy, J. and Little, J. 2001. Rural and urban differences in stage at diagnosis of colorectal and lung cancers. *Br J Cancer*. **84**(7), pp.910–914.

Campbell, N.C., Elliott, A.M., Sharp, L., Ritchie, L.D., Cassidy, J. and Little, J. 2000. Rural factors and survival from cancer: analysis of Scottish cancer registrations. *Br J Cancer*. **82**(11), pp.1863–1866.

Cancer Research UK 2021. Bowel cancer statistics. [Accessed 1 March 2022]. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer.

Canchola, A.J., Shariff-Marco, S., Yang, J., Albright, C., Hertz, A., Park, S.Y., Shvetsov, Y.B., Monroe, K.R., Le Marchand, L., Gomez, S.L., Wilkens, L.R. and Cheng, I. 2017. Association between the neighborhood obesogenic environment and colorectal cancer risk in the Multiethnic Cohort. *Cancer Epidemiology*. **50**, pp.99–106.

Carstairs, V. and Morris, R. 1991. *Deprivation and Health in Scotland*. Aberdeen: Aberdeen University Press.

Carstairs, V. and Morris, R. 1989. *Deprivation and mortality: an alternative to social class?* [Online]. Available from: https://academic.oup.com/jpubhealth/article/11/3/210/1497309.

Centre for Longitudinal Studies 2021. Our studies. [Accessed 1 September 2021]. Available from: https://cls.ucl.ac.uk/cls-studies/.

Chang, L.-C., Wu, M.-S., Tu, C.-H., Lee, Y.-C., Shun, C.-T. and Chiu, H.-M. 2014. Metabolic syndrome and smoking may justify earlier colorectal cancer screening in men. *Gastrointestinal Endoscopy*. [Online]. **79**(6), pp.961–969. Available from: https://linkinghub.elsevier.com/retrieve/pii/S001651071302631X.

Clark, S.D., Shute, B., Jenneson, V., Rains, T., Birkin, M. and Morris, M.A. 2021. Dietary patterns derived from UK supermarket transaction data with nutrient and socioeconomic profiles. *Nutrients*. **13**(5), pp.1–21.

Clark, T.G., Bradburn, M.J., Love, S.B. and Altman, D.G. 2003. Survival Analysis Part I: Basic concepts and first analyses. *British Journal of Cancer*. **89**(2), pp.232–238.

Coleman, M.P., Forman, D., Bryant, H., Butler, J., Rachet, B., Maringe, C., Nur, U., Tracey, E., Coory, M., Hatcher, J., McGahan, C.E., Turner, D., Marrett, L., Gjerstorff, M.L., Johannesen, T.B., Adolfsson, J., Lambe, M., Lawrence, G., Meechan, D., Morris, E.J., Middleton, R., Steward, J. and Richards, M.A. 2011. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the international cancer benchmarking partnership): An analysis of population-based cancer registry data. *The Lancet*. [Online]. **377**(9760), pp.127–138. Available from: http://dx.doi.org/10.1016/S0140-6736(10)62231-3.

Coleman, M.P., Rachet, B., Woods, L.M., Mitry, E., Riga, M., Cooper, N., Quinn, M.J., Brenner, H. and Estève, J. 2004. Trends and socioeconomic inequalities in cancer survival in England and Wales up to 2001. *British Journal of Cancer*. **90**(7), pp.1367–1373.

Cook, L., Griffiths, C. and Rooney, C. 2002. The implementation of ICD-10 for cause of death coding -- some preliminary results from the bridge coding study. *Health Statistics Quarterly*. (13), pp.31–41.

Daras, K., Green, M.A., Davies, A., Barr, B. and Singleton, A. 2019. Open data on health-related neighbourhood features in Great Britain. *Scientific Data*. [Online]. **6**(1), pp.1–10. Available from: http://dx.doi.org/10.1038/s41597-019-0114-6.

Davis, M.M., Renfro, S., Pham, R., Hassmiller Lich, K., Shannon, J., Coronado, G.D. and Wheeler, S.B. 2017. Geographic and population-level disparities in colorectal cancer testing: A multilevel analysis of Medicaid and commercial claims data. *Prev Med*. [Online]. **101**, pp.44–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28506715.

Dejardin, O., Bouvier, A.M., Faivre, J., Boutreux, S., De Pouvourville, G. and Launoy, G. 2008. Access to care, socioeconomic deprivation and colon cancer survival. *Alimentary Pharmacology & Therapeutics*. **27**(10), pp.940–949.

Dejardin, O., Jones, A.P., Rachet, B., Morris, E., Bouvier, V., Jooste, V., Coombes, E., Forman, D., Bouvier, A.M. and Launoy, G. 2014. The influence of geographical access to health care and material deprivation on colorectal cancer survival: Evidence from France and England. *Health & Place*. **30**, pp.36–44.

Department of Health 2007. *Cancer Reform Strategy* [Online]. Available from: https://www.nhs.uk/NHSEngland/NSF/Documents/Cancer Reform Strategy.pdf.

Department of Health 2009. *Tackling health inequalities: 10 years on a review of the developments in tackling health inequalities in England over the last 10 years.* [Online]. London. Available from: http://dera.ioe.ac.uk/id/eprint/11036.

Department of Health 2000. *The NHS Cancer Plan: a plan for investment, a plan for reform.*

Diez-Roux, A. V 1998. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *American Journal of Public Health*. **88**(2), pp.216–222.

Downing, A., Hall, P., Birch, R., Lemmon, E., Affleck, P., Rossington, H., Boldison, E., Ewart, P. and Morris, E.J.A. 2021. Data Resource Profile: The COloRECTal cancer data repository (CORECT-R). *International Journal of Epidemiology*. [Online]. Available from: https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyab122/6319547.

Duncan, C., Jones', K. and Moon, G. 1998. *Context, composition and heterogeneity: using multilevel models in health research*.

Duncan, C., Jones, K. and Moon, G. 1999. Smoking and deprivation: are there neighbourhood effects? *Social Science & Medicine*. **48**, pp.497–505.

Ellis, A. and Fry, R. 2010. Regional health inequalities in England. *Regional Trends*. **42**(1), pp.60–79.

Ellis, L., Coleman, M.P. and 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*. [Online]. **48**(2), pp.270–278. Available from: http://dx.doi.org/10.1016/j.ejca.2011.10.008.

Espelt, A., Borrell, C., Roskam, A.J., Rodríguez-Sanz, M., Stirbu, I., Dalmau-Bueno, A., Regidor, E., Bopp, M., Martikainen, P., Leinsalu, M., Artnik, B., Rychtarikova, J., Kalediene, R., Dzurova, D., Mackenbach, J. and Kunst, A.E. 2008. Socioeconomic inequalities in diabetes mellitus across Europe at the beginning of the 21st century. *Diabetologia*. **51**(11), pp.1971–1979.

Esri 2015. ArcMap (Version 10.4.1).

Eurostat 2013. *Revision of the European Standard Population - Report of Eurostat's task force*. Luxembourg.

Ezzati, M. and Riboli, E. 2013. Behavioral and Dietary Risk Factors for Noncommunicable Diseases. *New England Journal of Medicine*. **369**(10), pp.954–964.

Faggiano, F., Partanen, T., Kogevinas, M. and Boffetta, P. 1997.

Socioeconomic differences in cancer incidence and mortality *In*: M. Kogevinas, M. Pearce, M. Susser and P. Boffetta, eds. *Social Inequalities and Cancer*. Lyon: International Agency for Research on Cancer, pp.65–176.

Fearnhead, N.S., Wilding, J.L. and Bodmer, W.F. 2002. Genetics of colorectal cancer: Hereditary aspects and overview of colorectal tumorigenesis. *British Medical Bulletin*. **64**, pp.27–43.

Franklin, R. 2022. Quantitative methods I: Reckoning with uncertainty. *Progress in Human Geography*. **46**(2), pp.689–697.

Fransen, H.P., May, A.M., Beulens, J.W.J., Struijk, E.A., De Wit, G.A., Boer, J.M.A., Onland-Moret, N.C., Hoekstra, J., Der Schouw, Y.T.V., Bueno-De-mesquita, H.B. and Peeters, P.H.M. 2014. Association between lifestyle factors and quality- adjusted life years in the EPIC-NL cohort. *PLoS ONE*. **9**(11).

Fraser, L.K. and Edwards, K.L. 2010. The association between the geography of fast food outlets and childhood obesity rates in Leeds, UK. *Health and Place*. [Online]. **16**(6), pp.1124–1128. Available from: http://dx.doi.org/10.1016/j.healthplace.2010.07.003.

Fraser, L.K., Edwards, K.L., Cade, J. and Clarke, G.P. 2010. The geography of fast food outlets: A review. *International Journal of Environmental Research and Public Health*. **7**(5), pp.2290–2308.

Gartner, A., Farewell, D., Roach, P. and Dunstan, F. 2011. Rural/urban mortality differences in England and Wales and the effect of deprivation adjustment. *Social Science and Medicine*. [Online]. **72**(10), pp.1685–1694. Available from: http://dx.doi.org/10.1016/j.socscimed.2011.03.017.

GBD 2010 Diseases and Injuries Collaborators 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. **380**(9859), pp.2224–2260.

GBD 2019 Diseases and Injuries Collaborators 2020. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. [Online]. **396**(10258), pp.1204–1222. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673620309259.

Gibson, D.C., Prochaska, J.D., Yu, X. and Kaul, S. 2020. An examination between census tract unhealthy food availability and colorectal cancer incidence. *Cancer Epidemiology*. **67**.

Goon, S., Kim, H. and Giovannucci, E.L. 2021. Population attributable risk for colorectal and breast cancer in England, Wales, Scotland, Northern Ireland, and the United Kingdom. *AMRC Open Research*. [Online]. **3**, p.11. Available from: https://amrcopenresearch.org/articles/3-11/v1.

Green, M.A., Daras, K., Davies, A., Barr, B. and Singleton, A. 2018. Developing an openly accessible multi-dimensional small area index of 'Access to Healthy Assets and Hazards' for Great Britain, 2016. *Health and Place*. [Online]. **54**(August), pp.11–19. Available from: https://doi.org/10.1016/j.healthplace.2018.08.019.

Hagedoorn, P., Vandenheede, H., Vanthomme, K. and Gadeyne, S. 2018. Socioeconomic position, population density and site-specific cancer mortality: A multilevel analysis of Belgian adults, 2001–2011. *International Journal of Cancer*. **142**(1), pp.23–35.

van Ham, M. and Manley, D. 2012. Neighbourhood effects research at a crossroads. Ten challenges for future research Introduction. *Environment and Planning A*. **44**(12), pp.2787–2793.

van Ham, M., Manley, D., Bailey, N., Simpson, L. and Maclennan, D. (eds.). 2012. *Neighbourhood Effects Research: New Perspectives* [Online]. Dordrecht: Springer Netherlands. Available from: http://link.springer.com/10.1007/978-94-007-2309-2.

Hattersley, L., Creeser, R. and Great Britain. Office of Population Censuses and Surveys. 1995. *Longitudinal study 1971-1991 : history, organisation and quality of data*. H.M.S.O.

Henderson, R.H., French, D., Maughan, T., Adams, R., Allemani, C., Minicozzi, P., Coleman, M.P., McFerran, E., Sullivan, R. and Lawler, M. 2021. The economic burden of colorectal cancer across Europe: a population-based cost-of-illness study. *The Lancet Gastroenterology & Hepatology*. **6**(9).

Hiam, L., Harrison, D., McKee, M. and Dorling, D. 2018. Why is life expectancy in England and Wales 'stalling'? *Journal of Epidemiology and Community Health*. **72**(5), pp.404–408.

Hole, D J; McArdle, C.S. 2002. Impact of socioeconomic deprivation on outcome after surgery for colorectal cancer. *British Journal of Surgery*. **89**(5), pp.586–590.

Humphreys, K. and Carr-Hill, R. 1991. *Area Variations in Health Outcomes Artefact or Ecology* [Online]. Available from: https://academic.oup.com/ije/article/20/1/251/777033.

Incisive Health and Cancer Research UK 2015. *Saving Lives, Averting Costs*. London.

Ingleby, F.C., Belot, A., Atherton, I., Baker, M., Elliss-Brookes, L. and Woods, L.M. 2020. Assessment of the concordance between individual-level and area-level measures of socio-economic deprivation in a cancer patient cohort in England and Wales. *BMJ Open*. **10**(11).

International Agency for Research on Cancer 2021. *List of classifications by cancer sites with sufficient or limited evidence in humans, Volumes 1 to 129.*

Ionescu, M., Carey, F., Tait, I. and Steele, J. 1998. Socioeconomic status and stage at presentation of colorectal cancer. *The Lancet*. **352**, p.1439.

James, W.H.M., Lomax, N. and Birkin, M. 2019. Local level estimates of food, drink and tobacco expenditure for Great Britain. *Scientific Data*. [Online]. **6**(1), pp.1–14. Available from: http://dx.doi.org/10.1038/s41597-019-0064-z.

Janz, N. and Becker, M. 1984. The Health Belief Model: A Decade Later. *Health Education & Behavior*. **11**(1), pp.1–47.

Jivraj, S., Murray, E.T., Norman, P. and Nicholas, O. 2020. The impact of life course exposures to neighbourhood deprivation on health and well-being: A review of the long-term neighbourhood effects literature. *European*

Journal of Public Health. **30**(5), pp.922–928.

Jones, A.P., Haynes, R., Sauerzapf, V., Crawford, S.M., Zhao, H. and Forman, D. 2008a. Travel time to hospital and treatment for breast, colon, rectum, lung, ovary and prostate cancer. *European Journal of Cancer*. **44**(7), pp.992–999.

Jones, A.P., Haynes, R., Sauerzapf, V., Crawford, S.M., Zhao, H. and Forman, D. 2008b. Travel times to health care and survival from cancers in Northern England. *European Journal of Cancer*. **44**(2), pp.269–274.

Kajiwara Saito, M., Quaresma, M., Fowler, H., Benitez Majano, S. and Rachet, B. 2021. Socioeconomic gaps over time in colorectal cancer survival in England: flexible parametric survival analysis. *Journal of Epidemiology and Community Health*. [Online], jech-2021-216754. Available from: https://jech.bmj.com/lookup/doi/10.1136/jech-2021-216754.

Kaplan, E.L. and Meier, P. 1958. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Assocation*. **53**(282), pp.457–481.

Khaw, K.T., Wareham, N., Bingham, S., Welch, A., Luben, R. and Day, N. 2008. Combined impact of health behaviours and mortality in men and women: The EPIC-Norfolk prospective population study. *PLoS Medicine*. **5**(1), pp.0039–0047.

Kogevinas, M. and Porta, M. 1997. Socioeconomic differences in cancer survival: a review of the evidence *In*: M. Kogevinas, M. Pearce, S. M and M. Boffetta, eds. *Social Inequalities and Cancer*. Lyon: International Agency for Research on Cancer, pp.177–206.

Kunst, A.E., Bos, V., Lahelma, E., Bartley, M., Lissau, I., Regidor, E., Mielck, A., Cardano, M., Dalstra, J.A.A., Geurts, J.J.M., Helmert, U., Lennartsson, C., Ramm, J., Spadea, T., Stronegger, W.J. and Mackenbach, J.P. 2005. Trends in socioeconomic inequalities in self-assessed health in 10 European countries. *International Journal of Epidemiology*. **34**(2), pp.295–305.

Kwan, M.P. 2012. The Uncertain Geographic Context Problem. *Annals of the Association of American Geographers*. **102**(5), pp.958–968.

Lauby-Secretan, B., Scoccianti, C., Loomis, D., Grosse, Y., Bianchini, F. and Straif, K. 2016. *Body Fatness and Cancer-Viewpoint of the IARC Working Group*.

Laudicella, M., Walsh, B., Burns, E. and Smith, P.C. 2016. Cost of care for cancer patients in England: Evidence from population-based patient-level data. *British Journal of Cancer*. [Online]. **114**(11), pp.1286–1292. Available from: http://dx.doi.org/10.1038/bjc.2016.77.

Lejeune, C., Sassi, F., Ellis, L., Godward, S., Mak, V., Day, M. and Rachet, B. 2010. Socio-economic disparities in access to treatment and their impact on colorectal cancer survival. *International Journal of Epidemiology*. **39**(3), pp.710–717.

Leon, D.A., Jdanov, D.A. and Shkolnikov, V.M. 2019. Trends in life expectancy and age-specific mortality in England and Wales, 1970–2016, in comparison with a set of 22 high-income countries: an analysis of vital statistics data. *The Lancet Public Health*. **4**(11), pp.e575–e582.

Leyland, A.H. 2004. Increasing inequalities in premature mortality in Great Britain. *Journal of Epidemiology and Community Health*. **58**(4), pp.296–302.

Loopstra, R., McKee, M., Katikireddi, S.V., Taylor-Robinson, D., Barr, B. and Stuckler, D. 2016. Austerity and old-age mortality in England: a longitudinal cross-local area analysis, 2007–2013. *Journal of the Royal Society of Medicine*. **109**(3), pp.109–116.

Luben, R., Hayat, S., Khawaja, A., Wareham, N., Pharoah, P.P. and Khaw, K.T. 2019. Residential area deprivation and risk of subsequent hospital admission in a British population: the EPIC-Norfolk cohort. *BMJ open*. **9**(12), p.e031251.

Maas, J., Verheij, R.A., Groenewegen, P.P., De Vries, S. and Spreeuwenberg, P. 2006. Green space, urbanity, and health: How strong is the relation? *Journal of Epidemiology and Community Health*. **60**(7), pp.587–592.

