Diffuse large B-cell lymphoma: Impact of Socio-Demographic and Geographical Factors on Routes to Diagnosis, Disease Presentation and Outcomes of Patients in the Haematological Malignancy Research Network

Grace Isabelle Collum

PhD

University of York

Health Sciences

June 2025

Abstract

Background

An association between travel to healthcare and subsequent route to diagnosis (RTD), treatment and survival has not been previously reported for diffuse large B-cell lymphoma (DLBCL) patients in the UK. This thesis aims to address this gap using data from the Haematological Malignancy Research Network (HMRN).

Methods

Firstly, a literature review was conducted which showed a lack of research investigating the impact of travel time to healthcare on outcomes of cancer patients in the UK, particularly those with haematological malignancies, and informed the methods used in this thesis. RTDs were assigned using Hospital Episode Statistics and travel distances and times were calculated for journeys by car and public transport between the patient's home postcode at diagnosis and healthcare facilities (registered GP and diagnostic hospital) using TRACC (Basemap). Hospital activity in the two-years leading up to diagnosis was also calculated and compared to background rates using a nested case-control study. Associations with RTD and treatment were examined using logistic regression and survival was compared using standard time-to-event analyses.

Results

Journeys to healthcare across HMRN were generally short, with a median time to hospital by car of 10.8 minutes, however, corresponding journeys by public transport were four times longer, at 43.1 minutes. A total of 1292 patients (31.4%) were diagnosed following an EP, suggesting that many patients experience diagnostic delay and this result was supported by the hospital activity analysis which showed that patients with DLBCL had elevated hospital activity for two years before diagnosis compared to their matched controls. No statistically significant associations between travel to healthcare and RTD, treatment intent or survival were detected.

Conclusions

These findings reassuringly suggest that in the HMRN region, travel to healthcare does not influence RTD or survival of DLBCL and that patients receive equitable care from the NHS regardless of how close they live to healthcare services.

Contents

List of Tables

List of Figures

List of Abbreviations

A&E	Accident and Emergency
AOR	Adjusted Odds Ratio
CI	Confidence Interval
COR	Crude Odds Ratio
CWT	Cancer Waiting Times
DCO	Death Certificate Only
DLBCL	Diffuse large B-cell lymphoma
GP	General Practitioner/General Practice
HES	Hospital Episode Statistics
HILIS	HMDS integrated laboratory information system
HMDS	Haematological Malignancy Diagnostic Service
HMRN	Haematological Malignancy Research Network
HR	Hazard Ratio
IQR	Interquartile Range
NCDR	National Cancer Data Repository
NCRAS	National Cancer Registration and Analysis Service
NICE	National Institute for Health and Care Excellence
NHL	Non-Hodgkin Lymphoma
NHS	National Health Service
NS	Net Survival
OR	Odds Ratio
OS	Overall Survival
PCNSL	Primary Central Nervous System Lymphoma
RTD	Route To Diagnosis
TWW	Two-Week Wait

Acknowledgements

Firstly, I would like to thank the patients whose data has contributed to this research, without which this would not have been possible. I would also like to thank Cancer Research UK for my studentship and Basemap who generously provided me their travel-time analysis tool free of charge.

I am incredibly grateful to my PhD supervisors, Professor Alexandra Smith, Professor Debra Howell and Dr Maxine Lamb, who have provided me with abundant support in developing the skills needed to complete this research. Their supervision has been consistent, kind and honest and I very much appreciate all the time they have so generously given me. In addition to my supervisors, I have received helpful advice, suggestions and thoughtful insights from my Thesis Advisory Panel members, Professor Eve Roman and Dr Charlotte Kelly, who I would like to extend this thanks to. I have been very lucky to be part of the Epidemiology and Cancer Statistics Group for the last four years, every person in the team kindly welcomed me and offered help when they could provide it without hesitation - I cannot thank them enough for everything they have done for me. A special thanks to Tim and Lauren for sharing their statistical knowledge and expertise, Will for his IT support, Dan for his help with SQL and my fellow PhD students James and Ned for their reassurance and camaraderie.

Finally, I owe a special thanks to my husband Dom, parents, extended family and friends whose love and encouragement has been unwavering and without whom I would not have had the confidence to even apply for this PhD, let alone be handing it in.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

1 Introduction

1.1 Haematological malignancies

1.1.1 What is a haematological malignancy?

Haematological malignancies, commonly referred to as blood cancers, have traditionally been grouped into three broad categories: leukaemias, lymphomas and myelomas. In lay terms these are cancers of the blood, lymph systems and bone marrow, however, this is a simplistic representation of a unique, complex and heterogenous group of cancers that can affect people of any age, although incidence increases with age and varies between sub-types (Smith et al., 2011). They also have the potential to transform from one sub-type to another, for example Richter's syndrome, where chronic lymphocytic leukaemia transforms into an aggressive lymphoma, most commonly diffuse large B-cell lymphoma (DLBCL), for reasons not yet fully understood (Rossi, Spina and Gaidano, 2018). To understand the complex nature of these diseases, it is helpful to have a basic understanding of the formation of blood cells.

1.1.2 Haematopoiesis

All blood cells arise from a reservoir of multipotent haematopoietic stem cells (HSCs) by a process known as haematopoiesis (Ng and Alexander, 2017). HSCs reside in the bone marrow where a small number divide at any one time to replenish the supply of HSCs and/or create multipotent progenitor cells. Multipotent progenitor cells support the production of all blood cells by proliferating and beginning the process of differentiation down either the lymphoid or myeloid lineage. This is a hierarchical process and with each progressive division the cells become more numerous and more restricted in their capacity to differentiate until mature blood cells are formed (Bryder, Rossi and Weissman, 2006). The mature cells of the lymphoid lineage include plasma cells, lymphocytes and natural killer cells and the myeloid lineage includes erythrocytes, neutrophils and macrophages (Orkin, 2000).

Haematological malignancies result from an error in the differentiation process leading to uncontrollable proliferation of abnormal blood cells. The abnormal cells crowd the healthy blood cells which prevents them from carrying out their normal function. The complex differentiation process from HSCs to the many different types of blood cell derived from them, each with its own function, reflects the variety of malignancies associated with them.

1.1.3 Classification of haematological malignancies

Each subtype of haematological malignancy has its own distinct characteristics with regards to clinical presentation, management and treatment. Prior to 2001 there was a lack of consensus on the classification of haematological malignancies and most definitions were based on morphological findings only. This was problematic, as a classification system that is able to accurately differentiate between different subtypes is essential both clinically and in order to carry out meaningful research. However, in 2001 the World Health Organisation (WHO) released its 'Classification of Tumours. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues' which incorporated clinical features, immunophenotype and genetic abnormalities into its definitions (Jaffe *et al.* (Eds), 2001). This was subsequently revised in 2008, 2016 and 2022 in line with the latest scientific knowledge (Swerdlow et al., 2008; Swerdlow, 2016; WHO Classification of Tumours Editorial Board, 2024).

The new WHO classifications were adopted by the 3rd edition of the International Classification of Diseases for Oncology (ICD-O-3) which was last updated in 2022. The ICD-O-3 currently recognises over 100 different subtypes of haematological malignancy and shows the proportion of cases for those most commonly diagnosed between 2004-2015 (Figure 1:1). However, although adopted promptly in clinical practice, many non-specialist cancer registries still describe haematological malignancies according to the World Health Organisations (WHO) International Classification of Disease (ICD) -10 because of challenges collating the data required to assign ICD-O-3 codes due to diagnostic testing not being centralised in some regions.

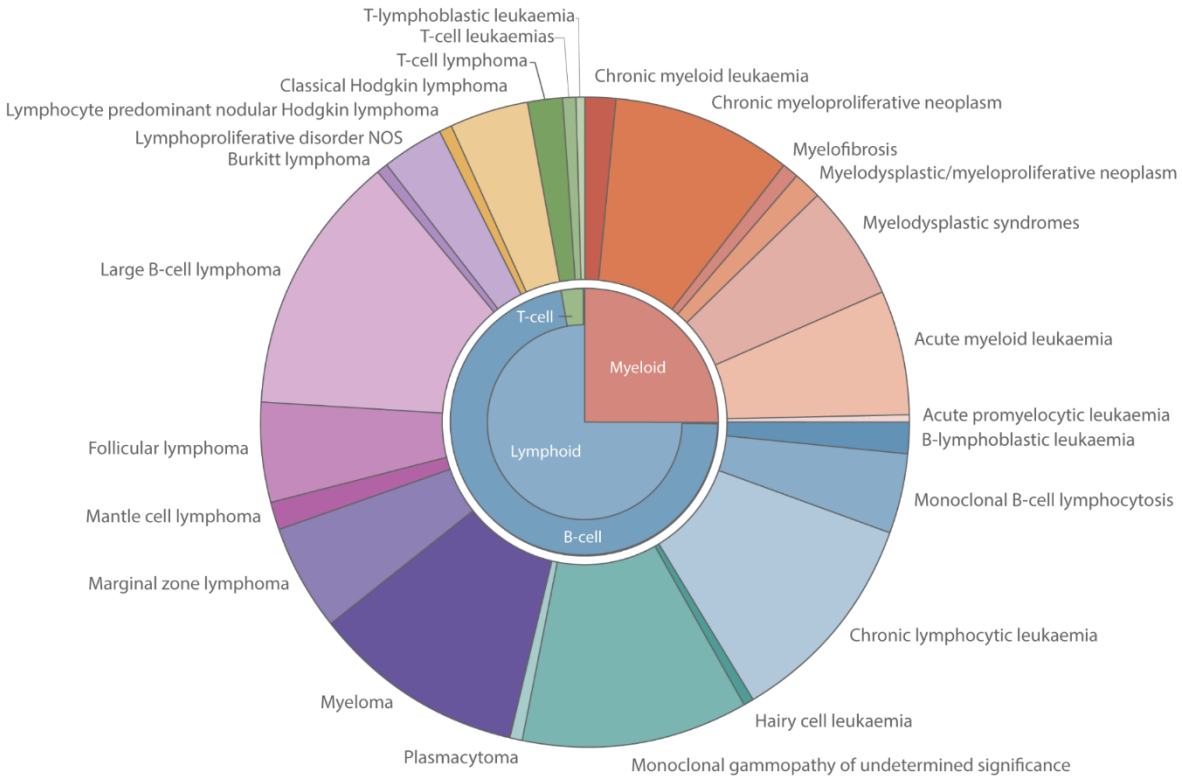


Figure 1:1. Diagnostic distribution of haematological malignancies and precursor conditions classified by ICD-O-3, Haematological Malignancy Research Network, 2004–2015.

1.1.4 Leukaemias

Leukaemia affects blood cells in the bone marrow, typically white blood cells but other cell types can also be affected. The classification of leukaemia is complex and is determined by a number of factors including the cells affected, which can be from either the lymphoid or myeloid lineage, and the genetic changes that have occurred in the affected cells, such as the presence or absence of the Philadelphia chromosome. It is a heterogenous group of diseases consisting of both acute diseases that are very aggressive and have poor survival, such as acute myeloid leukaemia (five-year net survival: 13.6% [95% confidence interval (CI): 11.8%-15.4%]) and indolent diseases with relatively good survival such as chronic lymphocytic leukaemia (CLL) (five-year net survival: 84.1% [95% CI: 81.7%-86.5%]) (HMRN, 2022b). The acute leukaemias are some of the most common childhood cancers (Steliarova-Foucher et al., 2017), but can be diagnosed at any age, while chronic leukaemias are typically diagnosed in older adults, for example the median age at diagnosis for a CLL patient is 71.7 years (Lamb et al., 2024).

1.1.5 Lymphomas

Lymphomas develop in the lymphatic system, a part of the immune system that is made up of tissues, vessels and organs including the spleen, bone marrow and lymph nodes. The cells affected in lymphoma are lymphocytes, commonly referred to as white blood cells, which help fight infection and disease. There are three main types of lymphocytes that originate from the proliferation and differentiation of HSCs down the lymphoid lineage; B cells, T cells and natural killer cells.

The classification of lymphomas is complex and depends on a number of factors including the type of lymphocyte affected and the location of the disease (WHO Classification of Tumours Editorial Board, 2024). However, they are often broadly divided into two groups; Hodgkin lymphoma (HL) and non-Hodgkin lymphoma (NHL). HL is characterised by the presence of large, abnormal cells sometimes containing more than one nucleus called Reed-Sternberg cells. All other types of lymphoma are classified as NHL, which is the 7th most commonly diagnosed cancer in the UK (Cancer Research UK, 2024). It is a heterogenous group that includes both aggressive and indolent sub-types such as DLBCL and follicular lymphoma respectively.

1.1.6 Myeloma

Myeloma develops in the bone marrow, which is the spongy tissue inside the bone where haematopoiesis takes place. The cells affected in myeloma are a type of lymphocyte called a plasma cell, which are derived from B-cells and their function is to produce immunoglobulins (also known as antibodies), which facilitate the detection and destruction of pathogens. The plasma cells of people with myeloma make abnormal immunoglobulins, called paraproteins, which do not function properly. Myeloma is incurable, but it can be managed with treatment for many years with approximately 50% of patients alive 5 years after diagnosis (HMRN, 2022c).

1.1.7 The burden of haematological malignancies

Individually many subtypes of haematological malignancy are rare, for example, the annual incidence rate of T-cell prolymphocytic leukaemia is 0.1 per 100,000 and the estimated UK incidence equates to 70 people newly diagnosed each year (HMRN, 2022a). However, collectively these cancers are common and as a group they are the fifth most commonly diagnosed cancer in economically developed regions of the world (Islami et al., 2021) and the third most common cause of cancer death in the UK (Cancer Research UK, 2022). They are often expensive to treat, with lengthy hospital stays and complex, expensive treatments,

making them the second most costly cancer with regards to total healthcare expenditure for European Union countries after breast cancer (Burns et al., 2016). The considerable burden of these diseases justifies in depth epidemiological research to better understand them and improve outcomes.

1.2 The Haematological Malignancy Research Network

The Haematological Malignancy Research Network (HMRN) was established in 2004. HMRN is an ongoing population-based cohort of patients newly diagnosed with a haematological malignancy in a catchment area that corresponds to two adjacent former Cancer Networks and covers a population of approximately four million people in Yorkshire and the Humber.

1.2.1 The purpose of HMRN

Although many non-specialist cancer registries collect data on haematological malignancies, they have historically struggled to fully capture the complexities of them as they group subtypes together, such as all NHLs, resulting in a gap in knowledge of specific haematological malignancies (Lamb et al., 2024). Recently, a few specialist registries and some national cancer registries have published additional statistics for distinct groups. For example, the National Cancer Registration and Analysis Service (NCRAS) has recently established the Get Data Out programme which reports statistics for cancers including haematological malignancies by ICD-O-3 codes (National Disease Registry Service, 2025). However, it is challenging for registries to complete data collection, as described in section 1.1.3, and as a consequence the ‘not otherwise specified’ groups are often large, meaning that differences between subtypes cannot be compared or controlled for. These differences can be considerable, for example, primary diffuse large B-cell lymphoma of the CNS (PCNSL), that affects the brain and spinal cord, has much worse survival than other subtypes of DLBCL. Furthermore, as mentioned in section 1.1.1., some blood cancers have the potential to transform to a more aggressive disease with poorer outcomes and tracking these cancers for progression to other subtypes is challenging for non-specialist registries.

The purpose of HMRN is to provide accurate, ‘real-world’ population-based data on haematological malignancies. HMRN diagnoses are coded according to WHO ICD-O-3 classification system, described in section 1.1.3, which is specific to oncology and better describes the different sub-types making research into specific sub-types possible. Patients in HMRN are also tracked for disease progression making the data unique and highly important for targeted and in-depth research into this diverse group of cancers.

1.2.2 HMRN study cohort

Currently, the HMRN cohort consists of approximately 55,000 individuals diagnosed with a blood cancer or related disorder and individuals are added to the cohort at a rate of approximately 2,500 per year (Roman et al., 2022). HMRN patients are all diagnosed at a

single haematopathology laboratory in Leeds, the Haematological Malignancy Diagnostic Service (HMDS) where all diagnostic tests are conducted and ICD-O-3 classification is applied. The sociodemographic structure of the cohort is broadly representative of the national population with regards to age, sex, ethnicity and levels of deprivation and patients are treated following national guidelines meaning that any findings using the HMRN cohort can be confidently extrapolated to the national population (Smith et al., 2018). Furthermore, the HMRN study region includes 14 hospitals (Figure 1:2) and covers an area that is geographically diverse including densely populated conurbations, such as Leeds and Bradford, as well as suburban and rural areas, such as those within the North York Moors.

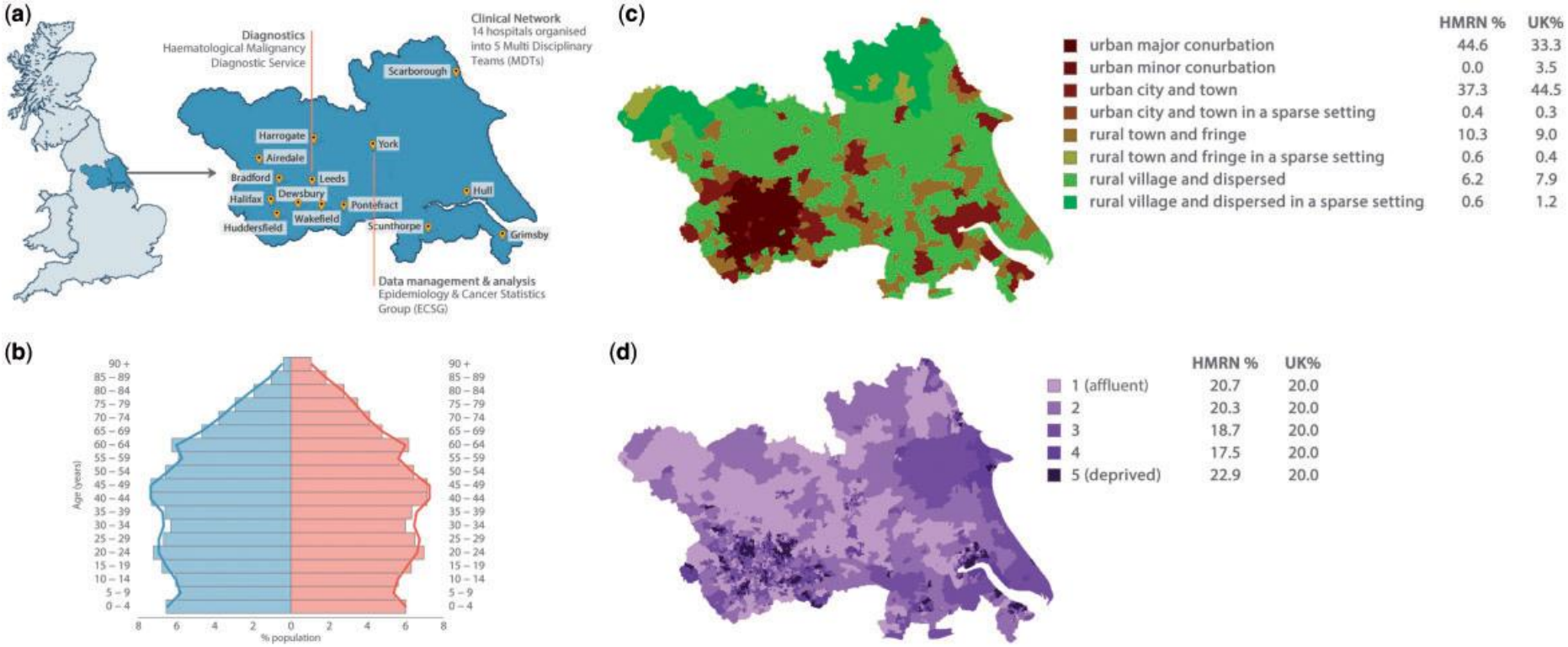


Figure 1:2. Haematological Malignancy Research Network (HMRN). A, Study location. B, Population age and sex distribution. C, Urban/rural distribution (Office of National Statistics definitions). D, Index of multiple deprivation (IMD): income domain.
Smith, A. et al. (2018). Cohort Profile: The Haematological Malignancy Research Network (HMRN): a UK population-based patient cohort. *International Journal of Epidemiology*, 47 (3), pp.700–700g. [Online]. Available at: doi:10.1093/ije/dyy044.

1.3 Diffuse large B-cell lymphoma

1.3.1 Definition and epidemiology of diffuse large B-cell lymphoma

DLBCL is a type of NHL and is the most common haematological malignancy with an estimated 4870 new cases per year in the UK (Smith et al., 2011). It is characterised by the clonal proliferation of malignant B-cells following genetic damage, which diffusely spread throughout the lymph nodes and in some cases the tissue that surrounds the lymph nodes, such as the thyroid. The disease can occur *de novo* or as a result of a transformation from a less aggressive malignancy (secondary), such as follicular lymphoma or chronic lymphocytic leukaemia.

DLBCL can be diagnosed at any age but occurs more frequently in the elderly with a median age at diagnosis of 70.3 years and like many haematological malignancies, is more common in males than females with a male-to-female incidence rate ratio of 1.3 (95% CI: 1.3-1.3) (Smith et al., 2018). Unlike many other cancers, such as lung cancer and its association with smoking, no strong aetiological drivers have been identified for developing DLBCL, other than male sex (Zhang et al., 2011). However, it has been observed that patients with certain comorbidities are at increased risk of developing DLBCL, for example, those with immunological disorders or who are chronically immunosuppressed by medication (Riaz et al., 2019; Gibson et al., 2014).

1.3.2 Diagnosing diffuse large B-cell lymphoma

Patients typically present with enlarged lymph nodes commonly in the neck, armpit or groin, although other lymph nodes can be affected, as shown in Figure 1:3. Following the biopsy of a suspicious lymph node, identified by clinical examination or radiography, a haematopathologist will examine the tissue under a microscope where malignant cells appear as large and atypical. Further diagnostic investigations are carried out to confirm a diagnosis of DLBCL, the stage of the disease and to provide a prognostic score. These include immunohistochemistry tests to detect B-cell antigens such as CD20, blood tests including a lactate dehydrogenase test, and a positron emission tomography scan combined with a computed tomography scan (PET-CT) which can accurately detect the spread of disease and identify bone marrow involvement without the need to perform an invasive and painful bone marrow biopsy.

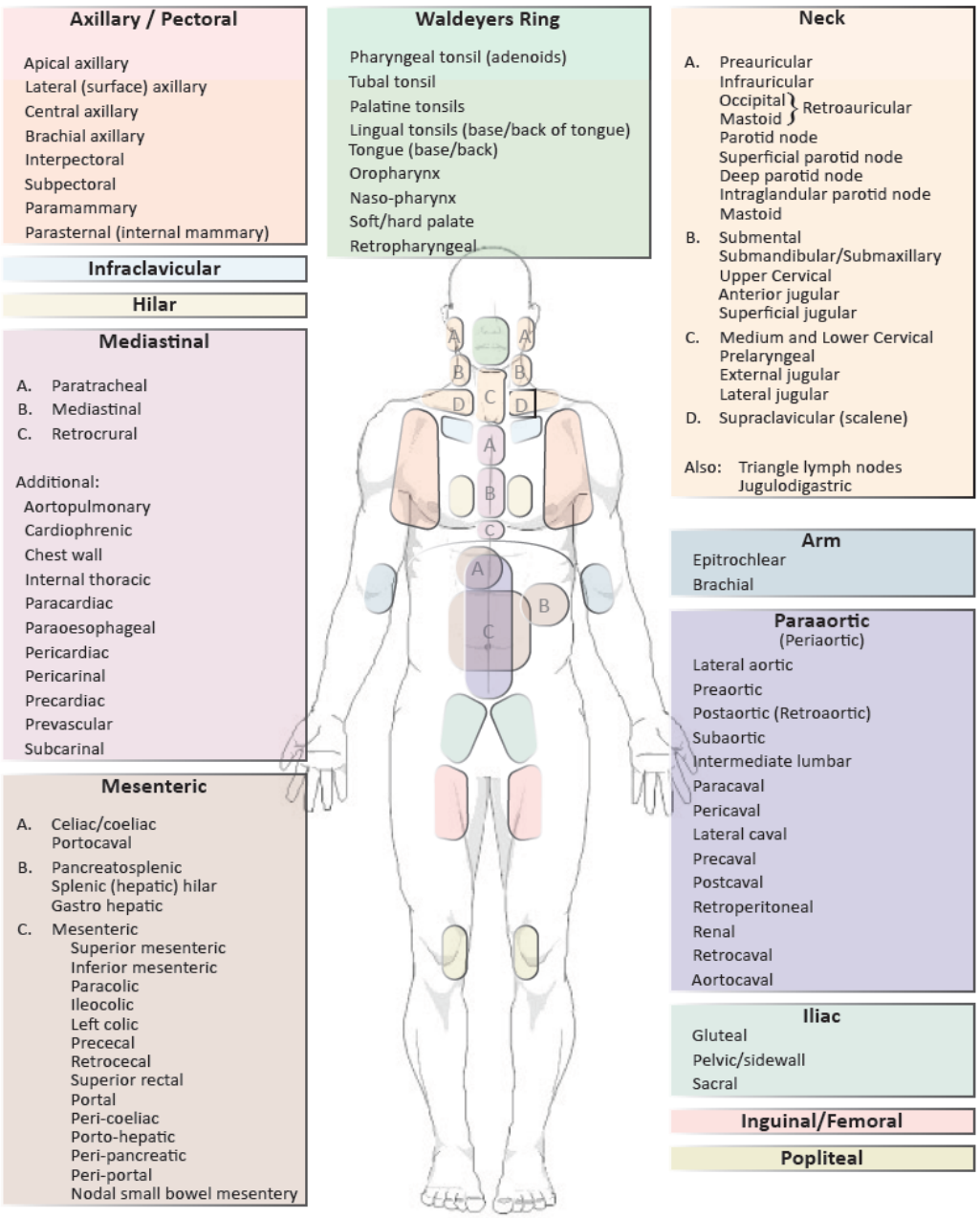


Figure 1:3. Anatomical distribution of lymph nodes.
From HMRN Data Collection Manual, version 10, September 2020. https://hmrn.org/resources/documentation.

1.3.3 Staging diffuse large B-cell lymphoma

Based on the results of the PET-CT scan, patients are staged using the modified Ann Arbor classification system (Lister et al., 1989). The Ann Arbor classification system was first introduced in the 1970s and is based primarily on the anatomy of a patient's disease as summarized in Table 1:1. This staging system was initially developed for patients with Hodgkin lymphoma and is used to stage all lymphomas, not specifically DLBCL.

Table 1:1. Ann Arbor staging system for lymphomas (Lister et al., 1989).

Stage	Description
I	Involvement of a single lymph node region or lymphoid structure (e.g., spleen, thymus, Waldeyer's ring).
IE	Involvement of a single extralymphatic organ or site.
II	Involvement of two or more lymph node regions on the same side of the diaphragm.
IIE	Involvement of a single extralymphatic site contiguous or proximal to a known nodal site plus involvement of one or more lymphoid region(s) or structure(s) the same side of the diaphragm.
III	Involvement of lymphoid regions or structures on both sides of the diaphragm.
IIIE	Localised involvement of a single extralymphatic site contiguous or proximal to known nodal site plus involvement of lymphoid regions or structures on both sides of the diaphragm.
IIIs	Localised splenic involvement plus involvement of lymphoid regions or structures on both sides of the diaphragm.
IIIsE	Fulfilling the definitions of both IIIE and IIIS.
IV	Diffuse or disseminated involvement of one or more extralymphatic organs or tissues with or without associated lymph node involvement. Liver involvement is always considered diffuse and therefore stage IV. Marrow involvement also indicates elevation to stage IV.

In the late 1980s, the Ann Arbor classification system was updated and these are referred to as the Cotswold modifications, which are;

- The addition of an 'X' suffix to denote bulky disease, which for most regions can be defined as enlargement of a single node or conglomerate nodal mass to 10cm or above.

- Further subdivision of stage III into III1 which is with or without splenic, hilar, celiac or portal nodes and III2 which is with para-aortic, iliac, mesenteric nodes.

Patients diagnosed at stage I or II are often described as having limited or early-stage disease and those diagnosed at stage III or IV as having advanced stage disease. The majority of patients with DLBCL will be diagnosed with stage IV or advanced stage disease (Smith et al., 2015).

1.3.4 B symptoms

In addition to staging patients, the absence or presence of B symptoms at diagnosis are also often recorded as well as their performance status which can be used to predict outcomes for patients and calculate a prognostic score. B symptoms are a set of symptoms often associated with advanced disease and 45% of DLBCL patients in HMRN are diagnosed with them (Kane et al., 2017). The absence or presence of B symptoms can be denoted in addition to a patient's stage using an A or B suffix respectively. They are:

- persistent or recurrent fever (>38°C)

- drenching night sweats
- weight loss (10% body weight over 6 months with no other explanation)

1.3.5 Performance status

Performance status estimates a person's ability to carry out everyday activities, such as getting dressed independently. The performance status of a patient is often assessed at diagnosis using the Eastern Cooperative Oncology Group (ECOG) scoring system which is described in Table 1:2 (Oken et al., 1982). The score can be used by the clinical team to predict if a patient will tolerate aggressive treatment and it contributes towards their prognostic score.

Table 1:2. Eastern Cooperative Oncology Group performance status scoring system (Oken et al., 1982).

Score	Description
0	Able to carry out all pre-disease performance without restriction.
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work.
2	Ambulatory and capable of all selfcare, but unable to carry out any work activities; up and about more than 50% of waking hours.
3	Capable of only limited selfcare; confined to bed or chair more than 50% of waking hours.
4	Completely disabled; cannot carry on any selfcare; totally confined to bed or chair.
Unknown	No information available to determine score.

1.3.6 Prognostic score

The International Prognostic Index (IPI) can be used to predict the prognosis of patients diagnosed with DLBCL using information including their stage and ECOG score (International Non-Hodgkin's Lymphoma Prognostic Factors Project, 1993). An overall score from 0 to 4 is calculated by adding together individual scores for each risk factor in the table below (Table 1:3), with a lower score predictive of better survival.

Table 1:3. Risk factors used to assign the IPI score (International Non-Hodgkin's Lymphoma Prognostic Factors Project, 1993).

Risk factor	Score +0	Score +1
Age (years)	≤60	>60
Ann Arbor stage	I or II	III or IV
Number of extranodal sites	0 or 1	>1
ECOG[a] score	0 or 1	>1
Serum lactate dehydrogenase*	Normal	Elevated

[a]ECOG= Eastern Cooperative Oncology Group

*An enzyme that is measured from a blood sample

1.3.7 Treatment for diffuse large B-cell lymphoma

The majority of patients diagnosed with DLBCL will receive three or six courses of R-CHOP as their first-line treatment, depending on their disease stage. R-CHOP is a chemo-immunotherapy regimen consisting of a targeted monoclonal antibody, rituximab, a steroid, prednisone, and three chemotherapy drugs; cyclophosphamide, doxorubicin and vincristine. R-CHOP has been the standard treatment for DLBCL for almost two decades, with the addition of rituximab in the early 2000's, which markedly improved survival (Coiffier et al., 2002; Wang, Li and Young, 2020). Rituximab works by binding to the CD20 antigen expressed on the surface of most B-cells, including abnormal B-cells produced by patients with DLBCL, triggering their destruction.

Patients not treated with standard R-CHOP as first-line therapy include those with PCNSL, patients aged under 18 years and very frail patients. PCNSL patients are typically treated with methotrexate as first-line therapy as this drug is able to cross the blood-brain barrier unlike most other chemotherapy drugs (Fox et al., 2019). Those diagnosed under the age of 18 typically tolerate chemotherapy well and are given aggressive regimens, however, the drugs used to treat paediatric DLBCL patients differ to those used in adult patients. (Sandlund and Martin, 2016). Patients who are too frail to tolerate standard R-CHOP may be given a less aggressive regimen such as R-CVP or mini R-CHOP, where the dose of the drugs is reduced, or it may be decided by their clinical team that they are unlikely to benefit from aggressive treatment and supportive/palliative care is in their best interest (Peyrade et al., 2011).

Since the introduction of R-CHOP as standard therapy, several clinical trials have been conducted to test the efficacy of other drugs for treating DLBCL such as the REMoDL-B trial (Maishman et al., 2013). REMoDL-B was a phase 3 trial that recruited patients between 2011 and 2015 and aimed to determine if progression-free survival is improved by the addition of a proteasome inhibitor, bortezomib, to the R-CHOP regimen (Davies et al., 2022). As of 2023, the National Institute for Health and Care Excellence (NICE) approved the use of polatuzumab vedotin in combination with rituximab, cyclophosphamide, doxorubicin and prednisolone (Pola-R-CHP) as first-line treatment for DLBCL (National Institute for Health and Care Excellence, 2023). This recommendation was made following evidence that DLBCL patients treated with Pola-R-CHP have reduced risk of disease progression compared to those treated with R-CHOP (Tilly et al., 2022). Polatuzumab vedotin is an antibody-drug conjugate that binds to CD79b, which is expressed on the surface of all mature B cells, including the abnormal B cells of DLBCL patients.