Macintyre, S., Ellaway, A. and Cummins, S. 2002. Place effects on health: how can we conceptualise, operationalise and measure them? *Social Science & Medicine*. [Online]. **55**(1), pp.125–139. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0277953601002143.

Mackenbach, J.P., Cavelaars, A.E.J.M., Kunst, A.E., Groenhof, F., Andersen, O., Borgan, J.K., Costa, G., Crialesi, R., Desplanques, G., Faggiano, F., Geurts, J.J.M., Grötvedt, L., Harding, S., Helmert, U., Junker, C., Lahelma, E., Lundberg, O., Matikainen, P., Matheson, J., Mielk, A., Minder, C., Mizrahi, AR, Mizrahi, A., Nolan, B., Rasmussen, N., Regidor, E., Giraldes, M.D.R., Spuhler, T., Vågerö, D. and Valkonen, T. 2000. Socioeconomic inequalities in cardiovascular disease mortality. An international study. *European Heart Journal*. **21**(14), pp.1141–1151.

Mackenbach, J.P., Stirbu, I., Roskam, A.-J.R., Schaap, M.M., Menvielle, G., Leinsalu, M., Kunst, A.E. and European Union Working Group on Socioeconomic Inequalities in Health 2008. Socioeconomic inequalities in health in 22 European countries. *The New England journal of medicine*. [Online]. **358**(23), pp.2468–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18525043.

Maheswaran, R., Pearson, T., Jordan, H. and Black, D. 2006. Socioeconomic deprivation, travel distance, location of service, and uptake of breast cancer screening in North Derbyshire, UK. *Journal of Epidemiology and Community Health*. **60**(3), pp.208–212.

Marmot, M., Allen, J., Boyce, T., Goldblatt, P. and Morrison, J. 2020. *Health equity in England: The Marmot Review 10 years on*. London: Institute of Health Equity.

Massarweh, N.N., Chiang, Y.J., Xing, Y., Chang, G.J., Haynes, A.B., You, Y.N., Feig, B.W. and Cormier, J.N. 2014. Association Between Travel Distance and Metastatic Disease at Diagnosis Among Patients With Colon Cancer. *JOURNAL OF CLINICAL ONCOLOGY*. **32**(9), pp.942-+.

McGregor, L.M., Bonello, B., Kerrison, R.S., Nickerson, C., Baio, G., Berkman, L., Rees, C.J., Atkin, W., Wardle, J. and Von Wagner, C. 2016. Uptake of bowel scope (Flexible sigmoidoscopy) screening in the english national programme: The first 14 months. *Journal of Medical Screening*. **23**(2),

pp.77–82.

McLeroy, K., Bibeau, D., Steckler, A. and Glanz, K. 1988. An ecological perspective on health promotion programs. *Health Education Quarterly*. **15**(4), pp.351–377.

Meade, M. and Earickson, R. 2000. *Medical Geography* 2nd ed. New York: Guilford Press.

Menvielle, G., Kunst, A.E., Stirbu, I., Strand, B.H., Borrell, C., Regidor, E., Leclerc, A., Esnaola, S., Bopp, M., Lundberg, O., Artnik, B., Costa, G., Deboosere, P., Martikainen, P. and Mackenbach, J.P. 2008. Educational differences in cancer mortality among women and men: A gender pattern that differs across Europe. *British Journal of Cancer*. **98**(5), pp.1012–1019.

Ministry of Housing Communities and Local Government 2019. English indices of deprivation 2019. [Accessed 1 April 2022]. Available from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019.

Mitchell, R. and Popham, F. 2008. Effect of exposure to natural environment on health inequalities: an observational population study. *The Lancet*. **372**(9650), pp.1655–1660.

Mitry, E., Rachet, B., Quinn, M.J., Cooper, N. and Coleman, M.P. 2008a. Survival from cancer of the colon in england and wales up to 2001. *British Journal of Cancer*. **99**, pp.S26–S29.

Mitry, E., Rachet, B., Quinn, M.J., Cooper, N. and Coleman, M.P. 2008b. Survival from cancer of the rectum in england and wales up to 2001. *British Journal of Cancer*. **99**, pp.S30–S32.

Møller, H., Sandin, F., Robinson, D., Bray, F., Klint, S., Linklater, K.M., Lambert, P.C., Phlman, L., Holmberg, L. and Morris, E. 2012. Colorectal cancer survival in socioeconomic groups in England: Variation is mainly in the short term after diagnosis. *European Journal of Cancer*. **48**(1), pp.46–53.

Morris, E., Quirke, P., Thomas, J.D., Fairley, L., Cottier, B. and Forman, D. 2008. Unacceptable variation in abdominoperineal excision rates for rectal cancer: Time to intervene? *Gut*. **57**(12), pp.1690–1697.

Morris, E.J.A., Goldacre, R., Spata, E., Mafham, M., Finan, P.J., Shelton, J., Richards, M., Spencer, K., Emberson, J., Hollings, S., Curnow, P., Gair, D., Sebag-Montefiore, D., Cunningham, C., Rutter, M.D., Nicholson, B.D., Rashbass, J., Landray, M., Collins, R., Casadei, B. and Baigent, C. 2021. Impact of the COVID-19 pandemic on the detection and management of colorectal cancer in England: a population-based study. *The Lancet Gastroenterology and Hepatology*. [Online]. **6**(3), pp.199–208. Available from: http://dx.doi.org/10.1016/S2468-1253(21)00005-4.

Morris, M.A., Hulme, C., Clarke, G.P., Edwards, K.L. and Cade, J.E. 2014. What is the cost of a healthy diet? Using diet data from the UK Women's Cohort Study. *Journal of Epidemiology and Community Health*. **68**(11), pp.1043–1049.

Munford, L., Khavandi, S., Bambra, C., Barr, B., Davies, H., Doran, T., Kontopantelis, E., Norman, P., Pickett, K., Sutton, M., Taylor-Robinson, D. and Wickham, S. 2021. *A year of COVID-19 in the North: Regional inequalities in health and economic outcomes*. Newcastle.

National Cancer Intelligence Network 2014. *Cancer by deprivation in England: Incidence, 1996-2010 & Mortality, 1997-2011* [Online]. United Kingdom. Available from: http://www.ncin.org.uk/view?rid=2691.

National Cancer Intelligence Network 2009. *Cancer Incidence and Survival By Major Ethnic Group, England, 2002-2006* [Online]. Available from: http://www.ncin.org.uk/publications/reports/reports_archive.

National Cancer Intelligence Network 2011. *The effect of rurality on cancer incidence and mortality* [Online]. Available from: http://www.ncin.org.uk/publications/data_briefings/rurality.

National Cancer Registration and Analysis Service 2022. Rapid Cancer Registrations Data. [Accessed 1 April 2022]. Available from: http://www.ncin.org.uk/collecting_and_using_data/rcrd.

NHS Digital 2019. *Health Survey for England 2018: Adults' health-related behaviours* [Online]. Available from: https://files.digital.nhs.uk/A3/2F03E4/HSE18-Adult-Health-Related-Behaviours-rep-v2.pdf.

Nomis 2021a. Mortality statistics - underlying cause, sex and age. [Accessed 12 November 2021]. Available from: https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=161.

Nomis 2021b. Population estimates - small area based by single year of age - England and Wales. [Accessed 14 July 2021]. Available from: https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=2010.

Nomis 2013. QS102EW - Population density. [Accessed 12 November 2021]. Available from: https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=143.

Norman, P. 2017. Area characteristics: Great Britain 1971 to 2011.

Norman, P. 2010. *Identifying change over time in small area socio-economic deprivation*.

Norman, P. and Darlington-Pollock, F. 2017. The Changing Geography of Deprivation in Great Britain: Exploiting Small Area Census Data, 1971 to 2011 *In*: J. Stillwell, ed. *Routledge Handbook of Census Resources, Methods and Applications*., pp.404–420.

Norman, P. and Riva, M. 2012. Population health across space and time: The geographical harmonisation of the office for national statistics longitudinal study for England and Wales. *Population, Space and Place*. **18**(5), pp.483–502.

OECD 1994. *Creating rural indicators for shaping territorial policy*. Paris: OECD Publications.

Office for Health Improvement and Disparities 2022. Public health profiles. [Accessed 1 April 2022]. Available from: https://fingertips.phe.org.uk/.

Office for National Statistics 2011. 2011 Census: boundary data (England and Wales) [data collection]. UK Data Service. SN:5819 UKBORDERS: Digitised Boundary Data, 1840- and Postcode Directories, 1980-.

Office for National Statistics 2021a. Accessing secure research data as an accredited researcher. [Accessed 1 November 2021]. Available from: https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme.

Office for National Statistics 2017. Adult smoking habits in the UK: 2015. [Accessed 13 November 2021]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2015.

Office for National Statistics 2019. Cancer registration statistics, England: 2017. [Accessed 22 June 2018]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2017.

Office for National Statistics 2020. Deaths involving COVID-19 by local area and socioeconomic deprivation: deaths occurring between 1 March and 30 June 2020. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsinvolvingcovid19bylocalareasanddeprivation/deathsoccurringbetween1marchand30june2020.

Office for National Statistics 2021b. *Health state life expectancies by national deprivation deciles, England: 2017 to 2019* [Online]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthinequalities/bulletins/healthstatelifeexpectanciesbyindexofmultipledeprivationimd/2017to2019.

Office for National Statistics 2018. Impact of registration delays on mortality statistics: 2016. [Accessed 12 November 2021]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/impactofregistrationdelaysonmortalitystatistics2016.

Office for National Statistics 2021c. Lower layer Super Output Area population estimates (supporting information). [Accessed 10 November 2021]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimates.

Office for National Statistics 2021d. Open Geography Portal. [Accessed 1 September 2021]. Available from: https://geoportal.statistics.gov.uk/.

Office for National Statistics 2021e. UK geographies. [Accessed 24 November 2021]. Available from: https://www.ons.gov.uk/methodology/geography/ukgeographies.

Oldroyd, R.A., Hobbs, M., Campbell, M., Jenneson, V., Marek, L., Morris, M.A., Pontin, F., Sturley, C., Tomintz, M., Wiki, J., Birkin, M., Kingham, S. and Wilson, M. 2021. Progress Towards Using Linked Population-Based Data For Geohealth Research: Comparisons Of Aotearoa New Zealand And The United Kingdom. *Applied Spatial Analysis and Policy*. **14**(4), pp.1025–1040.

Oliphant, R., Brewster, D.H. and Morrison, D.S. 2011. The changing association between socioeconomic circumstances and the incidence of colorectal

cancer: A population-based study. *British Journal of Cancer*. [Online]. **104**(11), pp.1791–1796. Available from: http://dx.doi.org/10.1038/bjc.2011.149.

Parsons, M.A. and Askland, K.D. 2007. Cancer of the colorectum in Maine, 1995-1998: Determinants of stage at diagnosis in a rural state. *Journal of Rural Health*. **23**(1), pp.25–32.

Pearce, J.R. 2018. Complexity and Uncertainty in Geography of Health Research: Incorporating Life-Course Perspectives. *Annals of the American Association of Geographers*. **108**(6), pp.1491–1498.

Pearce, J.R., Richardson, E.A., Mitchell, R.J. and Shortt, N.K. 2010. Environmental justice and health: The implications of the socio-spatial distribution of multiple environmental deprivation for health inequalities in the United Kingdom. *Transactions of the Institute of British Geographers*. **35**(4), pp.522–539.

Perchoux, C., Chaix, B., Cummins, S. and Kestens, Y. 2013. Conceptualization and measurement of environmental exposure in epidemiology: Accounting for activity space related to daily mobility. *Health and Place*. **21**, pp.86–93.

Pollock, A.M. and Vickers, N. 1997. Breast, lung and colorectal cancer incidence and survival in South Thames Region, 1987-1992: the effect of social deprivation. *Journal of Public Health Medicine*. **19**(3), pp.288–294.

Pollock, A.M. and Vickers, N. 1998. Deprivation and emergency admissions for cancers of colorectum, lung, and breast in south east England: Ecological study. *British Medical Journal*. **317**(7153), pp.245–252.

Pornet, C., Dejardin, O., Morlais, F., Bouvier, V. and Launoy, G. 2010. Socioeconomic determinants for compliance to colorectal cancer screening. A multilevel analysis. *J Epidemiol Community Health*. [Online]. **64**(4), pp.318–324. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19740776.

Public Health England 2020. *Cancer survival in England for patients diagnosed between 2014 and 2018, and followed up to 2019* [Online]. Available from: https://www.gov.uk/government/statistics/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-until-2019/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-to-2019.

Public Health England 2019. *Index of cancer survival for Clinical Commissioning Groups in England: adults diagnosed 2002 to 2017 and followed up to 2018* [Online]. Available from: https://www.gov.uk/government/statistics/cancer-survival-index-for-clinical-commissioning-groups/index-of-cancer-survival-for-clinical-commissioning-groups-in-england-adults-diagnosed-2002-to-2017-and-followed-up-to-2018.

Public Health England 2016. *National Diet and Nutrition Survey: Results from Years 5 and 6 (combined) of the Rolling Programme (2012/2013-2013-2014)*. London.

Quinn, M., Babb, P., Brock, A., Kirby, L. and Jones, J. 2001. *Cancer trends in England and Wales 1950-1999* SMPS No 66. London: Office for National Statistics.

Race Disparity Unit 2021. List of ethnic groups. [Accessed 11 October 2021]. Available from: https://www.ethnicity-facts-figures.service.gov.uk/style-guide/ethnic-groups.

Richardson, E.A., Mitchell, R., Shortt, N.K., Pearce, J. and Dawson, T.P. 2010. Developing summary measures of health-related multiple physical environmental deprivation for epidemiological research. *Environment and Planning A*. **42**(7), pp.1650–1668.

Roux, A.V.D. 2001. Investigating Neighborhood and Area Effects on Health. *American Journal of Public Health*. **91**(11), pp.1783–1789.

Schoenbach, V.J. and Rosamond, W.D. 2000. *Understanding the Fundamentals of Epidemiology: an evolving text*. Chapel Hill: University of North Carolina.

Schrijvers, C.T.M., Mackenbach, J.P., Lutz, J. -M, Quinn, M.J. and Coleman, M.P. 1995. Deprivation, stage at diagnosis and cancer survival. *International Journal of Cancer*. **63**(3), pp.324–329.

Schutze, M., Boeing, H., Pischon, T., Rehm, J., Kehoe, T., Gmel, G., Olsen, A., Tjonneland, A.M., Dahm, C.C., Overvad, K., Clavel-Chapelon, F., Boutron-Ruault, M.-C., Trichopoulou, A., Benetou, V., Zylis, D., Kaaks, R., Rohrmann, S., Palli, D., Berrino, F., Tumino, R., Vineis, P., Rodriguez, L., Agudo, A., Sanchez, M.-J., Dorronsoro, M., Chirlaque, M.-D., Barricarte, A., Peeters, P.H., van Gils, C.H., Khaw, K.-T., Wareham, N., Allen, N.E., Key, T.J., Boffetta, P., Slimani, N., Jenab, M., Romaguera, D., Wark, P.A., Riboli, E. and Bergmann, M.M. 2011. Alcohol attributable burden of incidence of cancer in eight European countries based on results from prospective cohort study. *BMJ*. [Online]. **342**, p.d1584. Available from: https://www.bmj.com/lookup/doi/10.1136/bmj.d1584.

Scoggins, J.F., Fedorenko, C.R., Donahue, S.M.A., Buchwald, D., Blough, D.K. and Ramsey, S.D. 2012. Is Distance to Provider a Barrier to Care for Medicaid Patients With Breast, Colorectal, or Lung Cancer? *Journal of Rural Health*. **28**(1), pp.54–62.

Sharpe, K.H., McMahon, A.D., Raab, G.M., Brewster, D.H. and Conway, D.I. 2014. Association between socioeconomic factors and cancer risk: A population cohort study in Scotland (1991-2006). *PLoS ONE*. **9**(2).

Sharples, K.J., Firth, M.J., Hinder, V.A., Hill, A.G., Jeffery, M., Sarfati, D., Brown, C., Atmore, C., Lawrenson, R.A., Reid, P.M.J., Derrett, S.L., Macapagel, J., Keating, J.P., Secker, A.H., De Groot, C., Jackson, C.G.C.A., Findlay, M.P.N., Macapagal, J., Keating, J.P., Secker, A.H., De Groot, C., Jackson, C.G.C.A. and Findlay, M.P.N. 2018. The New Zealand PIPER Project: colorectal cancer survival according to rurality, ethnicity and socioeconomic deprivation-results from a retrospective cohort study. *N Z Med J*. **131**(1476), pp.24–39.

Shelton, N., Marshall, C.E., Stuchbury, R., Grundy, E., Dennett, A., Tomlinson, J., Duke-Williams, O. and Xun, W. 2018. Cohort Profile: the Office for National Statistics Longitudinal Study (The LS). *International Journal of Epidemiology*. [Online], pp.1–9. Available from: https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyy243/5239953.

Sherk, A., Stockwell, T., Chikritzhs, T., Andréasson, S., Angus, C., Gripenberg, J., Holder, H., Holmes, J., Mäkelä, P., Mills, M., Norström, T., Ramstedt, M. and Woods, J. 2018. Alcohol consumption and the physical availability of take-away alcohol: Systematic reviews and meta-analyses of the days and hours of sale and outlet density. *Journal of Studies on Alcohol and Drugs*. **79**(1), pp.58–67.

Shortt, N.K., Tisch, C., Pearce, J., Richardson, E.A. and Mitchell, R. 2016. The density of tobacco retailers in home and school environments and relationship with adolescent smoking behaviours in Scotland. *Tobacco Control*. **25**(1), pp.75–82.

Shouls, S., Congdon, P. and Curtis, S. 1996. *Modelling inequality in reported long term illness in the UK: combining individual and area characteristics Compositional effects operate simply because of the varying distribution of types of people whose individual characteristics influence their*.

Shvetsov, Y.B., Shariff-Marco, S., Yang, J., Conroy, S.M., Canchola, A.J., Albright, C.L., Park, S.Y., Monroe, K.R., Le Marchand, L., Gomez, S.L., Wilkens, L.R. and Cheng, I. 2020. Association of change in the neighborhood obesogenic environment with colorectal cancer risk: The Multiethnic Cohort Study. *SSM - Population Health*. **10**.

Siegel, R.L., Sahar, L., Robbins, A. and Jemal, A. 2015. Where can colorectal cancer screening interventions have the most impact? *Cancer Epidemiology Biomarkers and Prevention*. **24**(8), pp.1151–1156.

Singh, G.K., Williams, S.D., Siahpush, M. and Mulhollen, A. 2011. Socioeconomic, rural-urban, and racial inequalities in US cancer mortality: Part I-All cancers and lung cancer and part II-Colorectal, prostate, breast, and cervical cancers. *Journal of Cancer Epidemiology*. **2011**.

Sloggett, A. and Joshi, H. 1994. Higher mortality in deprived areas: Community or personal disadvantage? *Bmj*. **309**(6967), p.1470.

Sloggett, A., Young, H. and 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**, pp.1–8.

Smith, G.D., Hart, C., Watt, G., Hole, D. and Hawthorne, V. 1998. *Individual social class, area-based deprivation, cardiovascular disease risk factors, and mortality: the Renfrew and Paisley study*.

Smith, J., Walshe, K. and Hunter, D. 2001. The “redisorganisation” of the NHS. *BMJ*. **323**, pp.1262–1263.

Smith, L., Downing, A., Norman, P., Wright, P., Hounsome, L., Watson, E., Wagland, R., Selby, P., Kind, P., Donnelly, D.W., Butcher, H., Huws, D., McNair, E., Gavin, A. and Glaser, A.W. 2020. Influence of deprivation and rurality on patient-reported outcomes of men living with and beyond prostate cancer diagnosis in the UK: A population-based study. *Cancer Epidemiology*. [Online]. **69**(September), p.101830. Available from: https://doi.org/10.1016/j.canep.2020.101830.

Smith, L., Norman, P., Kapetanstrataki, M., Fleming, S., Fraser, L.K., Parslow, R.C. and Feltbower, R.G. 2017. Comparison of ethnic group classification using naming analysis and routinely collected data: Application to cancer incidence trends in children and young people. *BMJ Open*. **7**(9), pp.1–8.

Stafford, M. and Marmot, M. 2003. Neighbourhood deprivation and health: Does it affect us all equally? *International Journal of Epidemiology*. **32**(3), pp.357–366.

Steel, N., Ford, J.A., Newton, J.N., Davis, A.C.J., Vos, T., Naghavi, M., Glenn, S., Hughes, A., Dalton, A.M., Stockton, D., Humphreys, C., Dallat, M., Schmidt, J., Flowers, J., Fox, S., Abubakar, I. and Aldridge, R.W. 2018. Changes in health in the countries of the UK and 150 English Local Authority areas 1990 – 2016 : a systematic analysis for the Global Burden of Disease Study 2016. , pp.1647–1661.

Steenland, K., Hu, S. and Walker, J. 2004. All-cause and cause-specific mortality by socioeconomic status among employed persons in 27 US States - 1984-1997. . **94**(6), pp.1037–1042.

Sturley, C., Newing, A. and Heppenstall, A. 2018. Evaluating the potential of agent-based modelling to capture consumer grocery retail store choice behaviours. *The International Review of Retail, Distribution and Consumer Research*. [Online]. **28**(1), pp.27–46. Available from: http://doi.org/10.1080/09593969.2017.1397046.

Syriopoulou, E., Morris, E., Finan, P.J., Lambert, P.C. and Rutherford, M.J. 2019. Understanding the impact of socioeconomic differences in colorectal cancer survival: potential gain in life-years. *British Journal of Cancer*. [Online]. (February), pp.1–7. Available from: http://www.nature.com/articles/s41416-019-0455-0.

Taylor-Robinson, D., Lai, E.T.C., Wickham, S., Rose, T., Norman, P., Bambra, C., Whitehead, M. and Barr, B. 2019. Assessing the impact of rising child poverty on the unprecedented rise in infant mortality in England, 2000-2017: time trend analysis. *BMJ Open*. **9**(10).

Textor, J., van der Zander, B., Gilthorpe, M.K., Liskiewicz, M. and Ellison, G.T.H. 2016. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology*. **45**(6), pp.1887–1894.

The King's Fund 2015. *Inequalities in life expectancy: changes over time and implications for policy* [Online]. Available from: file:///C:/Users/gy06ces/Downloads/Health state life expectancies by national deprivation deciles, England 2017 to 2019.pdf.

Townsend, P., Phillimore, P. and Beattie, A. 1988. *Health and Deprivation: Inequality and the North*. Andover: Croom Helm.

Tweed, E.J., Allardice, G.M., McLoone, P. and Morrison, D.S. 2018. Socio-economic inequalities in the incidence of four common cancers: a population-based registry study. *Public Health*. [Online]. **154**, pp.1–10. Available from: https://doi.org/10.1016/j.puhe.2017.10.005.

UK Colorectal Cancer Intelligence Hub 2021. CORECT-R. [Accessed 29 October 2021]. Available from: https://www.ndph.ox.ac.uk/corectr/corect-r.

UK Data Service 2021. What is the Five Safes framework? [Accessed 1 September 2021]. Available from: https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/.

UK National Screening Committee 2018. Bowel Cancer. [Accessed 1 September 2021]. Available from: https://view-health-screening-

recommendations.service.gov.uk/bowel-cancer/.

University College London 2021a. CeLSIUS. [Accessed 29 May 2021]. Available from: https://www.ucl.ac.uk/epidemiology-health-care/research/epidemiology-and-public-health/research/health-and-social-surveys-research-group/studies-10.

University College London 2021b. CeLSIUS data dictionary. [Accessed 29 May 2021]. Available from: https://www.ucl.ac.uk/infostudies/silva-php-resources/researchProjects/celsius/standalone/index.php.

von Wagner, C., Baio, G., Raine, R., Snowball, J., Morris, S., Atkin, W., Obichere, A., Handley, G., Logan, R.F., Rainbow, S., Smith, S., Halloran, S. and Wardle, J. 2011. Inequalities in participation in an organized national colorectal cancer screening programme: Results from the first 2.6 million invitations in England. *International Journal of Epidemiology*. **40**(3), pp.712–718.

Walsh, D., McCartney, G., Minton, J., Parkinson, J., Shipton, D. and Whyte, B. 2020. Changing mortality trends in countries and cities of the UK: a population-based trend analysis. *BMJ Open*. [Online]. **10**(11), p.e038135. Available from: https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2020-038135.

Walters, S., Quaresma, M., Coleman, M.P., Gordon, E., Forman, D. and Rachet, B. 2011. Geographical variation in cancer survival in England, 1991-2006: An analysis by cancer network. *Journal of Epidemiology and Community Health*. **65**(11), pp.1044–1052.

White, A., Ironmonger, L., Steele, R.J.C., Ormiston-Smith, N., Crawford, C. and Seims, A. 2018. A review of sex-related differences in colorectal cancer incidence, screening uptake, routes to diagnosis, cancer stage and survival in the UK. *BMC Cancer*. **18**(1).

Whynes, D.K., Frew, E.J., Manghan, C.M., Scholefield, J.H. and Hardcastle, J.D. 2003. Colorectal cancer, screening and survival: the influence of socio-economic deprivation. *Public Health*. **117**(6), pp.389–395.

Wilkins, E., Morris, M., Radley, D. and Griffiths, C. 2019. Methods of measuring associations between the Retail Food Environment and weight status: Importance of classifications and metrics. *SSM - Population Health*. [Online]. **8**(February), p.100404. Available from: https://doi.org/10.1016/j.ssmph.2019.100404.

Williams, N., Hughes, L.J., Turner, E.L., Donovan, J.L., Hamdy, F.C., Neal, D.E., Martin, R.M. and Metcalfe, C. 2011. Prostate-specific antigen testing rates remain low in UK general practice: A cross-sectional study in six English cities. *BJU International*. **108**(9), pp.1402–1408.

Woods, L.M., Rachet, B. and Coleman, M.P. 2006. Origins of socio-economic inequalities in cancer survival: A review. *Annals of Oncology*. **17**(1), pp.5–19.

Woods, L.M., Rachet, B., Ellis, L. and Coleman, M.P. 2012. Full dates (day, month, year) should be used in population-based cancer survival studies. *International Journal of Cancer*. **131**(7).

World Cancer Research Fund 2018. Continuous Update Project Findings & Reports. [Accessed 22 June 2018]. Available from:

https://www.wcrf.org/dietandcancer/cancers.

World Health Organisation 2019. International Statistical Classifcation of Diseases and Related Health Problems 10th Revision. [Accessed 1 November 2021]. Available from: https://icd.who.int/browse10/2019/en.

Wrigley, H. 2003. Inequalities in survival from colorectal cancer: a comparison of the impact of deprivation, treatment, and host factors on observed and cause specific survival. *Journal of Epidemiology & Community Health*. [Online]. **57**(4), pp.301–309. Available from: http://jech.bmj.com/cgi/doi/10.1136/jech.57.4.301.

Zhang, D., Matthews, C.E., Powell-Wiley, T.M. and Xiao, Q. 2018. Ten-year change in neighborhood socioeconomic status and colorectal cancer. *Cancer*. [Online], pp.1–8. Available from: http://doi.wiley.com/10.1002/cncr.31832.

Appendix A

UNIVERSITY OF LEEDS RESEARCH ETHICS COMMITTEE APPLICATION FORM [1]

UNIVERSITY OF LEEDS

Please read each question carefully, taking note of instructions and completing all parts. If a question is not applicable please indicate so. The superscripted numbers (eg[8]) refer to sections of the guidance notes, available at http://ris.leeds.ac.uk/UoLEthicsApplication. Where a question asks for information which you have previously provided in answer to another question, please just refer to your earlier answer rather than repeating information.
Information about research ethics training courses: http://ris.leeds.ac.uk/EthicsTraining.

To help us process your application enter the following reference numbers, if known and if applicable:

Ethics reference number:	
Student number and/ or grant reference:	**200266806**

PART A: Summary

A.1 Which Faculty Research Ethics Committee would you like to consider this application?[2]

- ○ Arts, Humanities and Cultures (AHC)
- ○ Biological Sciences (BIOSCI)
- ○ Business, Environment and Social Sciences (AREA)
- ○ FS&N, Engineering and Physical Sciences (EPS)
- ○ Medicine and Health (Please specify a subcommittee):
  - ○ School of Dentistry (DREC)
  - ○ School of Healthcare (SHREC)
  - ◉ School of Medicine (SoMREC)
  - ○ School of Psychology (SoPREC)

A.2 Title of the research[3]

Spatial and social variations in mortality by colorectal cancer in comparison with other causes of death in England and Wales

A.3 Principal investigator's contact details[4]	
Name *(Title, first name, surname)*	Miss Charlotte Sturley
Position	Postgraduate Researcher
Department/ School/ Institute	School of Medicine
Faculty	Faculty of Medicine and Health
Work address *(including postcode)*	Leeds Institute for Data Analytics, Level 11 Worsley Building, Clarendon Way, Leeds, LS2 9NL
Telephone number	0113 34 30941
University of Leeds email address	gy06ces@leeds.ac.uk

A.4 Purpose of the research:[5] (Tick as appropriate)

- ☐ Research
- ☑ Educational qualification: ***Please specify:*** *_____PhD_____*
- ☐ Educational Research & Evaluation[6]
- ☐ Medical Audit or Health Service Evaluation[7]
- ☐ Other

A.5 Select from the list below to describe your research: (You may select more than one)

- ☐ Research on or with human participants
- ☐ Research which has potential adverse environmental impact.[8] ***If yes, please give details:***

  _____

- ☑ Research working with data of human participants
  - ☐ New data collected by qualitative methods
  - ☐ New data collected by quantitative methods
  - ☐ New data collected from observing individuals or populations
  - ☑ Routinely collected data or secondary data
  - ☑ Research working with aggregated or population data
  - ☐ Research using already published data or data in the public domain
- ☐ Research working with human tissue samples (*Please inform the relevant Persons Designate if the research will involve human tissue)*[9]

A.6 Will the research involve NHS staff recruited as potential research participants (by virtue of their professional role) or NHS premises/ facilities?

☐ Yes ☑ No

If yes, ethical approval must be sought from the University of Leeds. Note that approval from the NHS Health Research Authority may also be needed, please contact FMHUniEthics@leeds.ac.uk for advice.

A.7 Will the research involve any of the following:[10] (You may select more than one)

If your project is classified as research rather than service evaluation or audit and involves any of the following an application must be made to the NHS Health Research Authority via IRAS www.myresearchproject.org.uk as NHS ethics approval will be required. ***There is no need to complete any more of this form****. Further information is available at http://ris.leeds.ac.uk/NHSethicalreview and at http://ris.leeds.ac.uk/HRAapproval.*
You may also contact ***governance-ethics@leeds.ac.uk*** *for advice.*

- ☐ Patients and users of the NHS (including NHS patients treated in the private sector)[11]
- ☐ Individuals identified as potential participants because of their status as relatives or carers of patients and users of the NHS
- ☐ Research involving adults in Scotland, Wales or England who lack the capacity to consent for themselves[12]
- ☐ A prison or a young offender institution in England and Wales (and is health related)[14]
- ☐ Clinical trial of a medicinal product or medical device[15]
- ☐ Access to data, organs or other bodily material of past and present NHS patients[9]

- ☐ Use of human tissue (including non-NHS sources) where the collection is not covered by a Human Tissue Authority licence[9]
- ☐ Foetal material and IVF involving NHS patients
- ☐ The recently deceased under NHS care
- ☑ None of the above

You must inform the Research Ethics Administrator of your NHS REC reference and approval date once approval has been obtained.

The HRA decision tool to help determine the type of approval required is available at http://www.hra-decisiontools.org.uk/ethics. If the University of Leeds is not the Lead Institution, or approval has been granted elsewhere (e.g. NHS) then you should contact the local Research Ethics Committee for guidance. The UoL Ethics Committee needs to be assured that any relevant local ethical issues have been addressed.

A.8 Will the participants be from any of the following groups? (Tick as appropriate)

- ☑ Children under 16[16] ***Specify age group:*** ___0-16_____
- ☐ Adults with learning disabilities[12]
- ☐ Adults with other forms of mental incapacity or mental illness
- ☐ Adults in emergency situations
- ☐ Prisoners or young offenders[14]
- ☐ Those who could be considered to have a particularly dependent relationship with the investigator, eg members of staff, students[17]
- ☐ Other vulnerable groups
- ☐ No participants from any of the above groups

Please justify the inclusion of the above groups, explaining why the research cannot be conducted on non-vulnerable groups.

The research uses secondary data on all death registrations for all ages in England and Wales. Since the data is secondary, collected as part of the death certification and registration processes, people are not 'participants' as such.

It is the researcher's responsibility to check whether a DBS check (or equivalent) is required and to obtain one if it is needed. See also http://ris.leeds.ac.uk/healthandsafetyadvice and http://www.homeoffice.gov.uk/agencies-public-bodies/dbs.

A.9 Give a short summary of the research[18]

*This section must be completed in **language comprehensible to the lay person**. Do not simply reproduce or refer to the protocol, although the protocol can also be submitted to provide any technical information that you think the ethics committee may require. This section should cover the main parts of the proposal.*

This research will analyse trends in colorectal (bowel) cancer mortality in England and Wales by age, sex, deprivation level and geographic area over a 25 year period (1993-2018). This will enable variations in mortality rates by demographic, socio-economic and geographic characteristics to be identified along with how these relationships have changed over time. Trends in colorectal cancer mortality will be compared to that of other types of cancer and broad causes of death to identity similarities and differences in their mortality trajectories.

The study will utilise secondary data from the Office of National Statistics (ONS) containing individual death registrations in England and Wales. The dataset covers all registered deaths in England and Wales between 1993 and 2018. Variables contained in the dataset include age, sex, postcode of usual residence and cause of death. Based on the postcode, additional area-based measures of deprivation (Townsend deprivation score) and rurality (ONS 2011 rural-urban classification) will be linked to the death registrations.

Directly standardised mortality rates will be calculated by sex, deprivation quintile, geographic area, rural-urban

classification and cause of death for each year between 1993 and 2018. The mortality rates by geographic area will be mapped and spatial cluster analysis carried out to identify whether there are statistically significant geographic clusters of high or low mortality rates.

The research will provide up to date colorectal cancer mortality trends by geographic area and socio-economic deprivation. These can be used to inform area-based public health initiatives and targeting of services.

The study is part of a PhD project investigating social and spatial variations in colorectal cancer in England and Wales funded by Cancer Research UK (CRUK) and based in the UK Colorectal Cancer Intelligence Hub at the University of Leeds.

A.10 What are the main ethical issues with the research and how will these be addressed?[19]
Indicate any issues on which you would welcome advice from the ethics committee.

The main ethical issues with the research are ensuring the data is securely stored and that individuals cannot be identified from the results.

The study will use secondary data of individual death registrations from the ONS which will be accessed via the UK Data Service (UKDS) Secure Lab. The Secure Lab will be accessed remotely via a secure web-based interface from a PC owned and managed by the University of Leeds. The data will never be downloaded and all analyses will be carried out within the Secure Lab environment.