1.3.8 Diffuse large B-cell lymphoma survival

DLBCL is an aggressive disease which requires immediate treatment but is potentially curable with a standard regimen of R-CHOP or Pola-R-CHP and has a five-year net survival of 61.1% (95% CI: 58.9% - 63.4%) (HMRN, 2025). However, 20-40% of patients will have relapsed or refractory disease, where their lymphoma does not respond to treatment, and survival for these patients is poor (Crump et al., 2017; Pennings et al., 2024). Other factors associated with poorer survival include advanced stage disease, a higher ECOG score and older age. For example, five-year overall survival for DLBCL patients; with stage I is 71.5% compared to 35.0% for those with stage IV; with ECOG score zero is 75.0% compared to 3.2% with a score of three or four; for those aged 18-54 is 73.7% compared to 8.1% for those aged 85 or over (Smith et al., 2015).

1.4 Delayed diagnosis and cancer survival

1.4.1 Delayed diagnosis

Cancer survival rates in the UK are poor and lag behind other European countries by as much as 25 years according to recent analysis published by Macmillan, with survival better for women with colon cancer diagnosed in Sweden between 1997-2001 than it was for women diagnosed in England between 2016-2020 (Macmillan Cancer Support, 2024a). It has previously been estimated that thousands of deaths could be prevented each year in the UK if survival matched the European average (Abdel-Rahman et al., 2009). The reasons for poor cancer survival in the UK are complex, but an often-cited contributing factor is diagnostic delay, which is associated with advanced disease and lower odds of receiving treatment with curative intent (Richards, 2009). In addition to this, if delays in the diagnostic pathway result in patients starting treatment later, this can also negatively impact their outcomes. A review and meta-analysis of studies across seven common cancers (bladder, breast, cervical, colon, head and neck, rectum, lung) reported that delays starting treatment can negatively impact survival, with a delay as small as four weeks associated with an increased risk of death (Hanna et al., 2020).

Delayed diagnosis can be defined as an avoidable delay to a patient's cancer diagnosis, with potential opportunities for delay to occur at multiple points along the diagnostic pathway, as shown in Figure 1:4 (Rubin et al., 2015). Firstly, symptomatic patients may delay help-seeking for a number of reasons. For example, they may not recognise the seriousness of their symptoms or they may have concerns about accessing their GP due to transport difficulties (Scott and Hoskin, 2024). Delays can also occur after the patient has first presented to their GP; they may present multiple times before referral to secondary care or they may not adhere to their investigation plans, and appointments for no-shows may not be automatically re-scheduled (Lyratzopoulos et al., 2012; Lyratzopoulos, Vedsted and Singh, 2015). Minimising delays during the diagnostic pathway is important because longer diagnostic intervals have been shown to be associated with poorer survival (Tørring et al., 2013).

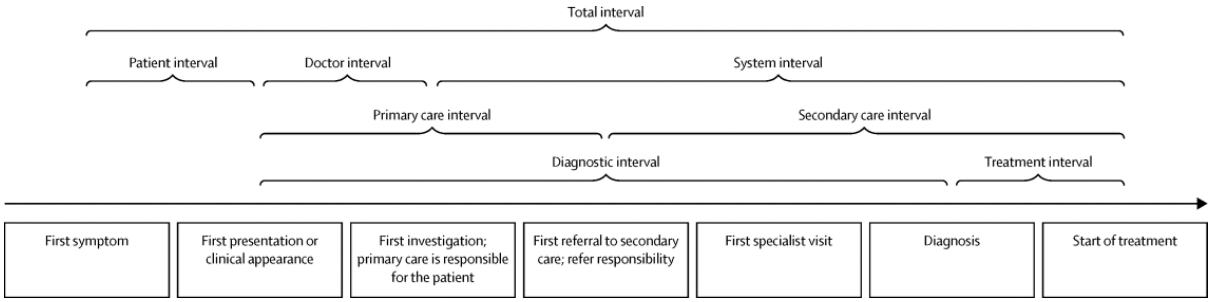


Figure 1:4. Overall milestones and time intervals from first symptom to start of treatment.
Rubin, G. et al. (2015). The expanding role of primary care in cancer control. *The Lancet Oncology*, 16 (12), pp.1231–1272. [Online]. Available at: doi:10.1016/S1470-2045(15)00205-3.

1.4.2 Policy on diagnostic delay

Achieving a greater proportion of cancer diagnoses at an early stage has been a key target of the NHS for some time - highlighted as an area where improvements could be made a quarter of a century ago in the 2000 NHS Cancer Plan (Department of Health, 2000) and again in the 2007 Cancer Reform Strategy (Department of Health, 2007). More recently, the 2019 NHS Long Term Plan (NHS, 2019) set a target of diagnosing three-quarters of cancers at an early stage (stage I or stage II) by 2028. In an effort to tackle diagnostic delay, the two-week wait (TWW) referral pathway was first introduced as part of the NHS Cancer Plan (2000), whereby patients who met the criteria for suspected cancer should be seen by a specialist within two weeks of referral by their GP (Department of Health, 2000). The referral guidelines have since been updated by the NICE to lower the threshold for urgent referrals for suspected cancer, with recommendations organised by symptoms and primary care investigations, in addition to cancer site (NICE, 2015, 2005). The TWW standard was in place for over two decades before being replaced in 2023 as part of the wider NHS England Plan by the Faster Diagnosis Standard, which states that patients urgently referred for suspected cancer should have cancer diagnosed or ruled out within 28 days. However, to date there is no evidence to suggest that these policies have resulted in a reduced time-to-diagnosis or improved survival for patients with cancer, including those with blood cancer (Neal et al., 2014; Price et al., 2020; Rachet et al., 2009).

1.4.3 Route to diagnosis

Before 2009, there was some information relating to patients' diagnostic pathways, however, this mostly focused on patients diagnosed through screening and people referred under the TWW pathway (Barrett and Hamilton, 2005; Blick et al., 2010). It was suggested that many patients diagnosed at a late stage of disease may not have entered the health care system through screening or following a TWW referral (Ellis-Brookes, 2015). However, little was

known about how many cancer patients were diagnosed following other GP referrals, through Accident and Emergency (A&E) or while they were in hospital for an unrelated condition. In response, an algorithm that uses routinely collected data to categorise the pathways in which cancer patients are diagnosed, their 'route to diagnosis' (RTD), was devised by a team at the National Cancer Intelligence Network, which later evolved into NCRAS and the methodology for this was published in the British Journal of Cancer in 2012 (Elliss-Brookes et al., 2012). These data were routinely collected and published between 2006-2020 and were available from NCRAS.

RTD can be defined as 'the sequence of interactions between the patient and the health care system, which led to a diagnosis of cancer' (Elliss-Brookes et al., 2012). There are a multitude of interactions with the healthcare system that can lead to a patient being diagnosed with cancer, but these unique pathways can be simplified and using the RTD algorithm, all cancer patients can be assigned to one of eight distinct diagnostic routes: Screen-Detected, TWW, Emergency Presentation, GP Referral, Inpatient Elective, Other Outpatient, Death Certificate Only and Unknown. A description of each of these can be found in Table 1:4, but they broadly represent the urgency of referral (Emergency, TWW, GP Referral), patients whose route started in secondary care (Inpatient Electives and Other Outpatients), those who were detected through screening (Screen-Detected), those diagnosed at death (Death Certificate Only) and patients with no useful information available on RTD (Unknown).

The Emergency Presentation route is associated with poor clinical outcomes and is widely used in health research as a proxy for delayed cancer diagnosis (Hamilton, 2012; McPhail et al., 2013). This assumes that most patients diagnosed as an Emergency were symptomatic and sought help from their GP beforehand, and that opportunities for earlier diagnosis were missed by primary care. There is evidence to suggest this could be the case, as it has been reported that most patients diagnosed via the Emergency Presentation route contacted their GP in the months beforehand (Mitchell et al., 2015; Renzi et al., 2016; Abel et al., 2017; Murchie et al., 2017). However, Murchie et al also reported that for six of the commonest cancers in Scotland, missed opportunities for earlier diagnosis in primary care only occurred for ~20% of Emergency Presenters and for some patients this route was the most appropriate diagnostic pathway (Murchie et al., 2017). It is also important to note that the proportion of patients diagnosed through each route varies considerably by cancer type, and that for a few aggressive, rapidly fatal cancers such as acute leukaemias, the Emergency Presentation route is the most appropriate route and is not a marker of diagnostic delay (McPhail et al., 2013).

Table 1:4. The eight 'routes to diagnosis' a cancer patient can be diagnosed via in England (Elliss-Brookes et al., 2012).

Route to Diagnosis	Description
Screen-Detected	Detected via one of the three national screening programmes; breast, cervical and bowel.
Two-Week Wait	An urgent referral with a suspicion of cancer from the GP.
Emergency Presentation	Diagnosed following an attendance at A&E[a], Emergency GP referral, Emergency transfer or Emergency admission. The Emergency must not be via an outpatient clinic and the patient must not have outpatient attendances before the Emergency in the six-month period prior to diagnosis.
GP Referral	Routine and urgent GP referrals excluding the Two-Week Wait referral route.
Inpatient Elective	Where the most recent HES[a] record to date of diagnosis is an Inpatient Elective record and there are no previous outpatient episodes in the six months prior to diagnosis.
Other Outpatient	An elective route that starts with an outpatient appointment; either self-referral, consultant to consultant, other or Unknown referral.
Unknown	No inpatient or outpatient HES[a] data available for the six months prior to diagnosis.
Death Certificate Only	No data available from inpatient or outpatient HES[a], CWT[a], Screening and with a death certificate diagnosis flagged by the registry in the NCDR[a].

[a]A&E=Accident and Emergency, HES=Hospital Episode Statistics, CWT=Cancer Waiting Times, NCDR= National Cancer Data Repository

Interestingly, diagnosis following an Emergency Presentation is associated with worse survival even after controlling for more advanced stage disease (Kane et al., 2017). This means that the differences observed in survival between patients diagnosed as an Emergency compared to non-Emergency routes cannot be fully explained by a higher proportion of patients in the Emergency group having late-stage disease, with patients diagnosed as an Emergency having worse survival than patients of the same stage diagnosed via non-Emergency routes.

1.4.4 The potential impact of diagnostic delay in patients with DLBCL

Some cancers are more amenable to early diagnosis than others, such as those: with a national screening programme, with 'alarming' or specific symptoms that may trigger patients to seek medical help or increase their likelihood of an urgent referral, and those which progress to an advanced stage more slowly. An example of a cancer where widespread early diagnosis has already been achieved is breast cancer, with 85% of women diagnosed at stage I or II (NHS England, 2023a). There are many factors that may facilitate this high proportion of early diagnoses for breast cancer. Principally, there is a national screening programme for women aged between 50-70 years old, patients with a lump or mass in the breast are likely to seek medical advice, and patients who present with a breast lump are eligible for an urgent suspected cancer referral (if the patient is aged 30 and over) under the

current NICE guidelines (NICE, 2015). In contrast to this, only 28% of women with NHL, which includes DLBCL, are diagnosed at an early stage (NHS England, 2023a).

Patients with DLBCL are particularly vulnerable to diagnostic delay for a number of reasons. Firstly, patients may delay help-seeking, and there is evidence to suggest this is the case as a study that used self-reported data estimated that patients with DLBCL wait an average of 98 days after symptom onset before presenting to primary care (Howell et al., 2013). This is likely because patients with DLBCL often experience vague symptoms including fatigue and weight loss, and it has been reported that non-specific symptoms like these are associated with delays in presentation to primary care (Forbes et al., 2014). Furthermore, the average age of diagnosis for a patient with DLBCL is 70 years old, an age when symptoms such as tiredness and weight loss may not be sufficiently alarming to prompt help-seeking, and patients may misattribute their symptoms as part of expected age-deterioration (Howell et al., 2019). There is also no screening programme to detect DLBCL, however, it is important to consider that screening is not appropriate for all cancers and the beneficial impact of current screening programmes is debated, with arguments against them centring on overdiagnosis. This happens when a cancer that would likely never have been clinically diagnosed is detected through screening, artificially inflating improvements in survival by giving the impression of prolonged survival (lead-time bias) and reducing patients overall wellbeing through distress caused by being told they have cancer (Marmot et al., 2013). Once patients with DLBCL present to healthcare, vague symptoms such as fatigue and infections (Howell, Smith and Roman, 2008; Howell et al., 2019), are unlikely to fulfil the requirements for an urgent referral for suspected cancer, unless in combination with other symptoms such as lymphadenopathy. Furthermore, GPs may not initially suspect cancer as symptoms are typically those of benign or self-limiting conditions and patients may be referred to multiple other non-cancer specialists before being referred to haematology, elongating their diagnostic pathway further and increasing the risk of diagnosis with advanced disease. It has been reported that patients with DLBCL often present to primary care with non-specific symptoms multiple times before being referred for further tests which extends their diagnostic pathway and increases their risk of being diagnosed as an emergency and with advanced stage disease (Howell et al., 2013). If the patient is referred routinely, they may also experience long referral wait times between primary and secondary care. If delayed diagnosis results in the patient being diagnosed with more advanced stage disease, survival may be impacted because DLBCL survival is better for those treated promptly and aggressively (Smith et al., 2015).

If some groups of patients with DLBCL are more likely to experience diagnostic delay than others, for example those with low socioeconomic status (SES), who live in rural areas and those who travel further to access healthcare, this could result in survival inequalities. It has been widely reported that patients with low SES are more likely to be diagnosed following an Emergency Presentation, a RTD synonymous with diagnostic delay and associated with worse survival (Gunnarsson, Ekholm and Olsson, 2013; Abel et al., 2015; Herbert et al., 2019a). For example, an association between increasing deprivation and the likelihood of being diagnosed with cancer via the Emergency route has been reported across 27 cancer types (Abel et al., 2015). This inequality contradicts the principle of the NHS, a universal healthcare provider free at the point of access where care should be equally available to all patients. Therefore, it is important to identify groups of patients vulnerable to diagnostic delay and to investigate the factors contributing to this, so that policy makers are well informed to implement necessary changes to reduce these inequalities.

These associations have been investigated extensively for other cancer types, particularly those with strong aetiological factors such as lung cancer, however, minimal research has focused on haematological malignancies and even less on DLBCL specifically, an aggressive cancer that has no strong lifestyle or environmental risk factors or screening programme (Zhang et al., 2011). Findings from research undertaken on other cancers, including other NHLs or all NHLs combined, may not reflect the outcomes of patients with DLBCL. Furthermore, many studies that include DLBCL patients do not distinguish between different NHL types in their analyses because they do not have access to data with sufficient granularity to do so. This presents a considerable gap in our knowledge of how these factors may influence outcomes of patients with DLBCL, which will be explored in this thesis using the high-quality data collected by HMRN.

1.5 Socio-demographic and geographical factors to be investigated for associations with diagnostic route and survival of DLBCL patients

1.5.1 Socioeconomic status

SES can be defined as an individual's combined social and economic position relative to others. A graded, positive association between higher SES and better health, akin to a dose-response relationship, has been frequently described suggesting that socioeconomic position, or the factors related to this, play a causal role in health status (Mackenbach et al., 2008; Marmot and Bell, 2012).

The risk of developing some cancers is strongly associated with exposures related to SES. For example, lung cancer is disproportionately diagnosed in individuals from more deprived areas and this can largely be attributed to higher rates of smoking in more deprived groups (Riaz et al., 2011). However, as mentioned above, no strong aetiological factors for developing DLBCL exist and consequently no association between DLBCL incidence and socioeconomic group has been reported. There is evidence to suggest that once the disease has developed, outcomes, such as how promptly a patient receives their diagnosis and their survival, differ between patients by SES (Smith et al., 2021a). This is unlikely to be fully explained by differential exposure to unhealthy lifestyle factors, such as smoking and poor diet, between different SES groups and suggests that the living environment may also influence patients' outcomes.

1.5.2 Measuring socioeconomic status

Traditionally, social class was based on a person's occupation, however, SES is complex and can change throughout the course of a person's lifetime meaning it cannot be accurately captured by occupation alone. Several methods have been developed since the 1980s to measure SES using different indicators such as income, level of education and access to services including healthcare.

Most routinely used measures of deprivation in health research are area-based which are beneficial in that that they enable the effect of SES to be estimated for an individual in the absence of individual data on SES being explicitly collected. This is important because data on individual income status is highly sensitive and seldom collected, for example, there is no income question on the census. However, area-based measures of deprivation are vulnerable to ecological fallacy, which is when the characteristics of an individual are assumed to be the same as the population in the area where they live. In the case of SES, an individual might live in a wealthy area particularly in a city, but may themselves be deprived or vice-versa.

1.5.2.1 Townsend Deprivations Score and Carstairs Index

Early methods of measuring area-based SES include the Townsend Deprivation Score (TDS) (Townsend, Phillimore and Alastair, 1988) and the Carstairs index (Carstairs and Morris, 1990) both of which are area-based measures of deprivation that use data from the census.

The TDS of an area is calculated using a combination of four equally weighted indicator variables:

- percentage of private household who are not owner occupied
- percentage of economically active residents who are unemployed
- percentage of private households who do not own a car or van
- percentage of overcrowded private households (more than one resident per room)

Following the log transformation of the unemployment and overcrowding variables to normalise their distributions, z scores of the percentages of each of these variables are calculated and added together to produce an overall score, with a higher score indicating more deprivation.

The Carstairs index is commonly used in research based in Scotland and is calculated in a similar way to the TDS using unweighted z scores of the following four variables:

- percentage of economically active male residents who are unemployed
- percentage of private households who do not own a car or van
- percentage of overcrowded private households (more than one resident per room)
- low occupational social class (class IV or V – partly skilled or unskilled workers)

However, there are limitations to both the Carstairs index and the Townsend score. For example, the association between car ownership and financial affluence may be influenced by other factors such as what type of area a person lives in, environmental concerns or choosing active transport such as walking or cycling for health benefits. For example, in dense, inner-city areas such as London, many affluent people may choose not to have a car despite being able to afford one because they can easily access everything they need using reliable alternative travel systems such as the London Underground, whereas people living

in rural areas are likely to own a car regardless of their economic situation because it is essential for them to get to work and other places. Furthermore, the reliance on census data which is collected once every ten years, means that these measures can soon become very outdated, especially in mobile populations such as inner cities.

1.5.2.2 The Index of Multiple Deprivation

In recent years, the most widely used measure of deprivation in England is the Index of Multiple Deprivation (IMD) which was last updated in 2019 (Ministry of Housing, Communities & Local Government, 2019). The IMD has been devised by the Social Disadvantage Research Centre at the University of Oxford since 2000. It is used by national and local organisations for a variety of purposes such as targeting resources to the most deprived areas and examining changes in relative deprivation between areas.

IMD is a relative, area-based measure of deprivation for small areas known as lower-layer super output areas (LSOAs) in England. LSOAs are census-based administrative spatial areas developed by the Office for National Statistics (ONS), with an average population size of 1500 people. LSOAs are small areas which makes them an appropriate unit for calculating the IMD because smaller areas are more likely to be homogenous in terms of level of deprivation than larger areas such as wards. Larger areas can be very socioeconomically diverse and deprivation scores calculated for these areas are unlikely to accurately represent the conditions of the people that live there.

What determines an individual's SES is complex and not simply related to access to financial resources. In an effort to capture these complexities, the overall IMD score for each LSOA is calculated from seven distinct domains of deprivation which are measured separately and then weighted and combined to give an overall score of multiple deprivation for each LSOA (Ministry of Housing, Communities & Local Government, 2019). The domains of the IMD and their weightings are:

- 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%).

Across the seven domains of the IMD there are 39 indicator variables that are domain specific. For example, the indicators of the 2019 income deprivation domain are:

- adults and children in Income Support families
- adults and children in income-based Jobseeker's Allowance families
- adults and children in income-based Employment and Support Allowance families
- adults and children in Pension Credit (Guarantee) families
- adults and children in Working Tax Credit and Child Tax Credit families, below 60% median income not already counted
- asylum seekers in England in receipt of subsistence support, accommodation support or both
- adults and children in Universal Credit families where no adult is 'Working – no requirements' conditionally regime.

The majority of indicators are measured as either the proportion or rate of the population in an area that are deprived according to the parameters of the indicator, using data collected from administrative records for the most recent time point available such as benefit records from the Department of Work and Pensions. IMD calculations only use data from the census on the rare occasion that the appropriate administration records are unavailable. The use of administration records instead of census data is beneficial as it means that IMD scores can be updated more frequently than once a decade and are more likely to accurately reflect the conditions of the people living in an area at any one time.

The indicators are then combined to generate each of the seven domain scores, which can be used individually or combined with the weightings described above as an overall composite score. When the domains are added together it is done in a way that is cumulative by exponentially transforming the ranked domain scores, so that a lack of deprivation in one domain does not cancel out deprivation in another domain. This means that the indices discriminate between and highlight deprived areas better than less deprived areas, with little information provided about the distribution of middle to high incomes.

IMD is a relative measure of deprivation with the scores used to rank each LSOA (~33,000 as of 2019) in the country from most (rank 1) to least deprived. In this way we can determine that the LSOA ranked as 100 is less deprived than LSOA ranked as 1, but not that it is 100 times less deprived. LSOAs with the same score are distinguished between based on population size with larger LSOAs assigned the lower rank because there are more people living in deprivation in them. The ranks themselves are seldom used in health research and the most common way of describing the IMD of an area is to use the decile or quintile of its rank, for example, the LSOA is in the most deprived 10% of the country.

1.5.3 Travel to healthcare

Transport can act as a barrier to accessing healthcare and it has previously been reported by the Social Inclusion Unit that over a 12-month period 1.4 million people missed, turned down or chose not to seek healthcare due to transport problems (Department of Health, 2006). There is evidence to suggest that, at least for some cancers, geographical access to healthcare impacts presentation, treatment and survival. For example, it has been reported that in Northern England, longer car journeys to the GP are associated with late-stage presentation of breast and colorectal cancer and poorer survival (Jones et al., 2008b). Other studies have shown that cancer diagnosis post-mortem, often interpreted as a marker of delayed diagnosis, was associated with increased travel to tertiary care but not GP access for some cancers (Jones et al., 2010). It has also been shown that increased travel to radiotherapy centre was associated with decreased odds of receiving radiotherapy treatment for cancer patients in Northern England (Jones et al., 2008a). For patients with lung cancer, although no independent effect of increased travel on the likelihood of receiving active treatment or surgery was detected, it was reported that increased travel amplified the effects of lower receipt of these treatments for the most deprived patients (Crawford et al., 2009).

Reorganisations of care that began with the NHS Five Year Forward View (NHS England, 2014) have resulted in considerable changes to primary care that may have resulted in longer travel to healthcare for many patients. For example, there has been a shift towards fewer but larger facilities staffed by multidisciplinary teams with fewer GPs and between 2013 and 2023 the number of general practices fell by 20%. In addition to potential inequalities in the accessibility of primary care, it has been demonstrated that primary care trusts (now obsolete) with more cases of breast, colorectal and lung cancer patients had longer average travel times to cancer management and treatment services, and longer travel time was also associated with poor survival rates, even after accounting for deprivation (Murage et al., 2016). Therefore, it is possible that access to care across the HMRN may not be equitable for patients with DLBCL.

If longer journeys to healthcare cause DLBCL patients to delay seeking medical advice, they may be at increased risk of Emergency Presentation, and/or with more advanced disease and poorer survival. This effect may be compounded by the fact that the symptoms are non-specific and could be wrongly attributed by the patient to other less serious and self-limiting conditions that don't require immediate medical intervention. Furthermore, long journeys to specialist cancer centres could deter some patients from receiving optimal treatment for their disease, which could also negatively impact survival.

1.5.4 Rurality

The influence of place of residence on health has been long observed. Until the 1940s, those that lived in squalid, over-crowded conditions in inner-city areas had increased mortality compared to those that lived in rural areas (Haines, 2001). In the UK, this association has persisted, with the Department for Environment, Food and Rural Affairs reporting the highest life expectancy for people born in 'Mainly Rural' areas and lower potential years of life lost from all cause death lower in 'Predominantly Rural' areas than 'Predominantly Urban' areas (Department for Environment Food & Rural Affairs, 2022).

However, the association between better health and rurality is not observed for cancer patients. For example, a systematic review of cancer survival in high-income countries reported that survival was generally worse for patients living in rural areas than those in urban areas, although none of the studies included were based in the UK (Afshar, English and Milne, 2019). Another systematic review and meta-analysis of cancer patients in predominantly developed countries estimated that rural patients had 5% lower survival than their urban counterparts after adjusting for potential confounders including SES (Carriere et al., 2018). Acknowledging that rural cancer patients face unique challenges, Macmillan and the Farming Community Network have established a Rural Communities Project to raise awareness of cancer in these communities (Macmillan Cancer Support, 2024b). This includes the 'Nip it in the Bud' campaign which encourages people living in rural communities, with a particular focus on farmers, to seek help for symptoms of cancer earlier.

The barriers to accessing healthcare and better outcomes for rural cancer patients are complex and driven by a number of patient, geographical and cultural factors. A qualitative study of Yorkshire patients who experienced symptoms of colorectal cancer reported many reasons why rural patients delay help seeking, including stoicism regarding health in rural populations, longer travel times to access primary care and more self-employed and seasonal workers with financial concerns regarding taking time off work (Dobson et al., 2023). A study of rural Australian lung cancer patients also identified travel to urban areas to

access healthcare and the financial burden associated with this as a barrier to accessing care (Ali et al., 2024). Furthermore, it has been reported that across many regions, rural communities have lower provision of healthcare services and rural patients have longer journeys to access specialist cancer services (Carriere et al., 2018).

As described above, HMRN consists of both urban and rural areas and it is possible that DLBCL patients living in rural areas of the Network may have worse outcomes including delayed diagnosis and poorer survival.

1.5.4.1 Office for National Statistics Rural/Urban Classification

Defining whether an area is rural or urban can be challenging because of different perceptions of what it means to be rural, such as a low population density, isolated location and agricultural landscape. However, it is important to be able to distinguish between rural and urban areas because the demographics, services available and challenges of the people living there can differ considerably.

The 2011 ONS rural-urban classification (RUC2011) was created by the Department of Town and Regional planning at the University of Sheffield and is a revised version of the 2001 release (Office for National Statistics, 2016). RUC2011 are generated at the Output Area (OA) level, the smallest geographical area for census statistics of between approximately 100 and 650 people, with classifications for larger units, including LSOAs, built from these. The ONS defines urban areas as ‘connected built up areas identified by Ordnance Survey mapping that have resident populations above 10,000 people (2011 census)’ and rural areas as ‘those areas that are not urban, i.e., consisting of settlements below 10,000 people or are open countryside’. Each OA in England is assigned a rural-urban classification based on whether it’s population-weighted centre is within an urban or rural area as defined above. Most people in England live in urban areas with the ONS reporting that in 2011, only 17.6% of the population of England lived in rural areas despite rural areas making up 85% of the land.

This dichotomous classification may be too simplistic to capture the diversity of rural and urban areas and these two classifications can be further subdivided into ten categories at the OA level based on settlement type, i.e., the population density of each 100m x 100m ‘square’, the density of the surrounding ‘squares’, and the wider context of each settlement, i.e., identify sparsely populated areas by looking at dwelling densities up to 30km away. However, it is important to note that these classifications are based on population density and settlement form alone and do not take into consideration the use or character of the

land. For example, an OA may be classified as urban despite covering a large area of open countryside if most of the population lives in urban settlements.

Settlements are more homogenous at the LSOA level and consequently eight settlement classifications are recognised from the ten at the OA level, these are:

- urban major conurbation
- urban minor conurbation
- urban city and town
- urban city and town in a sparse setting
- rural town and fringe
- rural town and fringe in a sparse setting
- rural village and dispersed
- rural village and dispersed in a sparse setting

.

2 Thesis Aims and Objectives

The overall aim of this thesis is to examine the relationship between travel to healthcare and diagnostic route and survival of patients with DLBCL in the Yorkshire and the Humber region. Patients with DLBCL may be vulnerable to diagnostic delay as evidenced by the high proportion of patients diagnosed following an Emergency Presentation which in turn means these patients are more likely to have advanced stage disease, and less likely to receive curative treatment so have worse survival (Kane et al., 2017). To date, few studies have investigated the impact of travel to healthcare on diagnostic route and survival of those with haematological malignancies and none have specifically examined this relationship in DLBCL patients in England, presenting a gap in knowledge about the disease. This is important as travel to healthcare could influence decision making at multiple points along the diagnostic and post-diagnostic pathway, such as when patients present to their GP and regarding treatment choice. In recent years novel, highly specialised and personalised treatments have been used increasingly to treat DLBCL, and these require delivery at specialist centres whereas the standard treatment for the last two decades of chemo-immunotherapy is delivered at all 14 hospitals in the Haematological Malignancy Research Network. If these novel therapies begin to be used more routinely, this could disadvantage patients with longer journeys to specialist centres. Therefore, it is important to understand if patients who travel further to access healthcare have worse outcomes, including survival. This thesis aimed to address this gap in the knowledge of DLBCL by examining the following hypotheses; if patients:

- who travel further to access healthcare are more likely to be diagnosed following an Emergency Presentation as a consequence of travel acting as a barrier to help-seeking

- who are diagnosed following an Emergency Presentation have increased hospital activity in the months leading up to their diagnosis compared to patients diagnosed via other routes

- who travel further to access healthcare are more likely to be diagnosed via an Emergency Presentation and less likely to be treated curatively

- have worse survival because they have further to travel to access healthcare and/or are diagnosed following an Emergency Presentation.

These hypotheses will be tested by completing the specific objectives of this thesis, which are to:

- synthesise the existing literature to; summarise and review previous research, identify gaps in the literature and inform the methods and analysis for this thesis
- describe travel times by both car and public transport to the GP and hospital and the socio-demographic characteristics of those with different travel times
- describe the pre-diagnostic hospital activity of DLBCL patients and compare this to the background rate of hospital activity using a nested case-control study
- describe the route to diagnosis of DLBCL patients and examine the association between travel time to patients' GP and hospital and socio-demographic characteristics with the Emergency Presentation route
- examine the impact of socio-demographic factors, travel to healthcare and route to diagnosis on receiving curative treatment
- examine the impact of socio-demographic factors, travel to healthcare and route to diagnosis and treatment intent on survival.

3 Literature Review

3.1 Outline

The purpose of this review is to summarise the current knowledge on the association between travel, rurality and socioeconomic status (SES), and route to diagnosis (RTD) and survival of cancer patients, with a focus on diffuse large B-cell lymphoma (DLBCL). This will provide a wider understanding of the context under which these associations exist and inform the methodology for this thesis. By compiling and critically evaluating the different methods used in the literature to quantify travel and define SES and rurality, the most appropriate methods for this thesis can be identified and selected. By summarising and evaluating what has already been done the review will also demonstrate how the research undertaken for this thesis is novel and adds to the current knowledge on the topic.

When comparing studies from different countries it is important to note that healthcare systems, which undoubtedly have an impact on patient outcomes, vary considerably between countries. In the UK and other European countries like Denmark, healthcare coverage is universal meaning that all residents have access to essential healthcare without incurring financial hardship. In contrast to this, there is no universal healthcare system in the USA and care is largely provided by the private sector and paid for by private insurance, out-of-pocket payments and public programs such as Medicaid and Medicare. Differences in insurance coverage in the USA can lead to inequalities in access to care and as of 2023 8% of the population had no private or public health insurance (USA Census Bureau, 2024). Another important difference between countries is the role of the General Practitioner (GP) and the relationship between primary and secondary care. For example, in the UK the GP acts as a gatekeeper to accessing hospital specialists but in other countries, including France and Germany, this is not the case and patients are able to self-refer to secondary care services without authorisation from the GP. These differences may impact the way in which patient's enter secondary care and are diagnosed and in doing so, influence their vulnerability to diagnostic delay.

Furthermore, the influence of SES on how patients access healthcare may differ between countries depending on their healthcare system. For example, in the UK healthcare is universal and free at the point of access and therefore the cost will not impact patients' decisions regarding help seeking and treatment. However, in the USA this is not the case and concerns relating to the cost of healthcare, may cause patients to delay help seeking. Furthermore, once patients are diagnosed a lack of financial resources may limit their treatment options and impact their survival. SES is also challenging to measure and even within countries several different ways to estimate it exist, including composite area-based

measures calculated using administrative data, individual-level data relating to income or education and more. It is important to remember that deprivation is relative to the society in which the person lives and therefore conditions described as deprived in one country may not be described as deprived in another.

It is also important to note that different definitions and methods of estimating diagnostic delay are used by studies within and outside of the UK, therefore comparing studies can be challenging and care must be taken when doing so. For example, in England a cancer diagnosis following an 'Emergency Presentation' is often used in health research as a proxy for diagnostic delay, however, other ways of estimating delayed diagnosis exist such as number of days between first presentation and diagnosis. Other countries define delayed diagnosis in ways that reflect their own healthcare system and cancer policies.

In addition to differences in the healthcare systems that make comparing results of studies from different regions difficult, geographical size varies markedly between countries such as the UK and USA which will likely impact average travel times to healthcare and geographical isolation associated with rurality. These differences in healthcare systems, geographical setting and definitions of diagnostic delay and SES are important to consider as they could result in inconsistent effects of travel, rurality and SES on RTD and survival.

This chapter is organised into three main sections, the first of which details the search strategy used for this review (section 3.2). The second section describes and critiques the different methods used to describe SES, rurality and quantify travel and includes summaries of the studies included in the review (section 3.3). The third section summarises and compares associations between travel, rurality, and SES and RTD and survival identified in the literature (section 3.4).

3.2 Search strategy

To comprehensively cover all potential areas of interest for this thesis, the review consisted of six main searches for literature relating to associations between:

- travel to healthcare and diagnostic route
- travel to healthcare and survival
- rural-urban location and diagnostic route
- rural-urban location and survival
- SES and diagnostic route
- SES and survival.

To identify the studies relevant to this thesis and because the amount of literature pertaining specifically to DLBCL is limited compared to that for all haematological malignancies and all cancers, each search was split into three categories; all cancers, all haematological malignancies and DLBCL. These are summarised in the following categories; investigating the influence of travel (Table 3:2), investigating the influence of rurality (Table 3:3) and investigating the influence of SES (Table 3:4) on diagnostic route and/or survival. Each table is organised by date of publication, with papers examining DLBCL at the top of the table.

Table 3:1. Summary of the inclusion/exclusion criteria of the literature review.

	Population	**Setting**	**Exposure**	**Outcome**	**Study**
Inclusion	Adults (aged 18 years and above) with cancer at baseline	Global north	Travel distance/time to healthcare; socioeconomic status; rurality	Delayed diagnosis; survival	Cohort; cross-sectional; case-control; clinical trials
Exclusion	Children (below 18 years of age)	Global south	Measures of socioeconomic status only using educational attainment.		Qualitative studies; case reports; conference posters; letters; commentaries; grey literature; published before 2013; published not in English

The cut-off start date for the search was 2013 so that the most recent and relevant research could be included while restricting the number of papers retrieved by the search to a practical and manageable quantity.

Studies of adults aged 18 years and older with a cancer diagnosis at baseline were included in the study. As some studies include cohorts of patients that include participants aged both under and over 18, where this was the case relevant studies were not excluded (see Table 3:1).

Only studies were participants resided in a global north country were included as these countries are more economically developed and typically have efficient healthcare systems and good quality cancer registries making the results of the studies more easily comparable to the analysis undertaken in this thesis.

Exposures were: how far participants travelled to access healthcare, the rurality of the participants' residential address and the participants SES. Travel to healthcare included both measures of distance and time taken to complete a journey to a primary, secondary or tertiary care facility. Rurality of participants was defined by each study and depended on factors such as geographical setting, with different countries having their own measures. The method used by each study to determine participants SES also varied according to factors such as geographical location. Individual-level measures of SES are seldom collected and therefore the majority of SES measures identified in this review were area-based. Studies that estimated SES based only on educational attainment were excluded.

The primary outcomes of this review were delayed diagnosis and survival and there were no exclusions for these outcomes.

Papers that were not written in English were excluded from this review as were qualitative studies, case reports, conference posters, letters, commentaries and grey literature including theses and government reports.

The PubMed database was searched for relevant studies in October 2023 and updated in February 2025 using the advanced search function. PubMed was chosen as it is a free resource that contains an extensive collection of health-related research papers from diverse regions of the world. Using the training received in writing and reviewing epidemiological papers as part of the Postgraduate Certificate from the London School of Hygiene and Tropical Medicine, search terms were created using keywords and phrases with truncation where appropriate, combined using Boolean operators. Some examples of the terms used include: "blood cancer" and "hematological malignanc*" for haematological malignancies;

"socioeconomic", "income" and "class" for SES; "travel" and "journey" for travel; "rural" and "remote" for rurality; "diagnostic delay" and "time to diagnosis" for delayed diagnosis; "survival" and "mortality" for survival. Searches were designed to be comprehensive, but not exhaustive.

The search results were uploaded into Covidence software, which automatically removes duplicates at import, for processing and study selection which was conducted in two phases; initially titles and abstracts alone were screened followed by full-text screening. In addition to reviewing the papers found by the search, the bibliographies were examined for relevant studies not retrieved by the initial search. 3881 records were identified in PubMed and after duplicate removal 2447 records remained for title and abstract screening, of which 2004 were excluded at this stage (see Figure 3:1). This left 443 records for full-text screening and examination of their bibliographies. The purpose of the review was to provide an over-view of and summarise the existing literature, so as to provide a background to the work undertaken in this thesis, as well as to critically evaluate the methods used and discuss the findings. All papers investigating DLBCL specifically were selected because this is the disease examined in this thesis and at least one paper representing each method encountered for measuring travel burden or defining rurality or SES were included. Other high-quality or landmark studies (such as those that are highly cited by others in the field) and those using data on haematological malignancies and UK populations were preferentially selected. This selection process resulted in a total of 76 records included in the review.

Relevant data, such as the study population and the setting, were extracted and summarised into tables which are included below (Table 3:2, Table 3:3, Table 3:4), restricted to studies included in the review.

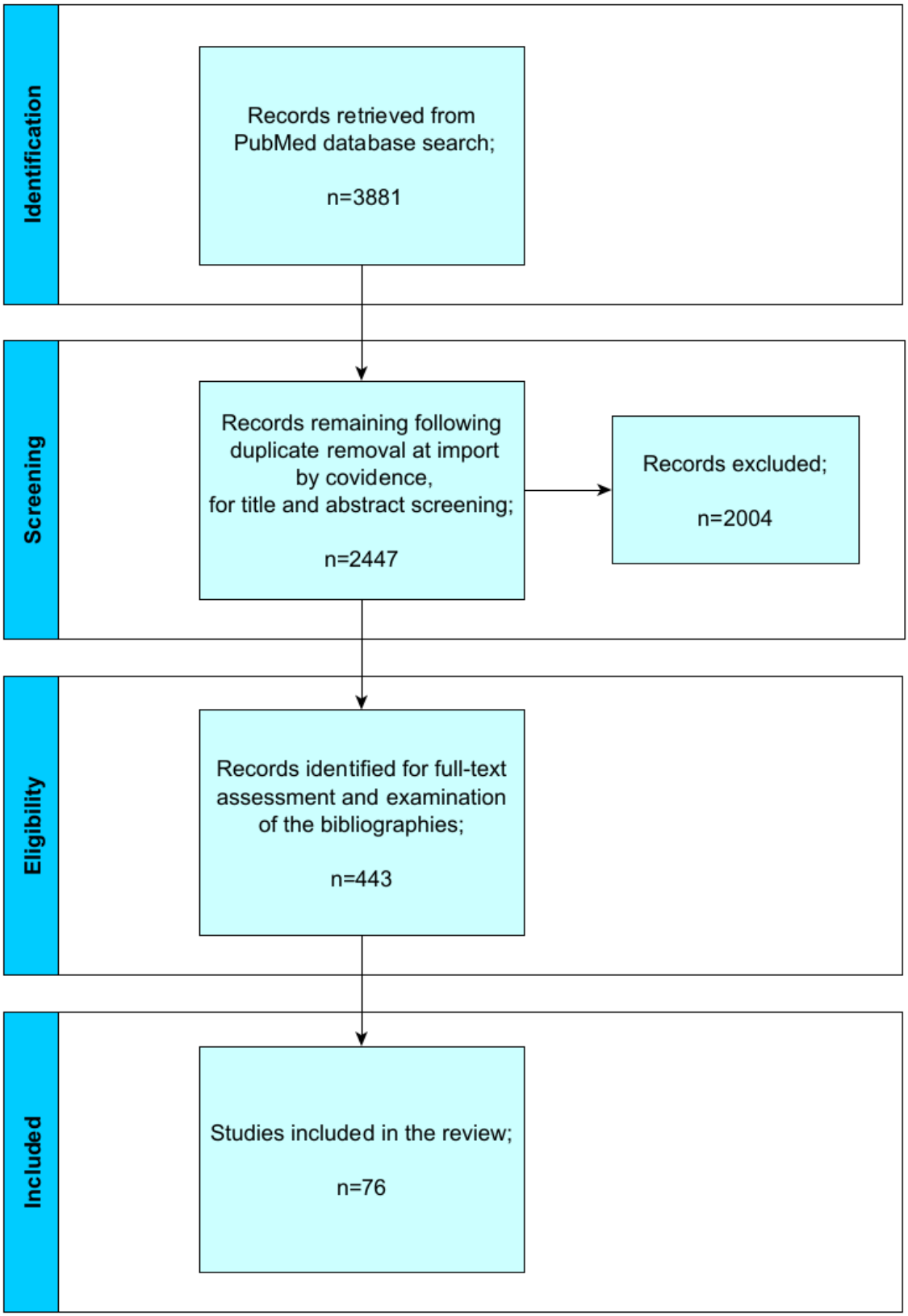


Figure 3:1. Flow chart of studies included in the review.

3.3 Methodology of studies

3.3.1 Quantifying travel

It is challenging to compare the results of studies that investigate the influence of travel to healthcare on outcomes of cancer patients because of heterogeneity in the populations, settings and cancer types in each study, as well as differences in the study methods. To calculate travel time or distance, a point must be chosen to represent the patient's origin, or starting point of their journey, and another point must be chosen as their destination, or where their journey ends. The researchers must then use a suitable calculation to estimate the time or distance between these two points. A summary of the methodology and findings of studies identified by this review are in Table 3:2.

Table 3:2. Studies included in the review that investigated the influence of travel to healthcare on diagnostic route or survival of cancer patients.

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Effect of place of residence and treatment on survival outcomes in patients with diffuse large B-cell lymphoma in British Columbia. Lee et al., 2014	1357 DLBCL patients. Canada.	Post code of patient's residential address. Post code of treatment centres corresponding to catchment area of patient's postcode.	Google Maps or MapQuest to estimate distance to chemotherapy (≤100,>100 km), radiotherapy (≤200,>200 km) and transplant centre (≤200,>200 km).	Null – no association between distance and survival (adjusted analyses).
Treatment selection and survival outcomes in early-stage diffuse large B-cell lymphoma: do we still need consolidative radiotherapy? Vargo et al., 2015	59,255 DLBCL patients. USA.	Not stated. Treatment facility.	Method to estimate distance not stated (≤30, >30 miles).	Negative – greater distance associated with inferior treatment (unadjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Factors related to the relative survival of patients with diffuse large B-cell lymphoma in a population-based study in France: does socioeconomic status have a role? Guyader-Peyrou et al., 2017	1165 DLBCL patients. France.	Geocoded patient's residential address. Nearest reference care centre.	ArcGIS to estimate time (≤15,16-44,45-118 mins).	Null – intermediate time associated with worse survival, no association between greatest time and survival (adjusted analyses).
The impact of structural factors on diagnostic delay in diffuse large B-cell lymphoma. Zurko et al., 2019	104 DLBCL patients. USA.	Not stated. Referral centre where they received care.	Method to estimate distance not stated (≤50, >50 miles and continuous).	Negative – greater distance associated with diagnostic delay (unadjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
The influence of geographical access to health care and material deprivation on colorectal cancer survival: evidence from France and England. Dejardin et al., 2014	7891 French and 32,722 English colorectal cancer patients. France and England	Patients residential address. Nearest cancer centre, hospital and radiotherapy unit.	ArcGIS (England) and MAPINFO (France) to estimate time to cancer centre (0-5, 5-20, 20-40,40-90, 90+ mins), radiotherapy centre (0-5, 5-20, 20-40,40-90, 90+ minutes) and hospital (0-5, 5-10,10-12,15-40, 40+ mins).	Null – no association between distance and survival (adjusted analyses). Greater distance associated with better survival in unadjusted analyses.
Association of distance to treatment facility on quality and survival outcomes after radical cystectomy for bladder cancer. Haddad et al., 2015	408 patients with bladder cancer. USA.	Patient's residential ZIP code. Treating facility.	Google maps to estimate distance (<50, 50-100,100.1-150, >150 miles).	Null – no association between distance and survival (adjusted).
Factors affecting receipt of expensive cancer treatments and mortality: evidence from stem cell transplantation for leukaemia and lymphoma. Mitchell and Conklin, 2015	5,722 leukaemia patients and 9,137 lymphoma patients. USA.	Patient's residential ZIP code. Nearest hospital that performs HSCT.	Method to estimate distance not stated (<10, 10- <35, 35-<70, >70 miles).	Negative – greater distance associated with less chance of receiving HSCT (adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Association of distance travelled for surgery with short- and long-term cancer outcomes. Wasif et al., 2016	169,650 colon, pancreatic, esophageal or liver cancer patients. USA.	Patient's residential ZIP code. ZIP code centroid of treating facility.	'Great circle' distance (<50, ≥50 miles and quartiles).	Positive – greater distance associated with lower mortality (unadjusted and adjusted analyses).
Distance from a comprehensive cancer center: a proxy for poor cervical cancer outcomes? Barrington et al., 2016	390 cervical cancer patients. USA.	Patient's residential address. Single comprehensive cancer centre.	'Geographic distance' was estimated (<100, ≥100 miles).	Negative – greater distance associated with worse survival.
Geographical disparities in access to cancer management and treatment services in England Murage et al. 2016	Did not use data on individual cancer patients. Instead calculated the travel times from all LSOA centroids to the nearest hospital and aggregated this data to Primary Care Trust (PCT) level. England.	LSOA population weighted centroids. Nearest hospital site offering treatment or management for the specified cancer.	ArcGIS to estimate mean travel time per PCT.	Negative – greater average travel times associated with worse survival (adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Impact of travel time and rurality on presentation and outcomes of symptomatic colorectal cancer. Murage et al., 2017	926 colorectal cancer patients. Scotland.	Patients' residential postcode. Postcode of GP at diagnosis.	ArcGIS to estimate time (continuous minutes for analysis and 25th,50th,75th ,99th percentiles for descriptive table).	Mixed – greater time not association with Emergency Presentation but is associated with better survival (adjusted analyses – no interaction term). With travel time-rurality interaction term added results differed e.g., greater travel time reduced odds of Emergency Presentation for urban residents.
Impact of travel distance to the treatment facility on overall mortality in USA patients with prostate cancer. Vetterlein et al., 2017	775,999 prostate cancer patients. USA.	Centroid of each patient's ZIP code. Street address of treating facility.	Haversine formula to estimate distance (<12.5,12.5-49.9,50-249.9 miles).	Positive – greater distance associated with lower mortality (adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
A cancer geography paradox? Poorer cancer outcomes with longer travelling times to healthcare facilities despite prompter diagnosis and treatment: a data-linkage study. Turner et al., 2017	12,339 colorectal, lung, breast, prostate, melanoma, oesophagogastric, cervical or ovarian cancer patients. UK.	Patients' residential postcode. Postcode of GP, diagnostic centre and treatment centre.	ArcGIS to estimate road and flight (for island residents) time to GP (<5, 5-9.9,10-14.9, >15 mins and separate category for island residents) and diagnosis and treatment centre (<15,15-29.9,30-59.9,>60 mins and separate category for island residents).	Mixed – greater time associated with prompter diagnosis and treatment but worse survival (unadjusted and adjusted analyses).
Cervical cancer care in rural Virginia: The impact of distance from an academic medical centre on outcomes & the role of non-specialized radiation centres. Rauh et al., 2018	180 cervical cancer patients. USA.	Patient's residential ZIP code. University of Virginia - served as a proxy for tertiary care travel.	Online mapping website (not stated) to estimate distance (<72, >72 miles).	Null – no association between distance and survival (adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Variation in geographical treatment intensity affects survival of non-small cell lung cancer patients in England. Tataru et al., 2018	144,357 lung cancer patients (excluding small cell tumours). England.	Patients' residential postcode. Nearest surgery and radiotherapy treatment centre(s).	ArcGIS to estimate time to surgical centre (0-15,16-25,26-35,36-55,>55 mins) and radiotherapy centre (0-10,11-20,21-30,31-40,>40 mins).	Negative – greater travel time associated with lower odds of surgery (adjusted and unadjusted).
Associations between distance, hospital volume, and outcomes following radical cystectomy in patients with muscle-invasive bladder cancer. Xia et al., 2018	6551 bladder cancer patients. USA.	Patient's residential ZIP code. Hospital that reported the case.	"Great circle" distance using the Haversine formula to estimate distance (quartiles for main analyses, continuous for sensitivity analyses).	Null – no association between travel distance and survival (adjusted analyses – addition of hospital volume). Greater distance associated with better survival in adjusted analyses – without hospital volume.
Racial and ethnic disparities in travel for head and neck cancer treatment and the impact of travel distance on survival. Graboyes et al., 2018	131,147 head and neck squamous cell carcinoma patients. USA.	Patients residential ZIP code centroid. ZIP code centroid of reporting hospital.	"Great circle" distance to estimate distance (<12.5,12.5-49.9,50-249.9 miles).	Positive – greater distance associated with better survival (adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Travel distance: Impact on stage of presentation and treatment choices in head and neck cancer. Ringstrom et al., 2018	6029 head and neck cancer patients. USA.	Longitude and latitude co-ordinates of exact street address or ZIP-code (dependent on quality of data source). Nearest otolaryngologist, radiation and academic medical centre.	ArcGIS to estimate distance and time (mean, median, range, quartiles).	Positive – greater distance to academic centres associated with better treatment (adjusted analyses).
Geographical access to GPs and modes of cancer diagnosis in England: a cross-sectional study. Murage et al., 2019	737, 495 breast, colorectal, cervical, lung, prostate, stomach, ovarian or brain cancer patients. England.	Patient's home. Patient's GP.	ArcGIS to estimate time (≤10,10.1-20,20.1-30,>30 mins).	Negative – greater distance associated with 'less desirable' RTD (adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Longer distance to specialized treatment centres does not adversely affect treatment intensity or outcomes in adult acute myeloid leukaemia patients. A Danish national population-based cohort study. Tøstesen et al., 2019	2992 acute myeloid leukaemia patients. Denmark.	City centre of habituation. Nearest specialized treatment centre offering remission-induction therapy and nearest centre offering stem cell transplantation.	Google maps to estimate distance to treatment centre (<10, 10-25,25-50,50-100, >100 km) and transplant centre (<10, 10-50,50-200, >200 km).	Null – no association between distance and treatment or survival (adjusted analyses).
Travel to a high-volume hospital to undergo resection of gallbladder cancer: does it impact quality of care and long-term outcomes? Beal et al., 2020	10,174 gallbladder cancer patients. USA.	Patient's residence. Hospital that reported on patients care.	“Great circle” distance using the Haversine formula to estimate distance (quartiles).	Null – no association between distance and treatment or survival (adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Effects of distance from academic cancer center on overall survival of acute myeloid leukemia: retrospective analysis of treated patients. Dhakal et al., 2020	449 acute myeloid leukaemia patients. USA.	Patient's residential address. Academic cancer centre.	Method to estimate distance not stated (<25,25-50,50-100,>100 miles and continuous).	Null – no association between distance and survival (unadjusted and adjusted analyses).
Travel time to care does not affect survival for patients with colorectal cancer in northern Sweden: A data linkage study from the Risk North database. Sjöström et al., 2020	3718 colorectal cancer patients. Sweden.	Patient's residential address. Nearest diagnostic hospital and operating hospital (for operated patients).	ArcGIS to estimate time (0-10, >10-20, >20-30,>30-40,>40-50,>50-60,>60-70,>70 mins).	Null – no association between distance and survival (unadjusted and adjusted analyses).
The impact of driving time, distance, and socioeconomic factors on outcomes of patients with locally advanced rectal cancer. Gotfrit et al., 2020	1064 rectal cancer patients. Canada.	Patient's residential postcode centroid. Nearest cancer centre address.	CDXZIPStream mapping software (using bing maps) and validated using Google distance matrix Application Programming Interface to estimate time (>1, <1 hour) and distance (>100, <100 km).	Negative – greater time associated with worse survival (unadjusted and adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Geographic impact on access to care and survival for non-curative esophagogastric cancer: a population-based study. Yee et al., 2021	10,228 oesophagogastric cancer patients. Canada.	Patient's residential postcode centroid. Nearest level 1, 2 or 3 cancer centre.	Straight-line distance (≤10,11-50,51-100, ≥101 km).	Mixed – greater distance associated with worse treatment. Moderate distance associated with worse survival (adjusted analyses).
Travel distance and overall survival in hepatocellular cancer care. Siegel et al., 2021	6860 hepatocellular cancer patients. USA.	Centre of the patient's ZIP code of residence. Street address of treating facility.	Straight line distance (<12.5, ≥50 miles).	Mixed – no association between distance and survival for patients treated at low-volume centres. Greater distance associated with worse survival for patients treated at high-volume academic centres (adjusted analyses).
Associations between geographic residence and USA adolescent and young adult cancer stage and survival. Johnson et al., 2021	178,688 cancer patients aged between 15-39. USA.	Patient's residential ZIP code centroid or city. Case reporting hospital.	Straight line distance (≤12.5, >12.5-<50, ≥50 miles).	Negative – greater distance associated with worse survival (unadjusted and adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Geographic variation in diagnostic and treatment interval, cancer stage and mortality among colorectal patients - An international comparison between Denmark and Scotland using data-linked cohorts. Murchie et al., 2021A.	4714 Danish and 1184 Scottish colorectal cancer patients. Denmark and Scotland.	Patients home address using population-weighted datazone centroids (derived from postcodes) in Scotland and street addresses in Denmark. GP and Hospital using postcodes in Scotland and street addresses in Denmark.	ArcGIS to estimate fastest travel times (categorical variable cut at the 25th, 50th, 75th and 90th percentile and continuous variables also used).	Mixed – moderate travel times associated with worse survival in Scotland and increased travel up to an hour associated with improved survival in Denmark. U-shaped association with time to GP (adjusted analyses).
Is place or person more important in determining higher rural cancer mortality? A data-linkage study to compare individual versus area-based measures of deprivation. Murchie et al., 2021B.	11,803 breast, ovarian, cervical, prostate, melanoma, lung, oesophagastric or colorectal cancer patients. Scotland.	Patients' residential postcode. Postcode of GP and treatment centre.	ArcGIS to estimate time to GP (<5,5-9.9,10-14.9,>15 mins) and treatment centre (<15, 15-29.9, 30-59.9, >60 mins) unless patient was an island dweller.	Negative – greater distance to treatment centre associated with worse survival (adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Impact of travel burden on clinical outcomes in lung cancer. Stoyanov et al., 2022.	9240 lung cancer patients. Bulgaria.	City of residence. Treating facility.	Google maps to estimate distance to treatment centre (same city, <50km, ≥50km) and time to treatment centre (same city, <60 mins, ≥)	Negative – greater distance and time to treatment centre associated with worse survival (adjusted analyses). Null - no association between distance and stage at diagnosis (adjusted analyses).
Impact of travel burden on the treatment of stage I and II breast cancer: A National Cancer Database analysis. Perry et al., 2023	293,318 breast cancer patients. USA.	Patient's residential ZIP code. Treating hospital.	Straight line distance (≤ 20, >20 miles).	Null - no association between distance and survival (adjusted analyses).
Impact of travel distance on outcomes for clinical trial patients: the Kinghorn Cancer Centre experience. Lim et al., 2023	173 solid-organ malignancy patients. Australia.	Patient's home address. Cancer centre (clinical trial).	Method to estimate distance not stated (0-10,10-50,>50 km).	Null - no association between distance and survival (adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
Cancer survival and travel time to nearest reference care center for 10 cancer sites: an analysis of 21 French cancer registries. Gardy et al., 2023.	160,634 patients with; breast, colon-rectum and anal, pancreas, prostate, skin melanoma, bladder, head and neck, kidney or liver cancer. France.	Residential address (X, Y coordinates). Nearest University Hospital or Cancer Control Centre.	ArcGIS and a road-Network database to estimate travel time (<30, 30-59, 60-89, >90 minutes for descriptive results and continuous for survival models).	Negative – greater time to cancer centre associated with worse survival; lung cancer, melanoma, breast cancer (adjusted analyses). Reverse U-Shape – greater time to cancer centre associated with worse survival for those in the 'middle'; liver cancer, colon-rectum cancer in males, head and neck cancer in females (adjusted analyses). Null – no association between greater time to cancer centre and survival; pancreas cancer, kidney cancer, bladder cancer, head and neck cancer in men and colon-rectum cancer and anal cancer in women (adjusted analyses). Positive - greater time to cancer centre associated with better survival; prostate cancer (adjusted analyses).

Title, author and year of publication	Population and setting	Origin and destination point	Method to estimate travel time or distance	Direction of association
The impact of travel time to cancer treatment centre on post-diagnosis care and mortality among cancer patients in Scotland. Turner et al., 2023.	17, 639 patients diagnosed with; breast, cervical, colorectal, lung, melanoma, ovarian, prostate or oesophageal cancer. Scotland.	Patients' residential postcode. Main regional cancer treatment centre.	Google API to estimate travel time (<15, 15-29.9, 30.0-59.9 and >60.0 minutes and an 'island-dwellers' category).	Negative – those with the furthest to travel (island dwellers) had worse survival (adjusted analyses).
Access to primary care and mortality in excess for patients with cancer in France: Results from 21 French Cancer Registries. Gardy et al., 2024.	151,984 cases diagnosed with; breast, colon-rectum and anal, lung, pancreas, prostate, skin melanoma, bladder, head and neck, kidney or liver cancer. France.	Patients residential address using X and Y coordinates. Primary care resources.	Two indices that consider travel-distance to the facilities and accessibility of resources based on supply and demand.	Negative – those living in areas with lower access to primary care had worse survival for four of the ten cancers studied (adjusted for socioeconomic status)

3.3.1.1 Choice of origin

There is a lack of consistency in the literature regarding the origin used for travel calculations. Most studies based in the UK used the patient's residential postal code (Murage et al., 2017; Turner et al., 2017; Tataru et al., 2018; Murchie et al., 2021b) and many studies based in other countries used a similar measure, such as ZIP code for the USA (Lee et al., 2014; Mitchell and Conklin, 2015; Graboyes et al., 2018; Rauh et al., 2018; Xia et al., 2018; Siegel et al., 2021; Perry et al., 2023). This is the most accurate origin point to choose and if available should be used, however, postal code is highly identifiable and because of data confidentiality is not always available.

Some studies that did not use postal code or ZIP code used the centroid of small areas such as their Lower Layer Super-Output Areas (LSOA) (UK) (Murage et al., 2016) or residential ZIP code (USA) (Vetterlein et al., 2017; Graboyes et al., 2018; Gotfrit et al., 2020; Yee et al., 2021). These are convenient as this data is more readily available, however, using centroids of small areas as the origin point could introduce inaccuracies into travel calculations. This is because the boundaries of these areas are typically chosen based on number of people or households in an area rather than geographical size, for example, an LSOA covers approximately 1000 to 3000 people as opposed to a certain number of square-metres. Consequently, the size of an area represented by a single centroid can vary substantially depending on the population density and travel estimates for those living in sparsely populated rural areas are more likely to be inaccurate compared to their urban counterparts as the distance between the centroid and the subject’s actual residence could be much greater. Another issue with using aggregate level data to estimate distance is the possibility of ecological fallacy, which occurs when the characteristics of a group are assumed to be the same as the individuals within it (Mizen et al., 2015). In this case, if the centroid of an administrative area is used as the origin point for calculating travel distance to a destination, such as a hospital, the same distance will be calculated for everyone who resides in that area. If the hospital is close to the centroid this would result in everyone from this area being assigned a short travel time to the hospital, but this might not be the case for some patients. This could lead to erroneous conclusions being drawn, for example, if cancer patients from the administrative area have poor survival compared to other areas, it could be concluded that short travel time is associated with worse survival. This may not be the case and in reality, it could be those who live furthest away from the hospital that are driving the poor survival. Therefore, care must be taken when using such measures and the smallest unit of aggregation available should be chosen to minimise bias.

A study that investigated the impact of travel time on survival of patients with lung cancer in Bulgaria used the patients 'city of residence' as the origin point for their journey to treatment centre (Stoyanov et al., 2022). The authors went to considerable effort to make this a robust estimate, choosing to run the calculations in google maps as they believed this would yield the most accurate travel estimates with an up-to-date road Network and consideration of the mountainous terrain. They also repeated the calculations several times under different conditions and used the average values to group patients into broad categories for distance and time. However, the usefulness of these travel estimates is questionable, as with such a broad origin point misclassification seems likely and the ability to discern meaningful differences between groups diminished.

3.3.1.2 Choice of destination

The choice of destination is influenced by each study's hypothesis. If the hypothesis is that distance is acting as a barrier to help-seeking and diagnosis, the researchers will probably choose a primary care facility as the destination (Murage et al., 2017, 2019; Turner et al., 2017). For example, a study based in England that investigated the association between travel time to healthcare and RTD for patients with eight different cancer types used the patients General Practitioner (GP) as the destination for the travel calculations (Murage et al., 2019). However, if the hypothesis is that travel to healthcare is associated with survival, the hospital is likely to be chosen as the destination (Graboyes et al., 2018; Beal et al., 2020; Gardy et al., 2023). For example, a study in the USA investigated the hypothesis that travel to treating facility influenced the treatment that patients with prostate cancer received and subsequently their mortality (Vetterlein et al., 2017). Another possible destination point is nearest healthcare facility, as opposed to the actual facility attended by the subject. A study that investigated the impact of distance on stage of presentation of patients with head and neck cancer in the USA used nearest facilities with the justification that some patients will choose to travel further to receive care at hospitals they perceive as better (Ringstrom et al., 2018).

3.3.1.3 Estimating distance or time between origin and destination points

The two most widely used methods for calculating travel distance or time between origin and the destination points is to calculate the straight-line distance, commonly referred to as 'as the crow flies' (e.g., using the Haversine formula), or use Geographical Information Systems (GIS) to estimate the road network distance between.

Many of the studies based in the USA use National Cancer Database (NCDB) data which includes the 'great circle' distance, the straight-line distance between two points on a sphere,

between the centroid of a patient's residential ZIP code and the centroid of their reporting facility's ZIP code (Beal et al., 2020; Vetterlein et al., 2017; Xia et al., 2018; Yee et al., 2021). This is a simplistic method that fails to consider the features of the road network and land that may influence travel distance or time. However, a study found that there was high correlation between straight-line distance and road distance for non-Emergency travel to community hospitals using data from the USA, Columbia and Puerto Rico (Boscoe, Henry and Zdeb, 2012).

Other studies use GIS software, such as ArcGIS, to estimate travel by private car between a set of origin and destination points (Dejardin et al., 2014; Guyader-Peyrou et al., 2017; Murage et al., 2017; Turner et al., 2017; Ringstrom et al., 2018; Tataru et al., 2018; Murchie et al., 2021a). Although many of these softwares are sophisticated and provide good estimates of journey distance and times, for example by adjusting the estimated times based on the time of day, they are not capable of calculating the actual times experienced by patients which will depend on many factors including the route taken and traffic conditions on a specific day. However, there is evidence of good concordance between reported times and those estimated by GIS (Haynes et al., 2006). Another issue with using private car estimates generated in GIS to investigate the influence of travel on patients' outcomes is the assumption that no patients have used public transport, patient transport, such as those provided by the health service or charities, or active transport, which includes walking and cycling. Making a journey by public transport may take substantially more time than by private car, with some journeys potentially requiring multiple buses, trains or trams. A study reported that estimated journey time between residential centroids and public libraries using public transport in Helsinki was over twice the estimate for private car in their model (Salonen and Toivonen, 2013). Another UK study found that using public transport considerably increased travel time to treatment facility for cancer patients, for example, in Yorkshire and the Humber 8.3% of journeys by private car exceeded one hour compared to 39.8% for public transport (Han et al., 2024). Therefore, using only private car estimates could introduce bias into a study if certain groups, such as those with low SES and the elderly, are more likely to use public transport because they do not own a car, the cost of fuel or parking, difficulty driving in city centres or other reasons. Unfortunately, it is unlikely that patient travel method is recorded in the datasets used by many studies, however, Haynes et al. reported that 87% of cancer patients travel by car while only 5% used the bus, making private car estimates the most appropriate measure in lieu of travel method data (Haynes et al., 2006).

Few studies used web-based mapping methods, such as Google maps (Lee et al., 2014; Haddad et al., 2015; Tøstesen et al., 2019; Stoyanov et al., 2022), to calculate distance or time. This is acceptable if using non-identifiable data, such as centroids of large census areas, but is not appropriate when using highly identifiable data such as postal code of residential address.

A novel approach to estimate the impact of accessibility to primary care resources on mortality for patients with ten commonly diagnosed cancers in France was to use two indices that consider both travel-distance to the facility and accessibility of resources based on supply and demand; the Spatial aCcessibility multiscALar index (SCALe) and the Accessibilité Potentielle Localisée (APL). SCALe is defined as the average distance by road that the population in a geographical area has to travel to access care, weighted by the theoretical pressures on these facilities and APL is a measure of accessibility to GPS that considers supply and demand in the local area and surrounding areas (Gardy et al., 2024).

3.3.1.4 Analysis of travel data

Studies also differ in how they use travel data in their analyses. Few studies kept distance or time as a continuous variable (Murage et al., 2017; Zurko, Wade and Mehta, 2019; Dhakal et al., 2020; Murchie et al., 2021a; Gardy et al., 2023) and the majority used arbitrary cut-off points to convert the continuous travel variables into categorical variables, probably because this makes the results easier to interpret (Lee et al., 2014; Haddad et al., 2015; Mitchell and Conklin, 2015, 2015; Vargo et al., 2015; Guyader-Peyrou et al., 2017; Rauh et al., 2018; Tataru et al., 2018). Many studies assigned patients to a category based on their estimated travel distance, e.g., short (<=12.5 miles), intermediate (>12.5 & <50 miles) and long (>=50 miles) and compared outcomes between these groups (Johnson et al., 2021). Unfortunately, there is a lack of consensus on what these cut-offs should be, which makes comparison of studies challenging. Furthermore, there is no standardised method for how they should be determined and those selected may vary in accordance with what the researchers themselves perceive to be an acceptable distance, which will also vary by region. For example, individuals who live in large countries such as the USA and Australia may be used to travelling further for services than those living in the UK and therefore may have different expectations of the distances they must travel for care.