Direct identifiers such as name and address are not contained within the data set, however, it is possible that the combination of age, sex, postcode and cause of death could enable individuals to be identified. Security measures will be observed when using the data, for example locking of the screen when leaving the PC unattended and not allowing screens to be visible to people other than Secure Lab users working on the same project. The security measures are outlined in the signed Secure Access User Agreement (attached). It is a requirement of the Secure Lab that users complete mandatory training and obtain ONS Accredited Researcher status. Charlotte Sturley and Paul Norman (who will also have access to the Secure Lab project) are both Accredited Researchers.

Access to individual death registrations is required due to problems encountered when using the mortality by cause data openly published by the ONS (on Nomisweb.co.uk). In the published data, small counts for areas below region level are rounded (counts of 0, 1 and 2 are rounded to 0, while counts of 3, 4 and 5 are rounded to 5). This led to problems when calculating directly standardised mortality rates for colorectal cancer due to some small counts of deaths from colorectal cancer by age group and geographic area. In order to address the problem of small numbers when calculating the mortality rates, the count of deaths will be aggregated by time period, age groups, geographic areas and/or deprivation quintiles. The unrounded counts are required to do this aggregation. A key consideration will be which geographic unit to use that will provide adequate detail for analysis while ensuring reliable results. Some experimentation will be required to identify the appropriate level of geography. This will involve calculation of mortality rates at both Local Authority (LA) and Middle Layer Super Output Area (MSOA).

To ensure that no individual can be identified from statistical results, they will be subject to a statistical disclosure control check by UK Data Service staff before they are released from the secure data environment. This includes a threshold rule that no cell within a table or graph should contain less than 10 individuals. Furthermore, only outputs of publishable quality that contain sufficient explanation of the results will be released.

PART B: About the research team

B.1 To be completed by students only[20]	
Qualification working towards (eg Masters, PhD)	PhD
Supervisor's name (Title, first name,	Dr Amy Downing

surname)	
Department/ School/ Institute	Leeds Institute of Medical Research (LIMR)
Faculty	Faculty of Medicine and Health
Work address (including postcode)	Leeds Institute for Data Analytics, Level 11 Worsley Building, Clarendon Way, Leeds, LS2 9NL
Supervisor's telephone number	0113 34 30308
Supervisor's email address	a.downing@leeds.ac.uk
Module name and number (if applicable)	

B.2 Other members of the research team (eg co-investigators, co-supervisors) [21]	
Name *(Title, first name, surname)*	Dr Michelle Morris
Position	Co-supervisor
Department/ School/ Institute	Leeds Institute of Medical Research (LIMR)
Faculty	Faculty of Medicine and Health
Work address *(including postcode)*	Leeds Institute for Data Analytics, Level 11 Worsley Building, Clarendon Way, Leeds, LS2 9NL
Telephone number	0113 34 30883
Email address	m.morris@leeds.ac.uk

Name (Title, first name, surname)	Dr Paul Norman
Position	Co-supervisor
Department/ School/ Institute	School of Geography
Faculty	Faculty of Environment
Work address (including postcode)	School of Geography, University of Leeds, Leeds, LS2 9JT
Telephone number	0113 34 38199
Email address	p.d.norman@leeds.ac.uk

Part C: The research

C.1 What are the aims of the study?[22] (Must be in language comprehensible to a lay person.)

The aims of this research are to:

- Describe trends in colorectal cancer mortality in England and Wales over a 25 year period (1993-2018) by geographical location and area-level socio-economic deprivation
- Compare colorectal cancer mortality trends to those of other types of cancer and broad causes of death
- Identify whether there are statistically significant hot spots of colorectal cancer mortality across England and Wales and whether these have changed over time

C.2 Describe the design of the research. Qualitative methods as well as quantitative methods should be included. (Must be in language comprehensible to a lay person.)

It is important that the study can provide information about the aims that it intends to address. If a study cannot answer the questions/ add to the knowledge base that it intends to, due to the way that it is designed, then wasting participants' time could be an ethical issue.

The research will use data from the Office for National Statistics (ONS) comprising individual death registrations in England and Wales between 1993 and 2018. The data set includes the age, sex, postcode of usual residence and cause of death for each individual. Area-based measures of deprivation (Townsend deprivation score) and rurality (ONS 2011 rural-urban classification) will be linked to the death registrations based on the individuals postcode.

Directly standardised colorectal cancer mortality rates (per 100,000 population) will be calculated for each year by sex, deprivation quintile, geographic area, rural-urban classification and cause of death. Additional aggregation, such as pooling data over multiple years and using broader age groups, may be required when calculating the mortality rates to handle small number problems which may arise when the number of deaths is split by sex, age group, cause of death and deprivation quintile. Spatial clustering analysis techniques will be used to identify whether there are statistically significant clusters of high or low colorectal cancer mortality rates in England and Wales and how these patterns have changed over time.

Mortality rates for other types of cancer and broad cause of death will be calculated and compared to those for colorectal cancer to identify similarities and differences in the mortality trajectories.

C.3 What will participants be asked to do in the study?[23] *(e.g. number of visits, time, travel required, interviews)*

N/A. The study involves the use of secondary data only.

C.4 Does the research involve an international collaborator or research conducted overseas?[24]

☐ Yes ☑ No

If yes, describe any ethical review procedures that you will need to comply with in that country:

Describe the measures you have taken to comply with these:

Include copies of any ethical approval letters/ certificates with your application.

C.5 Proposed study dates and duration

Research start date (DD/MM/YY): _01/11/2020_____ Research end date (DD/MM/YY): _30/11/2021_____

Fieldwork start date (DD/MM/YY): _N/A_____ Fieldwork end date (DD/MM/YY): _N/A_____

C.6. Where will the research be undertaken? (i.e. in the street, on UoL premises, in schools)[25]

On University of Leeds premises (Leeds Institute for Data Analytics).

RECRUITMENT & CONSENT PROCESSES

C.7 How will potential participants in the study be identified, approached and recruited?[26]
How will you ensure an appropriately convened sample group in order to meet the aims of the research? Give details for subgroups separately, if appropriate. How will any potential pitfalls, for example dual roles or potential for coercion, be addressed?

N/A. The research does not involve recruitment of participants.

C.8 Will you be excluding any groups of people, and if so what is the rationale for that?[27]
Excluding certain groups of people, intentionally or unintentionally may be unethical in some circumstances. It may be wholly appropriate to exclude groups of people in other cases

N/A. The research will use secondary data on all deaths registered in the study period.

C.9 How many participants will be recruited and how was the number decided upon?[28]
It is important to ensure that enough participants are recruited to be able to answer the aims of the research.

N/A. The research does not involve recruitment of participants.

If you have a formal power calculation please replicate it here.

Remember to include all advertising material (posters, emails etc) as part of your application

C10 Will the research involve any element of deception?[29]
If yes, please describe why this is necessary and whether participants will be informed at the end of the study.

No.

C.11 Will informed consent be obtained from the research participants?[30]
☐ Yes ☑ No
If yes, give details of how it will be done. Give details of any particular steps to provide information (in addition to a written information sheet) e.g. videos, interactive material. If you are not going to be obtaining informed consent you will need to justify this.

If participants are to be recruited from any of potentially vulnerable groups, give details of extra steps taken to assure their protection. Describe any arrangements to be made for obtaining consent from a legal representative.

Will research participants be provided with a copy of the Privacy Notice for Research? If not, explain why not.
Guidance is available at https://dataprotection.leeds.ac.uk/information-for-researchers.

☐ Yes ☑ No

The research uses secondary data.

***Copies of any written consent form, written information and all other explanatory material should accompany this application**. The information sheet should make explicit that participants can withdraw from the research at any time, if the research design permits. Remember to use meaningful file names and version control to make it easier to keep track of your documents.*
Sample information sheets and consent forms are available from the University ethical review webpage at http://ris.leeds.ac.uk/InvolvingResearchParticipants.

C.12 Describe whether participants will be able to withdraw from the study, and up to what point (eg if data is to be anonymised). If withdrawal is not possible, explain why not.
Any limits to withdrawal, eg once the results have been written up or published, should be made clear to participants in advance, preferably by specifying a date after which withdrawal would not be possible. Make sure that the information provided to participants (eg information sheets, consent forms) is consistent with the answer to C12.

N/A. The research uses secondary data on death registrations.

C.13 How long will the participant have to decide whether to take part in the research?[31]
It may be appropriate to recruit participants on the spot for low risk research; however consideration is usually necessary for riskier projects.

N/A. The research does not involve recruitment of participants.

C.14 What arrangements have been made for participants who might have difficulties understanding verbal explanations or written information, or who have particular communication needs that should be taken into account to facilitate their involvement in the research?[32] *Different populations will have different information needs, different communication abilities and different levels of understanding of the research topic. Reasonable efforts should be made to include potential participants who could otherwise be prevented from participating due to disabilities or language barriers.*

N/A. The research does not involve recruitment of participants.

C.15 Will individual or group interviews/ questionnaires discuss any topics or issues that might be sensitive, embarrassing or upsetting, or is it possible that criminal or other disclosures requiring action could take place during the study (e.g. during interviews or group discussions)?[33] *The information sheet should explain under what circumstances action may be taken.*

☐ Yes ☑ No *If yes, give details of procedures in place to deal with these issues.*

C.16 Will individual research participants receive any payments, fees, reimbursement of expenses or any other incentives or benefits for taking part in this research?[34]

☐ Yes ☑ No

If Yes, please describe the amount, number and size of incentives and on what basis this was decided.

RISKS OF THE STUDY

C.17 What are the potential benefits and/ or risks for research participants in both the short and medium-term?[35]

N/A. The research does not involve recruitment of participants.

C.18 Does the research involve any risks to the researchers themselves, or people not directly involved in the research? *Eg lone working*[36]

☐ Yes ☑ No

If yes, please describe: _____

Is a risk assessment necessary for this research?
If you are unsure whether a risk assessment is required visit http://ris.leeds.ac.uk/HealthAndSafetyAdvice or contact your Faculty Health and Safety Manager for advice.

☐ Yes ☑ No If yes, please include a copy of your risk assessment form with your application.

RESEARCH DATA

C.19 Explain what measures will be put in place to protect personal data. E.g. anonymisation procedures, secure storage and coding of data. Any potential for re-identification should be made clear to participants in advance.[37] *Please note that research data which appears in reports or other publications is not confidential, even if it is fully anonymised. For a fuller explanation see http://ris.leeds.ac.uk/ConfidentialityAnonymisation. Further guidance is available at http://ris.leeds.ac.uk/ResearchDataManagement.*

As described in section A.10, the data will be accessed via the UK Data Service Secure Lab. The Secure Lab is a secure, web-based interface which will be accessed remotely from a named PC at the University of Leeds. The data does not contain direct identifiers of individuals and strict security measures will be observed to prevent the disclosure of potentially identifiable information. The data will never be downloaded from the secure environment and any outputs are subject to statistical disclosure control checks before they can be released. The checks ensure no individuals can be identified from the research outputs.

C.20 How will you make your research data available to others in line with: the University's, funding bodies' and publishers' policies on making the results of publically funded research publically available. Explain the extent to which anonymity will be maintained. *(max 200 words)* Refer to http://ris.leeds.ac.uk/ConfidentialityAnonymisation and http://ris.leeds.ac.uk/ResearchDataManagement for guidance.

Results of the statistical analysis will be written up as a PhD thesis chapter and submitted for peer-review journal publication. The results will be subject to statistical disclosure control checks before they will be released from the secure research environment. The research data cannot be downloaded from the Secure Lab and will therefore not be made publically available.

C.21 Will the research involve any of the following activities at any stage (including identification of potential research participants)? (Tick as appropriate)

- ☐ Examination of personal records by those who would not normally have access
- ☐ Access to research data on individuals by people from outside the research team
- ☐ Electronic surveys, please specify survey tool: _____ (further guidance)
- ☐ Other electronic transfer of data
- ☐ Use of personal addresses, postcodes, faxes, e-mails or telephone numbers
- ☐ Use of audio/ visual recording devices (NB this should usually be mentioned in the information for participants)
- ☐ FLASH memory or other portable storage devices

Storage of personal data on, or including, any of the following:

- ☐ University approved cloud computing services
- ☐ Other cloud computing services
- ☐ Manual files
- ☐ Private company computers
- ☐ Laptop computers
- ☐ Home or other personal computers (not recommended; data should be stored on a University of Leeds server such as your M: or N: drive where it is secure and backed up regularly: http://ris.leeds.ac.uk/ResearchDataManagement.)

Unclassified and Confidential University data must be kept on the University servers or in approved cloud services such as Office 365 (SharePoint or OneDrive). The N: Drive or Office 365 should be used for the storage of data that needs to be shared. If Highly Confidential information is kept in these shared storage areas it must be encrypted. Highly Confidential data that is not to be shared should be kept on the M: Drive. The use of non-University approved cloud services for the storage of any University data, including that which is unclassified, is forbidden without formal approval from IT. Further guidance is available via http://ris.leeds.ac.uk/ResearchDataManagement.

C.22 How do you intend to share the research data? (Indicate with an 'X) *Refer to http://library.leeds.ac.uk/research-data-deposit for guidance.*

- ☐ Exporting data outside the European Union
- ☐ Sharing data with other organisations
- ☐ Publication of direct quotations from respondents
- ☐ Publication of data that might allow identification of individuals to be identified

- ☐ Submitting to a journal to support a publication
- ☐ Depositing in a self-archiving system or an institutional repository
- ☐ Dissemination via a project or institutional website
- ☐ Informal peer-to-peer exchange
- ☐ Depositing in a specialist data centre or archive
- ☐ Other, please state: _____.
- ☑ No plans to report or disseminate the data

C.23 How do you intend to report and disseminate the results of the study? (Indicate with an 'X) Refer to http://ris.leeds.ac.uk/ResearchDissemination and http://ris.leeds.ac.uk/Publication for guidance.

- ☐ Conference presentation
- ☑ Peer reviewed journals
- ☑ Publication as an eThesis in the Institutional repository
- ☐ Publication on website
- ☐ Other publication or report, please state: _____
- ☐ Submission to regulatory authorities
- ☐ Other, please state: _____.
- ☐ No plans to report or disseminate the results

C.24 For how long will data from the study be stored? Please explain why this length of time has been chosen.[38] Refer to the RCUK Common Principles on Data Policy and http://ris.leeds.ac.uk/info/71/good_research_practice/106/research_data_guidance/5.
***Students**: It would be reasonable to retain data for at least 2 years after publication or three years after the end of data collection, whichever is longer.*

N/A Research data will be accessed remotely via the UKDS Secure Lab and will not be stored locally at the University of Leeds.

_____ years, _____ months

CONFLICTS OF INTEREST

C.25 Will any of the researchers or their institutions receive any other benefits or incentives for taking part in this research over and above normal salary or the costs of undertaking the research?[39]

☐ **Yes** ☑ **No**

If yes, indicate how much and on what basis this has been decided

_____

C.26 Is there scope for any other conflict of interest?[40] *For example, could the research findings affect the any ongoing relationship between any of the individuals or organisations involved and the researcher(s)? Will the research funder have control of publication of research findings? Refer to http://ris.leeds.ac.uk/ConflictsOfInterest.*

☐ Yes ☑ No

If so, please describe this potential conflict of interest, and outline what measures will be taken to address any ethical issues that might arise from the research.

C.27 Does the research involve external funding? (Tick as appropriate)

☑ Yes ☐ No ***If yes, what is the source of this funding?*** _Cancer Research UK_____

NB: If this research will be financially supported by the US Department of Health and Human Services or any of its divisions, agencies or programmes please ensure the additional funder requirements are complied with. Further guidance is available at http://ris.leeds.ac.uk/FWAcompliance and you may also contact your FRIO for advice.

PART D: Declarations

Declaration by Principal Investigators

1. The information in this form is accurate to the best of my knowledge and belief and I take full responsibility for it.
2. I undertake to abide by the University's ethical and health & safety guidelines, and the ethical principles underlying good practice guidelines appropriate to my discipline.
3. If the research is approved I undertake to adhere to the study protocol, the terms of this application and any conditions set out by the Research Ethics Committee (REC).
4. I undertake to seek an ethical opinion from the REC before implementing substantial amendments to the protocol.
5. I undertake to submit progress reports if required.
6. I am aware of my responsibility to be up to date and comply with the requirements of the law and relevant guidelines relating to security and confidentiality of patient or other personal data, including the need to register when necessary with the University's Data Protection Controller (further information available via http://ris.leeds.ac.uk/ResearchDataManagement).
7. I understand that research records/ data may be subject to inspection for audit purposes if required in future.
8. I understand that personal data about me as a researcher in this application will be held by the relevant RECs and that this will be managed according to the principles established in the Data Protection Act.
9. I understand that the REC may choose to audit this project at any point after approval.

Sharing information for training purposes: *Optional – please tick as appropriate:*

☐ I would be content for members of other Research Ethics Committees to have access to the information in the application in confidence for training purposes. All personal identifiers and references to researchers, funders and research units would be removed.

Principal Investigator:

Signature of Principal Investigator: ....C Sturley.....
(This needs to be an actual signature rather than just typed. Electronic signatures are acceptable)

Print name: ....Charlotte Sturley..... Date: (dd/mm/yyyy): ...22/09/2020.....

Supervisor of student research:

I have read, edited and agree with the form above.

Supervisor's signature: .A Downing.....
(This needs to be an actual signature rather than just typed. Electronic signatures are acceptable)

Print name: ...Amy Downing..... Date: (dd/mm/yyyy): .23/09/2020.....

Please submit your form by email to the FREC or School REC's mailbox.

Remember to include any supporting material such as your participant information sheet, consent form, interview questions and recruitment material with your application.