Another way in which the impact of travel on outcomes was investigated was to focus on subjects with the most extreme distances or times (Siegel et al., 2021). For example, Siegel at el., compared those living less than 12.5 miles from their treating facility to those over 50 miles away. This is a useful method if most subjects have short journeys and the hypothesis of the study is that only those with the most extreme travel will have different outcomes.

3.3.2 Defining rural-urban location

As described in section 1.5.4., defining whether an area is rural or urban can be challenging because of different perceptions of what it means to be rural. Furthermore, a criterion that is suitable to define rurality in the UK, may not be appropriate for use in other countries such as the USA, making it difficult to compare different regions. However, even studies in the same region display considerable variation in the definitions and the methods used to measure rurality, as illustrated in Table 3:3.

The influence of rurality on cancer outcomes has been investigated more extensively in the USA and Australia than the UK, which is likely a reflection of the size of these countries and the greater number of residents residing in rural areas. Consequently, most of the studies identified in this review are not based in the UK.

Table 3:3. Studies included in the review that investigated the influence of rurality on diagnostic route of cancer patients.

Title, author and year of publication	Population and setting	Method used to define urban-rural residence	Direction of association
Effect of place of residence and treatment on survival outcomes in patients with diffuse large B-cell lymphoma in British Columbia. Lee et al., 2014	1,357 DLBCL patients. Canada.	Used cut-offs based on population density as defined by Statistics Canada.	Mixed – no difference in survival between patients in rural and large urban areas, but those in small and medium urban areas had worse survival (adjusted analyses).
Rural and urban patients with diffuse large B-cell and follicular lymphoma experience reduced overall survival: A National Cancer DataBase study. Ritter et al., 2019	83,108 DLBCL patients and 43292 follicular lymphoma patients. USA.	Rural-Urban Continuum Codes (RUCC).	Mixed – urban and rural residence associated with worse survival compared to metro residence (unadjusted and adjusted analyses).
Prostate cancer in Scotland: does geography matter? An analysis of incidence, disease characteristics and survival between urban and rural areas. Laing et al., 2014.	3308 prostate cancer patients. Scotland.	Based on health board – patients under the care of NHS Highland and Western Isles classified as rural and patients under the care of NHS Lothian classified as urban.	Negative – rurality associated with worse survival (did not reach statistical significance).

Title, author and year of publication	Population and setting	Method used to define urban-rural residence	Direction of association
Survival of cancer patients in urban and rural areas of Germany--a comparison. Nennecke et al., 2014	817,182 cancer patients (16 cancer types included). Germany.	District types of settlement structure defined by the Federal Institute for Research on Building, Urban Affairs and Spatial Development.	Mixed – some cancers showed differences in survival by district type, e.g., female breast cancer patients living in cities had better survival than other districts, but no association was detected for most cancers, including NHL (adjusted analyses).
Impact of travel time and rurality on presentation and outcomes of symptomatic colorectal cancer: a cross-sectional cohort study in primary care. Murage et al., 2017	926 colorectal cancer patients. Scotland.	Scottish government urban rural classification.	Mixed – rurality not associated with Emergency Presentation but associated with better survival (adjusted analyses).
Rural–Urban Disparities in Time to Diagnosis and Treatment for Colorectal and Breast Cancer. Bergin et al., 2018	433 colorectal cancer patients and 489 breast cancer patients. Australia.	Australian Statistical Geography Standard-Remoteness Areas index.	Mixed - rural residence was associated with longer total intervals for colorectal but not breast cancer; with most disparities post presentation (adjusted analyses).

Title, author and year of publication	Population and setting	Method used to define urban-rural residence	Direction of association
A population-based analysis of urban–rural disparities in advanced pancreatic cancer management and outcomes. Canale et al., 2018	659 pancreatic cancer patients. Canada.	Arbitrary cut-offs based on travel distance.	Null – no association between rurality and treatment or survival (adjusted analyses).
Geographic Distribution and Survival Outcomes for Rural Patients with Cancer Treated in Clinical Trials. Unger et al., 2018	36,995 patients from 17 different cancer-specific cohorts. USA.	RUCC	Null – rurality associated with worse survival in 1 of 17 cohorts included in the study (adjusted analyses).
Impact of rurality on processes and outcomes in melanoma care: results from a whole-Scotland melanoma cohort in primary and secondary care. Murchie et al., 2018	9519 melanoma patients. Scotland.	Scottish government urban rural classification.	Null – no association between rurality and survival (adjusted analyses).

Title, author and year of publication	Population and setting	Method used to define urban-rural residence	Direction of association
Disparity of colon cancer outcomes in rural America: making the case to travel the extra mile. Raman et al., 2019	647,949 colon cancer patients. USA.	RUCC	Negative – rurality associated with worse survival, although for patients who travelled to high-volume centres this association did not exist (adjusted analyses)
Differences in cancer survival by remoteness of residence: an analysis of data from a population-based cancer registry. Afshar et al., 2020	331,302 incident cancer cases (30 cancer types included). Australia.	Australian Bureau of Statistics remoteness structure.	Mixed – direction of association depended on cancer type, e.g., melanoma patients had better survival outside major cities but stomach cancer had better survival outside major cities (adjusted analyses).
Impact of geography on Scottish cancer diagnoses in primary care: Results from a national cancer diagnosis audit. Murchie et al., 2020	1905 cancer patients. Scotland.	Scottish Government 2-fold and 6-fold urban rural classification.	Mixed – rurality not significantly associated with diagnostic route or longer primary care intervals (date of first presentation to primary care with relevant symptoms to referral date) but rural patients had longer diagnostic intervals (date of presentation to date of diagnosis).

Title, author and year of publication	Population and setting	Method used to define urban-rural residence	Direction of association
Analysing the impact of living in a rural setting on the presentation and outcome of colorectal cancer. A prospective single centre observational study. MacVicar et al., 2020	2463 colorectal cancer patients. Scotland.	Scottish Government 2-fold and 6-fold and urban rural classification.	Null – more cancers detected through the bowel screening programme for rural patients and rurality not associated with survival.
Epidemiology of acute myeloid leukemia in Virginia: Excellent survival outcomes for patients in rural Appalachia. Isaac et al., 2021	163 AML patients. USA.	Rural-Urban Commuting Area (RUCA).	Null – no association between rurality and survival (adjusted analyses).
Disparity of ovarian cancer survival between urban and rural settings. Ulmer et al., 2022	111,627 ovarian cancer patients. USA.	National Center for Health Statistics urban–rural classification scheme for counties.	Negative – rurality associated with worse survival (adjusted analyses).

Title, author and year of publication	Population and setting	Method used to define urban-rural residence	Direction of association
Assessing the impact of rurality on oesophagogastric cancer survival in the North-East of Scotland- a prospective population cohort study. Griffin et al., 2023	1038 gastric cancer patients. Scotland.	Scottish Government urban rural classification.	Positive – rurality associated with better survival (adjusted analyses).
The impact of rurality on patient experience and diagnostic pathway intervals in Scotland's cancer patients: Further results from a national cancer diagnosis audit. Maxwell et al., 2023.	4309 cancer patients. Scotland.	Scottish Government 2-fold urban rural classification.	Mixed – rurality not associated with Emergency Presentation but rural patients had longer primary care intervals (date of first presentation to primary care with relevant symptoms to referral date) and longer diagnostic intervals (date of presentation to date of diagnosis).
Disparities in cancer stage of diagnosis by rurality in California, 2015-2019. Oh et al., 2024.	497,559 patients with; breast, prostate, lung, colorectal cancer or melanoma. USA.	Census tract aggregation zones.	Negative – increasing rurality associated with increased odds of late-stage diagnosis for breast cancer, lung cancer and melanoma (male only).

Title, author and year of publication	Population and setting	Method used to define urban-rural residence	Direction of association
Estimating the Impact of Rurality in Disparities in Cancer Mortality. Kenzik et al., 2024.	757,655 patients with; breast, lung, cervical, colorectal, prostate, endometrial, kidney, pancreas, ovarian, stomach, liver, oesophagus, thyroid, meningioma, oral cavity, larynx, or pharynx cancer. USA.	RUCC	Negative – increasing rurality associated with worse survival.

3.3.2.1 Travel-distance based

Some studies used travel distance to healthcare to define rurality. For example, a study that investigated urban-rural disparities in advanced pancreatic cancer in Canada defined an urban residential location as being 100 km or less away from a treatment centre and a rural location as being over 100 km away (Canale, Cho and Cheung, 2018). This is a simplistic measure of rurality that fails to fully capture all the potential barriers associated with geographical isolation, such as fewer public transportation options and worse infrastructure, which may compound the effect of having to travel further to access healthcare.

3.3.2.2 Scottish Government Urban Rural Classification

The Scottish Government Urban Rural Classification system defines rural areas in Scotland based on two factors; accessibility, based on drive time analysis, and population (Geographic Information Science & Analysis Team and Rural & Environment Science and Analytical Services Division, 2022). Two formats are available, a six-level format and an eight-level format, both of which comprise of four main groups; large urban areas, other urban areas, small towns and rural areas. The small towns and rural areas groups are further divided into accessible and remote for the six-level version and accessible, remote and very remote for the eight-level version.

This is a popular method to assign rurality based on the patient's home address postcode and is used by many Scottish studies (Murage et al., 2017; Murchie et al., 2018; MacVicar et al., 2020; Griffin et al., 2023; Maxwell et al., 2023). The categories are often collapsed into two, rural and urban, for example a paper that described the impact of rurality on outcomes of patients with colorectal cancer used the 2003-2004 version of this system to classify patients' residence as either rural or urban (Murage et al., 2017).

3.3.2.3 Health boards

One Scottish study used health boards, of which there are 14 in Scotland, to create an urban and rural cohort of patients and the outcomes of patients belonging to the 'rural' and 'urban' health boards were compared (Laing et al., 2014).

3.3.2.4 Rural Urban Classification – England

Although no studies included in this review used the classification system, it would be appropriate to use the Office for National Statistics Rural-Urban Classifications, described in section 1.5.4.1, to define rurality in studies based in England. This classification can be used for different levels of census output areas, including LSOAs. Eight different types of settlements are recognised at the LSOA level, characterised by their form and population

density; urban major conurbation, urban minor conurbation, urban city and town, urban city and town in a sparse setting, rural town and fringe, rural town and fringe in a sparse setting, rural village and dispersed, rural village and dispersed in a sparse setting. These eight classifications can be broadly split in two; urban and rural.

3.3.2.5 Rural-Urban Commuting Area – USA

The Rural-Urban Commuting Area (RUCA) codes were used by one study in this review to define rurality (Isaac et al., 2021). They were devised by the United States Department of Agriculture and use data from the census and American Community Survey on population density, daily commuting and urbanisation to categorise census tracts, which are small areas that represent approximately 4000 inhabitants, into ten codes (U.S. Department of Agriculture, Economic Research Service, 2023a). Instead of using all ten codes, Isaac et al., 2021 used a collapsed version of the codes to compare characteristics and outcomes of 'rural' and 'urban' patients. A limitation of RUCA codes is that they are updated once per decade in line with the census, so given how rapidly urbanisation can occur, may not accurately reflect the characteristics of an area for the entire census period, which could lead to the misclassification of individuals.

3.3.2.6 Rural-Urban Continuum Codes – USA

The Rural-Urban Continuum Codes (RUCC) can be used to define the rurality of patients in studies based in the USA (Unger et al., 2018; Raman et al., 2019; Ritter et al., 2019; Kenzik et al., 2024). RUCC classifies all USA counties into one of nine categories; three metro groups based on population size and six non-metro groups based on both population size and adjacency to metro areas (U.S. Department of Agriculture, Economic Research Service, 2023b). Despite a large number of categories RUCC remains a crude measure of rurality for individual patients because they are assigned at the county-level, some of which can have extremely large populations and cover large geographical regions. For example, San Bernardino County has a population of over 2 million people and covers an area of ~20,000 square miles, but has a single RUCC, the 'Metro – counties in metro areas of 1 million population or more'. Furthermore, many studies group together similar RUCC codes to create fewer categories for their analyses which exacerbates this problem by further reducing the granularity in the categories. For example, a study examining differences in survival between rural and urban DLBCL patients collapsed the nine RUCCs into 3 groups; metro, urban and rural (Ritter et al., 2019).

3.3.2.7 National Center for Health Statistics Urban–Rural Classification Scheme – USA

Another method akin to the RUCC system is the National Center for Health Statistics Urban–Rural Classification Scheme for Counties. This classifies counties in the USA into six groups that can be collapsed into broader groups if required based on the Office of Management and Budget standards for defining metropolitan statistical areas and population-based data from the census (Ulmer et al., 2022). Again, this is a simplistic measure of assigning rurality at an individual level and is susceptible to ecological fallacy. However, it does have the advantage over RUCCs of distinguishing between locations within large metropolitan areas with populations over one million, e.g., central areas or the suburbs. This is an important distinction for health research that the RUCC codes fail to make, as the environment of patients that reside in these two areas will likely be quite different.

3.3.2.8 Census tract aggregation zones – USA

A novel approach to assigning rurality adopted by a study of cancer patients in the USA was to use population-weighted census-tract level rurality data from the US Census to calculate rurality for zones (Oh et al., 2024). Zones had a minimum population of 50,000 and were generated using software that identified geographical areas which were similar in terms of ethnicity, poverty and rurality. Zones were categorized into seven levels, depending on the proportion of residents in rural blocks within them, beginning at 0% (not rural) and increasing in 10% increments from 0% to 50% and above.

3.3.2.9 Australian Statistical Geography Standard Remoteness Structure

The Australian Bureau of Statistics classifies Australia into a hierarchy of statistical areas, with levels that directly aggregate to the next. Remoteness is assigned at the ‘statistical area level 1’ level (~400 people) and comprises five categories; major cities, inner regional, outer regional, remote and very remote (Australian Bureau of Statistics, 2021). A dichotomised version of this system was used to compare cancer survival between those that live in and outside of major cities (Afshar et al., 2020).

3.3.2.10 Population Centre and Rural Area Classification – Canada

A study based in Canada that investigated the effect of place of residence on treatment and survival of patients with DLBCL used the Canadian Population Centre and Rural Area Classification to categorise patients' residence into large urban, medium urban, small urban and rural areas (Lee et al., 2014). This uses data from the census and Statistics Canada’s Business Register to organise areas into these four groups based on their population size, population density and employment density.

3.3.3 Describing socioeconomic status

Socioeconomic status is a construct that we use to describe how disadvantaged one person or group of people are relative to others. There are many different ways in which researchers attempt to capture the SES of patients in their study, which can be broadly separated into two categories, individual measures and area-based measures, as summarised in Table 3:4. Individual measures use data that directly relates to each person in the study, for example their personal income, whereas area-based measures relate an individual's SES to the characteristics of the area they live in; in the UK this is often a LSOA.

Table 3:4. Studies included in the review that investigated the influence of socioeconomic status on diagnostic route or survival of cancer patients.

Title, author and year of publication	Population and setting	Individual or Area -based measure	Method used to describe SES	Direction of association
Predictors of delay in diagnosis and treatment in diffuse large B-cell lymphoma and impact on survival. Nikonova et al., 2014	278 DLBCL patients. Canada.	Area-based.	Median income for postal code using census data.	Null – no association between deprivation and diagnostic or treatment delay (unadjusted and adjusted analyses)
Socioeconomic disparities in mortality after diffuse large B-cell lymphoma in the modern treatment era. Tao et al., 2014	33,032 DLBCL patients. USA.	Area-based.	Composite neighbourhood SES score based on patients' residential census-block group (Yost index).	Negative – greater deprivation associated with worse survival (adjusted analyses).
Impact of age and socioeconomic status on treatment and survival from aggressive lymphoma: a UK population-based study of diffuse large B-cell lymphoma. Smith et al., 2015	2137 DLBCL patients. England.	Area-based.	Income domain of IMD.	Null – no association between deprivation and receiving treatment with curative intent or survival (adjusted analyses).

Title, author and year of publication	Population and setting	Individual or Area -based measure	Method used to describe SES	Direction of association
Emergency admission and survival from aggressive non-Hodgkin lymphoma: A report from the UK's population-based Haematological Malignancy Research Network. Kane et al., 2017	1660 DLBCL patients. England.	Area-based.	Income domain of IMD.	Null – no association between deprivation and Emergency Presentation (unadjusted analyses).
Factors related to the relative survival of patients with diffuse large B-cell lymphoma in a population-based study in France: does socioeconomic status have a role? Guyader-Peyrou et al., 2017	1,312 DLBCL patients. France.	Area-based.	French ecological European Deprivation Index (EDI).	Null – no association between deprivation and survival (adjusted analyses).
No outcome disparities in patients with diffuse large B-cell lymphoma and a low socioeconomic status. Boslooper et al., 2017	343 DLBCL patients. Netherlands.	Area-based.	Composite score provided by “Sociaal Cultureel Planbureau (SCP)”.	Null – no association between deprivation and treatment or survival (adjusted analyses).

Title, author and year of publication	Population and setting	Individual or Area -based measure	Method used to describe SES	Direction of association
Investigating the inequalities in route to diagnosis amongst patients with diffuse large B-cell or follicular lymphoma in England. Smith et al., 2021A.	15,551 follicular lymphoma and 30,078 DLBCL patients. England.	Area-based.	IMD.	Negative – greater deprivation associated with Emergency Presentation (unadjusted and adjusted analyses).
Association between multimorbidity and socioeconomic deprivation on short-term mortality among patients with diffuse large B-cell or follicular lymphoma in England: a nationwide cohort study. Smith et al., 2021B.	14,043 follicular lymphoma and 27,379 DLBCL patients. England.	Area-based.	IMD.	Negative – greater deprivation associated with greater mortality (unadjusted and adjusted analyses).
Socioeconomic inequalities in treatment and relative survival among patients with diffuse large B-cell lymphoma: a Hong Kong population-based study. Lee et al., 2021	4017 DLBCL patients. Hong Kong.	Household level.	Based on the need for welfare assistance.	Negative – greater deprivation associated with reduced use of chemotherapy and/or rituximab and worse survival (unadjusted and adjusted analyses).

Title, author and year of publication	Population and setting	Individual or Area -based measure	Method used to describe SES	Direction of association
Association of socioeconomic status with autologous hematopoietic cell transplantation outcomes for lymphoma. Hong et al., 2016	154 Hodgkin lymphoma, 262 DLBCL, 102 follicular lymphoma, 99 mantle cell lymphoma, 52 T-cell lymphoma and 18 other non-Hodgkin lymphomas. USA	Area-based.	Median annual household income by ZIP-code from census data.	Negative – greater deprivation associated with worse survival (unadjusted and adjusted analyses).
Affluence and private health insurance influence treatment and survival in non-Hodgkin's lymphoma. Comber et al., 2016	2973 NHL patients. Ireland.	Area-based.	Pobal HP area-based deprivation score.	Negative -greater deprivation associated with greater frequency of Emergency Presentation and indirectly increased mortality (through late stage and Emergency Presentation).
Impact of race, ethnicity, and socioeconomic status over time on the long-term survival of adolescent and young adult hodgkin lymphoma survivors. Berkman et al., 2021	15,899 Hodgkin lymphoma patients. USA.	Area-based.	County-level composite score for SES using method described by Truong et al.	Negative – greater deprivation associated with worse survival (adjusted analyses).

Title, author and year of publication	Population and setting	Individual or Area -based measure	Method used to describe SES	Direction of association
Association of individual low-income status and area deprivation with mortality in multiple myeloma. Hong et al., 2023	7637 multiple myeloma patients. USA.	Area-based and Individual.	Area-based - county-level Social Deprivation Index was used (Robert Graham Center). Individual-based - low income was defined as patients that were dually eligible and enrolled in Medicare and Medicaid.	Mixed – mixed interactions between individual deprivation and area-based deprivation (adjusted analyses).
Emergency presentation and socioeconomic status in colon cancer. Gunnarsson, Ekholm and Olsson., 2013	12,293 colon cancer patients. Sweden.	Individual.	Information on patients' highest attained levels of education and annual income was obtained from a continuously updated longitudinal integrated database on labour market research (LISA).	Negative – greater deprivation (income) associated with greater odds of Emergency Presentation (unadjusted and adjusted analyses).
Time trends in socioeconomic inequalities in cancer mortality: results from a 35-year prospective study in British men. Ramsay et al., 2014	7489 men followed up for 35 years and 1484 cancer deaths occurred. UK.	Individual.	Longest-held occupation of subjects at study entry was used to define social class using the Registrar Generals' Social Class Classification.	Negative – greater deprivation (manual vs. non-manual social class) associated with greater cancer mortality (adjusted analyses).

Title, author and year of publication	Population and setting	Individual or Area -based measure	Method used to describe SES	Direction of association
No socioeconomic inequalities in ovarian cancer survival within two randomised clinical trials. Abdel-Rahman et al., 2014	1290 ovarian cancer patients. England and Wales.	Area-based.	Carstairs index for those diagnosed before 1996. Income domain of IMD for those diagnosed from 1996.	Null – no association between deprivation and survival (unadjusted and adjusted analyses).
Cancer-specific variation in Emergency Presentation by sex, age and deprivation across 27 common and rarer cancers. Abel et al., 2015	749,645 cancer patients (27 different types). England.	Area-based.	Income domain of IMD.	Negative – greater deprivation associated with greater odds of Emergency Presentation for 24/27 cancers (unadjusted and adjusted analyses).
The role of receipt and timeliness of treatment in socioeconomic inequalities in lung cancer survival: population-based, data-linkage study. Forrest et al., 2015	22,967 lung cancer patients. England	Area-based.	Income domain of IMD.	Negative – greater deprivation associated with worse treatment and survival (adjusted analyses).

Title, author and year of publication	Population and setting	Individual or Area -based measure	Method used to describe SES	Direction of association
Impact of marital status, insurance status, income, and race/ethnicity on the survival of younger patients diagnosed with multiple myeloma in the United States. Costa et al., 2016	10,161 patients with multiple myeloma. USA.	Area-based.	Individual SES variables for the county of residence.	Negative – greater deprivation associated with increased mortality (unadjusted and adjusted analyses).
Impact of treatment and insurance on socioeconomic disparities in survival after adolescent and young adult Hodgkin lymphoma: A population-based study. Keegan et al., 2016	9353 Hodgkin lymphoma patients. USA.	Area-based.	Composite neighbourhood SES score based on patients' residential census-block group (Yost index).	Negative – greater deprivation associated with worse survival (adjusted analyses).
Incidence, socioeconomic deprivation, volume-outcome and survival in adult patients with acute lymphoblastic leukaemia in England. Maheswaran and Morley., 2018	2921 adult ALL patients. England.	Area-based.	Income domain of IMD.	Negative – greater deprivation associated with worse survival (adjusted analyses).

Title, author and year of publication	Population and setting	Individual or Area -based measure	Method used to describe SES	Direction of association
Socioeconomic differences in selection for liver resection in metastatic colorectal cancer and the impact on survival. Vallance et al., 2018	13,656 colorectal cancer patients. England.	Area-based.	IMD.	Negative – greater deprivation associated with lower rates of resection and worse survival (adjusted analyses).
The impact of socioeconomic factors on treatment choice and mortality in chronic myeloid leukaemia. Larfors et al., 2017	980 chronic myeloid leukaemia patients. Sweden.	Individual.	Individual SES variables from the integrated database for labour market research (LISA).	Null – no association between deprivation and survival (adjusted analyses). (Greater deprivation associated with worse survival in unadjusted analyses.)
Does geodemographic segmentation explain differences in route of cancer diagnosis above and beyond person-level sociodemographic variables? Bright et al., 2020	36,194 lung cancer and 32,984 colorectal cancer patients. UK	Area-based and household-level.	IMD linked at LSOA level and Mosaic consumer classification group at household or postcode level.	Negative – greater deprivation associated with greater odds of Emergency Presentation, some differences in odds of Emergency Presentation between mosaic groups (unadjusted and adjusted analyses).

Title, author and year of publication	Population and setting	Individual or Area -based measure	Method used to describe SES	Direction of association
Socioeconomic deprivation is associated with decreased survival in patients with acute myeloid leukemia. Le Floch et al., 2020	209 acute myeloid leukaemia patients. France.	Individual.	Composite SES score from the EPICES (Evaluation de la Précarité et des Inégalités de santé dans les Centres d'Examens de Santé – translated as 'Evaluation of Deprivation and Inequalities in Health Examination Centres') questionnaire.	Negative – greater deprivation associated with worse survival (unadjusted and adjusted analyses).
Impact of race, ethnicity, and socioeconomic status over time on the long-term survival of adolescent and young adult hodgkin lymphoma survivors. Berkman et al., 2021	15,899 Hodgkin lymphoma patients. USA.	Area-based.	County-level composite score for SES using method described by Truong et al.	Negative – greater deprivation associated with worse survival (adjusted analyses).
An investigation of cancer survival inequalities associated with individual-level socioeconomic status, area-level deprivation, and contextual effects, in a cancer patient cohort in England and Wales. Ingleby et al., 2022	9276 breast, prostate or colorectal cancer patients. UK.	Area-based and Individual.	Income domain of IMD linked at LSOA level and individual-level data from the census.	Mixed – mixed associations by cancer type and sex. Some cancers exhibited an interaction between individual and area-based measures of deprivation (adjusted analyses).

3.3.3.1 Individual measures

Individual-based measures of SES are advantageous because the data directly relates to each person in the study, however, they are seldom used because it is difficult to acquire data relating to SES at an individual level. Few studies identified in this review used individual measures of SES but those that did were either based in countries such as Sweden where this type of data is routinely collected or relied on additional methods to estimate individual SES such as the EPICES questionnaire (Gunnarsson, Ekholm and Olsson, 2013; Le Floch et al., 2020; Larfors et al., 2017). A UK study examined cancer survival inequalities using both the IMD and individual-level data relating to socioeconomic status which they obtained from the Census (Ingleby et al., 2022).

In the past, simplistic individual measures of SES based on the occupation of the primary earner in the household such as that devised by the Registrar General's social class scheme were used to describe 'social grade' (Carr-Hill and Pritchard, 1992). This measure fails to fully capture the many facets of SES and only a single paper in this review used social class to estimate socioeconomic status (Ramsay et al., 2014). More sophisticated measures that use a combination of different factors such as income, health and educational attainment are more popular in health research, as described below.

3.3.3.2 Composite area-based measures

Area-based measures are convenient as the data is widely available, however, they are vulnerable to ecological fallacy, which can lead to the misclassification of subjects in the study and erroneous conclusions.

Many papers based in the UK in the 1990's and early 2000's used either the Townsend or the Carstairs method for determining SES, which are described in section 1.5.2.1. These measures are now rarely used and only one paper in this review utilised the Carstairs index (Abdel-Rahman et al., 2014).

The most widely used measure of SES in recent UK papers identified in this review is the Index of Multiple Deprivation (IMD) which is described in section 1.5.2.2 (Abel et al., 2015; Bright et al., 2020; Forrest et al., 2015; Smith et al., 2021a; Vallance et al., 2018, 2018). Studies that use the IMD to define SES can use the composite score for all domains combined or choose the domain most pertinent to their research question. Many health research papers use the income domain of the IMD (Abel et al., 2015; Smith et al., 2015; Maheswaran and Morley, 2018). This is likely because inclusion of other domains,

particularly the health domain, could potentially invalidate their analyses by artificially inflating the association between deprivation and the health outcome they are interested in.

Similar area-based composite measures of SES that utilise administrative data are used in studies in the USA and elsewhere. This includes the Yost index (USA) which was used by one paper in this review (Tao et al., 2014) and has seven indicator variables including 'median household income' and 'proportion older than 16 in the workforce without a job' (Yost et al., 2001). The French European Deprivation Index is also similar, combining ten weighted census-derived variables (Pornet et al., 2012) and was used by a study that examined the influence of socioeconomic status on the outcomes of patients with DLBCL in France (Guyader-Peyrou et al., 2017).

To further complicate matters, there is some evidence to suggest that area-level and individual-level effects act separately to influence cancer outcomes (Woods et al., 2022). Therefore, it could be argued that to fully capture the influence of SES on outcomes, data relating to a patient's personal SES and that of the area they reside should be used. It is also possible that in addition to acting separately, there is interaction between the SES of the individual and the SES of the place they live. A lack of healthcare services in deprived areas, for example, may disproportionately affect the most deprived individuals who may not have the resources to travel to seek help elsewhere. Evidence from multiple myeloma patients in the USA suggests that there is an interaction between individual low-income status and area deprivation on mortality, although this relationship was not linear (Hong et al., 2023).

3.3.3.3 Geodemographic segmentation

A novel approach to describe SES used in a paper investigating sociodemographic differences in RTD is Experian's mosaic customer segmentation groups (Bright et al., 2020). This classifies households in the UK into 15 broad lifestyle groups, which can be further split into 66 more detailed groups, based on a combination of factors including age, home-ownership and lifestyle choices. These include 'Vintage Value' which are elderly people who mostly live alone and have low incomes and 'Aspiring Homemakers' who are younger households who typically live in affordable housing in private suburbs.

3.4 Associations between travel, socioeconomic status and urban-rural location with diagnostic route and survival

3.4.1 Travel and diagnostic route

The association between travel to healthcare and the diagnostic route of patients with cancer has been inconsistently reported. A large, cross-sectional study in England reported associations between longer journeys to GP and a 'less desirable' route to diagnosis (Emergency Presentation and Death Certificate Only) across eight cancer sites (Murage et al., 2019). For example, they found that for every ten-minute increase in travel time, there was a statistically significant increase in the relative risk of being diagnosed via the Emergency route for breast cancer patients. The dose-response relationship between increasing travel and Emergency Presentation supports the suggestion of a causal relationship according to principle five of the Bradford-Hill criteria for causality - biological gradient. This states that a dose-response relationship between an environmental exposure and an outcome strengthens the argument that an association truly exists between them (Hill, 1965). However, they also reported that increased travel was associated with reduced likelihood of being diagnosed via the Screen-Detected and Two-Week Wait routes, which are associated with good prognosis and therefore are not 'less-desirable' routes. Furthermore, the analyses were not adjusted for stage at diagnosis, a likely confounder of this relationship.

A study that examined geographic variation in the diagnostic interval and stage of patients with colorectal cancer in Scotland and Denmark reported no association with diagnostic interval (time between GP referral and cancer diagnosis) or advanced stage disease and travel time to GP (Murchie et al., 2021a). However, a reverse U-shaped relationship between increasing travel-time to hospital and advanced stage disease was detected for Scottish patients only, with an increase in odds for those travelling up to 40 minutes and then a reduction in odds for those with longer travel times, compared to those with a travel-time of 15 minutes. They also reported that increased travel time to the hospital was associated with a shortened diagnostic interval in both countries. Another Scottish study reported that compared to those who lived within 15 minutes of their cancer centre, patients who travelled from an island to the mainland or who travelled further on the mainland to access their treatment centre were more likely to be treated within 62 days of GP referral (Turner et al., 2017).

As described in section 1.4.4, patients with non-specific symptoms, such as those with DLBCL, are more likely to be diagnosed at a later stage of disease and via an Emergency

Presentation (Pearson et al., 2020). This relationship may be compounded by the need to travel further to access care, however, a limited number of studies have examined the relationship between RTD of DLBCL patients and travel to healthcare and more studies are needed to understand the nature of this relationship. A study that included data on 104 DLBCL patients in the USA examined whether increased distance to tertiary care was associated with delayed diagnosis, which they defined as more than three months between first symptom to diagnosis (Zurko, Wade and Mehta, 2019). They reported no association between the mean time in months between first symptom and diagnosis for patients living less than or more than 50 miles from the tertiary care unit. However, they did find that patients who had delayed diagnosis lived further from tertiary care and they also detected a modest but statistically significant increased odds of delayed diagnosis per mile travelled to tertiary care centre (odds ratio: 1.01 [95% CI: 1.001-1.02]). Another paper that used data on 278 Canadian DLBCL patients reported that distance travelled to the treatment centre was not a predictor of diagnostic or treatment delay (Nikonova et al., 2015). However, the applicability of these findings to UK patients is uncertain because the characteristics of the populations and environment are different, which presents a gap in our current knowledge.

3.4.2 Travel and survival

A likely hypothesis for the relationship between travel to healthcare and cancer survival is that increasing travel is associated with worse survival, however, contrary to this assumption, many studies report that increased travel is associated with better cancer survival (Wasif et al., 2016; Vetterlein et al., 2017; Graboyes et al., 2018; Siegel et al., 2021). It seems improbable that the act of travelling more to access cancer care is directly associated with better survival and more likely that some characteristic of the patients who travel further, such as higher SES, or the treating facilities they are travelling to, such as specialist centres with better equipment, is driving this association. For example, in England many of the most deprived patients live in inner-city areas where healthcare facilities tend to be located and it has been extensively reported that patients living in deprived areas of high-income countries persistently have worse cancer survival than their more affluent counterparts (Coleman et al., 2004; Rachet et al., 2010; Singh and Jemal, 2017). The confounding effect of SES on travel to healthcare is exemplified by a study of colorectal cancer patients in England that reported better survival for those living further away from healthcare in unadjusted analyses, but no association when adjusted for deprivation (Dejardin et al., 2014). Similarly, opposite associations between travel times and survival for breast and colorectal cancer at a population level were reported in an unadjusted and adjusted model (Murage et al., 2016). In the unadjusted models, longer travel times were

associated with better survival, but in those adjusted for deprivation, longer travel was associated with worse survival. In the USA, a study reported that bladder cancer patients who travelled further to hospital had better survival in unadjusted analyses, but this association disappeared when hospital volume (the number of procedures to remove bladder cancer per hospital per year) was included (Xia et al., 2018). It has been reported that patients living in the USA, including those with DLBCL, treated at high-volume centres or academic centres have better outcomes than those treated at local hospitals (Ermann et al., 2020; Lidsky et al., 2017; Vardell et al., 2023). Therefore, it is possible that some patients may choose to travel further for treatment at centres they perceive to be better, and those able to do so are likely to be more affluent and fitter.

Other studies have reported an association in the opposite direction, that patients who travel further for care are at a survival disadvantage (Barrington et al., 2016; Stoyanov et al., 2022). A study based in the USA found that patients with rectal cancer who had a drive time of over an hour to their nearest cancer centre had worse survival (Gotfrit et al., 2020). Another study reported that patients with leukaemia and lymphoma were less likely to receive a stem cell transplant the further they travelled. Receiving a stem cell transplant improved mortality considerably and therefore patients with greater travel were found to be at a survival disadvantage (Mitchell and Conklin, 2015). This suggests that some of the association between travel and mortality observed for cancer patients could be explained by differences in treatment received between the travel groups.

Finally, some papers reported no association between travel and survival for cancer patients (Haddad et al., 2015; Dhakal et al., 2020; Sjöström et al., 2020; Lim et al., 2023; Perry et al., 2023). For example, no difference in survival was detected by distance to specialised treatment centres for acute myeloid leukaemia patients in Denmark (Tøstesen et al., 2019). Similarly, a USA study that dichotomised cervical cancer patients into two groups using the median travel distance to academic medical centre reported no difference in survival between the two groups (Rauh et al., 2018). Interestingly, a study based in Scotland found no difference in one-year mortality by travel time for patients living on the mainland, but did find that patients living on the Shetland or Orkney Islands, referred to as ‘island-dwellers’, had worse one-year mortality (Turner et al., 2017). This difference in survival could not be explained by delays in diagnosis or commencement of treatment and the authors suggested that this may instead be explained by a lower incidence of outpatient oncology appointments. It is also possible that in addition to increased distance to healthcare facilities, these patients are susceptible to other factors, such as limited infrastructure and intra-island travel, that could make accessing healthcare more difficult and influence their outcomes.

Interestingly, a study that investigated the association between travel time to nearest care centre and survival for the 10 most common solid cancers in France reported different effects of travel for different cancer sites, and for some cancer sites different effects by sex (Gardy et al., 2023). These were: increased travel associated with worse survival, increased travel associated with improved survival, no association between travel and survival, and a 'Reverse U-Shape Pattern' where survival was worse for those with medium length journeys. This reverse U-shape association was also observed between travel time to hospital and one-year mortality for patients with colorectal cancer in Scotland, with the highest mortality for patients with mid-range travel times, whereas for patients in Denmark, increased travel time to hospital up to 60 minutes was associated with lower one-year mortality (Murchie et al., 2021a). Interestingly, the same paper reported that the shape of the association between travel time to GP and mortality was the opposite for travel time to hospital for Scottish patients, instead displaying a U-shaped pattern, with those with mid-range travel times having the lowest mortality.

A limited number of studies have specifically investigated the association between DLBCL survival and travel. A study of DLBCL patients in Canada that included travel in a multivariable analysis investigating the effect of place of residence and treatment on survival outcomes found that distance to treatment did not have any individual prognostic value (Lee et al., 2014). Another study based in France described an association between distance to the nearest reference centre and survival, with those that lived intermediate distances having the worst survival (Guyader-Peyrou et al., 2017). However, this was one of many factors investigated in the paper.

The inconsistency in the direction of association between travel and survival is indicative of its complex nature. This includes varying associations by cancer type and setting and its vulnerability to confounding by many factors including the SES, fitness and age of patients. Therefore, confounders such as SES should be accounted for in analyses investigating the effect of travel on survival and care must be taken when drawing comparisons between studies investigating different cancer types and in different settings.

3.4.3 Rurality and diagnostic route

The influence of rurality on diagnostic route differs between cancer types. For example, a large study in the USA reported that rurality was associated with late-stage disease at diagnosis for some cancer types, but not others (Oh et al., 2024). Similarly, a study in Australia reported that rural residence was associated with longer time to diagnosis for colorectal cancer patients', but for breast cancer patients rural-urban differences were

minimal (Bergin et al., 2018). These differences could be explained by patients presenting to healthcare with a breast lump being more likely to be referred for diagnostic tests than patients with colorectal cancer, who often present with non-specific symptoms.

Two Scottish studies that used data from the national cancer diagnosis audit reported that there was no difference in the proportion of Emergency Presentations between rural and urban patients, but that rural patients had more GP consultations prior to diagnosis (Murchie et al., 2020; Maxwell et al., 2023). Furthermore, longer time between presentation and referral to secondary care (primary care interval) for rural patients was reported by Maxwell at al., 2023, and longer time between presentation and diagnosis (diagnostic interval) for rural patients was reported by both studies. These results imply that rural patients experience more diagnostic delay than urban patients, however, the authors also reported that rural patients did not have increased odds of 'prolonged' presentation or diagnostic intervals, over 60 or 90 days respectively. They suggested that the longer intervals were unlikely to be clinically significant and could be explained by an increase in investigations carried out by the GP prior to referral and instead of indicating diagnostic delay may actually reflect appropriate 'gatekeeping'. Another Scottish study reported that rural colorectal cancer patients were more likely than urban patients to be detected through screening, a diagnostic route often described as desirable, however, this did not translate to a survival advantage for these patients (MacVicar et al., 2020).

Interestingly, a study of colorectal cancer patients in Scotland reported that the association between travel time to healthcare and Emergency Presentation differed significantly between urban and rural patients (Murage et al., 2017). For example, longer travel time was associated with lower odds of Emergency Presentation for urban patients, but the opposite was the case for rural patients.

There is limited literature regarding differences in RTD or delayed diagnosis by rurality for patients with haematological malignancies and this search returned no papers that investigated this association for DLBCL patients specifically.

3.4.4 Rurality and Survival

The findings of studies examining the influence of rurality on survival are mixed; some report no differences between urban and rural dwellers, some report better outcomes for rural patients, and others report worse outcomes for rural patients. The inconsistency in the direction of association illustrates the complexity of the relationship, which is likely confounded by other factors, principally SES.

An Australian study reported that for 11 cancer types, excess five-year mortality was significantly greater for patients who did not reside in major cities and this association decreased but persisted for most cancers when adjusted for SES (Afshar et al., 2020). This indicates that only some of the difference in survival associated with rurality can be explained by differences in SES. Interestingly, the direction of the association between outcomes of non-Hodgkin lymphoma (NHL) patients and rurality changed when the analysis was adjusted for SES, so that those who lived outside major cities had better survival. This suggests that in this population, the association between rurality and survival for NHL patients can be explained by SES. Another study that used data from 111,627 patients diagnosed with ovarian cancer in the USA reported that rural residence was associated with worse survival compared to living in urban and metropolitan areas in unadjusted and adjusted analyses (Ulmer et al., 2022). The authors suggested that the independent association between rural residence and worse survival was compounded by other factors identified in the analysis, such as insurance status and comorbidities, that together explain the disparity in survival of urban and rural patients. Another large study based in the USA that examined the effect of rurality on mortality across multiple cancer sites reported that rural residence was associated with a 2.4% absolute increase in five-year mortality (Kenzik et al., 2024). Furthermore, they reported that highly deprived rural patients had 6% higher mortality compared to low-deprivation urban patients, illustrating the interaction between rurality and deprivation in the USA. Australia and the USA are both much larger than the UK and the geographical remoteness associated with rurality may be more extreme than that experienced by patients residing in the UK. A survival disadvantage for rural prostate cancer patients has been reported in Scotland, although this association was weak (Laing et al., 2014).

Contrary to the findings discussed above, two Scottish studies reported better survival for rural cancer patients (Murage et al., 2017; Griffin et al., 2023). Interestingly, another study based in Scotland examining the influence of rurality on outcomes of melanoma patients found no evidence of a trend for worse survival with increasing rurality (Murchie et al., 2018). Similarly, a study that used data from 17 cancer-specific cohorts representing 36,995 patients enrolled in clinical trials across the USA found no evidence of statistically significant differences in survival for any patient cohorts by urban-rural location other than a single breast cancer cohort (Unger et al., 2018). This suggests that when treatment and care is equal, such as in the setting of a clinical trial, residential location is not associated with survival and the authors proposed that other papers that have reported a survival disadvantage for rural patients failed to adequately control for differences in access to care.

A German study that included data from NHL patients reported no consistent improvement in survival with increasing urbanisation (Nennecke et al., 2014).

There are a small number of studies that have investigated the influence of rurality on survival of patients with DLBCL. Lee et al examined data from 1357 DLBCL patients in Canada and found that rural patients had comparable survival to those who resided in large urban areas, whereas those residing in small and medium urban areas had worse survival and this observation could not be explained by longer driving distances to care or lower income in these groups (Lee et al., 2014). Another study that included over 83,000 DLBCL patients from across the USA reported that rural and urban patients had worse overall survival compared to those who lived in metropolitan areas when confounders including age, stage and insurance status were controlled for (Ritter et al., 2019).

3.4.5 Socioeconomic status and diagnostic route

It has been repeatedly observed that patients with low SES who are diagnosed as Emergencies have worse survival than more affluent patients diagnosed as Emergencies, even after adjusting for confounders (McPhail et al., 2013). It has also been reported that the most deprived patients have a greater risk of Emergency Presentation across many cancer types, including NHL (Abel et al., 2015). Therefore, it has been suggested that reducing the proportion of diagnoses made via the Emergency route could narrow the deprivation gap in cancer survival.

Haematological malignancies are frequently diagnosed following an Emergency attendance, for example, it has been estimated that 27% of NHLs in England are diagnosed via the Emergency route compared to 24% of all cancers (Elliss-Brookes et al., 2012). A study based on data from the Haematological Malignancy Research Network (HMRN) (described in section 1.2) reported that 40% of DLBCL patients in their cohort were diagnosed as an Emergency and that survival was worse for these patients compared to those diagnosed via other routes (Kane et al., 2017). However, they did not report any difference in the proportion of patients presenting as an Emergency by deprivation.

An Irish study found that first hospital admission as an Emergency instead of a planned admission was less common for affluent NHL patients compared to deprived NHL patients (Comber et al., 2016). However, they did not distinguish between the different types of NHL and the healthcare system in Ireland differs to that of the UKs, meaning these observations may not be generalisable to UK patients with DLBCL. Another study based in Canada reported that income and education, two factors commonly used to describe SES, were not predictive of diagnostic delay in DLBCL patients (Nikonova et al., 2015).

3.4.6 Socioeconomic status and survival

The majority of papers included in this review report that increasing deprivation is associated with worse survival (Berkman et al., 2021; Costa, Brill and Brown, 2016; Forrest et al., 2015; Hong et al., 2016; Keegan et al., 2016; Le Floch et al., 2020; Lee et al., 2021; Maheswaran and Morley, 2018; Ramsay et al., 2014; Smith et al., 2021b, 2021a; Tao et al., 2014; Vallance et al., 2018). The observation that for many types of cancer, including some of the most common, such as breast, prostate and colorectal cancer, survival is worse for those who are most deprived is not recent (Coleman et al., 2004; Kogevinas et al., 1991). Some, but not all, of the differences in survival by SES can be explained by the lifestyle factors associated with SES, such as smoking, and policy to reduce such inequalities has been introduced, including the NHS Cancer Plan (2000) (Afshar, English and Milne, 2021). However, the survival gap between the most and least deprived patients is a persistent problem and in the years following the introduction of the NHS Cancer Plan (2000), there has been no considerable improvement in reducing socioeconomic inequalities in cancer survival (Exarchakou et al., 2018; Herbert et al., 2019a).

Interestingly, despite the frequently reported persistent nature of survival inequalities for cancer patients, a study of ovarian cancer patients in a clinical trial in England and Wales reported no inequalities in survival by SES (Abdel-Rahman et al., 2014). In clinical trials, patients in any particular arm have equal access to treatment and care and therefore these findings suggest that some of the socioeconomic inequalities in survival observed in population-based studies are the result of unequal access to care and treatment. Although, it is worth considering that trial participants are not typically representative of the disease population, for example, they are usually younger patients with fewer comorbidities (Elting et al., 2006). Other non-trial studies have reported similar findings, such as an English study that reported that liver cancer patients who were more deprived had worse 3-year survival (Vallance et al., 2018). However, when the survival analysis was restricted to only those who received liver resection surgery, no socioeconomic differences were observed. Similarly, a study investigating socioeconomic inequalities in survival of lung cancer patients in England found that inequalities remained after adjusting for patient, tumour and healthcare system factors but disappeared when treatment was accounted for (Forrest et al., 2015). The same study also reported that patients belonging to the least deprived categories were more likely to receive a resection even after adjusting for differences in comorbidities, suggesting that undertreatment of more deprived patients could be contributing to the lower survival observed in this group. Furthermore, a study of 20,000 women with breast cancer found that pre-existing health status and primary care consultation history did not explain differences in

cancer survival by socioeconomic group and reported that it is the post-diagnosis period that is most susceptible to socioeconomic disparities (Woods et al., 2022). These results suggest that treatment inequalities contribute to the deprivation gap in survival for some cancer patients, which is interesting as all these studies are based in the UK, where treatment is provided by the NHS and care should be equitable across SES groups. Therefore, it seems probable that other barriers exist that prevent patients from low SES groups accessing optimum healthcare.

It has been reported that despite an overall improvement in 1-, 5- and 10- year survival over time, inequalities between the affluent and deprived remained largely unchanged for patients with NHL (Rachet et al., 2008). Another English study found that deprivation was a strong predictor of 1-year mortality among DLBCL patients (Smith et al., 2021a). However, a paper using HMRN data, a study based in the north of England, that investigated DLBCL alone found no evidence of a difference in survival between different socioeconomic groups (Smith et al., 2015). Findings from other countries are mixed, with some reporting no effect of SES and others reporting worse outcomes for low SES patients with DLBCL. For example, studies based in France (Guyader-Peyrou et al., 2017) and the Netherlands (Boslooper et al., 2017) found that SES was not associated with survival, while a study in the USA found that survival disparities were associated with neighbourhood SES (Tao et al., 2014). Tao et al., 2014, also reported that the survival gap by neighbourhood SES increased after the introduction of rituximab, which further implies that treatment is an important contributor to SES disparities in survival. The obvious suggestion for what might be causing these contradictory findings is the difference in healthcare systems between the USA and European countries. However, disparities in survival for DLBCL patients by SES group have also been reported in Hong Kong, which has a public health system akin to that in the UK (Lee et al., 2021).

3.5 Summary and next steps

This review has summarised the relevant literature examining the effect of travel to healthcare, rurality and SES on route to diagnosis and survival of cancer patients. In doing so, it has facilitated the selection of the appropriate methods for this thesis, which are described in more detail in the Materials and Methods chapter.

Most studies that examined the influence of SES on diagnostic route or survival of cancer patients reported that those living in more deprivation had worse outcomes. This relationship was observed across many different cancer types and multiple countries, highlighting the persistent nature of this association. Where no association with SES and survival was reported, this was often in settings where treatment was equitable across all SES groups, such as in a clinical trial, or where treatment had been included as a confounder. These results suggest that cancer patients from low SES groups face barriers to accessing optimum healthcare, even in countries where healthcare is universal, such as the UK.

The review showed that studies used a variety of different methods to assign both deprivation and rurality to participants and these were typically measures specifically developed for the country the study was based in. For example, rurality was often assigned using national datasets which is appropriate because what is described as rural in a large country such as Australia is likely different to the definition that would be applied in a much smaller country like the UK. Therefore, despite no studies included in this review using the ONS rural-urban classification to assign rurality, it was decided that because this was the national dataset that covered our region, this classification would be used in this thesis. Similarly, the way in which studies assigned deprivation varied between regions, with many studies based in England using the IMD which is the deprivation measure that HMRN routinely links to and was used in this thesis to ensure that results were easily comparable to what has been published already. Some studies used the composite measure of the IMD which uses a variety of data sources across seven distinct domains, while others opted to use the income domain of the IMD in isolation. For this thesis only the income domain of the IMD was used due to concerns that data used in the calculation of the health domain could artificially inflate associations between deprivation and the outcomes of this thesis.

The associations between rurality and travel to healthcare on cancer patients' route to diagnosis and survival were less consistent than those reported for SES: some studies reported no association, some reported better outcomes for those who travelled further, and others reported worse outcomes for those with the longest journeys. Furthermore, it is challenging to compare the results of studies due to inconsistencies in the methods used to

measure and analyse travel data; also, findings from regions with different healthcare provision and geographical landscapes may not be applicable to UK patients. The effect of travel and rurality on diagnostic route and survival of patients with DLBCL has been examined by a small number of studies, none of which were based in England, highlighting a gap in the knowledge of DLBCL.

The review demonstrated that currently there is no standardised method for estimating travel time to healthcare for cancer patients. For example, there was variation in the way in which travel distance/time was calculated, with both straight-line or 'as the crow flies' distances and shortest journey based on the road network between the origin and destination points reported, as well as differences in the origin and destination points used in the calculations, with some studies using postcode (or equivalent) while others used centroids of census-based areas. Furthermore, most estimates reported in the literature were for journeys made by private car, with little information about journeys made by other modes of transport such as public transport available. The review also showed that there is no standardised approach in which the travel data generated in these calculations was used in the statistical analyses, with some studies keeping travel as continuous variables and others converting it into a categorical variable using a number of different cut-offs. After reviewing the methods used in these studies and considering the data available in HMRN, it was decided that calculating travel distance and time by both car and public transport using the road network and using postcodes for the origin and destination points would yield the most accurate estimates. This was done using a travel time analysis software (TRACC by Basemap) to calculate shortest journeys from each patient's residential address at diagnosis to their registered General Practitioner and diagnostic hospital (see section 4.5). The decision on how to use this travel data in the analysis and was made following discussion at my Thesis Advisory Panel meetings, as discussed in section 4.8.2.

There are multiple ways in which diagnostic delay can be measured, as highlighted by the review, and because different regions have different healthcare systems these also vary depending on the geographical location of the study. Using the Emergency Presentation route to distinguish cancer patients who have experienced diagnostic delay is a popular method used in many studies based in the UK and so to ensure this work was comparable with what had already been published relating to delayed cancer diagnosis in England, this method was chosen for this thesis.

4 Materials and Methods

A summary of the data sources used in this thesis to examine the impact of socio-demographic and geographical factors on routes to diagnosis, disease presentation and outcome of patients with diffuse large B-cell lymphoma (DLBCL) in the Haematological Malignancy Research Network (HMRN) can be seen below in Table 4:1. All statistical analyses were conducted in Stata 18 (StataCorp,Texas).

Table 4:1. Summary of data sources used in this thesis.

Data	Source
Patients' disease, clinical presentation and treatment data.	HMRN routinely collected data
Hospital Episode Statistics	NHS England
Mortality	NHS England
Rural-urban classifications	Office for National Statistics
Index of Multiple Deprivation	Ministry of Housing, Communities and Local Government
GP postcodes (for importing into mapping software only)	NHS England (EPRACCUR file)
All postcodes in the HMRN region (for importing into mapping software only)	NHS postcode directory (August 2022)
Road Network	Ordnance Survey (Meridian 2)
Public transport timetables	DataCutter tool in TRACC (Basemap)
Life tables	NHS England

4.1 Study population

Patients from the HMRN cohort newly diagnosed with DLBCL between January 2005 and August 2019 were included in the analysis. 2019 was chosen as the cut-off point for inclusion in this thesis for data completeness and due to the impact of the coronavirus pandemic on patients' diagnostic experience and survival from 2020 onwards. Only patients with a *de novo* diagnosis of DLBCL within that time frame were included, that is, any patient recorded as having been diagnosed with another haematological malignancy prior to their DLBCL were excluded. A small number of patients treated at hospitals outside of the Network were excluded as well as patients whose medical records could not be accessed and therefore did not have complete follow-up (section 5.1). All patients with a haematological malignancy in the area and beyond are diagnosed at a single NHS laboratory regardless of whether they have private healthcare and this facilitates tracking private patients for entry into the NHS.

4.2 Haematological Malignancy Research Network case ascertainment and data collection

Patients diagnosed with a new, ICD-O-3 coded haematological malignancy or precursor condition by the Haematological Malignancy Diagnostic Service (HMDS) in Leeds since September 2004 are included in the HMRN patient cohort. HMDS is a specialist diagnostic laboratory providing a fully integrated diagnostic pathway that includes information on histology, immunohistochemistry, gene expression profiling and more, meaning that all HMRN patients are diagnosed at a single site. Patients are entered into the cohort on the day they are first diagnosed with a haematological malignancy or precursor condition, with newly diagnosed patients ascertained on a weekly basis via NHS systems.

Under section 251 of the NHS Act 2006, HMRN has permission to collect and store core patient data without explicit patient consent. The core data that HMRN collects consists of laboratory data, clinical data, data from linked national datasets and self-reported data from consenting patients (Figure 4:1).

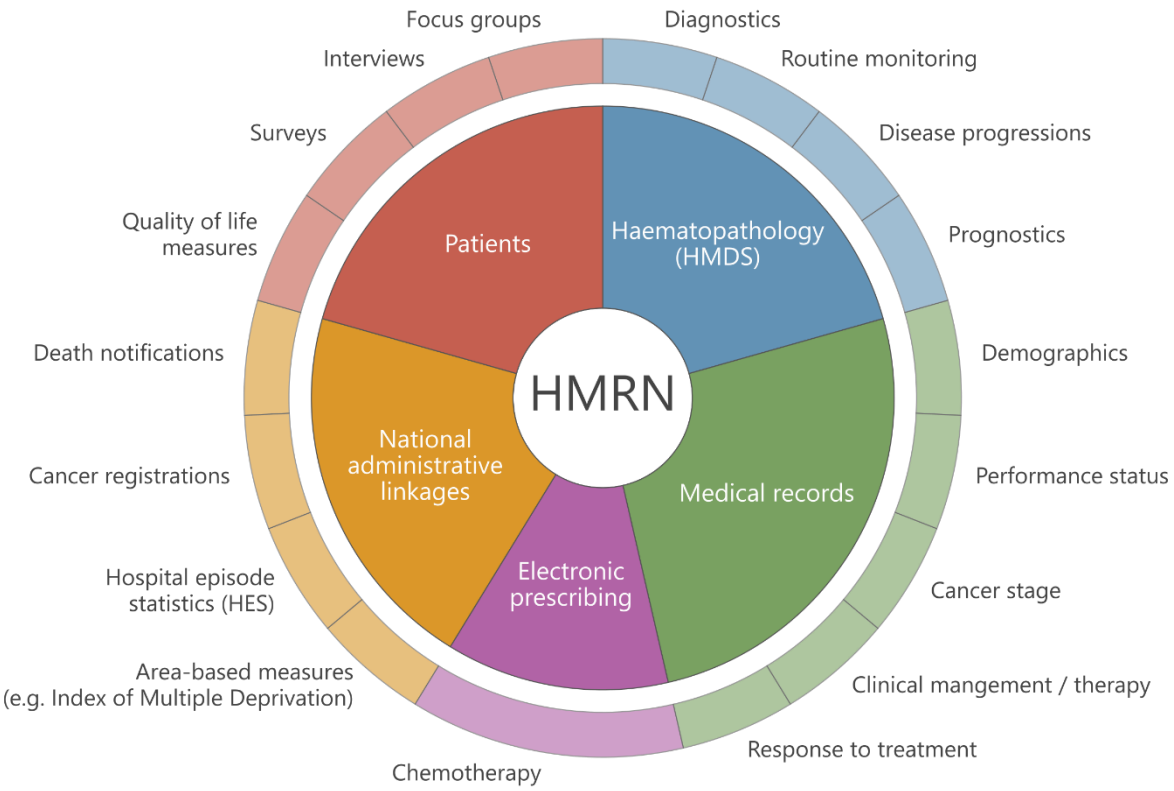


Figure 4:1. Haematological Malignancy Research Network data sources.

4.2.1 Primary data collection

Figure 4:2 illustrates the flow of primary data collected by HMRN and stored on the HMRN central database, which is managed by the Epidemiology and Cancer Statistics Group (ECSG). ECSG is the research group that I am part of and is a multi-disciplinary team of researchers, data managers, research nurses, clinical research assistants, IT and support professionals, based at the University of York.

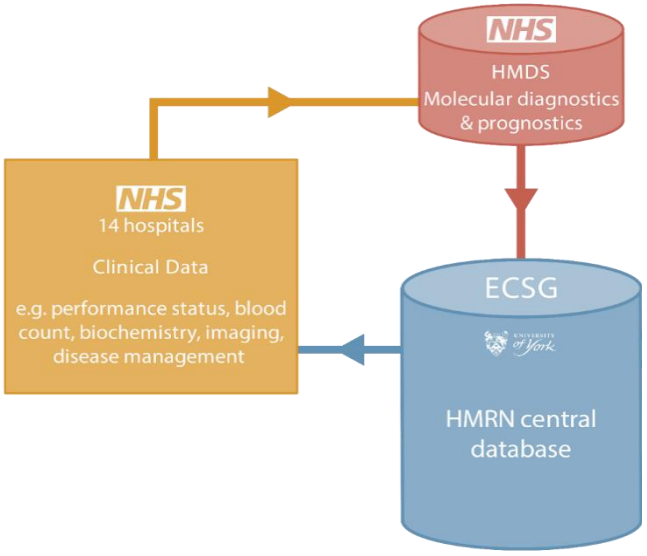


Figure 4:2. Haematological Malignancy Research Network (HMRN) data collection and linkages.

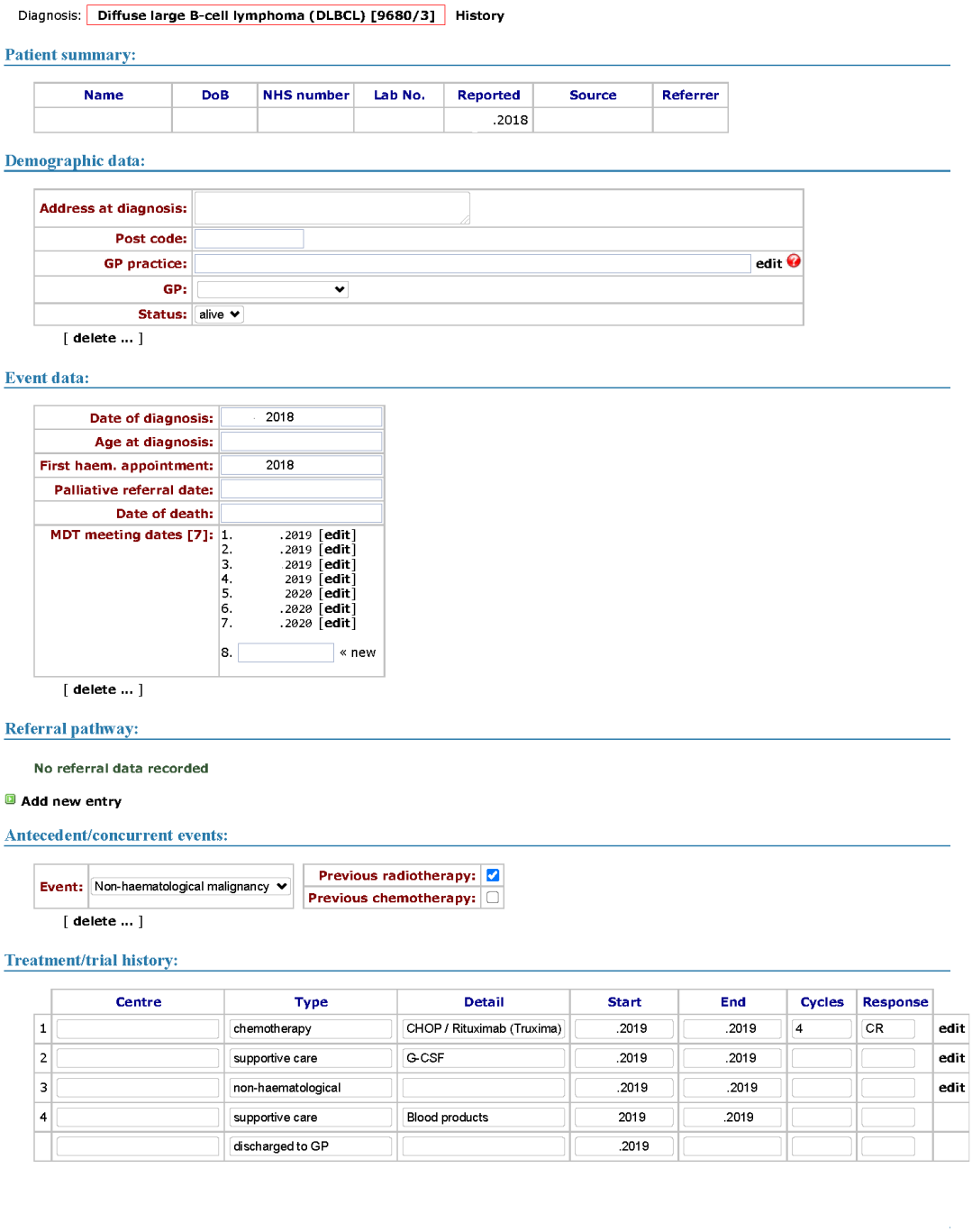
Demographic and clinical data are abstracted from the patients' medical records by a team of trained research nurses using data collection forms that are specific to the patient's diagnosis. This information is initially collected approximately a year after the patient enters the HMRN cohort and then repeated at specific follow-up points, ending at death. This includes the patient's home address and registered General Practice (GP) as well as information relating to relevant medical conditions such as Human Immunodeficiency Virus and Down syndrome which could be related to their haematological malignancy diagnosis. For patients with a previous cancer diagnosis, any chemotherapy or radiotherapy treatment resulting from this is also recorded, to identify therapy-related myeloid cases.

Presentation data is collected using results of tests performed around the date of diagnosis and before treatment starts including biopsies, blood tests, scans and clinical notes.

Disease-specific presentation information is also recorded, which for patients with DLBCL includes the presence or absence of B symptoms, sites of disease involvement and stage of disease using the Ann Arbor classification. The Ann Arbor stage is also calculated independently using primary data variables from the patients' medical records within an in-house algorithm. Sometimes it is not possible to collect all of this data, either because it was not recorded in the patient's medical records or a relevant test was not performed and if this is the case the result is recorded as 'not known' or 'not done' respectively. For some diagnoses, there are additional forms to record results of tests undertaken specifically for that disease, such as Positron Emission Tomography (PET) scans for lymphoma patients. Figure 4:3 and Figure 4:4 demonstrate data entry on HILIS for a patient with DLBCL, with sensitive data redacted to preserve confidentiality.

Treatment data is collected by the research nurses at abstraction of the medical records and is ongoing until death, even if a patient receives aggressive treatment and is effectively 'cured' of their disease. These data include the start and end dates of any treatments, the hospital where treatment was administered and the specific drugs, e.g., R-CHOP. Any other relevant information is also recorded as free text, such as dose attenuation and the reason for this. The purpose of the treatment is also recorded, e.g., for palliative reasons only. If the patient was treated with aggressive chemotherapy their response to treatment, e.g., complete response (CR) or progressive disease (PD), is inferred from the clinical notes and scans, supported by laboratory data.

Diagnostic data comes directly from HMDS via their integrated laboratory information system, HILIS, which is used for specimen tracking and reporting. This includes dates and results of tests relating to the patient's initial diagnosis, such as histology and immunohistochemistry results, as well as from any tests undertaken as part of ongoing monitoring for disease progression, transformation and response to treatment.



Diagnosis: Diffuse large B-cell lymphoma (DLBCL) [9680/3] History

Patient summary:

Name	DoB	NHS number	Lab No.	Reported	Source	Referrer
				.2018		

Demographic data:

Address at diagnosis:	
Post code:	
GP practice:	edit
GP:	
Status:	alive

[delete ...]

Event data:

Date of diagnosis:	2018
Age at diagnosis:	
First haem. appointment:	2018
Palliative referral date:	
Date of death:	
MDT meeting dates [7]:	1. .2019 [edit] 2. .2019 [edit] 3. 2019 [edit] 4. 2019 [edit] 5. 2020 [edit] 6. .2020 [edit] 7. .2020 [edit] 8. « new

[delete ...]