To help speed up the review of your application:

- ❑ Answer the questions in plain English, avoid using overly technical terms and acronyms not in common use.
- ❑ Answer all the questions on the form, including those with several parts (refer to the guidance if you're not sure how to answer a question or how much detail is required).
- ❑ Include any relevant supplementary materials such as
  - ❑ Recruitment material (posters, emails etc)
  - ❑ Sample participant information sheet
  - ❑ Sample consent form. Include different versions for different groups of participants eg for children and adults, clearly indicating which is which.
  - ❑ Signed risk assessment (If you are unsure whether a risk assessment is required visit http://ris.leeds.ac.uk/HealthAndSafetyAdvice or contact your Faculty Health and Safety Manager for advice).

  Remember to include use version control and meaningful file names for the documents.
- ❑ If you are not going to be using participant information sheets or consent forms explain why not and how informed consent will be otherwise obtained.
- ❑ If you are a student it is essential that you discuss your application with your supervisor.
- ❑ Submit a signed copy of the application, preferably electronically. Students' applications need to be signed by their supervisors as well.

Appendix B

UK Data Service

DEA Research Project Accreditation Application

Public
17 December 2019
Version: 01.00

T +44 (0)1206 872143
E help@ukdataservice.ac.uk
ukdataservice.ac.uk

Background on changes to the form

The ONS is changing the way it deals with applications for controlled access to ONS data.

The UK Statistics Authority has issued a Research Code of Practice and Accreditation Criteria concerning the disclosure, processing, holding or use of personal information under the Digital Economy Act (DEA).

The form that follows is for **Project accreditation** that will be submitted to the new Research Accreditation Panel (RAP) to independently accredit projects. The project application form will be going online in 2020, but a Word copy will be used until the online version is available.

Completing the form

The new form found below replaces the UK Data Service's *ONS Research Project Application* form, version 02.00, November 2017.

The project lead should complete this, using the excellent guidance available for DEA Research Project Applications and send to the UK Data Service by email to secure.applications@ukdataservice.ac.uk.

Please ensure that the people listed in section 3 (Research team) *correspond exactly* with the members of your UK Data Service Project for which controlled data has been requested.

UK Statistics Authority

Research Project Accreditation Application

Chapter 5 of Part 5 of the Digital Economy Act (DEA) 2017 facilitates the linking and sharing of datasets held by public authorities (as defined in the Act) for research purposes, across the UK. The provisions aim to broaden the capacity of research to deliver direct and indirect public benefits, including the production of valuable new research insights about UK society and the economy. To provide clarity and transparency about how the research power will operate, the UK Statistics Authority has issued a Research Code of Practice and Accreditation Criteria concerning the disclosure, processing, holding or use of personal information under the DEA.

The Code sets out the criteria for the accreditation of processors, research projects and researchers and the Authority has established the Research Accreditation Panel (RAP) to independently accredit each of these.

You need to complete this form if you wish to apply for access to unpublished data for your research.

Guidance on completing this form can be found on the UK Statistics Authority website.

1 Application Type:

☒ Research project ☐ Exploratory analysis

2 Project Lead:

2.1 *Personal details and contact information of project lead.*

Last name*	Sturley		
First name*	Charlotte		
Middle names (if any)			
What is your role on the project?	☒ Accredited Researcher[1] ☐ Peer Reviewer[2] with access to secure data ☐ Peer Reviewer with access to cleared outputs only		
Do you have an Accredited Researcher (AR) Number?	Yes ☒	No ☐	AR Number: 32621

[1] *To become an Accredited Researcher, please complete the Accredited Researcher application form available on the UK Statistics Authority website.*

[2] A Peer Reviewer may be:

- A researcher who accesses data classified as personal information in a secure setting at a DEA Accredited Processor to validate previous analysis and who has access to cleared outputs from the research. The researcher will need to be accredited for this Peer Reviewer role (if not accredited, please submit separate Accredited Researcher application), or
- A researcher who only accesses cleared non-disclosive outputs (researcher does not require accreditation to perform this role).

2.2 *If you do not yet have an Accredited Researcher number, please complete the following table about yourself.*[3]

Organisation name	
Organisation address	
Organisation postcode	
Work telephone no.	
Work email	
Date of birth	

3 Researcher Team:

3.1 *If you are the Research Project Lead for the research project, please provide the names and details of all members of the team. Please add more tables if required.*

Last Name*	Norman		
First Name*	Paul		
Middle names (if any)			
What is their role on the project?	☒ Accredited Researcher ☐ Peer Reviewer with access to secure data ☐ Peer Reviewer with access to cleared outputs only		
Do you authorise this person to deputise for you as Project Lead?	Yes ☒		
Do they have an AR Number?	Yes ☒	No ☐	AR Number: 30745

3.2 *If the person named does not have an Accredited Researcher number, please complete the following section with their details.*

Organisation name	
Organisation address	
Organisation postcode	

[3] *To become an Accredited Researcher, please complete the Accredited Researcher application form available on the UK Statistics Authority website.*

* *Fields marked with an asterisk are mandatory and will be published in the record of DEA Accredited Researchers and their Accredited Research projects on the UK Statistics Authority website, to meet transparency requirements.*

Work telephone no.	
Work email	

4 Research Sponsor:

4.1 *Have you been commissioned to perform this research for another organisation?*

☐ Yes ☒ No

4.2 *If you answered 'Yes', please provide the details of this organisation below and the work details of your contact there.*

Sponsor Organisation name*	
Last Name of your main contact in that organisation	
First Name of your main contact in that organisation	
Contact's address	
Contact's postcode	
Contact's telephone no.	
Contact's email	

5 Title of the research project: *

Spatial and social variations in mortality by colorectal cancer in comparison with other causes of death in England and Wales

Estimated duration of research project:

Start Date: 01/08/2020	**Publish Date: 31/07/2021**

6 Research Theme:

6.1 *Select the theme below which best describes your research project.*

Births and mortality

6.2 *If asked to specify additional theme information, add it in this box*

7 **Abstract of the research project: ***

Include a short description of the project and its benefits, in no more than 100 words.

The research project will analyse colorectal cancer mortality trends in England and Wales by age, sex, geographic area, deprivation level and rurality over a 25 year period. Variations in mortality by demographic and socio-economic characteristics will be investigated along with how these relationships have changed over time.

Trends in colorectal cancer mortality will be compared to those for other types of cancer and broad causes of death to identify similarities and differences in their mortality trajectories.

The project will provide up-to-date information on trends and inequalities in colorectal cancer mortality which could inform area-based targeting of public health initiatives.

8 **Purpose of Research Project: ***

Provide a detailed description of the purpose for which the data are requested, describing the aims of the study/research in no more than 500 words. Where research is part of a larger programme please include details below.

The aim of the research project is to investigate variations in colorectal cancer mortality rates in England and Wales over time by geographic location, area-based deprivation and rurality.

Following a steady decline in mortality rates during the 20th Century in England and Wales, since 2010/11 improvements in mortality have slowed down or stalled for both men and women. This study will investigate whether this pattern persists for different causes of death, in particular colorectal cancer and how this compares to other types of cancer and broad causes of death.

Furthermore, the project will investigate whether colorectal cancer mortality trends vary by area type (including level of deprivation, rurality and population density) and whether inequalities in mortality have reduced or widened over time.

This research is part of a wider PhD project investigating spatial and social variations in colorectal cancer incidence, survival and mortality in England and Wales funded by Cancer Research UK and based in the UK Colorectal Cancer Intelligence Hub at the University of Leeds.

9 **Research Methodology:**

Provide details of the research protocol or methodology (e.g. data linkage or matching, web scraping etc) and how you intend to use the data, in no more than 1000 words.

Individual mortality records held by the ONS will be linked to Townsend deprivation scores, the ONS rural-urban classification and population density categories based on the small area of residence of each individual.

Directly standardised mortality rates will be calculated by sex, geographic area, deprivation quintile, rural-urban classification, population density and cause of death for each year between 1993 and 2018. The European Standard Population 2013 will be used as the standard. An innovation of this work will be to contrast the stratification of rates by deprivation and urban-rural types with trajectories of geographic areas grouped by classification techniques which will allow trends to emerge which otherwise might be masked by more conventional area groupings.

The mortality rates by sex, deprivation quintile, rural-urban classification and trajectory groups will be visualised on graphs and the rates by geographic area will be mapped. Hot spot analysis will be carried out to identify geographic clusters of high or low mortality rates. These will be analysed over both space and time to identify whether the location of clusters have changed over the time period.

A key consideration when calculating the mortality rates will be which geographic units to use that will provide adequate detail for analysis whilst ensuring reliability of results. Some experimentation will be required to identify the appropriate level of geography. This will include calculation of rates at both Local Authority and Middle Layer Super Output Area (MSOA) level. Additional techniques, such as pooling of data over multiple years and using broader age categories, could also be implemented to handle small number problems that may arise when the number of deaths is split by sex, age group deprivation quintile and cause of death.

10 Data Required[4]:

10.1 *List the title(s) and years/releases of the dataset(s)[5] you will use. Make sure you include enough detail to uniquely identify the data you want.*

Death Registrations in England and Wales, 1993-2018: Secure Access

10.2 *If you intend to bring in any data for your project, give details of the data including who the owner is and provide evidence that the owner has given permission for their data to be used by you for this research.*

[4] If you request access to ONS Longitudinal Study data sets or Civil Registration data sets you must complete the appropriate supplementary forms on the ONS website.
[5] If you request access to data within the ONS Secure Research Service you should refer to the Data Catalogue on the ONS website (and use the same names).

10.3 *Explain why access to legally protected (unpublished) data is needed. Please state what other data sources have been considered and why they are not sufficient for your purposes.*

Access to individual mortality records is required due to problems with small counts when using the mortality by cause data published on Nomisweb.

As part of the study we wish to investigate time trends in colorectal cancer (C18 Malignant neoplasm of colon and C19-C21 Malignant neoplasm of rectosigmoid junction, rectum and anus) compared with all cancers (C00-D48 Neoplasms) and all causes. The rounding to 0 and 5 in the published data leads to problems in calculating the mortality rates for colorectal cancer due to low counts by age group and geographic area.

10.4 *Does your project include any linking of data sources (as defined within the application guidance)?*

☒ Yes ☐ No

If yes, provide the following details below:

- *description of the data sources(s) to be linked;*
- *summary of the key variables;*
- *summary of the linking methodology; and*
- *the justification for the linking.*

Townsend deprivation scores and quintiles, population density categories and the ONS rural-urban classification for 2011 geographies are to be appended to mortality records based on each individual's small-area of residence. This will enable analysis of mortality trends by area deprivation and rurality/urbanity.

11 Software Required:

Select the software that you will require for your research project.

☐	Adobe Acrobat Reader DC	☐	Notepad++	☒	SPSS
☐	Anaconda 3	☒	QGis	☐	Spyder
☐	ArcGIS	☐	Qtconsole	☐	STATA
☐	Jupyter Notebook	☐	R for Windows	☐	Winzip
☐	Microsoft Office Professional Plus	☒	R Studio		
☐	ML-WIN	☐	SAS	☐	Other (please specify)

12 Ethics:

You will need to provide evidence of an ethics consideration for your research project.

Have you had ethical approval for this project from your organisation or elsewhere?

☐ Yes ☒ No

If no, you may wish to use and submit a completed UK Statistics Authority Data Ethics self-assessment form available on our website.

If yes, please give details of the ethical approval body and any caveats, attaching any relevant documentation.

13 Public Good:

13.1 Please describe how your research project will provide a public good. Complete all the sections that apply.

Public Good	Describe how this research project will provide this public good
Provide an evidence base for public policy decision-making	
Provide an evidence base for public service delivery	The research will provide evidence for area based targeting of initiatives to address health inequalities.
Provide an evidence base for decisions which are likely to significantly benefit the UK economy, society or quality of life of people in the UK	
To replicate, validate or challenge Official Statistics	
To replicate, validate or challenge existing research	
To significantly extend understanding of social or economic trends or events by improving knowledge or challenging widely accepted analyses	This study will provide up to date mortality trends by deprivation and rurality for specific causes of death.
To improve the quality, coverage or presentation of existing statistical information	

13.2 Have any risks to public benefit been identified? What are they and how have they been mitigated?

14 Duration of access:

What is your best estimate of the last time you will need access to the unpublished data?

Note: if applying for exploratory analysis, access will be granted for a maximum of 12 months.

31/07/2021

15 Location of data access:

It is a requirement of the Digital Economy Act that any linking or matching of data and the provision of access to the data are carried out by a DEA Accredited Processor.

15.1 If you require data to be linked, which processor service do you want to use to **prepare the data**. A list of DEA Accredited Processors is available on the UK Statistics Authority website.

*15.2 Which processor service do you want to use to **access the data** in a secure environment provided by a DEA Accredited Processor.* A list of DEA Accredited Processors is available on the UK Statistics Authority website.

UKDS Secure Lab

16 Publications:

In order to access unpublished data for research purposes, you must promise that your findings will be made publicly available. Once published, you must notify ONS Research Support about where that publication can be found. Exemptions may only be granted in exceptional circumstances with the approval of the Research Accreditation Panel.

Note: If you are applying for exploratory analysis, no publications are permitted.

***16.1** Do you wish to apply for an exemption from publishing your findings?*

☐ Yes ☒ No

If yes, skip to question 16.5.

***16.2** How do you intend to make your research available to the public?*

Results will be written up as a chapter in PhD thesis and for publication in a scientific journal.

16.3 *Which specific journals, websites or reports do you intend to use to publish this research?*

Health & Place; Demographic Research; or Journal of Epidemiology & Community Health

16.4 *What is your best estimate of the project publication date.*

31/07/2021

16.5 *What, if any, are the circumstances that mean you need an exemption from making your results publicly available?*

16.6 *Please indicate if you would be willing to work with ONS to showcase the impact of your research findings. Your response here will not affect whether your project is accredited.*

Yes ☒	No ☐

Relevant information in the fields marked * will be published on our website as a public record of all Accredited Researchers and their research projects.

Data is at the heart of what we do, so it's important to us that your data is held and used securely, responsibly and transparently. Read our privacy policy here.

Once completed, please return this application (and the UK Statistics Authority Data Ethics self-assessment form if needed) by email to:
secure.applications@ukdataservice.ac.uk

Appendix C

Office for National Statistics

Microdata Release Panel Research Project Application

1 Application Type:

☒ Full project application ☐ Exploratory analysis

2 Lead Researcher:

2.1 *Personal details and contact information of lead researcher*

Surname*	Sturley		
First Names*	Charlotte Elizabeth		
Do you have an Accredited Researcher (AR) Number?	Yes ☐	No ☒	AR Number:

2.2 *If you do not have an Accredited Researcher Number, please complete the following section.*[1]

Institution or Organisation	Leeds Institute for Data Analytics, University of Leeds
Full Address	Level 11 Worsley Building, University of Leeds, Leeds, LS2 9NL.
Telephone No	
Email	gy06ces@leeds.ac.uk
Date of Birth	04/05/1988
Nationality	British

3 Researcher Team:

3.1 *If you are leading a research team, please provide the names and details of all members of the team. Please add more tables if required.*

Surname*	Downing		
First Names*	Amy		
Do you have an AR Number?	Yes ☐	No ☒	AR Number:

[1] *To become an Accredited Researcher, please complete the Accredited Researcher application form on:*
https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme

3.2 *If the researcher does not have an Accredited Researcher Number, please complete the following section.*

Institution or Organisation	University of Leeds
Telephone No	0113 3430308
Email	a.downing@leeds.ac.uk
Date of Birth	
Nationality	British

Surname*	Norman		
First Names*	Paul		
Do you have an AR Number?	Yes ☒	No ☐	AR Number: ONSF20074

Surname*	Morris		
First Names*	Michelle		
Do you have an AR Number?	Yes ☐	No ☒	AR Number:

Institution or Organisation	University of Leeds
Telephone No	0113 3430883
Email	m.morris@leeds.ac.uk
Date of Birth	
Nationality	British

4 Research Sponsor:

4.1 *Are you carrying out this project on behalf of a third party organisation?*

☐ Yes ☒ No

4.2 *If you are working on behalf of a third party organisation, please provide the details of this organisation below.*

Sponsor	
Institution or Organisation*	
Relationship to Institution or Organisation	
Address	

Telephone No	
Email	

5 Title of the research proposal:*

Variations in bowel cancer survival by individual characteristics and area type.

Estimated duration of full research project:

Start Date: 01/05/2018	**Publish Date:** 31/08/2020

6 Research Theme:

Please select the theme below which best suits your research project.

- ☒ Births and Mortality
- ☐ Business Change, e.g. growth
- ☐ Crime and Justice
- ☐ Economic Output and Productivity
- ☐ Education and Skills
- ☐ Employment
- ☒ Health, Social Care and Wellbeing
- ☐ Labour Market
- ☐ Migration
- ☐ Personal and Household Finances
- ☐ Population
- ☐ Sector Specific (please specify):
- ☐ UK Economy
- ☐ Other (*please specify*):

7 Abstract of the research proposal:*

Please include a short description of the project and its benefits, in no more than 100 words.

This project will utilise data from the ONS Longitudinal Study to carry out a statistical analysis of variations in bowel cancer registrations and survival by individual characteristics (age, sex, education, employment, social grade, etc.) and area type over a long follow-up period. The analysis will compare bowel cancer registrations to the general population. This updates previously published research using the ONS LS.

8 Purpose of Research Proposal:

Please provide a detailed description of the purpose for which the data are requested, describing the aims of the study/research in no more than 500 words. Where research is part of a larger programme please include details below.

Bowel cancer is the third most common cancer in both men and women in the UK and overall is the second most common cause of cancer death. There are around 41,000 new cases of bowel cancer diagnosed in the UK each year and 16,000 deaths from the disease (Cancer Research UK, 2017). Variations in incidence and survival exist in relation to age, sex, socio-economic status and ethnicity as well as between geographic regions within the UK. This study aims to determine whether socio-economic factors are involved in the incidence of and survival from bowel cancer.

Previous research using the ONS Longitudinal Study (LS) by Brown et al. (1998) found the trend in socio-economic differentials in bowel cancer incidence was not consistent over time. In the period 1986-90, incidence was highest among women in IV/V (unskilled and semi-skilled manual occupations) and among men in social classes IIIN and IIIM (non- manual and manual skilled occupations). Using an earlier measure of social class, higher incidence was found among women in the non-manual classes. Continued research is therefore required to investigate whether the pattern of colorectal cancer incidence is changing.

This study will update research by Brown et al. (1998) by utilising the latest data from the ONS LS. The ONS LS links census and vital event information (including cancer registrations) to the records of study members across successive censuses. Thus it provides an ideal data source to investigate the relationship between socio-economic position and cancer survival over a long follow-up period.

Registrations of bowel cancer among LS members will be examined by individual and household measures of socio-economic position, including occupational social grade, housing tenure, car ownership, educational level and employment status. Individual-level indicators such as these are not routinely collected by healthcare providers so this type of linkage is only possible through studies such as the ONS LS. In addition, area-level indicators including area deprivation and geodemographic classification will be appended to LS members based on their place of residence. The characteristics of bowel cancer sufferers will be compared to those of the general population. The length of survival for bowel cancer sufferers by individual and area level indicators will be estimated.

This study is part of a PhD project investigating social and spatial variations in bowel cancer incidence and survival. The PhD is funded by Cancer Research UK via the Bowel Cancer Intelligence Unit at the University of Leeds. Further studies will explore alternative data sources to obtain additional information about individuals with bowel cancer. These include data from Public Health England relating to cancer registrations and hospital admissions, lifestyle information from the UK Biobank and new forms of data such as consumer and social media data. Measures of area type will be kept consistent with those used in the analysis of the ONS LS.

References:
Brown, J., Harding, S., Bethune, A. & Rosato, M. (1998) Longitudinal study of socio-economic differences in the incidence of stomach, colorectal and pancreatic cancers. Population Trends 94: 35-41.

9 Research Methodology:

Please provide details of the research protocol or methodology (e.g. data linkage or matching, web scraping etc) and how you intend to use the data, in no more than 500 words.

The ONS LS links cancer registrations and mortality data to the records of study members via the National Health Service Central Register. Records will be extracted for LS members with a diagnosis of colorectal cancer. As of April 2015 there were 14,000 unique cases of colorectal cancer within the ONS LS. In addition, deprivation scores and quintiles and area classification will be appended to the record of each LS member based on their place of residence using a lookup table.

Descriptive statistics will be produced detailing the number of bowel cancer cases and split by measures of socio-economic status (social class, housing tenure, education level etc.).

Logistic regression will be used (with death from colorectal cancer as the outcome of interest) to quantify the probability of dying from bowel cancer by socio-economic status and area type.

Survival analysis will be performed using the cox regression method (with time from data of diagnosis of colorectal cancer to death from colorectal cancer the outcome of interest). Hazard ratios will be calculated in order to compare the risk of death from colorectal cancer by socio-economic status and area type.

In addition, the impact of changes in individual and/or household characteristics and area type over time on colorectal cancer survival will be investigated. For example, the impact of social mobility (change in social class) and migration (change in geographical area).

10 Data Required[2]:

10.1 *Please provide the title(s) of the dataset(s)[3] and the releases required.*

ONS Longitudinal Study

10.2 *Please explain why access to legally protected data is needed. Please state what other data sources have been considered and why they are not sufficient for your purposes.*

---

[2] If the applicant requests access to ONS Longitudinal Study data sets or Civil Registration data sets they must complete the appropriate supplementary forms on the ONS website.
[3] If the applicant requests access to data within the Virtual Microdata Laboratory they should refer to the VML Data Catalogue on the ONS website for this section.
https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme

This study seeks to investigate variations in cancer survival by individual socio-economic status. Therefore individual-level data on LS members' socio-economic position is required.

The ONS LS is the only data source through which this cancer registration-survival work is possible. Cancer registry data and hospital episodes statistics data do not record detailed individual level data on the socio-economic characteristics of patients. Alternative large survey data sources were considered. The UK Biobank links individual participant records to national cancer registries, however, the follow-up period is short compared with the ONS LS and there are many fewer cases of colorectal cancer among Biobank participants. Furthermore, the UK Biobank has a bias geography so the data set may not be representative of the general population of the UK.

10.3 *Does your project proposal include any matching of data sources as defined within the application guidance?*

☐ Yes ☒ No

If yes, please provide the following details below:

- *a description of the data sources(s) to be matched to the ONS data;*
- *a summary of the key variables; and*
- *a summary of the matching methodology.*

10.4 *Does your project proposal include any linking of data sources as defined within the application guidance?*

☒ Yes ☐ No

If yes, please provide the following details below:

- *a description of the data sources(s) to be matched to the ONS data;*
- *a summary of the key variables;*
- *a summary of the matching methodology; and*
- *the justification for the linking.*

Townsend deprivation scores and quintiles are to be appended to each LS member based on the LSOA of their place of residence. This will enable analysis of cancer survival by area deprivation.

11 Ethics:

Does your project require ethics approval from your Organisation, Institution or Sponsor?

☐ Yes ☒ No

If yes, please give details of the applications, attaching any relevant documentation.

If no, your project may be considered by the ONS ethics committee prior to approval.

12 Public Good:

Please describe how your project meets the required benefit to public good, as outlined in section 3 of the Approved Researcher Scheme.

Public Benefit	How project will achieve public benefit
Provide an evidence base for public policy decision-making	
Provide an evidence base for public service delivery	The research will provide evidence of cancer incidence and survival patterns for sub-groups of the population which could inform demand for cancer treatment services.
Provide an evidence base for decisions which are likely to significantly benefit the UK economy, society or quality of life of people in the UK	
To replicate, validate or challenge Official Statistics	
To replicate, validate or challenge existing research	The study will update previously published research on socio-economic variations in the incidence of colorectal cancer using the ONS LS by Brown et al. (1998).
To significantly extend understanding of social or economic trends or events by improving knowledge or challenging widely accepted analyses	

To improve the quality, coverage or presentation of existing statistical information	

13 Duration of access:

Please indicate how long access to data is likely to be required.

Note: if applying for exploratory analysis, your access will granted for a maximum of 12 months only.

From: 01/05/2018 To: 31/08/2020

14 Location of data access:

Please state where you intend to access the data for research (e.g. Secure Research Service (SRS), Secure Lab etc).

Secure Research Service

15 Publications:

ONS expects that research undertaken through the Approved Researcher Scheme will be published, other than in exceptional circumstances.

Note: If applying for exploratory analysis, no publications are permitted.

***15.1** Do you have exceptional circumstances which will prevent you from publishing the results of your project once completed?*

☐ Yes ☒ No *If yes, skip to question 15.6.*

***15.2** How will you make the results of your research available?*

Results will be written up as a chapter in PhD thesis and for publication in a journal paper.

***15.3** Where do you plan to publish your analysis/results?*

Peer-reviewed journal.

15.4 *Please provide an estimated timescale for publication.*

Expected thesis submission 31/08/2020.

15.5 *Please outline any intended future use for products (such as linked or matched data sets or tools) produced as a result of the research, and how they will be accessed.*

15.6 *Please explain the exceptional circumstances for not publishing your results once the project is complete. Please note that refusing to publish your research outputs may result in a rejection of your project application.*

If the project is to be undertaken under the Approved Researcher scheme, then relevant information in the fields marked * will be published on the ONS website as a public record of all Accredited Researchers and their research projects.

Once completed, please return this application to ONS by email:
Research.Support@ons.gov.uk

Appendix D

UK Colorectal Cancer Intelligence Hub, project reference: (for Hub use)

Protocol Title: Developing an area-based index of colorectal cancer risk for England to investigate the association between neighbourhood environment and colorectal cancer incidence

Short Title: Neighbourhood environment and CRC risk

Project reference: (for Hub use) 0059

Version number: 0.5 Date: 24/05/2021

Project Type:

X	Select all that apply with an 'X'
X	Analysis
	New data linkage
X	Standard CORECT-R extract
	New derived variable(s)
	Commercial*

*If any pharmaceutical company, insurance company, or management consultancy is involved please inform the Hub directly at CRCHub@leeds.ac.uk

Principal investigator (PI)

Name	Charlotte Sturley
Institution	University of Leeds
Contact address	Leeds Institute for Data Analytics, University of Leeds LS2 9JT
Email address	gy06ces@leeds.ac.uk
Telephone number	0113 34 30941

Co-investigators

Name	Role within project	Institution

	(such as clinical oversight, data analysis, patient representation, etc.)	
Dr Amy Downing	PhD supervision	University of Leeds
Dr Michelle Morris	PhD supervision	University of Leeds
Dr Paul Norman	PhD supervision	University of Leeds

Charlotte Sturley and Dr Amy Downing will need data access.

Funder: Cancer Research UK

Plain language summary of proposed work:

It is estimated that over 50% of bowel cancer cases in the UK can be prevented. Previous research has tended to focus on individual behaviours which have been found to increase the risk of developing bowel cancer. These include being overweight, smoking, lack of exercise, drinking alcohol, lack of fibre in the diet, and eating processed meat. However, rather than being lifestyle 'choices', these factors may be the result of socio-economic factors, such as poverty or the environment. This project will take a local neighbourhood approach and look for links between the local area and the rate of bowel cancer. Factors such as the number of fast food outlets, health services and access to green spaces will all be considered. The aim is to produce an area-based index of bowel cancer risk to help inform efforts to prevent the disease.

Patient and public involvement and engagement (PPIE):

The applicant has presented proposals from her PhD studies to the Bowel Cancer Intelligence UK (BCI UK) Patient-Public Group (PPG) and they are very supportive of research aimed at understanding how bowel cancer varies by location and circumstance; the PPG see the value in identifying where cancer prevention and treatment can be improved. Preliminary findings will be presented to the PPG and their input will be sought on the most effective ways of disseminating the results

Background and rationale:

It is estimated that over 50% of colorectal cancer cases in the UK are preventable (Brown et al. 2018) based on population exposure to modifiable risk factors. The latest research estimates that preventable cases could be as high as 67% of cases in men and 60% of cases in women (Goon et al. 2021).

A number of risk factors are associated with colorectal cancer. The International Agency for Research on Cancer and World Cancer Research Fund classify the following factors as having a 'sufficient' or 'convincing' evidence of a causal association with colorectal cancer risk:

cigarette smoking, alcoholic drinks, body fatness/BMI, processed meat, insufficient dietary fibre, and insufficient physical activity (IARC, WCRF, Lauby-Secretan et al. 2016).

Differences in the proportion of cancer cases attributed to modifiable risk factors have been observed between countries in the UK (Brown et al., Goon et al.) due to different prevalence of exposure to risk factors. This in turn is affected by demographic differences, such as higher levels of socio-economic deprivation, which drive differences in exposure to theoretically avoidable lifestyle factors. Furthermore, many lifestyle 'choices' are driven by environmental and social factors, such as food availability and pricing, and susceptibility to these factors varies by socio-economic position (McGill et al. 2015).

While previous research has focused on individual-level exposure to modifiable cancer risk factors, this study will investigate how the local 'neighbourhood' environment is associated with exposure to these risk factors. The broader geographical determinants of health have been considered within the health geography literature. The location of different types of retail outlets is one area of interest. For example, research has shown a positive relationship between the number of fast food outlets in an area and rates of childhood obesity (Fraser and Edwards, 2010). Other studies have shown the density of alcohol-related outlets to be associated with alcohol-related harms (Sherk et al. 2018) and tobacco outlets with risk of smoking (Shortt, et al. 2016). Studies have also investigated the impact of the natural environment on health. For example, accessibility to green space has been shown to be positively associated with physical and mental wellbeing (Mitchell and Popham, 2008).

Research on the role of the neighbourhood environment and colorectal cancer risk is limited. Studies examining neighbourhood-level factors have tend to include only socio-economic status. A recent cohort study by Canchola et al. (2017) in the US investigated the association between a set of neighbourhood obesogenic attributes (socio-economic status, population density, restaurant and retail food environments, numbers of recreational facilities and businesses, commute patterns, traffic density and street connectivity) and risk of CRC. A modest association between higher traffic density and increased CRC risk was found, possibly indicating a less walkable environment which could have a negative impact on physical activity, but no other associations were observed. A similar study based on the same cohort (Schetsov et al. 2020) found that changes over time in neighbourhood environment attributes were associated with the risk of colorectal cancer. A recent ecological analysis by Gibson et al. (2020) examined the relationship between the local food environment and colorectal cancer incidence in Texas. Colorectal cancer incidence was not strongly associated with census tracts with higher unhealthy food availability, but the study did not examine other aspects of the built environment.

This study will investigate the relationship between the neighbourhood environment and CRC incidence in a UK context. Neighbourhood environment attributes will be sourced from the Access to Healthy Assets and Hazards (AHAH) index (open source data). It is important to understand this relationship in order to prioritise and target interventions and policies to reduce cancer incidence.

Study aim:

1. To investigate the association between the neighbourhood environment and CRC risk in England.
2. To develop an area-based CRC risk index.

Study objectives

Objective	Description
1	Describe the study population by the characteristics of the local area in which they reside.
2	Develop an area-based index at Lower Layer Super Output Area (LSOA) level of environment attributes associated with CRC risk, encompassing accessibility to health-related features in three domains: the retail environment, health services, and the physical environment.
3	Examine the association between CRC incidence and the CRC risk index (and sub-domains) at area (LSOA) level.
4	Model the relationship between the number of CRC cases per LSOA and the CRC risk index (and sub-domains) adjusting for demographic and socioeconomic characteristics of the areas.

Study design:

The study is a retrospective observational study using population-based data available from the UK Colorectal Cancer Intelligence Hub's COloRECTal cancer data Repository (CORECT-R).

Methodology & data:

Data covering the period 1 January 2000 – 31 December 2018 will be requested from the colorectal cancer data repository (CORECT-R) for all patients who have been diagnosed with ICD10 C18-21. These data will include patient characteristics (age, sex, ethnic group) and geographic information (LSOA, rurality, area-based socio-economic status). SEE DATA SPEC.

Incidence rates will be derived by combining counts of colorectal cancers by sex and age-band in each LSOA with population estimates for LSOAs from the Office for National Statistics (ONS).

Additional variables will be appended to the CORECT-R data based on LSOA. LSOAs are a small-area statistical geography (mean population 1500) routinely used by policy makers. Neighbourhood attributes will be sourced from the Access to Healthy Assets and Hazards (AHAH) index, an open source multi-dimensional index of the accessibility to health-related features of the environment for LSOAs across Great Britain (Green et al. 2018). Indicators identified as relevant to CRC risk in the literature will be selected across the AHAH index's three domains of accessibility.

Retail Environment:

- Accessibility to fast food outlets (processed meat consumption)

- Accessibility to off-licences (alcoholic drinks)
- Accessibility to tobacconists (cigarette smoking)

Health services:
- Accessibility to GP practices (diagnosis)
- Accessibility to leisure services (physical activity)

Environmental quality:
- Accessibility to green spaces (physical activity)

To meet objective 1 descriptive statistics will be calculated showing the number and proportion of those with CRC in the England by patient characteristics (age, sex, ethnic group, socio-economic status) and neighbourhood attributes (rurality, socio-economic status and health-related environment features listed above). The neighbourhood attributes will be categorised based on quintiles or deciles of their distribution across England.

To meet objective 2 variables from the AHAH index will be appended to the CORECT-R data at LSOA level. The variables will be combined to create three domain scores (retail domain, health domain and physical environment domain) as well as an overall index score using the same methodology adopted by Green et al. (2018) to create the AHAH index.

To meet objective 3 Pearson's correlation will be used to examine the association of CRC incidence by area to the overall index and each domain score. A series of choropleth maps will be produced to visualise the distribution of the risk index (and domains) for England.

To meet objective 4 Poisson regression models will be constructed to calculate the incidence rate ratio for colorectal cancer by quartile of the overall index score and of each domain score.

Project Timelines

The bulk of the analysis is likely to be completed within the next nine months but the end date reflects the possibility of thesis corrections and analysis in support of journal submissions.

Start date: May 2021 End date: December 2022

Intended outputs:

- Table(s) containing the area-based CRC risk index at LSOA level
- Table(s) of Pearson correlation coefficients of the risk index and domain scores to CRC incidence
- Choropleth maps – one for each domain of the risk index and one for the overall index
- Table(s) of incidence rate ratios of the risk index and domain scores
- The work will form a PhD chapter and associated journal article
- The work will be presented at health geography meetings/conferences

Is there potential to identify outlier providers?

No

Ethics:

The project is covered by the Establishing a UK Colorectal Cancer Intelligence Hub Research Ethics approval (18/SW/1034) granted by the South West – Central Bristol Research Ethics Committee on the 1st June 2018.

References:

Brown, K.F., Rumgay, H., Dunlop, C. et al. 2018. The fraction of cancer attributable to known risk factors in England, Wales, Scotland, Northern Ireland, and the UK overall in 2015. *British Journal of Cancer*. **118**(8), pp.1130-1141.

Goon, S., Kim, H. and Giovannucci, E.L. 2021. [Pre-print]. Population attributable risk for colorectal and breast cancer in England, Wales, Scotland, Northern Ireland, and the United Kingdom. *AMRC Open Research*.