Referral pathway:

No referral data recorded

Add new entry

Antecedent/concurrent events:

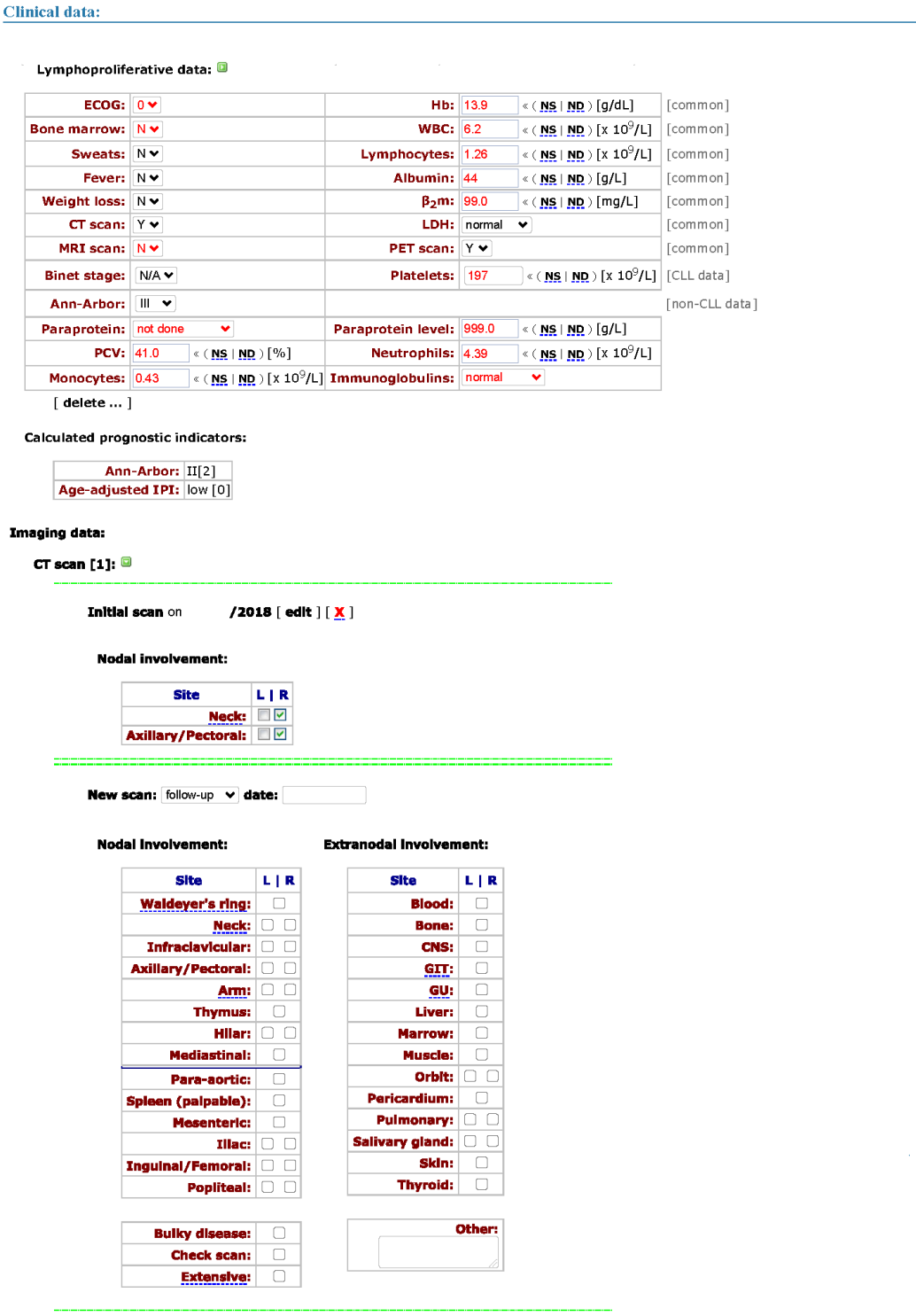
Event:	Non-haematological malignancy	Previous radiotherapy:	☑
		Previous chemotherapy:	☐

[delete ...]

Treatment/trial history:

	Centre	Type	Detail	Start	End	Cycles	Response	
1		chemotherapy	CHOP / Rituximab (Truxima)	.2019	.2019	4	CR	edit
2		supportive care	G-CSF	.2019	.2019			edit
3		non-haematological		.2019	.2019			edit
4		supportive care	Blood products	2019	.2019			
		discharged to GP		.2019				

Figure 4:3. Example of data entry on HILIS with sensitive data redacted for a patient diagnosed with diffuse large B-cell lymphoma in HMRN.



Clinical data:

Lymphoproliferative data:

ECOG:	0	Hb:	13.9 « (NS | ND) [g/dL]	[common]
Bone marrow:	N	WBC:	6.2 « (NS | ND) [x 10^9/L]	[common]
Sweats:	N	Lymphocytes:	1.26 « (NS | ND) [x 10^9/L]	[common]
Fever:	N	Albumin:	44 « (NS | ND) [g/L]	[common]
Weight loss:	N	β_2m:	99.0 « (NS | ND) [mg/L]	[common]
CT scan:	Y	LDH:	normal	[common]
MRI scan:	N	PET scan:	Y	[common]
Binet stage:	N/A	Platelets:	197 « (NS | ND) [x 10^9/L]	[CLL data]
Ann-Arbor:	III			[non-CLL data]
Paraprotein:	not done	Paraprotein level:	999.0 « (NS | ND) [g/L]	
PCV:	41.0 « (NS | ND) [%]	Neutrophils:	4.39 « (NS | ND) [x 10^9/L]	
Monocytes:	0.43 « (NS | ND) [x 10^9/L]	Immunoglobulins:	normal	

[delete ...]

Calculated prognostic indicators:

Ann-Arbor:	II[2]
Age-adjusted IPI:	low [0]

Imaging data:

CT scan [1]:

Initial scan on **/2018** [**edit**] [**X**]

Nodal involvement:

Site	L | R
Neck:	☐ ☑
Axillary/Pectoral:	☐ ☑

New scan: follow-up **date:**

Nodal Involvement:

Site	L | R
Waldeyer's ring:	☐
Neck:	☐ ☐
Infraclavicular:	☐ ☐
Axillary/Pectoral:	☐ ☐
Arm:	☐ ☐
Thymus:	☐
Hilar:	☐ ☐
Mediastinal:	☐
Para-aortic:	☐
Spleen (palpable):	☐
Mesenteric:	☐
Iliac:	☐ ☐
Inguinal/Femoral:	☐ ☐
Popliteal:	☐ ☐

Extranodal Involvement:

Site	L | R
Blood:	☐
Bone:	☐
CNS:	☐
GIT:	☐
GU:	☐
Liver:	☐
Marrow:	☐
Muscle:	☐
Orbit:	☐ ☐
Pericardium:	☐
Pulmonary:	☐ ☐
Salivary gland:	☐ ☐
Skin:	☐
Thyroid:	☐

Bulky disease:	☐
Check scan:	☐
Extensive:	☐

Other:

Figure 4:4. Example of data entry on HILIS with sensitive data redacted for a patient diagnosed with diffuse large B-cell lymphoma in HMRN.

4.2.2 Patient self-reported data

With permission from their clinical team, patients in the HMRN cohort are asked to consent to further contact. This includes a request to complete a self-reported questionnaire that is sent to patients soon after diagnosis (approximately six weeks later) to prevent recall bias. The documentation for this can be accessed on the HMRN webpage (https://hmrn.org/resources/documentation). This questionnaire supplements the data which is routinely collected by HMRN and includes questions relating to:

- ethnicity, height and weight
- smoking status
- any symptoms experienced before being diagnosed with their condition
- when the patient first sought help for their condition
- quality of life.

Information from these questionnaires is not used in this thesis, however, the collection of this data highlights HMRN's commitment to hearing and understanding patients' experiences, including along the diagnostic pathway.

4.3 Linked national data

In addition to collecting primary data, HMRN links to a number of national databases including death notifications, cancer registrations and Hospital Episode Statistics (HES) as shown in Figure 4:5. The time periods available for each of these datasets is also shown in Figure 4:5. Patients can also be assigned to area-based population indices such as the Index of Multiple Deprivation (IMD) (described in section 4.3.2) using the postcode of their home address at diagnosis, which is recorded as part of the initial data collection process described in section 4.2.1.

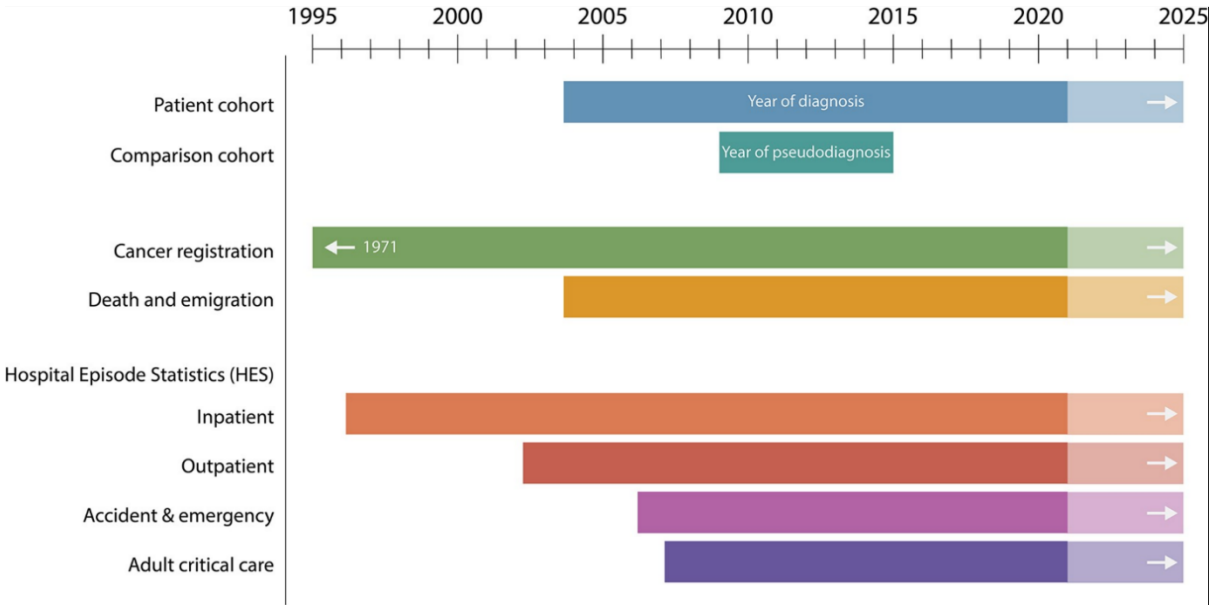


Figure 4:5. National data availability: Haematological Malignancy Research Network patient cohort and comparison cohort.

4.3.1 Hospital Episode Statistics

Administrative and clinical data are routinely collected by healthcare providers to support patient care and are submitted to NHS England for the purpose of payment for, and the planning and monitoring of, healthcare. This includes records of all patients admitted to NHS hospitals in England including private patients treated in NHS hospitals, patients who do not live in England and care delivered by centres funded by the NHS (including those in the independent sector). This data is compiled, cleaned and processed before being released as HES on a monthly basis, which includes clinical, patient, administrative and geographical information relating to inpatient care, outpatient appointments and Accident and Emergency attendances and is available in four datasets, two of which were used in this thesis:

- Hospital Episode Statistics Admitted Patient Care (HES APC) – available from 01/04/1989
- Hospital Episode Statistics Outpatients (HES OP) – available from 01/04/2003

Although not the primary purpose of HES data, it is used extensively in health research. Periodically, HMRN applies to NHS England to obtain matched HES data for patients in the cohort by their NHS numbers. This data is then linked to the patient's unique study ID and can be used in conjunction with other data available in HMRN for a variety of analyses including estimating the proportion of patients diagnosed through each route to diagnosis and examining rates of hospital activity prior to diagnosis.

4.3.2 Index of Multiple Deprivation

No individual-level measure of socioeconomic status is currently collected by HMRN, however, because HMRN has access to patient's residential address, this information can be used to assign each patient an IMD score based on their Lower Super Output Area (LSOA). Between 2005 and 2019, the timeframe in which patients included in this thesis were diagnosed, there have been five versions of the IMD (2004, 2007, 2010, 2015 and 2019). The 2015 release of the IMD was chosen for this thesis because it approximately reflects the median date of diagnosis for this sample of patients. Consequently, for some patients, particularly those diagnosed at the start of the registry, other releases of the IMD are closer to their diagnostic date. However, this is of limited concern because of the high concordance between the different releases of the IMD, for example, 88% of neighbourhoods in the most deprived decile in the 2019 release were also in the most deprived decile in the 2015 release (Ministry of Housing, Communities & Local Government, 2019).

IMD is a composite measure of deprivation made up of seven distinct domains as described in section 1.5.2.2. For this thesis, the 'Income Deprivation Domain' was used in isolation to estimate deprivation because of concerns that inclusion of other domains could lead to overestimations of the associations between some of the variables in the analysis. For example, the 'Health Deprivation and Disability Domain' of the IMD uses years of potential life lost as one of its indicator variables. Cancer patients generally have reduced survival compared to the general population and are therefore likely to contribute towards a higher deprivation score for the Health Deprivation and Disability Domain of the IMD. It would be wrong to conclude that a higher deprivation score for this domain was contributing to lower survival of cancer patients because the explanatory variable (deprivation) and outcome variable (cancer survival) are at least in part measuring the same thing. Similarly, the 'Barriers to Housing and Services Domain' includes distance to GP practice as an indicator variable which is problematic when estimating differences in travel distance to healthcare services by patient characteristics such as deprivation.

For this thesis, the IMD data was categorised into quintiles using the rank of the Income Deprivation Domain for each LSOA in the country, where quintile 1 contains the 20% least deprived LSOAs and quintile 5 the 20% most deprived LSOAs.

4.3.3 Death notifications

Mortality data for patients registered in the HMRN cohort is sent from NHS England and the last update to this data was August 2023. This means that date of death is only available for patients who die after August 2023 when it is recorded by the study nurses at data collection and input into the HMRN database before the official death certificate is sent by NHS England.

4.3.4 Comparison cohort

HMRN patients diagnosed between 2009 and 2015 (n=18,127) are matched by NHS England (formerly NHS Digital) on age, sex and residency in the study area to ten randomly selected controls, who have not been diagnosed with a haematological malignancy or related condition, from the population-based National Health Service Central Register. Each control is linked to their matched case by a unique study identification code and the diagnosis date of each case is assigned as a pseudo-diagnosis date to their matched controls. In the same manner as the patient cohort, controls are linked to their own data in national databases such as HES, not to that of their matched case.

The purpose of the control cohort is to provide estimates of background healthcare activity and comorbidities of the general population to compare with that of patients diagnosed with a haematological malignancy. In this thesis, data from the controls was used to estimate the background rate of hospital activity and compare this to the rate of hospital activity of the cases (HMRN DLBCL patients) in the two years before- and one year after- diagnosis. This shows how many months before diagnosis on average patients with DLBCL are attending hospital and in combination with information relating to route to diagnosis helps inform if they experienced diagnostic delay.

4.4 Processing primary data for analysis

4.4.1 Treatment data

Any treatment that a patient in HMRN receives as part of their cancer care, including supportive and palliative care, is abstracted from their medical records and recorded in the HMRN database. Information relating to the regimen type (chemotherapy, radiotherapy, supportive care, etc.), the date that the treatment started and ended, the name of the regimen, such as R-CHOP, and a list of the drugs used (e.g., Rituximab, Cyclophosphamide, Doxorubicin, Vincristine, Prednisolone) is recorded for each treatment that the patient receives.

Data relating to all treatments a patient received following their DLBCL diagnosis were extracted for analysis. To identify patients treated with curative intent, the TreatLine ado-file developed by ECSG (Stata code in Appendix 2) was firstly executed selecting only 'active' treatments to produce 'active' lines. A treatment was considered to be active if the regimen type was described in the HMRN Treatment table as chemotherapy, CAR-T therapy, radiotherapy or a stem cell transplant. The command was executed a second time for patients who had no 'active lines' identified by TreatLine, with an expanded list of regimen types selected, such as discharged to GP and steroids. This was done to ensure that for patients who may have received a non-active treatment prior to their active treatment, the active treatment was prioritised and identified as the first-line treatment.

After the treatment data had been organised into lines and checked and cleaned for any errors, a new variable called 'curative intent' was generated, which was coded as 'yes' or 'no' for each patient based on the treatment they received as first-line, to identify patients who received aggressive treatment for their disease. Most patients in this analysis received the standard regimen for DLBCL as first-line therapy, R-CHOP, an aggressive chemo-immunotherapy and were therefore coded as 'yes' for the curative intent variable. However, this therapy is not suitable for all DLBCL patients and alternative therapies may administered with curative intent, for example, patients with either suspected or diagnosed primary diffuse large B-cell lymphoma of the CNS (PCNSL) are typically treated with methotrexate, a single-agent chemotherapy drug, as first-line treatment with the intention to cure them of their disease. A comprehensive list of the different regimens received as first-line and whether they were coded as curative can be found in Table 4:2.

Table 4:2. First-line chemotherapy regimens and treatment intent: HMRN patients newly diagnosed with diffuse large B-cell lymphoma between January 2005-August 2019.

Regimen name	Drugs	Curative Intent
AraC, Methotrexate, Rituximab	AraC, methotrexate, rituximab	Yes
Chlorambucil, Dexamethasone, Rituximab	Chlorambucil, dexamethasone, rituximab	Yes
Chlorambucil, Prednisolone	Chlorambucil, prednisolone	No
Chlorambucil, Rituximab	Chlorambucil, rituximab	Yes
CHO, Rituximab	Cyclophosphamide, doxorubicin, vincristine (+/-rituximab)	Yes
CHOP	Rituximab, cyclophosphamide, doxorubicin, vincristine, prednisolone	Yes
CHOP, Velcade	Rituximab, cyclophosphamide, doxorubicin, vincristine, prednisolone, velcade	Yes
CODOX-M	Cyclophosphamide, cytarabine, vincristine, doxorubicin, methotrexate	Yes
CODOX-M, IVAC	Cyclophosphamide, doxorubicin, vincristine, methotrexate/ifosfamide, etoposide, cytarabine	Yes
COP	Cyclophosphamide, vincristine, prednisolone, methotrexate, hydrocortisone	Yes
COPAD	Cyclophosphamide, vincristine, prednisolone, doxorubicin, GCSF	Yes
CRUK HD-MTX in PCNSL	HD methotrexate, glucarpidase	Yes
CRUK R-GCVP	Rituximab, gemcitabine, cyclophosphamide, vincristine, prednisolone	Yes
CVP	Cyclophosphamide, vincristine, prednisolone	Yes
CYM/AraC, Methotrexate, Rituximab	AraC, methotrexate (+/- rituximab)	Yes
DHAP/R-DHAP	Dexamethasone, AraC, cisplatin (+/- rituximab)	Yes
EPOC/ R-EPOC	Etoposide, prednisolone, vincristine, cyclophosphamide (+/- rituximab)	Yes
EPOCH, Rituximab	Etoposide, prednisolone, vincristine, cyclophosphamide, doxorubicin, rituximab	

Regimen name	Drugs	Curative Intent
Etoposide, Dexamethasone	Etoposide, dexamethasone	No
Etoposide, Prednisolone	Etoposide, prednisolone	No
EURO-LB 02 Induction 1a (protocol)	Vincristine, daunorubicin, asparaginase, cyclophosphamide, cytarabine, mercaptopurine, methotrexate, prednisolone	Yes
Gemcitabine, Dexamethasone, Cisplatin, Rituximab	Gemcitabine, dexamethasone, cisplatin, rituximab	Yes
Gemcitabine, Prednisolone	Gemcitabine, prednisolone	No
Gem-RCVP/GCVP/ R-GCVP	Gemcitabine, cyclophosphamide, vincristine, prednisolone, (+/- rituximab)	Yes
Ibrutinib, Rituxamab	Ibrutinib, rituximab	Yes
IdaRAM/ IdaRAM, Rituximab	Idarubicin, AraC, methotrexate, dexamethasone (+/- rituximab)	Yes
IVAC	Ifosfamide, etoposide, cytarabine	Yes
IVAC-M / Rituximab	Ifosfamide, etoposide, cytarabine, methotrexate, rituximab	yes
MATRIX/ MATRIX, R-ICE	Methotrexate, cytarabine, thiotepa, rituximab (+/- ifosfamide,etoposide)	Yes
MCOP, Rituximab	Rituximab, cyclophosphamide,mitoxantrone, vincristine, prednisolone	Yes
Methotrexate, AraC, Hydrocortisone	Methotrexate, AraC, hydrocortisone	Yes
Methotrexate, Rituximab	Methotrexate, rituximab	Yes
Methotrexate, Cytarabine, Thiotepa, Rituximab	Methotrexate, cytarabine, thiotepa, rituximab	Yes
Mini-CHOP, Rituximab	Same as R-CHOP but lower doses	Yes
PMitCEBO/ PMitCEBO, Rituximab	Prednisolone, mitoxantrone, cyclophosphamide, etoposide, bleomycin, vincristine (+/- rituximab)	Yes

Regimen name	Drugs	Curative Intent
R-CEOP	Rituximab, cyclophosphamide,etoposide, vincristine, prednisolone	Yes
R-CHOP	Rituximab, cyclophosphamide, doxorubicin, vincristine, prednisolone	Yes
R-CHOP, Acalabrutinib	Rituximab, cyclophosphamide, doxorubicin, vincristine, prednisolone, acalabrutinib	Yes
R-CHOP, AraC	Rituximab, cyclophosphamide, doxorubicin, vincristine, prednisolone, AraC	Yes
R-CHOP, Enalapril	Rituximab, cyclophosphamide, doxorubicin, vincristine, prednisolone, enalapril	Yes
R-CHOP, Methotrexate	Rituximab, cyclophosphamide, doxorubicin, vincristine, prednisolone, methotrexate	Yes
R-CHOP, Methotrexate, AraC	Rituximab, cyclophosphamide, doxorubicin, vincristine, prednisolone, methotrexate, AraC	Yes
R-CHOP, Velcade	Rituximab, cyclophosphamide, doxorubicin, vincristine, prednisolone, velcade	Yes
R-CODOX-M	Rituximab, cyclophosphamide, doxorubicin, vincristine, methotrexate	Yes
R-CODOX-M, IVAC	Rituximab, cyclophosphamide, doxorubicin, vincristine, methotrexate/ifosfamide, etoposide, cytarabine	Yes
R-CVP	Rituximab, cyclophosphamide, vincristine, prednisolone,	Yes
Rituximab, Dexamethasone	Rituximab, dexamethasone	No
Rituximab, Methylprednisolone	Rituximab, methylprednisolone	No
Rituximab, Prednisolone	Rituximab, prednisolone	No
Thalidomide, Dexamethasone	Thalidomide, dexamethasone	No
VP, Rituximab/Vincristine, Prednisolone, Rituximab	Vincristine, prednisolone, rituximab	Yes
Vincristine, Dexamethasone	Vincristine, dexamethasone	No
Vincristine, Prednisolone	Vincristine, prednisolone	No

4.4.2 Route to Diagnosis

The algorithm published by Elliss-Brookes et al, described in section 1.4.3, uses a combination of Hospital Episode Statistics (HES), National Cancer Data Repository (NCDR), National Cancer Waiting Times (CWT) and National Breast and Bowel Screening Programme data to assign a Route to Diagnosis (RTD) to cancer patients (Elliss-Brookes et al., 2012). The process of assigning a RTD for each patient begins by identifying the 'end point' of the route from their HES records, i.e., their last healthcare interaction prior to receiving their cancer diagnosis. The algorithm then searches backwards through the patients' HES records from this end point to identify the 'start point' of the route and the information from this HES record is used to assign a route to diagnosis.

In the original algorithm, RTD is initially assigned using HES data and then may be changed to the Two-Week Wait (TWW) or Screen-Detected route after examination of the Cancer Waiting Times and National Screening Programme data. However, HMRN does not have access to the National Screening Programme data or CWT data, meaning that the Screen-Detected route could not be assigned and the TWW step could not be completed in this way. Consequently, some adaptations to the Ellis-Brookes algorithm have been made so that a RTD can be assigned to patients in HMRN using only the HES data, and the HES variables used and adaptations made are summarised in Table 4:3, Table 4:4 and Figure 4:6.

Table 4:3. Description of Hospital Episode Statistics (HES) data variables used to assign routes to diagnosis.

HES Variable	Source	Description
ATTENDED	HES APC	Indicates whether or not the patient attended the scheduled appointment
ADMIDATE	HES APC	Admission date (recorded for all episodes within a spell)
ADMIDATE_CIPS	HES APC	Admission date for start of spell
APPTDATE	HES OP	Date when the appointment was scheduled
ADMIMETH	HES APC	Method of admission (recorded for all episodes within a spell)
ADMIMETH_CIPS	HES APC	Method of admission for start of spell
REFSOURC	HES OP	Source of referral
PRIORITY	HES OP	The priority of a request for services

HES=Hospital Episode Statistics, APC=Admitted Patient Care, OP=Outpatients

Table 4:4. Hospital Episode Statistics (HES) data codes and their priorities for assigning diagnostic route for HMRN patients.

Priority	Original ADMIMETH Codes (Admission Method)	Additional ADMIMETH codes	Original REFSOURC codes (Referral Source)	Additional REFSOURC codes	Route to Diagnosis
1	21,22,23,28	24,2A, 2B,2D	1,4,10	-	Emergency Presentation
2	N/A	-	3,12	AND PRIORITY variable in HES data = 3	Two-Week Wait
3	N/A	-	3,12	AND PRIORITY variable in HES data ≠ 3	GP Referral
4	N/A	-	2,5,6,7,8,11,13,14,15, 16,92,93,97,99	-	Other Outpatient
5	31,32,82,83,84,89	11,12,13,81	N/A	-	Inpatient Elective

Currently, there is no screening programme in place in the UK to detect haematological malignancies and therefore the lack of National Screening Programme data was not detrimental to this work as it is unlikely that many patients, if any at all, in this cohort would have been diagnosed through this route. However, although unlikely, it is possible that a patient with DLBCL might be diagnosed as a result of attending a National Screening Programme, for example, they might have an enlarged lymph node in the axilla detected in a mammogram (National Breast Screening Programme) and then be referred to Haematology and in this case Screen-Detected would be the most appropriate RTD to assign. To identify patients where this could be the case, all outpatient HES records occurring within six months of diagnosis where the source of referral was recorded as being from a National Screening Programme (REFSOURC [hereinafter referred to as Referral Source] code 17) were identified and these patients' HES data and HMRN data collection forms were individually examined to determine if the Screen-Detected route was the most appropriate RTD to assign. However, no such instances were identified in this cohort of patients and these

referrals were all either to ophthalmology following diabetic eye screening, appointments relating to other suspected cancers or a coding error in the HES data and consequently the Screen-Detected route was omitted.

It is reasonable to expect a considerable proportion of DLBCL patients to be diagnosed through the TWW route, which is assigned in the original algorithms using the CWT data. Without the CWT data, TWW patients are likely to be assigned to the GP Referral route if the HES algorithms detect the outpatient appointment made following the referral under TWW made by the patients' GP as the start point of the RTD. Although these two routes to diagnosis begin in the same way, it is important to distinguish between them as TWW patients have been urgently referred to a specialist by their GP, nurse or dentist because they have suspected cancer and represent a distinct group. Therefore, in the absence of the CWT data an extra step was added to the original algorithms (see Appendix 1 for original Elliss-Brookes et al., 2012 algorithms) to identify patients belonging to the TWW route in HMRN, which is described in more detail in section 4.4.2.1.

The Death Certificate Only (DCO) route is used by cancer registries to represent the small proportion of cancer patients where the first mention of their cancer is on the death certificate, i.e., there are no laboratory tests or clinical decisions recorded in the patients' medical records regarding the cancer diagnosis. When this occurs, the GP is contacted for any additional information about the cancer to ascertain if it was genuinely only detected at death. Although HMRN does not receive notification of DCO patients, HMDS may notify HMRN of patients whose diagnostic sample was taken post-mortem, but this number is minimal. It was checked if these patients had any HES activity prior to diagnosis and if not, these patients were assigned to the Unknown route.

The data for this analysis is taken from the HES Admitted Patient Care (also known as inpatient care) and HES Outpatient datasets (described in section 4.3.1) and the variables used can be found in Table 4:3. For each patient, all inpatient and outpatient HES records occurring up to six months prior to their 'date of sample' were extracted for analysis, excluding records for outpatient appointments that were cancelled by, or on behalf, of the patient (ATTENDED code 2). 'Date of sample' is the date that the sample which led to the patient's DLBCL diagnosis was taken and is obtained from the HILIS data (section 4.2), hereinafter referred to as date of diagnosis or diagnosis. This is often used by cancer registries as the diagnostic date and was used in this thesis in lieu of the formal diagnostic date because diagnosing haematological malignancies is complex and the time taken to complete this process can vary considerably, whereas date of sample is consistent across all patients. Consequently, some DLBCL patients may be effectively 'diagnosed' in clinic and

have appointments regarding disease management before they receive a formal confirmation of their diagnosis from HMDS and it would be inappropriate to include HES records relating to these appointments when determining the patient's diagnostic route.

If there was no HES activity for the six months prior to diagnosis the patient was assigned to the 'Unknown' route. Patients who had at least one HES episode in the six months before they were diagnosed were sorted into either the 'Inpatient HES Step' (as described in section 4.4.2.2) or 'Outpatient HES step' (as described in section 4.4.2.3), depending if the last record before diagnosis, the end point of their route, was an inpatient or outpatient episode. Some patients have multiple HES episodes on this date, each of which could be identified as the end point of the route but which could result in different RTDs being assigned. Therefore, it is important that records are organised consistently, first by date, then by 'priority order' and lastly for inpatient records only using a HMRN variable called CIPS_ID which uniquely identifies each inpatient record in chronological order. The priority order used in this analysis is an adapted version of the order described in the original Elliss-Brookes et al., 2012 paper (Table 4:4) which prioritises activity using ADMIMETH codes (hereinafter referred to as Admission Method) and Referral Source codes based on the RTDs associated with them. This ordering should not be confused with the PRIORITY (hereinafter referred to as Priority) variable in the outpatient HES data which is used to identify HES activity that resulted from TWW referrals.

If there was both inpatient and outpatient records on the date with HES activity closest to their diagnosis the patient was assigned to the 'Outpatient HES step'. If this was not the case and the patient had one or multiple HES episodes within 28 days of date of diagnosis and one of these was an inpatient episode they were assigned to the 'Inpatient HES step'. If they only had outpatient HES activity within the 28 days preceding their diagnosis they were assigned to the 'Outpatient HES step'. If a patient did not have both inpatient and outpatient activity on the same day or a HES episode in the 28 days before diagnosis but did have at least one HES record within the six months prior to diagnosis, they were assigned to either the 'Inpatient HES step' or 'Outpatient HES step' depending on whether the HES record the shortest amount of time before their date of diagnosis was for an inpatient or outpatient episode.

4.4.2.1 Additional Two-Week Wait step

In this step, information from the Referral Source and Priority variables in outpatient HES records were combined to identify patients referred to secondary care under the TWW pathway in the 62 days prior to- or 14 days after- their date of diagnosis. The Referral Source code was used to identify patients referred to secondary care by a GP and the

Priority code indicated the urgency of this referral; 1 for ‘routine’, 2 for ‘urgent’, 3 for ‘two-week wait’ and 9 for ‘not known’. However, the TWW code of the Priority variable is only available for HES episodes from 2008 onwards meaning that for episodes occurring before 2008 there is no way to distinguish between routine and TWW referrals made by GPs in this timeframe. Therefore, if such an episode was identified as the start point of a patient’s RTD, they would be assigned to the GP Referral route regardless of whether the referral was routine or made under the TWW pathways.

In response to concerns that this could lead to underestimation of the TWW route, the route was prioritised in the algorithm and any HMRN patient identified as having had a TWW referral in the 62 days leading up to diagnosis or 14 days after diagnosis were assigned to this route. Furthermore, TWW referrals were only considered by the original algorithm if a ‘decision to treat date’ was within 62 days before- or 31 days after- diagnostic date, however, this rule was not applied to this algorithm as HMRN does not have access to the ‘decision to treat date’ which is part of the CWT dataset. Therefore, it is possible that some patients may be incorrectly assigned to the TWW RTD if they had a referral to secondary care under the TWW pathway in the 62 days leading up to diagnosis or 14 days after diagnosis that did not directly lead to treatment for their DLBCL. However, it seems unlikely that many TWW referrals made this close to a patient’s diagnosis would be unrelated to their DLBCL.

4.4.2.2 Using inpatient data to identify the start point of the route

The ‘Inpatient HES step’ starts by characterising the admission method of the episode identified as the end point of the route. If the admission method was a transfer (see Table 4:4 for admission methods codes) the HES data was searched iteratively for a previous episode within the six-month time-frame where the admission method was not a transfer. If no such episode could be identified the patient was assigned to the ‘Inpatient Elective’ RTD.

The data was searched for patients where the admission method for the HES record identified as the end point of their route indicated that they were admitted to hospital following an emergency attendance (see Table 4:4 for admission method codes) and these patients were assigned to the ‘Emergency Presentation’ route. If the admission code of the HES record identified as the end point of the route indicated that a patient was admitted following an emergency attendance via a consultant clinic (code 24), the HES records were searched for an outpatient appointment that preceded this inpatient HES record. If a previous outpatient record could be identified, the patient was redistributed to the ‘Outpatient HES step’; if no previous outpatient attendance for the six-month time frame existed the patient was assigned to the ‘Emergency Presentation’ route.

If the admission method of the end point of the route indicated that a patient was admitted ante-partum (code 31), the patient was assigned to the ‘Inpatient Elective’ route. If any of the other codes associated with the ‘Inpatient Elective’ route described in Table 4:4 (excluding code 81) were used to describe the admission method, the patients HES records were searched for an earlier outpatient appointment. This is because these codes represent either being on a waiting list, having a booked admission, or a planned admission and therefore prior outpatient HES activity relevant to their RTD is anticipated. If a previous outpatient appointment was identified, the patient was redistributed to the ‘Outpatient HES step’, otherwise they were assigned to the ‘Inpatient Elective’ route category.

4.4.2.3 Using outpatient data to identify the start point of the route

The ‘Outpatient HES step’ starts by characterising the referral source of the episode identified as the end point of the route. If the referral source of this episode indicated that the appointment was made following a consultant-to-consultant referral (code 5) the patients HES records were checked iteratively for a previous outpatient appointment that was not a consultant referral. If no such previous record existed, the patient was assigned to the ‘Other Outpatient’ route. If a previous outpatient appointment was found, this was determined to be the start point of the route and the RTD was assigned based on the referral source code attached to this appointment (as described in Table 4:4). If the HES record determined to be the end point of the route was not a consultant referral, the RTD was assigned based on the referral source code attached to this appointment.

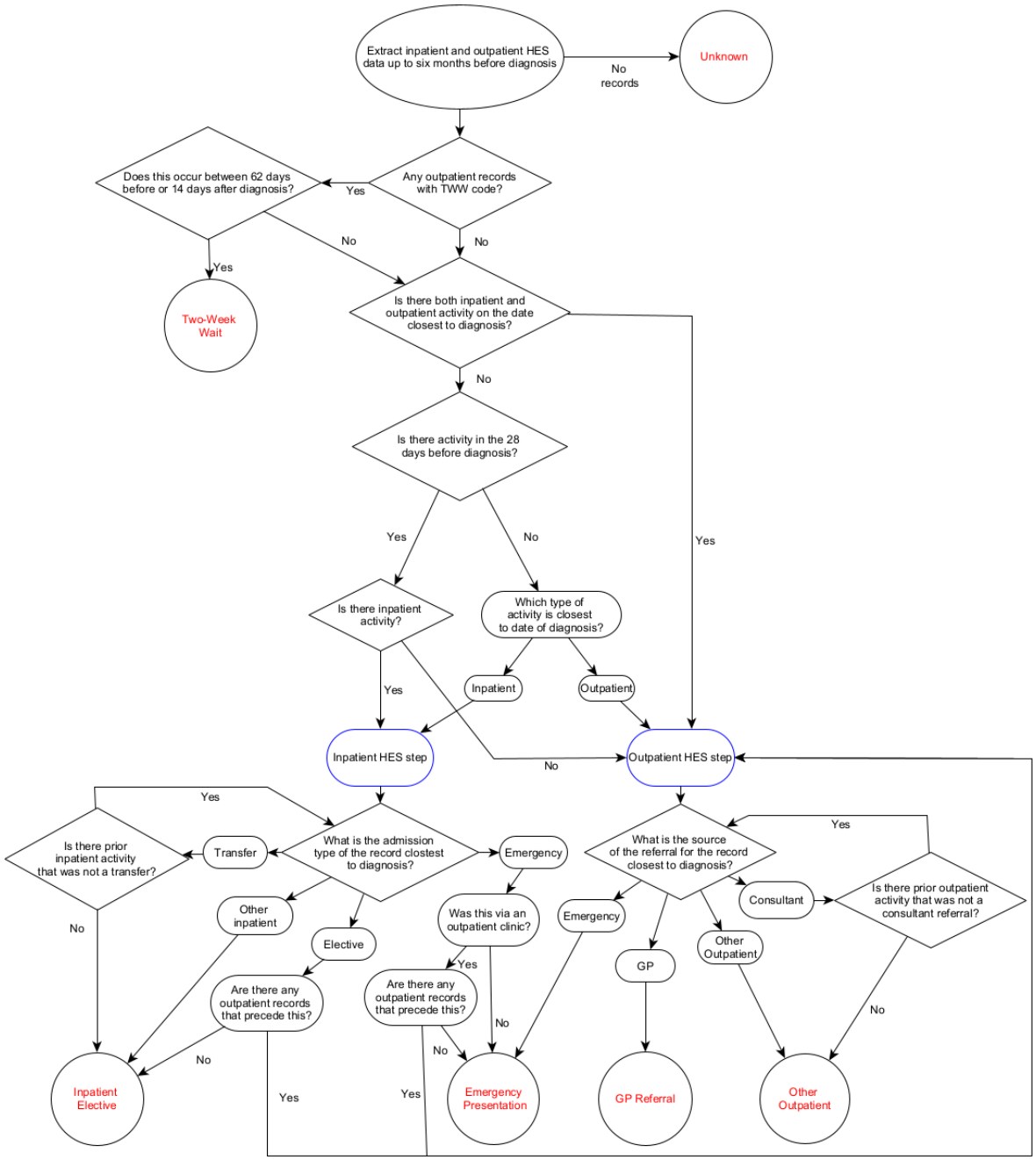


Figure 4:6. Flow chart illustrating how route to diagnosis was assigned for HMRN patients.

4.5 Travel data

4.5.1 TRACC

Travel data were generated using the desktop version of TRACC 2.1.4 (Basemap). The distance travelled and time taken to complete journeys between different origin and destination points were calculated by private car and public transport separately. There are a number of settings which can be adjusted when preparing calculations in TRACC. Notably, all estimated journeys were outbound (i.e., from the origin point to the destination) and the maximum distance an origin point could be from a road for the calculation to run was 800m. The maximum distance between an origin and destination point as the crow flies was set at 100km. For journeys by car, the time taken was estimated for morning peak traffic conditions and the average speeds used to calculate these journey times were:

- Minor road - 35km/hr
- A road – 75km/hr
- B road – 65km/hr
- Motorway – 100km/hr
- Walking speed – 4.8km/hr

For journeys made by public transport the time frame that the journey had to be completed within was set at between 7-10am on a Monday, which was extended by an hour from the default setting of 7-9am to reflect the long distances some patients in the Network might need to travel to access healthcare. Monday morning is the default setting in TRACC and this was not changed because many patients are likely to have had morning appointments during the working week, when congestion is typically worse than other times of the day or on a weekend. The calculations were not restricted to a maximum number of changes per journey but the maximum distance to the first stop was set at 2km and the maximum internal connection (i.e., the distance between getting off at a stop and walking to another stop to make the next leg of the journey) was set at 500m.

4.5.2 Origin point

The patient's postcode at the time of diagnosis was used as their origin point in the travel calculations. Postcode data is highly identifiable so to preserve patient confidentiality all postcodes covered by the HMRN region, instead of only patients' postcode, were mapped to all destination points in TRACC. The postcodes for the HMRN region were found by taking

the postcodes that belong to the Cancer Networks that mapped to our study region from the August 2022 NHS postcode directory.

To produce contour maps, a shapefile covering the HMRN region was generated in TRACC using the DataCutter tool, with a 250m grid overlaid. Each point on the grid was treated as an origin point, instead of postcodes, to produce smoother contours on the maps.

4.5.3 Destination points

For each patient in the study there were two destination points; the postcode of their GP practice at diagnosis and the hospital where the patient was diagnosed. The postcodes of the hospitals were taken from the contact section of their websites and the postcodes of the GP practices were taken from the EPRACCUR data set which is created by NHS Digital and lists all GP practices addresses and postcodes.

4.5.4 Defunct postcodes

For a small number of home addresses (n=11) and GP practices (n=53) the postcode recorded in the HMRN database could not be located using the OS Code Point function in TRACC either because the postcode was no longer in use (defunct) or was outside the HMRN Network. Consequently, travel distances or times could not be calculated for some patients.

To minimise the amount of missing data the northings and eastings of defunct postcodes were used to find the nearest active postcode in the NHS postcode directory so that the travel distances and times for these nearest active postcodes could be used as proxies for the values of the retired postcodes.

4.5.5 Road network data

The road network data used in this analysis was Meridian 2 data which was downloaded as a shapefile from the DataCutter tool in TRACC. Meridian 2 data are derived from the Ordnance Survey Roads database and has a resolution of 1m. It contains data for motorways, major roads and minor roads as of 2016 but is not fully comprehensive, for example, cul-de-sacs less than 200m are not represented unless they terminate at a roundabout or leave an isolated developed land use area and the data does not contain information on drive restrictions such as one-way streets.

4.5.6 Public transport data

The timetables for the bus, tram and train services in operation in the Yorkshire and Humber region in 2017 were downloaded into TRACC using the DataCutter tool. To download public transport data from the DataCutter tool you first select the region which you require public transport data for. Because services often run across regions, the DataCutter tool does not truncate journeys at the perimeter of the region selected and instead includes in the download the full timetable of the service for any journey that enters the region. For example, the region selected for this analysis was Yorkshire and the Humber, so if a train journey terminated in Leeds and started in Manchester, the entire route would be downloaded because part of the journey is in the selected region.

These digital public transport timetables are available in quarterly updates and for this analysis the 2017 dataset was downloaded because it was the oldest, complete data set available in the DataCutter tool at the time of the download. However, some patients in this thesis were diagnosed over a decade before 2017 and consequently the public transport network may have been quite different when they travelled to access healthcare regarding their DLBCL diagnosis. The limitations of this data are discussed in more detail in the Discussion chapter.

4.5.7 Measuring distance and time to GP, diagnostic hospital and treating hospital by private car and public transport

Using TRACC, travel distances and times by both car and public transport between all origin points to all GP practices in the HMRN region and to all HMRN hospitals were calculated meaning that, where possible, travel data for all origin-destination combinations in the HMRN region was generated. For each origin-destination combination, TRACC calculates the fastest route for morning peak travel conditions using the Dijkstra shortest path algorithm. The travel data generated was then exported as CSV files into a secure table in the HMRN database and patients were linked to this data using their home address postcode in combination with their GP's postcode and the postcode of their diagnostic hospital. Any identifiable data, such as the patient's home address postcode, was deleted before the data was saved and exported for analysis in Stata.

4.5.8 Heat maps

To generate heat maps illustrating the accessibility of hospitals across the HMRN region, a shapefile was created outlining the boundary of HMRN, with a 250m grid overlaid. Travel times by car and public transport from each point on the grid to the nearest HMRN hospital

were generated and heat maps illustrating the time to complete this journey by car and public transport were produced separately.

Using the grid as the origin points for the calculations, instead of individual patient postcodes is beneficial because the heat map produced covers the entire Network whereas using patient postcodes would produce a sparser heat map because there are more postcodes in the Network than patients and consequently large areas of the Network may not be represented by a patient included in this thesis. Using the grid instead of postcodes also does not risk exposing the identities of patients who live in less populated areas. Sometimes the calculation was not possible if the origin point was over 800m away from a public road or for public transport calculations if the origin point was more than 2km from a train station, bus stop or tram stop. These areas are displayed as 'blank' areas on the heat maps with no colour overlaid.

4.6 Descriptive statistics

Descriptive statistics were undertaken to explore the variation in and distribution of variables and to understand the relationship between the covariates and the outcome variables prior to undertaking regression analyses. Summary statistics of characteristics of the study participants were described first including sex, age at diagnosis, disease subtype and stage, using standard descriptive methods including counts and percentages. The distribution of each of the continuous variables was examined using histograms and box and whisker plots prior to undertaking any statistical analysis to ensure that the most appropriate tests were used. Due to the non-normal distribution of many of the variables used in this thesis medians and inter-quartile ranges were used to describe the average values for continuous variable and non-parametric tests such as Wilcoxon rank-sum and Kruskal-Wallis were used to test for differences in these variables by covariates. To test differences in observed and expected values of categorical outcome variables, such as route to diagnosis, between groups of patients' chi-squared tests were used.

4.7 Hospital activity rates in the pre- and post- diagnostic period

To examine the pattern of hospital activity for patients with DLBCL in HMRN compared to that in an age-sex matched cohort from the region (as described in section 4.3.4), hospital activity rates for HMRN patients (cases) and their matched controls for the two years before- and year after- diagnosis (pseudo-diagnosis for matched controls) were calculated. The purpose of this was to determine if patients, particularly those diagnosed via the Emergency Presentation diagnostic route, attended hospital more frequently in the months leading up to their diagnosis, which could indicate that opportunities for earlier diagnosis were missed. All

inpatient and outpatient HES activity for cases and their matched controls that occurred in the 24 months before- and 12 months after- diagnosis/pseudo-diagnosis were extracted. Cases and their matched controls were excluded if the case had no HES records corresponding to this period of time or they had died over a month before diagnosis. Each unique HES record was counted as an activity and for individual cases and controls, a total number of activities was counted for each month starting at 24 months before diagnosis/pseudo-diagnosis and ending 12 months after diagnosis/pseudo-diagnosis. The total number of activities per month for all cases and controls was summed separately and from this, an average activity rate per 100 people was calculated by dividing the total activity for that month for cases/controls by the total number of cases/controls multiplied by 100. All cases and controls were included in the denominator for each month regardless of whether they contributed any activity to that month, on the condition that they are alive at the start of the month. Rate ratios with 95% confidence intervals comparing the activity of cases and controls were also produced.

4.7.1 Patients with missing travel data

There are a small number of patients (n=337, 8.2%) for whom it was not possible to link to some or any travel data because: the patient's residential postcode at diagnosis was not recorded in the database, the patient's GP was not recorded in the database, they lived outside the study region or the postcode could not be located in TRACC.

4.8 Measures of association

Binary logistic regression is a commonly used statistical technique that predicts the likelihood of an outcome with two possibilities, based on one (univariate) or more (multivariate) independent variables/covariates. The 'logit' command and sub-commands in Stata were used to calculate unadjusted and adjusted odds ratios using univariate and multivariate logistic regression respectively to examine if certain groups of patients, such as those with longer travel times to healthcare, were more likely to experience the outcome of interest, such as whether a patient received treatment with curative intent (yes/no).

Continuous variables (age at diagnosis and travel variables) were converted into categorical variables after likelihood-ratio tests (using the 'lrtest' command in Stata) revealed no difference in model fit between the continuous and categorical variables as demonstrated by the p-values which were all non-significant at the 95% confidence level. The largest group was generally chosen as the reference category, except where categories were similar in size and one group aided interpretation. For example, the largest age group category was

chosen as the reference group but for IMD, quintile one (least deprived) was chosen as the reference group.

All variables examined in the univariate models were also included in the multivariate models, unless they displayed collinearity with other variables in the model. Multicollinearity was assessed using Spearman's correlation coefficient ('correlate' command in Stata) with an *a priori* cut-off of r=±0.4.

4.8.1 Outcomes

The outcome variables were: route to diagnosis and treatment intent.

The data sources and method used for assigning route to diagnosis are described in section 4.4.2. The algorithm used to assign route to diagnosis categorises patients into five distinct routes which were collapsed into two groups for the regression analysis: Emergency Presentation and Elective Referrals (which comprised of all other routes excluding the Unknown route). This decision was made due to concerns about how accurately the TWW route could be assigned using the data available to HMRN (described in more detail in section 6.2.4). Additionally, as outlined in the aims and objectives (2), one of the questions this thesis aimed to answer is whether particular groups of HMRN patients with DLBCL are more likely to be diagnosed following an Emergency Presentation, which is associated with poor outcomes, and therefore this is the route we were primarily interested in distinguishing.

The data sources and method for determining if a patient was treated with curative intent or not is described in section 4.4.1. This was an important outcome variable as DLBCL is an aggressive disease which left untreated has a very poor prognosis, therefore it is important to know if certain groups of patients are less likely to receive curative treatment, such as those with longer journeys to healthcare as described in the aims and objectives (2). In the logistic regression where treatment intent was the outcome variable, route to diagnosis was included as a confounder in the adjusted model.

4.8.2 Exposures

The exposure variables in these analyses were: deprivation (IMD quintile), rurality (rural-urban classification) and travel to healthcare.

A deprivation score was assigned to individuals using the IMD, which is released by the Ministry of Housing, Communities & Local Government and described in more detail in sections 1.5.2.2 and 4.3.2. Every LSOA in England is assigned a rank from most to least deprived based on their IMD score and for this thesis a deprivation score of 1 (least

deprived) to 5 (most deprived) was assigned to each individual depending on the quintile their home address at diagnosis belonged to according to the ranks, i.e., those assigned a score of 5 belonged to an LSOA that ranked in the 20% most deprived in the country.

HMRN patients were assigned a rural-urban classification based on the 2011 Office for National Statistics Rural/Urban Classification (RUC2011) (as described in section 1.5.4) of the LSOA of their home address at diagnosis. At the LSOA level, eight rural-urban categories can be assigned but, in this thesis, this was collapsed into a binary variable with individuals assigned to either the 'Rural' (rural town and fringe, rural town and fringe in a sparse setting, rural village and disperse, rural village and dispersed in a sparse setting) or 'Urban' (urban major conurbation, urban minor conurbation, urban city and town, urban city and town in a sparse setting) group.

Travel variables were made using the travel data generated in TRACC as described above in section 4.5. As shown in Figure 4:7 and Figure 4:8, the distributions of the continuous travel variables are skewed right with the majority of patients making short journeys to access their GP and hospital, and outliers representing a relatively small group of patients that had considerably longer journeys to access these services. Due to the right-skewed distributions, the 25th percentile, 75th percentile and 95th percentile were chosen as the cut-off values to generate categorical variables with four levels respectively: low, medium, high and highest. The 95th percentile was chosen to capture those with the most extreme travel distances and times as it is plausible that this group of patient's outcomes may differ. An additional category was also created for those with missing data ('not known') and a 'timed out' category for the public transport variables which indicates where the calculation was possible but could not be completed within the set time, as described in section 4.5.1. These cut-offs, which have been used previously to distinguish those with the most extreme travel (Haynes, Pearce and Barnett, 2008), were chosen *a priori* following extensive discussion at my Thesis Advisory Panel meetings regarding the most appropriate way to use the travel data. Furthermore, sensitivity analysis where the travel variables were arranged into categories using the 25th, 50th and 75th percentiles was conducted and there was no significant difference to the odds ratios calculated.

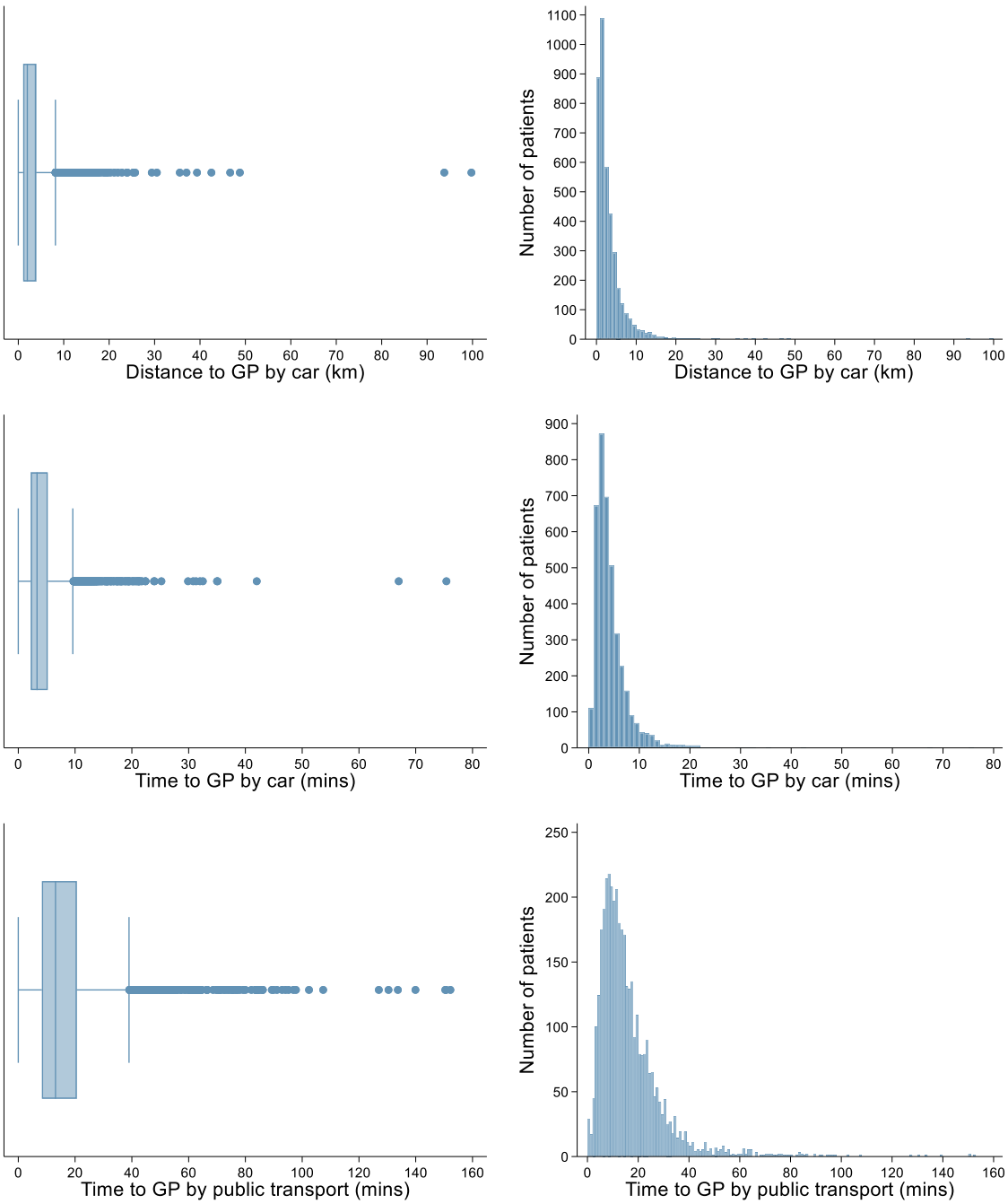


Figure 4:7. Box plots and histograms showing the distribution of travel distance in kilometres (km) and time in minutes (mins) by private car and public transport to patients' registered GP.

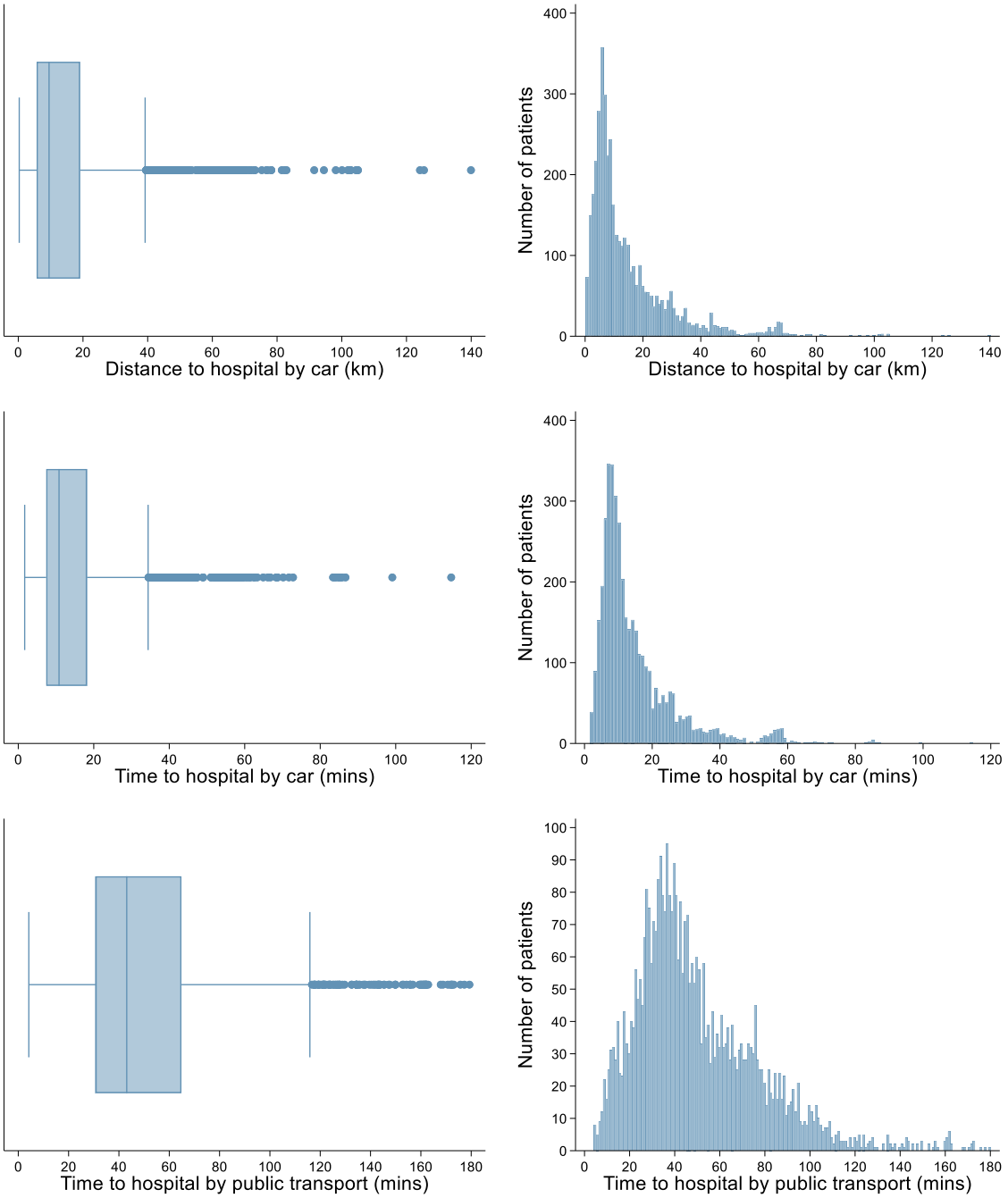


Figure 4:8. Box plots and histograms showing the distribution of travel distance in kilometres (km) and time in minutes (mins) by private car and public transport to patients' diagnostic hospital.

The four travel time variables can be thought of as two pairs; journey times to GP and journey times to hospital. The correlation between these pairs of variables is clearly illustrated in Figure 4:9, with those who have shorter journeys by car to their healthcare destination also generally having shorter journeys by public transport. These variables are highly, but not perfectly, correlated and although many patients will be in the same categories for the two pairs of variables, some will be in different categories. For example, for journeys to GP, two patients in the ‘highest’ category for time by car are in the ‘medium’ category for time by public transport and a further 3 are in the ‘timed out’ category. This shows that the variables for journeys by car and journeys by public transport are similar but distinct and justifies examining them individually.

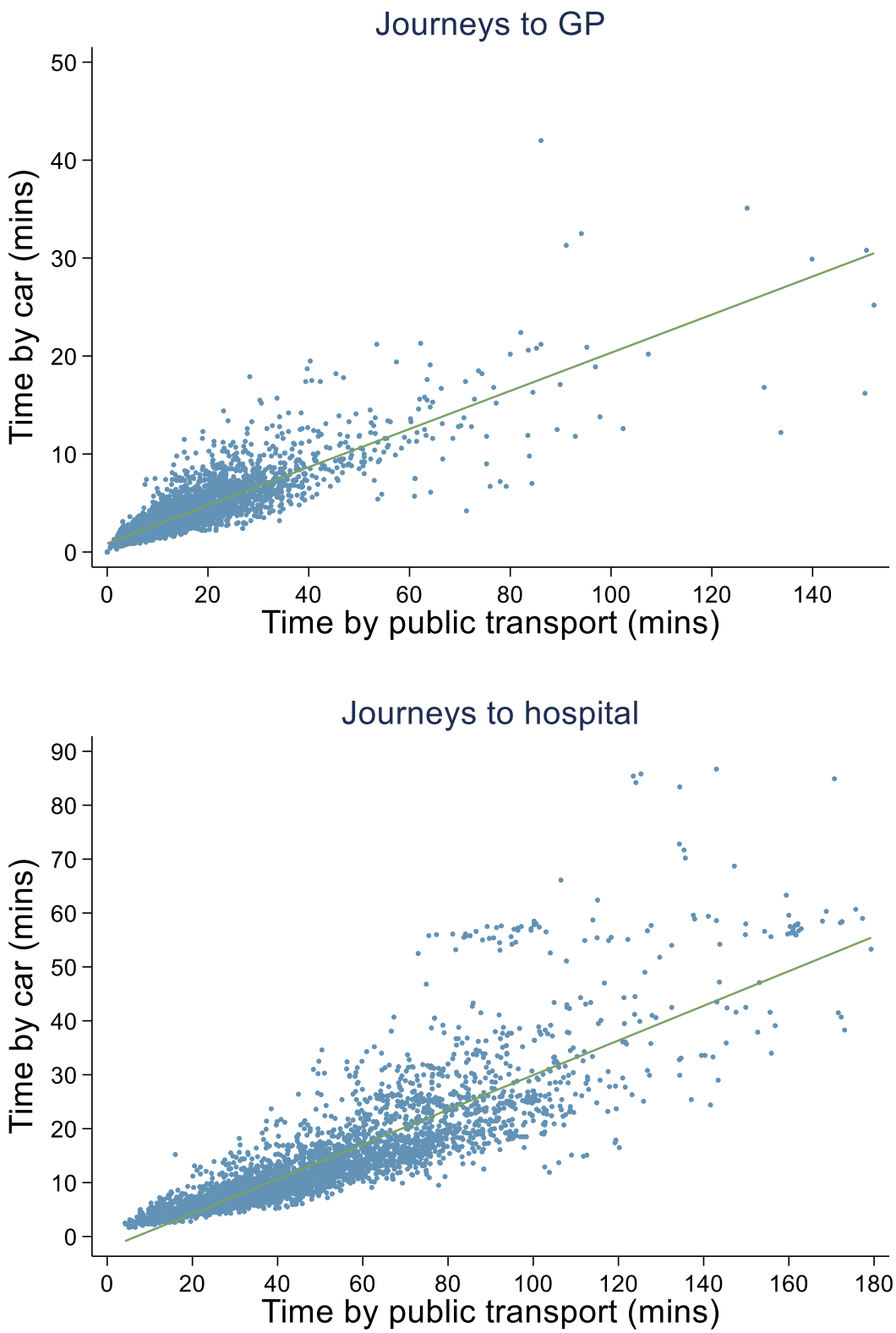


Figure 4:9. Scatter plots showing the relationship between journey time by public transport and car in minutes (mins) to access GP and hospital

4.8.3 Covariates

Covariates were selected *a priori* if they were of interest or were clinically meaningful. The covariates chosen for these analyses were: sex, age at diagnosis, diagnostic subtype (PCNSL or non-PCNSL), stage, B symptoms, performance status (Eastern Cooperative Oncology Group [ECOG] score), diagnostic hospital and Cancer Alliance. These are important because they may impact how patients are diagnosed, their likelihood of receiving curative treatment and survival. For example, PCNSL is an aggressive subtype which may need to be adjusted for in some analyses or have results presented separately. Patients with PCNSL may experience different symptoms to other DLBCL patients, such as headaches and blurred vision, which could alter their susceptibility to diagnostic delay. These patients are also more likely to be referred to St James's Hospital in Leeds for their diagnostic work up (as this is a specialist centre in the Network) which could result in them being exposed to longer travel to hospital than other patients. Similarly, socio-demographic and geographical factors such as Cancer Alliance are important to consider. For example, the areas covered by Humber, Coast & Vale Cancer Alliance are more geographically isolated and have fewer transport connections compared to areas belonging to the West Yorkshire & Harrogate alliance which are more urbanised, which could influence patients' outcomes.

Information on sex, age at diagnosis, stage at diagnosis, diagnostic subtype, B symptoms, performance status and diagnostic hospital were obtained from data routinely collected by HMRN. Age at diagnosis is collected for all HMRN patients and is their age in years at the time they received their DLBCL diagnosis. For the regression analyses, this was converted into a categorical variable made up of six age groups (<18, 18-39, 40-59, 60-69, 70-79, ≥80). Sex was a binary variable (male and female) and represents the patient's biological sex, obtained from their medical records. Stage at diagnosis is Ann Arbor stage (as described in section 1.3.3) of disease at presentation and is derived from Computed Tomography (CT) and Positron Emission Tomography (PET) scans. For the regression analyses, diagnostic subtype was condensed into a binary variable which was coded as yes/no for PCNSL, which is an important subtype to identify because it is particularly aggressive and outcomes for these patients are generally poor. B symptoms are recorded in the HMRN data as 'yes' if it is noted in the medical records that the patient had at least one B symptom (fever, weight loss, night sweats), 'no' if none of these were mentioned in the records and 'null' if it was non-calculable (as described in section 1.3.4). Performance status is taken from the patient's medical notes and recorded in HMRN as 0,1,2,3 or 4 as per the standard ECOG scores (described in section 1.3.5) and a 'not known' group for patients where ECOG status could not be abstracted from the medical records.

Diagnostic hospital is the hospital where the first diagnostic sample was taken and was used in this thesis instead of treating hospital (which is defined as the hospital where the first treatment was given) because not all patients in the cohort are treated and therefore treating hospital had a higher proportion of missing values. It was also reasoned that most patients would be diagnosed at the hospital the shortest distance away from their home address and as all hospitals in the HMRN are capable of delivering the standard treatment for DLBCL, any patient not treated at their diagnostic hospital may have chosen to travel further for care and this longer journey would not represent poor accessibility to care as intended.