Fraser, L. and Edwards, K. 2010. The association between the geography of fast food outlets and childhood obesity rates in Leeds, UK. *Health Place*. **16**(6), pp.1124-1128.

Green, M., Daras, K., Davies, A., Barr, B. and Singleton, A. Developing an openly accessible multi-dimensional small area index of 'Access to Healthy Assets and Hazards' for Great Britain, 2016. 2018. *Health and Place*. **54**, pp.11-19.

International Agency for Research on Cancer. List of classifications by cancer sites with sufficient or limited evidence in humans, Volumes 1 to 119. Accessed September 2017.

Lauby-Secretan B, Scoccianti C, Loomis D, et al. 2016. Body Fatness and Cancer - viewpoint of the IARC Working Group. *New England Journal of Medicine*. **375**(8), pp.794-798.

McGill, R. et al. 2015. Are interventions to promote healthy eating equally effective for all? Systematic review of socioeconomic inequalities in impact. *BMC Public Health*. **15**(457).

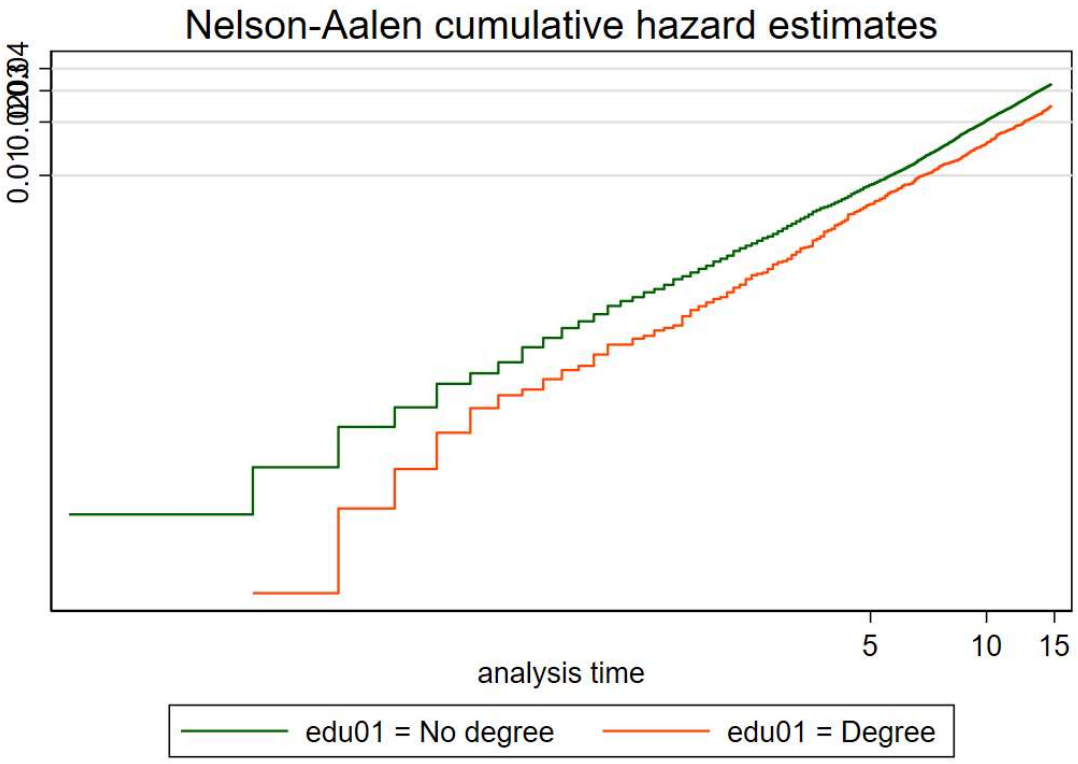
Mitchell, R., Popham, F., 2008. Effect of exposure to natural environment on health inequalities: an observational population study. *Lancet*. **372**(9650), pp.1655–1660.

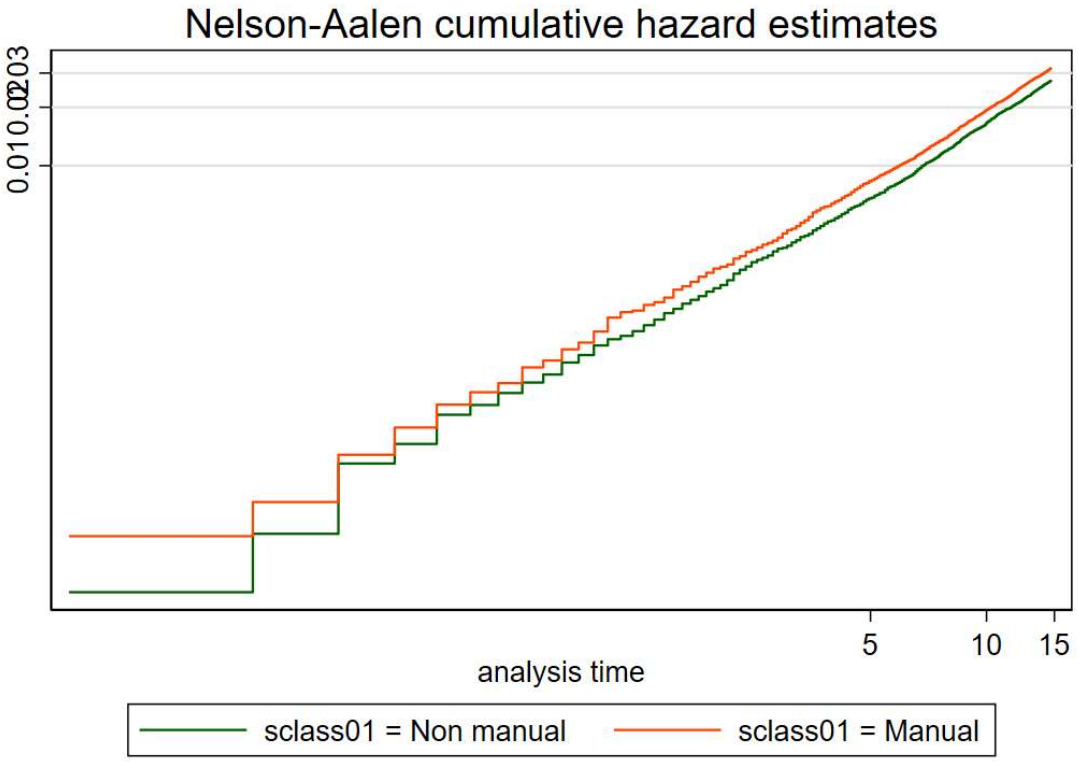
Sherk, A., et al., 2018. Alcohol consumption and the physical availability of take-away alcohol: systematic reviews and meta-analyses of the days and hours of sale and outlet density. *Journal of Studies on Alcohol and Drugs*. **79**(1), pp.58–67.

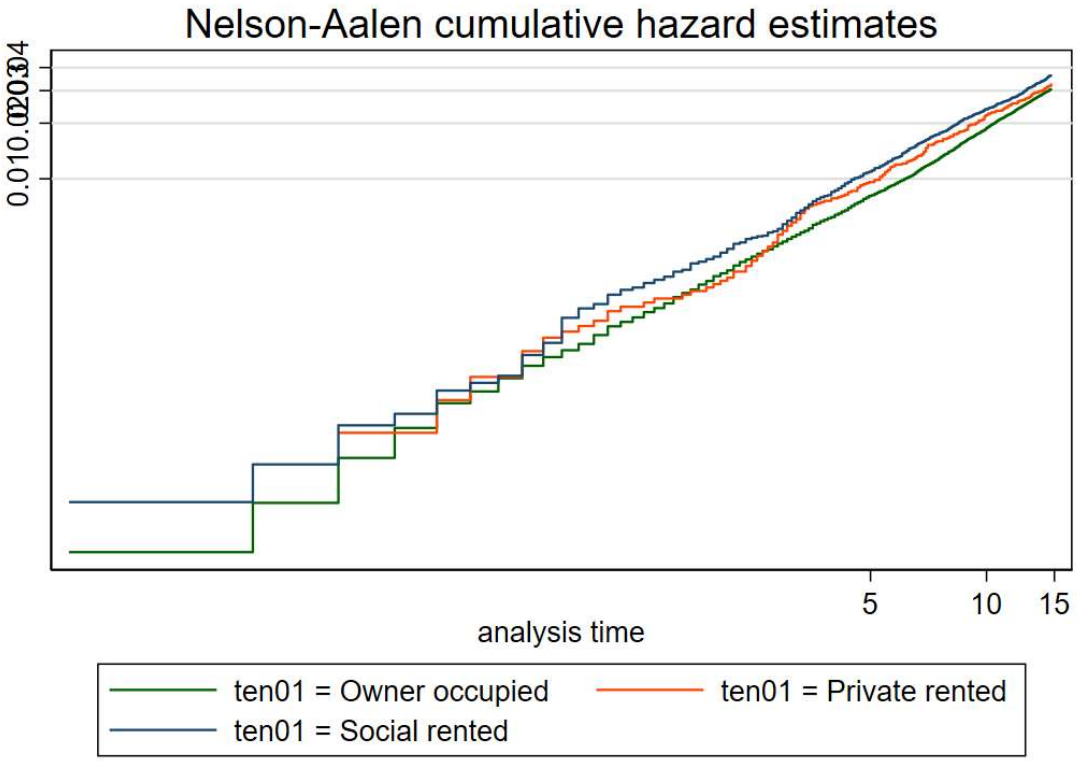
Shortt, N., et al., 2016. The density of tobacco retailers in home and school environments and relationship with adolescent smoking behaviours in Scotland. *Tobacco Control*. **25**, pp.75–82.

World Cancer Research Fund & American Institute for Cancer Research. 2018. Continuous Update Project. Diet, Nutrition, Physical Activity and the Prevention of Cancer. Summary of evidence. [Online]. [Accessed 1 April 2021]. Available from: wcrf.org/matrix

Appendix E


Nelson-Aalen cumulative hazard estimates
0.01
0.02
0.03
0.04
5
10
15
analysis time
edu01 = No degree
edu01 = Degree


Nelson-Aalen cumulative hazard estimates
0.01
0.02
0.03
5
10
15
analysis time
sclass01 = Non manual
sclass01 = Manual



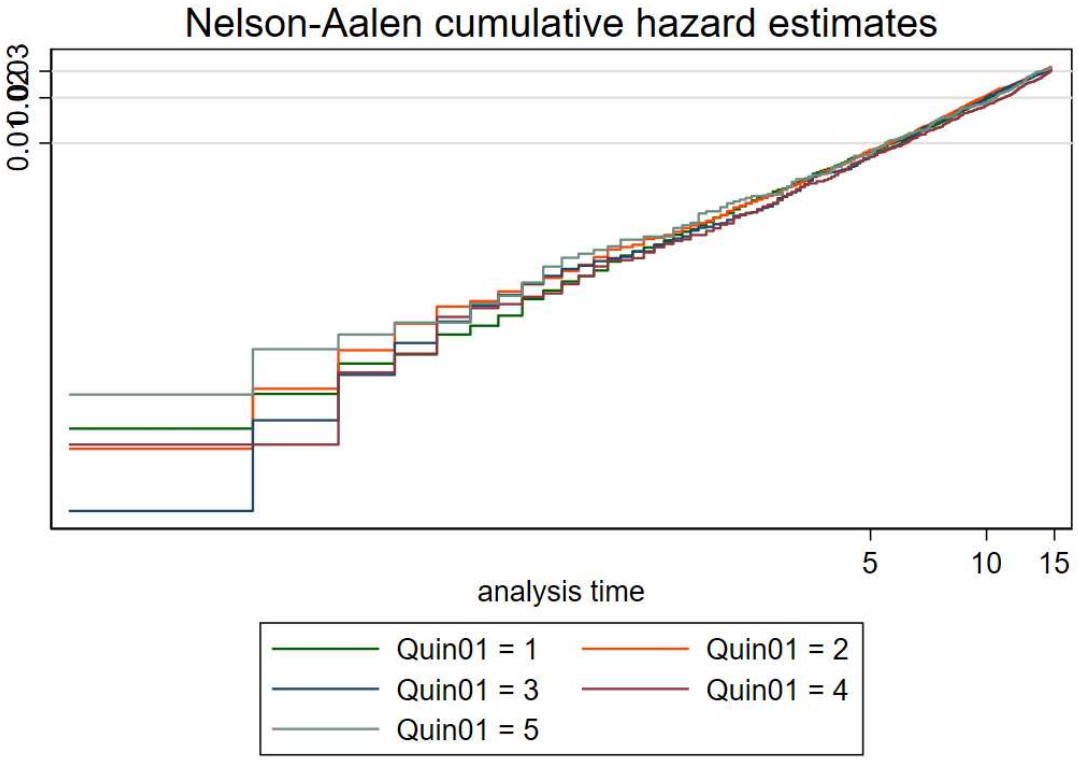
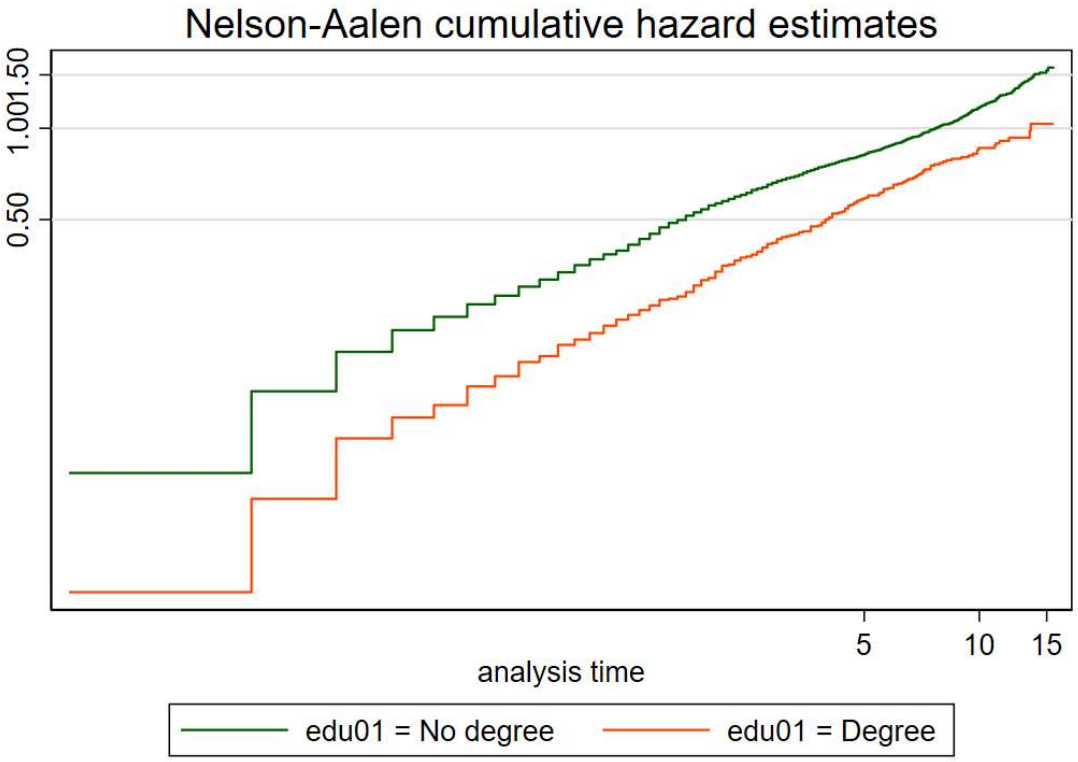
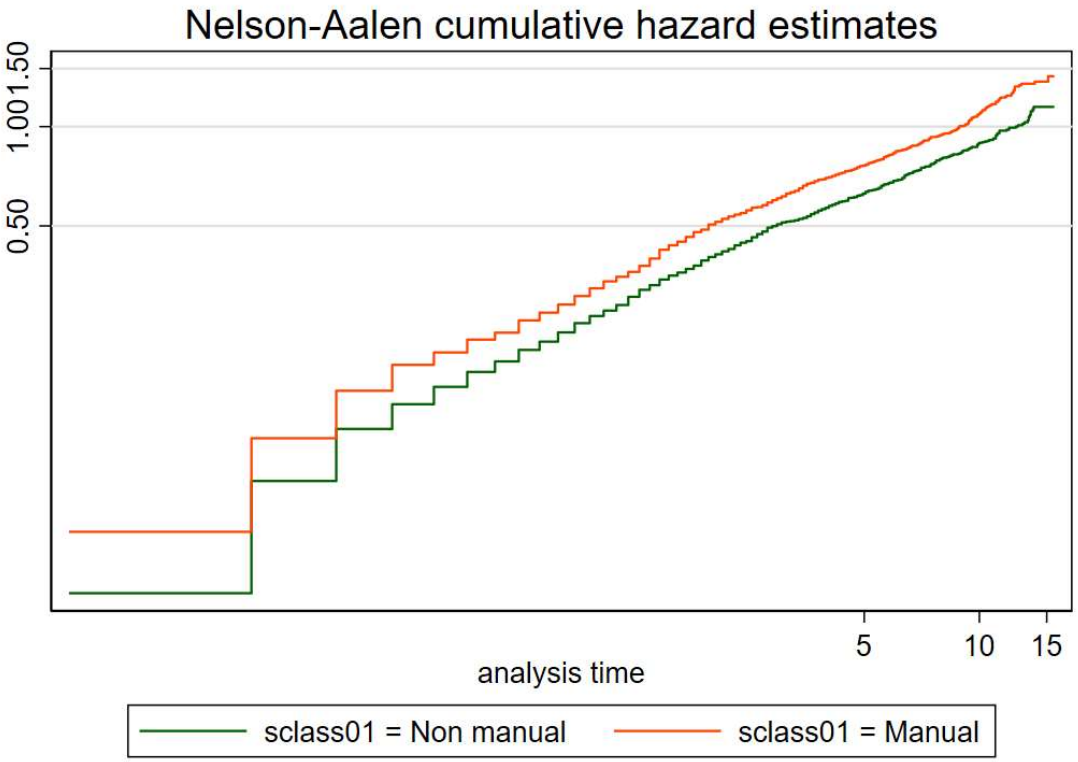
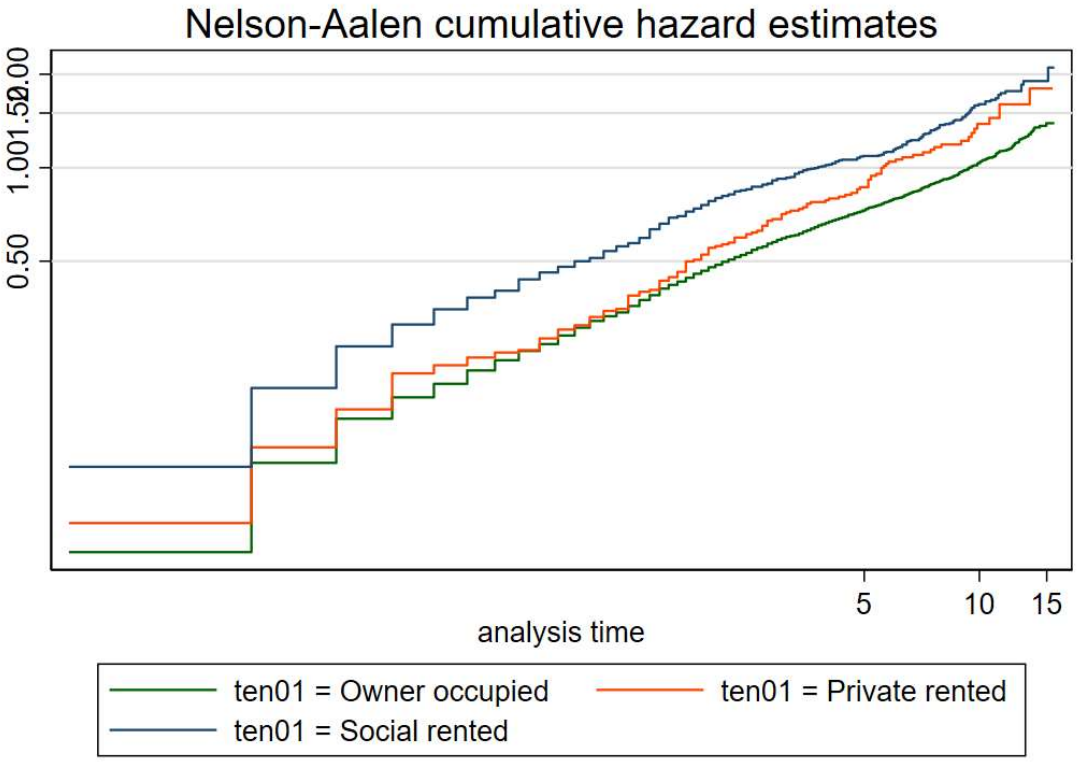


Figure E.1 Log-log plots of probability of colorectal cancer registration by: a) educational attainment; b) social class; c) housing tenure; d) area deprivation (Data source: ONS LS)


Nelson-Aalen cumulative hazard estimates
0.50
1.00
1.50
5
10
15
analysis time
edu01 = No degree
edu01 = Degree


Nelson-Aalen cumulative hazard estimates
0.50
1.00
1.50
5
10
15
analysis time
sclass01 = Non manual
sclass01 = Manual



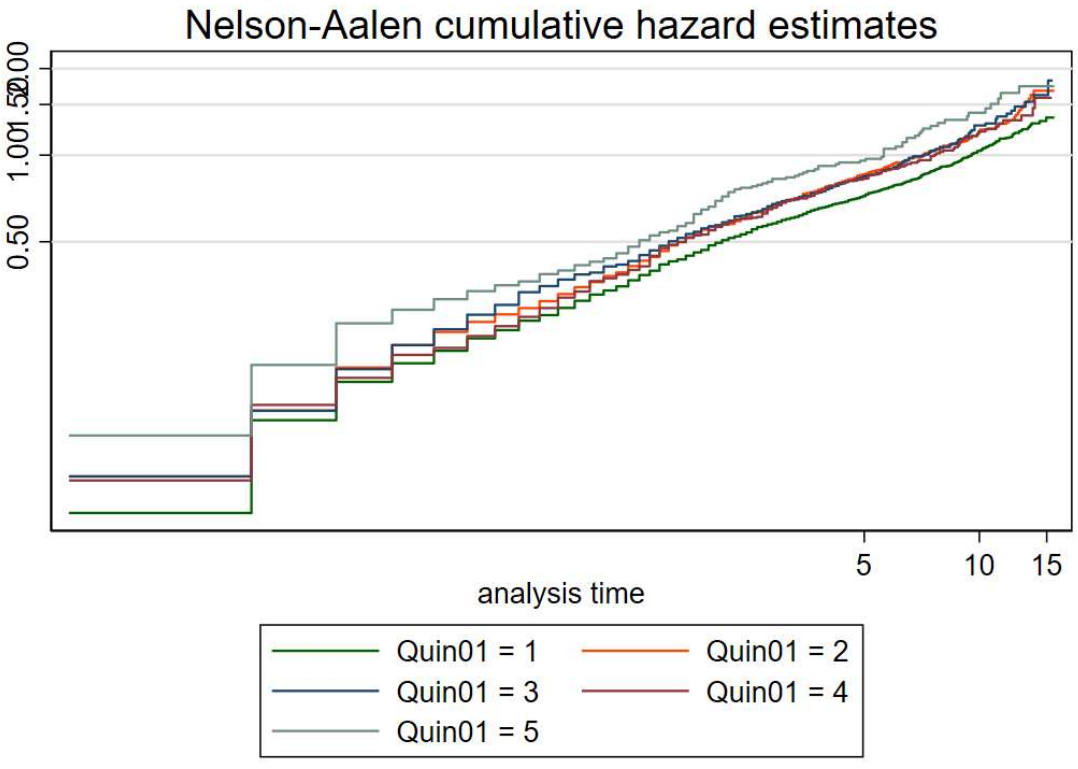
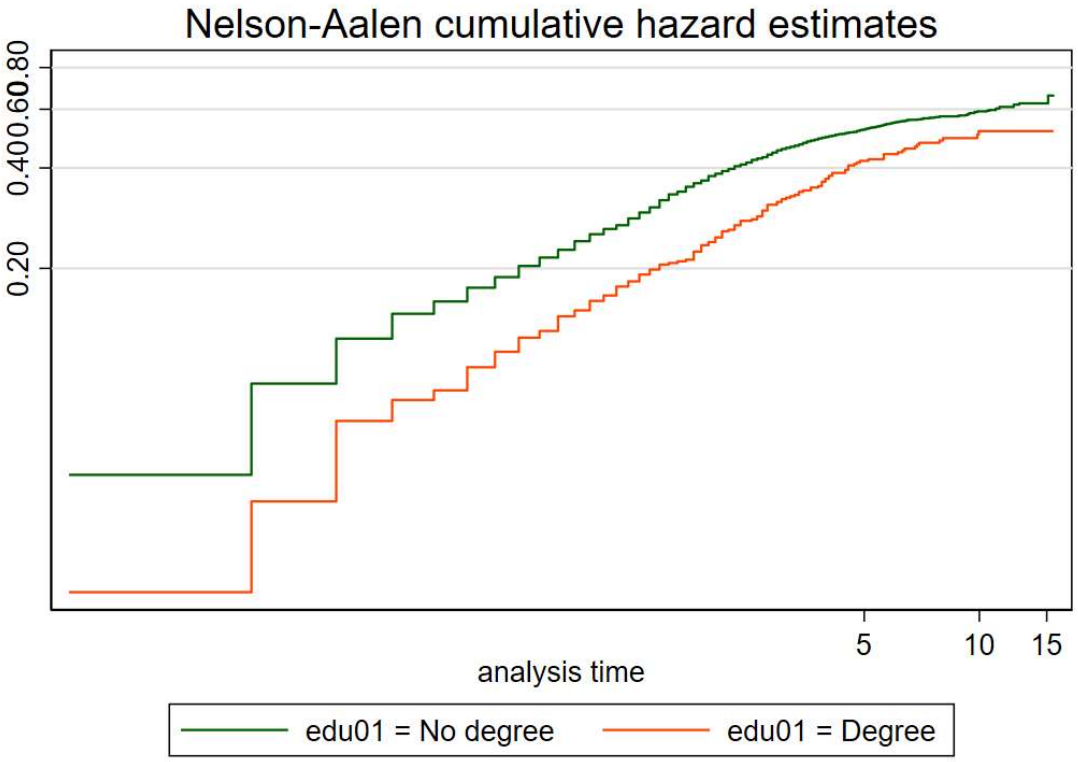
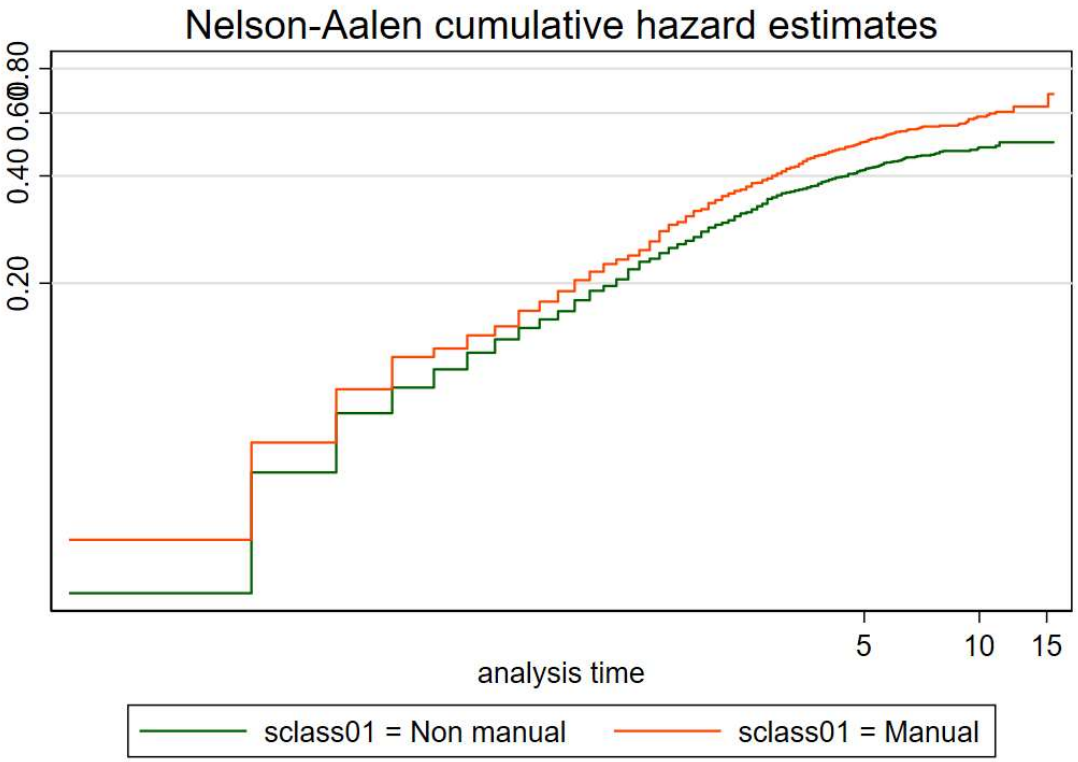
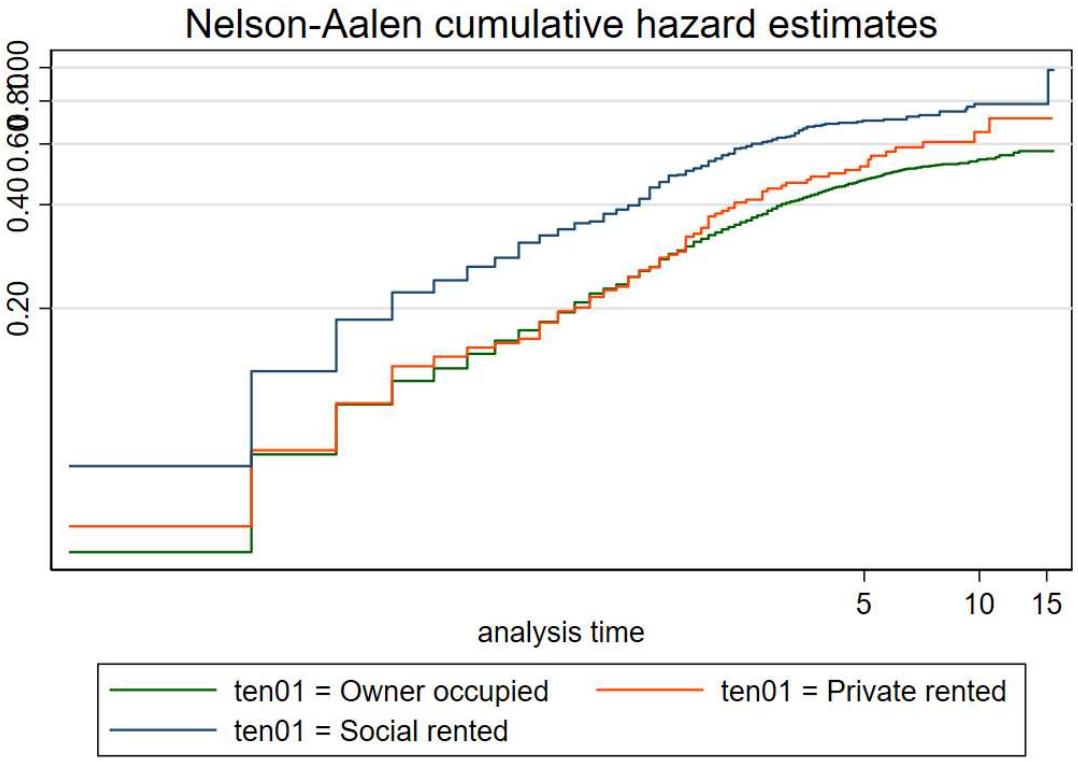


Figure E.2 Log-log plots for probability of death from all-causes by: a) educational attainment; b) social class; c) housing tenure; d) area deprivation (Data source: ONS LS)


Nelson-Aalen cumulative hazard estimates
0.20
0.40
0.60
0.80
5
10
15
analysis time
edu01 = No degree
edu01 = Degree


Nelson-Aalen cumulative hazard estimates
0.20
0.40
0.60
0.80
5
10
15
analysis time
sclass01 = Non manual
sclass01 = Manual



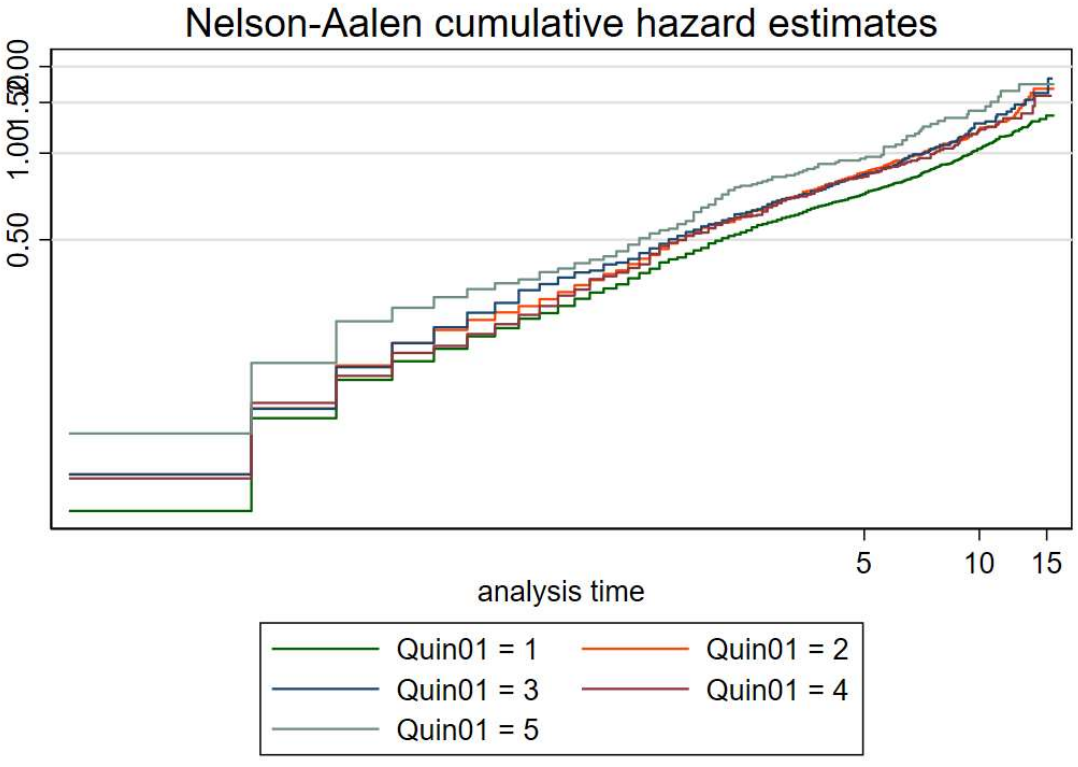


Figure E.3 Log-log plots for probability of colorectal cancer death by: a) educational attainment; b) social class; c) housing tenure; d) area deprivation (Data source: ONS LS)