The Cancer Alliance assigned to each patient was determined using their diagnostic hospital. There are 21 Cancer Alliances in England, two of which serve the HMRN region; West Yorkshire and Harrogate Cancer Alliance and Humber, Coast and Vale Cancer Alliance. The West Yorkshire and Harrogate Cancer Alliance includes: Airedale NHS Foundation Trust, Bradford Teaching Hospitals NHS Trust, Calderdale and Huddersfield NHS Foundation Trust, Mid Yorkshire Teaching NHS Trust and Leeds Teaching Hospitals NHS Trust. The Humber, Coast and Vale Cancer Alliance includes: Hull and East Yorkshire Hospitals NHS Trust, North Lincolnshire and Goole Hospitals NHS Foundation Trust and York Teaching Hospitals NHS Foundation Trusts. Cancer Alliances are partnerships of hospital trusts, health and social care organisations that manage the care of patients with cancer locally. They were introduced from 2016 as part of the NHS Cancer Strategy and replaced the former Cancer Networks (The Independent Cancer Taskforce, 2015).

Cancer Alliance was not included in multivariate regression analyses because it was completely determined by diagnostic hospital and including both variables would have introduced redundancy and multicollinearity into the models.

4.9 Estimating impact on survival

Survival analysis, or time-to-event analysis, is concerned with examining both if and when an event, in this case death, occurs. In this thesis, overall survival, net survival and hazard ratios were calculated by patients: sex, age group, diagnostic subtype, stage at diagnosis, B symptoms, performance status (ECOG score), deprivation (IMD quintile), rurality (rural-urban classification), travel to healthcare, treatment intent and route to diagnosis.

4.9.1 Overall survival

Overall survival is a crude estimate of survival that compares the number of patients alive at a defined start point, usually diagnoses, to those still alive at another point in time, for example, one year after diagnosis. The ‘sts’ command and subcommands in Stata were

used to calculate the survivor function, which is the probability of surviving to time *t* or beyond, using Kaplan-Meier estimates. Kaplan-Meier estimates survival using:

- Serial time - amount of time each individual was observed from the start of the study until an event
- Event status – whether the event, in this case death, occurred or if the patient was censored
- Group – which group that patient belongs to (e.g., age group), if we are estimating survival for different groups separately

4.9.2 Net survival

To estimate the impact of DLBCL alone on survival of patients, an estimate that is not influenced by other causes of death, often referred to as competing risks, is useful. Net survival is not influenced by other causes of death and was estimated using the 'stns' command and sub-commands in Stata, which produces Pohar Perme estimators. Population mortality estimates were taken from lifetables for the Yorkshire and the Humber region stratified by calendar year of diagnosis, age, sex and IMD quintile (NHS England, 2023b).

4.9.3 Cox proportional-hazards model

Cox proportional hazard (PH) models offer a multivariable approach to time-to-event analysis that allows for adjustment for several potential confounders and also the magnitude of the effect of each factor. This is advantageous over univariate analysis, such as Kaplan-Meier, when comparing survival between groups as it is likely that multiple factors, or covariates, will impact patients' survival. The 'stcox' command in Stata was used to produce hazard ratios. The proportional hazards assumption was assessed using the 'estat phtest' command, which is based on Schoenfeld residuals, using a p-value of >0.05, alongside log-log plots generated using the 'stphplot' command; stage at diagnosis and performance status (ECOG score) did not meet this assumption. However, after conducting a sensitivity analysis using stage at diagnosis and performance status as time-varying covariates, hazard ratios and significance levels remained consistent with the final models.

RTD and treatment intent also violated the proportionality assumption and because these were important outcomes this was handled by splitting the survival follow-up time and displaying results for only those treated with curative intent. The plots showed that for RTD, survival was divergent in the first year following diagnosis, with the Emergency Presentation group having a much steeper survival curve than the Elective Referral group between date

of diagnosis and one-year post-diagnosis, followed by the survival curves becoming almost parallel for the two RTD groups between years one- to five- post-diagnosis (see Figure 5:15). To account for this, survival estimates were presented separately for these two time periods: date of diagnosis to one-year post-diagnosis, and one- to five-years post-diagnosis.

5 Results

5.1 Study population baseline characteristics

A total of 4178 patients were newly diagnosed with diffuse large B-cell lymphoma (DLBCL) in the Haematological Malignancy Research Network (HMRN) region between January 2005 and August 2019 (Figure 5:1). Of these, 4115 patients had medical records available for abstraction of demographic and clinical data. Of the 4115 patients with complete data collection: postcode of home address at diagnosis could not be obtained for five patients, postcode of registered General Practice (GP) at diagnosis was not available for 100 patients and 63 patients could not be linked to Hospital Episode Statistics (HES) data.

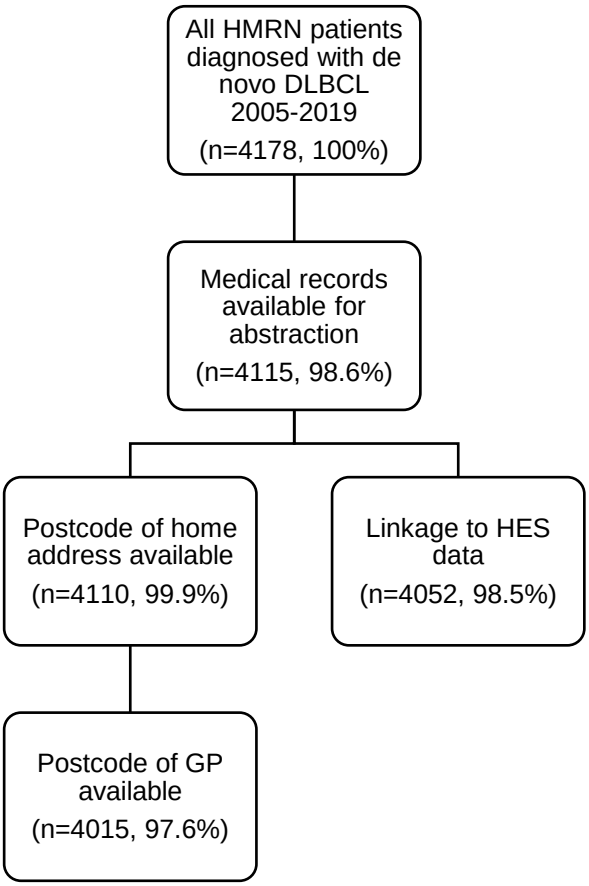


Figure 5:1. Completeness of data: HMRN patients newly diagnosed with DLBCL between January 2005-August 2019.
DLBCL=diffuse large B-cell lymphoma, HMRN=Haematological Malignancy Research Network, HES=Hospital Episode Statistics

More males were diagnosed with DLBCL (n=2179, 53.0%) than females (Table 5:1). Cases occurred in all age categories, but more cases were diagnosed in older patients compared to younger patients as illustrated by the median diagnostic age of 70.4 years and over half of

diagnoses occurred in patients aged over 70, compared to less than one percent of cases in patients aged under 18. The median was chosen to represent the diagnostic age of the typical patient in this cohort due to the skewed distribution of age at diagnosis described above. The majority of patients were diagnosed with diffuse large B-cell lymphoma, not otherwise specified (NOS) (n=3620, 88.0%), with the next largest subtype being primary diffuse large B-cell lymphoma of the CNS (PCNSL) (n=141, 3.4%). The largest proportion of patients were diagnosed with stage IV disease (n=1740, 42.3%) and some patients were not fully staged (n=498, 12.1%). Only 122 patients (3.0%) were assigned the highest Eastern Cooperative Oncology Group (ECOG) score of 4 compared to 1543 (37.5%) patients with a score of 1 and 1196 (29.1%) with a score of 0 and the majority of patients presented without B symptoms (n=2294, 55.7%). The most common route to diagnosis (RTD) was the GP route (n=1334, 32.4%), closely followed by Emergency Presentation (n=1292, 31.4%). The percentage of patients belonging to each quintile of the Index of Multiple Deprivation (IMD) Income Domain ranged from 20.1% (n=827) in quintile one (least deprived) to 21.0% (n=865) in quintile five (most deprived quintile), and almost 80% (n=3260) of patients lived in urban areas. Patients were diagnosed at hospitals across the HMRN region, with more patients under the care of the West Yorkshire & Harrogate Cancer Alliance (n=2469, 60%) than Humber, Coast & Vale Cancer Alliance, and the greatest percentage diagnosed at St James's Hospital Leeds (13.1%).

Table 5:1. Baseline clinical and socio-demographic characteristics and diagnostic route: HMRN patients newly diagnosed with diffuse large B-cell lymphoma between January 2005-August 2019.

	All Diagnoses N (%)
Total	4115 (100.0)
Sex:	
Male	2179 (53.0)
Female	1936 (47.0)
Age at diagnosis (years):	
<18	27 (0.7)
18-39	225 (5.5)
40-59	786 (19.1)
60-69	976 (23.7)
70-79	1245 (30.3)
≥80	856 (20.8)
Median (IQR)	70.4 (59.9-78.6)
Diagnostic subtype:	
Diffuse large B-cell lymphoma (DLBCL), NOS	3620 (88.0)
High grade B-cell lymphoma[a]	77 (1.9)
Large B-cell lymphoma, other[b]	40 (1.0)
Plasmablastic lymphoma	38 (0.9)
Primary diffuse large B-cell lymphoma of the CNS (PCNSL)	141 (3.4)

	All Diagnoses N (%)
Primary mediastinal large B-cell lymphoma	111 (2.7)
T-cell/histiocyte-rich large B-cell lymphoma	88 (2.1)
Stage:	
I	526 (12.8)
II	696 (16.9)
III	655 (15.9)
IV	1740 (42.3)
Not fully staged[c]	498 (12.1)
B Symptoms:	
Yes	1818 (44.2)
No	2294 (55.7)
Not known	3 (0.1)
ECOG score:	
0	1196 (29.1)
1	1543 (37.5)
2	823 (20.0)
3	365 (8.9)
4	122 (3.0)
Not known	66 (1.6)
Route to Diagnosis:	
Emergency presentation	1292 (31.4)
Two-week wait	745 (18.1)
GP	1334 (32.4)
Other outpatient	466 (11.3)
Inpatient Elective	86 (2.1)
Unknown/Diagnosed at death	192 (4.7)
IMD income quintile:	
1 (least deprived)	827 (20.1)
2	889 (21.6)
3	819 (19.9)
4	710 (17.3)
5 (most deprived)	865 (21.0)
Not known	5 (0.1)
Rural-urban classification:	
Rural	850 (20.7)
Urban	3260 (79.2)
Not known	5 (0.1)
Cancer Alliance:	
West Yorkshire & Harrogate	2469 (60.0)
Humber, Coast & Vale	1646 (40.0)
Diagnostic Hospital:	
Airedale	144 (3.5)
Bradford Royal	373 (9.1)
Calderdale	324 (7.9)
Castle Hill, Humberside	317 (7.7)
Dewsbury & District	243 (5.9)
Grimsby	184 (4.5)
Harrogate	208 (5.1)
Huddersfield Royal Infirmary	69 (1.7)
Hull Royal Infirmary	363 (8.8)

	All Diagnoses N (%)
Leeds General Infirmary	334 (8.1)
Pinderfields General	199 (4.8)
Pontefract General	36 (0.9)
Scarborough	122 (3.0)
Scunthorpe General	183 (4.4)
St. James's, Leeds	539 (13.1)
York	477 (11.6)

[a]High grade B-cell lymphoma NOS, high grade B-cell lymphoma MYC and BCL2 and/or BCL6 rearrangements
[b]Intravascular large B-cell lymphoma, large B-cell lymphoma with IRF4 rearrangement and primary cutaneous DLBCL, leg type
[c]Further staging scan and/or bone marrow assessment not performed
HMRN=Haematological Malignancy Research Network, IQR=Interquartile Range, ECOG= Eastern Cooperative Oncology Group, IMD=Index of Multiple Deprivation

5.2 Travel to healthcare

5.2.1 Completeness of travel data

In total, 91.8% (n=3778) of patients had complete travel data and only 16 patients (0.4%) could not be assigned any travel data, as shown in Figure 5:2. Some patients had missing data for some, but not all variables, for example, they have a travel distance and time to GP by car, but not a travel time by public transport to GP because the journey was not possible within the parameters set for a public transport journey as described in section 4.5.1.

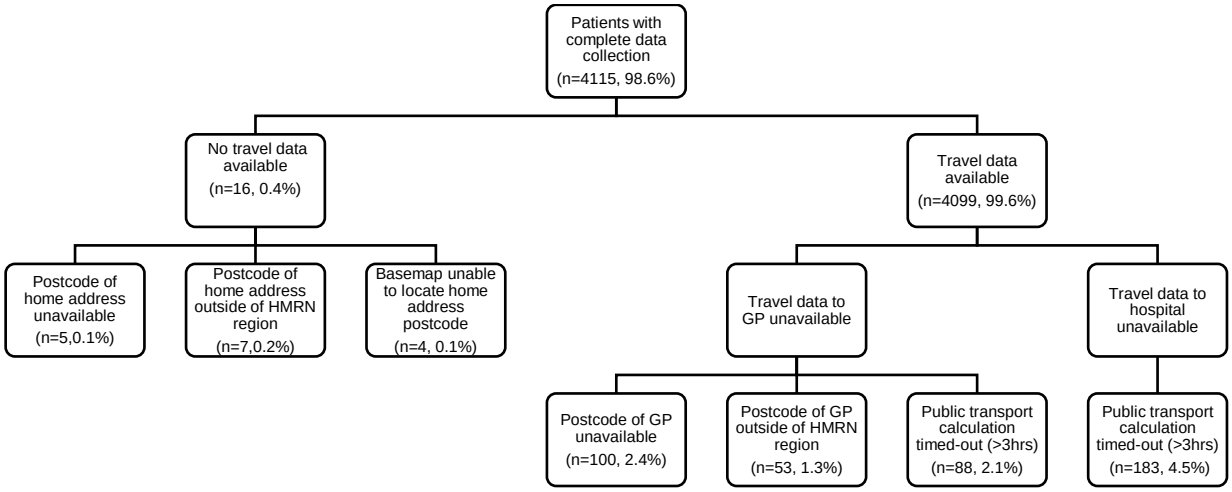


Figure 5:2. Completeness of travel data.

The median travel distance between patient's home address and registered GP at diagnosis was 2.0 km and the median time of this journey by private car and public transport

respectively was 3.3 minutes and 13.1 minutes (Table 5:2). Median travel distance between patient's home address and diagnostic hospital was considerably greater at 9.5 km and the median time of this journey by private car and public transport respectively was 10.8 minutes and 43.1 minutes.

Table 5:2. Travel in kilometres (km) and minutes (mins) by car and public transport to healthcare.

	All Diagnoses N (%)	**Range**
Total	*4115 (100.0)*	**-**
Distance to GP:		
Median (IQR) (km)	2.0 (1.1 - 3.9)	-
Low	988 (24.0)	0-1.07 km
Medium	1974 (48.0)	1.07-3.91 km
High	787 (19.1)	3.92-9.34 km
Highest	197 (4.8)	9.35-99.75 km
Not known	169 (4.1)	-
Time to GP by car:		
Median (IQR) (mins)	3.3 (2.2 - 5.2)	-
Low	1057 (25.7)	0-2.2 mins
Medium	1866 (45.3)	2.3-5.1 mins
High	822 (20.0)	5.2-10.4 mins
Highest	201 (4.9)	10.5-75.4 mins
Not known	169 (4.1)	-
Time to GP by public transport:		
Median (IQR) (mins)	13.1 (8.3 - 20.6)	-
Low	979 (23.8)	0-8.3 mins
Medium	1921 (46.7)	8.4-20.6 mins
High	764 (18.6)	20.7-37.7 mins
Highest	194 (4.7)	38.0-152.2 mins
Timed out	88 (2.1)	-
Not known	169 (4.1)	-
Distance to hospital:		
Median (IQR) (km)	9.5 (5.7 - 19.1)	-
Low	1025 (24.9)	0-5.67 km
Medium	2050 (49.8)	5.68-19.11 km
High	819 (19.9)	19.12-45.28 km
Highest	205 (5.0)	45.33-139.94 km
Not known	16 (0.4)	-
Time to hospital by car:		
Median (IQR) (mins)	10.8 (7.4 - 18.2)	-
Low	1029 (25.0)	0-7.4 mins
Medium	2051 (49.8)	7.5-18.2 mins
High	812 (19.7)	18.3-38.7 mins
Highest	207 (5.0)	38.8-114.7 mins
Not known	16 (0.4)	-
Time to hospital by public transport:		
Median (IQR) (mins)	43.1 (30.7 - 64.8)	-
Low	982 (23.9)	0-30.7 mins

	All Diagnoses N (%)	Range
Medium	1955 (47.5)	30.8-64.8 mins
High	784 (19.1)	64.9-100.9 mins
Highest	195 (4.7)	101.0-179.3 mins
Timed out	183 (4.4)	-
Not known	16 (0.4)	-

IQR=Interquartile range

5.2.2 Travel to GP

The median time of a journey to the GP by public transport was almost four times as long as the equivalent journey by car at 13.1 minutes (interquartile range (IQR): 8.3-20.7 mins) compared to 3.3 minutes (IQR: 1.1-3.9 mins) (Table 5:3).

There is some evidence that travel distance and time by private car and public transport varies by some patient and socio-demographic characteristics, but not clinical characteristics or diagnostic route. Female patients travelled less distance (Wilcoxon rank sum test=2.48, p=0.01) and spent less time travelling by car (Wilcoxon rank sum test=2.19, p=0.03) and public transport (Wilcoxon rank sum test=1.89, p=0.06) compared to male patients. Younger patients also had shorter journeys, for example, patients aged under 18 years travelled over half a kilometre less to access their GP (median: 1.4 km, IQR: 0.9-2.7 km) compared to patients aged 60 and over (median: 2.0 km, IQR: 1.1-4.0 km). No differences by subtype, stage at diagnosis, presence of B symptoms or ECOG score were detected. Similarly, there was no difference in travel distance or time to primary care between patients diagnosed as an Elective Referral compared to those diagnosed as an Emergency Presentation, although those with Unknown diagnostic route travelled further. There were also differences in distance and time spent travelling by IMD quintile (distance to GP by private car: Kruskal-Wallis test=200.60, p<0.001; time to GP by private car: Kruskal-Wallis test=185.41, p<0.001; time to GP by public transport: Kruskal-Wallis test=114.30, p<0.001). For example, patients living in areas belonging to IMD quintile 5 (most deprived) travelled over one kilometre less to their GP than those in IMD quintile 1 (least deprived). Unsurprisingly, compared to those living in urban areas those living in rural areas also had further to travel (Wilcoxon rank sum test=10.23, p<0.001) and spent more time travelling (time to GP by private car: Wilcoxon rank sum test=10.66, p<0.001; time to GP by public transport Wilcoxon rank sum test=5.19, p<0.001). Travel times and distances also varied by diagnostic hospital, for example, patients diagnosed at Scunthorpe General Hospital travelled twice as far to access their GP (median: 3.0 km, IQR: 1.2-5.6 km) compared to patients diagnosed at Huddersfield Royal Infirmary (median: 1.5 km, IQR: 0.8-3.0 km).

Table 5:3. Baseline clinical and socio-demographic characteristics and diagnostic route distributed by travel distance in kilometres (km) and time in minutes (mins) by car and public transport to registered GP practice at diagnosis.

	Patients with travel to GP by car data available	Distance to GP by private car (km)	Time to GP by private car (mins)	Patients with travel to GP by public transport data available	Time to GP by public transport (mins)
	N (%)	Median (IQR)	Median (IQR)	N (%)	Median (IQR)
Overall	*3946 (100.0)*	*2.0 (1.1-3.9)*	*3.3 (2.2-5.2)*	*3858 (100.0)*	*13.1 (8.3-20.7)*
Sex:					
Males	2094 (53.1)	2.1 (1.1 - 4.1)	3.4 (2.2 - 5.4)	2037 (52.8)	13.3 (8.4 - 21.0)
Females	1852 (46.9)	1.9 (1.1 - 3.7)	3.2 (2.1 - 5.1)	1821 (47.2)	12.9 (8.1- 19.8)
Wilcoxon rank-sum test (p-value)	-	2.48 (0.01)	2.19 (0.03)	-	1.89 (0.06)
Age at diagnosis (years):					
<18	25 (0.6)	1.4 (0.9 - 2.7)	2.8 (1.9 - 4.1)	25 (0.6)	12.9 (7.7 - 20.6)
18-39	207 (5.2)	1.7 (1.0 - 3.7)	3.1 (2.1 - 4.9)	206 (5.3)	11.7 (7.7 - 19.9)
40-59	760 (19.3)	1.9 (1.1 - 3.7)	3.2 (2.2 - 5.1)	746 (19.3)	12.8 (8.2 - 19.9)
60-69	937 (23.7)	2.0 (1.1 - 4.0)	3.4 (2.1 - 5.3)	902 (23.4)	12.9 (8.2 - 20.1)
70-79	1192 (30.2)	2.0 (1.1 - 4.1)	3.4 (2.2 - 5.3)	1165 (30.2)	13.4 (8.5 - 21.3)
≥80	825 (20.9)	2.0 (1.1 - 4.0)	3.3 (2.2 - 5.3)	814 (21.1)	13.2 (8.5 - 20.9)
Kruskal-Wallis test (p-value)	-	5.91 (0.32)	5.06 (0.41)	-	5.03 (0.41)
Primary CNS lymphoma:					
Yes	135 (3.4)	2.1 (1.0 - 4.1)	3.5 (2.1 - 5.3)	130 (3.4)	14.2 (8.7 - 23.4)
No	3811 (96.6)	2.0 (1.1 - 3.9)	3.3 (2.2 - 5.2)	3728 (96.6)	13.1 (8.3 - 20.5)
Wilcoxon rank-sum test (p-value)	-	0.53 (0.60)	0.37 (0.71)	-	0.98 (0.33)
Ann-Arbor stage:					
I	510 (12.9)	2.2 (1.2 - 4.0)	3.5 (2.2 - 5.2)	495 (12.8)	13.4 (8.4 - 21.0)
II	660 (16.7)	1.9 (1.1 - 3.9)	3.3 (2.2 - 5.2)	647 (16.8)	13.2 (8.3 - 20.9)

	Patients with travel to GP by car data available	Distance to GP by private car (km)	Time to GP by private car (mins)	Patients with travel to GP by public transport data available	Time to GP by public transport (mins)
	N (%)	Median (IQR)	Median (IQR)	N (%)	Median (IQR)
III	629 (15.9)	2.0 (1.0 - 4.2)	3.4 (2.1 - 5.6)	611 (15.8)	12.8 (8.3 - 20.8)
IV	1666 (42.2)	2.0 (1.0 - 3.8)	3.3 (2.1 - 5.2)	1632 (42.3)	13.1 (8.2 - 20.3)
Not fully staged	481 (12.2)	1.9 (1.1 - 4.0)	3.2 (2.3 - 5.3)	473 (12.3)	13.0 (8.6 - 20.3)
Kruskal-Wallis test (p-value)	-	2.31 (0.68)	0.88 (0.93)	-	1.20 (0.88)
B Symptoms:					
Yes	1737 (44.0)	2.0 (1.1 - 3.9)	3.3 (2.2 - 5.2)	1698 (44.0)	13.1 (8.3 - 20.3)
No	2206 (55.9)	2.0 (1.1 - 3.9)	3.3 (2.2 - 5.2)	2157 (55.9)	13.1 (8.2 - 20.8)
Not known	3 (0.1)	2.3 (1.7 - 5.2)	3.5 (2.6 - 6.5)	3 (0.1)	18.7 (11.6 - 32.5)
Wilcoxon rank-sum test (p-value)[a]	-	-0.27 (0.79)	0.08 (0.94)	-	0.24 (0.81)
ECOG score:					
0	1145 (29.0)	2.0 (1.1 - 4.0)	3.4 (2.2 - 5.2)	1120 (29.0)	13.2 (8.0 - 20.6)
1	1480 (37.5)	2.0 (1.0 - 4.0)	3.3 (2.1 - 5.3)	1445 (37.5)	12.8 (8.4 - 21.1)
2	799 (20.2)	2.0 (1.1 - 3.8)	3.3 (2.2 - 5.1)	783 (20.3)	13.0 (8.3 - 19.9)
3	348 (8.8)	2.0 (1.1 - 3.7)	3.4 (2.2 - 5.0)	338 (8.8)	13.1 (8.6 - 20.0)
4	112 (2.8)	1.9 (1.0 - 4.7)	3.4 (2.2 - 5.4)	111 (2.9)	14.3 (8.9 - 22.0)
Not known	62 (1.6)	2.1 (1.3 - 3.8)	3.5 (2.3 - 5.3)	61 (1.6)	13.3 (9.5 - 20.8)
Kruskal-Wallis test (p-value)	-	0.88 (0.97)	0.92 (0.97)	-	2.88 (0.72)
Route to Diagnosis:					
Emergency presentation	1236 (31.3)	1.9 (1.0 - 3.9)	3.3 (2.2 - 5.2)	1213 (31.4)	13.3 (8.1 - 21.0)
Elective referrals	2527 (64.0)	2.0 (1.1 - 3.9)	3.3 (2.2 - 5.2)	2468 (64.0)	12.9 (8.3 - 20.1)
Unknown	183 (4.6)	2.4 (1.3 - 4.6)	3.7 (2.3 - 6.1)	177 (4.6)	14.0 (8.8 - 23.0)
Kruskal-Wallis test (p-value)	-	4.61 (0.10)	3.41 (0.18)	-	3.30 (0.19)
IMD income quintile:					

	Patients with travel to GP by car data available	Distance to GP by private car (km)	Time to GP by private car (mins)	Patients with travel to GP by public transport data available	Time to GP by public transport (mins)
	N (%)	Median (IQR)	Median (IQR)	N (%)	Median (IQR)
1 (least deprived)	805 (20.4)	2.8 (1.4 - 5.2)	4.2 (2.6 - 6.7)	769 (19.9)	16.2 (9.8 - 24.2)
2	858 (21.7)	2.4 (1.2 - 4.6)	3.6 (2.3 - 5.8)	828 (21.5)	14.1 (8.9 - 22.4)
3	777 (19.7)	2.0 (1.1 - 4.0)	3.3 (2.2 - 5.2)	757 (19.6)	12.9 (8.4 - 19.9)
4	682 (17.3)	1.6 (0.9 - 3.1)	2.9 (1.9 - 4.5)	681 (17.7)	11.7 (7.2 - 17.8)
5 (most deprived)	824 (20.9)	1.6 (0.9 - 2.7)	2.9 (1.9 - 4.2)	823 (21.3)	11.9 (7.4 - 17.2)
Kruskal-Wallis test (p-value)	-	200.60 (<0.001)	185.41 (<0.001)	-	114.30 (<0.001)
Rural-urban classification:					
Rural	798 (20.2)	3.2 (1.2 - 6.6)	4.5 (2.3 - 7.6)	738 (19.1)	14.7 (9.0 - 23.8)
Urban	3148 (79.8)	1.9 (1.1 - 3.5)	3.2 (2.1 - 4.7)	3120 (80.9)	12.8 (8.2 - 19.6)
Wilcoxon rank-sum test (p-value)	-	10.23 (<0.001)	10.66 (<0.001)	-	5.19 (<0.001)
Cancer Alliance:					
West Yorkshire & Harrogate	2413 (61.2)	1.8 (1.0 - 3.3)	3.1 (2.0 - 4.6)	2389 (61.9)	12.0 (7.7 - 18.4)
Humber, Coast & Vale	1533 (38.8)	2.4 (1.2 - 5.1)	3.9 (2.4 - 6.6)	1469 (38.1)	15.0 (9.5 - 23.4)
Wilcoxon rank-sum test (p-value)	-	-9.47 (<0.001)	-10.92 (<0.001)	-	-9.82 (<0.001)
Diagnostic hospital:					
Airedale	143 (3.6)	1.7 (0.9 - 3.3)	3.1 (1.9 - 5.2)	136 (3.5)	11.1 (7.3 - 17.9)
Bradford Royal	362 (9.2)	1.6 (1.0 - 2.8)	2.9 (2.0 - 4.2)	362 (9.4)	12.1 (6.9 - 17.4)
Calderdale	311 (7.9)	1.7 (1.0 - 3.2)	3.0 (1.9 - 4.4)	310 (8.0)	10.6 (7.0 - 18.7)
Castle Hill, Humberside	302 (7.7)	2.4 (1.2 - 4.9)	4.0 (2.5 - 6.5)	292 (7.6)	14.3 (10.3 - 21.9)
Dewsbury & District	239 (6.1)	1.9 (1.1 - 3.5)	3.3 (2.1 - 4.5)	239 (6.2)	12.7 (8.1 - 18.4)
Grimsby	146 (3.7)	2.9 (1.4 - 5.6)	4.0 (2.6 - 6.5)	146 (3.8)	16.4 (10.0 - 26.5)
Harrogate	203 (5.1)	2.2 (1.2 - 4.0)	3.5 (2.3 - 5.2)	190 (4.9)	14.6 (9.2 - 21.6)
Huddersfield Royal Infirmary	69 (1.7)	1.5 (0.8 - 3.0)	2.9 (2.1 - 4.1)	69 (1.8)	11.2 (7.0 - 17.6)

	Patients with travel to GP by car data available	Distance to GP by private car (km)	Time to GP by private car (mins)	Patients with travel to GP by public transport data available	Time to GP by public transport (mins)
	N (%)	Median (IQR)	Median (IQR)	N (%)	Median (IQR)
Hull Royal Infirmary	338 (8.6)	2.0 (1.2 - 4.4)	3.8 (2.3 - 6.1)	323 (8.4)	14.0 (9.2 - 21.1)
Leeds General Infirmary	325 (8.2)	1.7 (0.9 - 3.3)	3.0 (1.9 - 4.4)	323 (8.4)	11.3 (7.3 - 19.4)
Pinderfields General	197 (5.0)	1.9 (1.0 - 3.4)	3.2 (2.1 - 4.5)	197 (5.1)	13.1 (8.3 - 17.1)
Pontefract General	36 (0.9)	1.8 (1.2 - 3.6)	2.8 (1.9 - 5.9)	36 (0.9)	12.3 (8.0 - 17.1)
Scarborough	120 (3.0)	2.1 (1.0 - 4.4)	3.0 (1.8 - 5.2)	112 (2.9)	13.2 (7.6 - 21.1)
Scunthorpe General	166 (4.2)	3.0 (1.2 - 5.6)	4.1 (2.4 - 6.7)	162 (4.2)	15.4 (9.7 - 23.6)
St. James's, Leeds	528 (13.4)	1.9 (1.1 - 3.6)	3.2 (2.1 - 4.8)	527 (13.7)	12.1 (8.2 - 18.6)
York	461 (11.7)	2.6 (1.2 - 5.5)	4.0 (2.5 - 7.3)	434 (11.2)	16.8 (9.8 - 26.6)
Kruskal-Wallis test (p-value)	-	123.66 (<0.001)	152.87 (<0.001)	-	131.72 (<0.001)

[a]Excluding those with ‘Not known’ values

IQR=Interquartile Range, ECOG= Eastern Cooperative Oncology Group, IMD=Index of Multiple Deprivation

5.2.3 Travel to hospital

Similar to travel time to GP, the median time of a journey to the hospital by public transport was almost four times as long as the equivalent journey by private car at 10.8 minutes (IQR: 7.4-18.2 mins) compared to 43.1 minutes (IQR: 30.7-64.8 mins) (Table 5:4). There is also evidence that travel distance and time by private car and public transport varies by some clinical and socio-demographic characteristics.

There is some evidence to suggest that female patients travel less to access their hospital than males, although, the absolute differences are minimal. Similarly, differences exist by age for travel distance and time, with this most pronounced for patients aged under 18 years old, who travelled a median distance of 18.6 km (IQR: 9.7-27.4 km) to their hospital whereas all other age groups travelled between 8.9 km and 9.6 km. Interestingly, this is the opposite of what was observed for travel to GP, where younger patients had shorter journeys. However, this result is not a surprise as paediatric DLBCL patients are treated differently to adult patients and are all referred to St. James's, Leeds, regardless of whether this is their nearest hospital in the Network. Patients diagnosed with PCNSL travelled one and a half times as far as other DLBCL patients and this journey also took more time by car and public transport. As described in section 4.8.3, PCNSL patients are diagnosed and treated differently to other DLBCL patients and therefore are likely to be referred to specialised centres in the Network, which explains their longer journeys. No differences in travel distance or time by car or public transport were detected by stage at diagnosis or ECOG score, and no differences in distance or time by car by diagnostic route were detected. However, there is some difference in the median time of journeys by public transport by route to diagnosis, with those diagnosed as emergencies having the shortest journeys (Wilcoxon rank-sum test=5.65, p=0.06). There were also differences in the distance (Wilcoxon rank-sum test=-3.12, p=0.002) and time by car (Wilcoxon rank-sum test=-2.61, p=0.010) and public transport (Wilcoxon rank-sum test=-3.00, p=0.003) travelled by patients who presented with or without B Symptoms. Median journeys also varied by IMD quintile, for example, the most deprived patients lived ~4.5 km closer to their diagnostic hospital compared to the least deprived and had a 10-minute shorter journey by public transport. There were also significant differences in journey distance and time for those living in rural compared to urban areas, for example, rural patients travelled a median distance of 18.7 km (IQR: 10.2-29.6 km) compared to 8.3 km (IQR: 5.3-15.3 km) for urban patients and differences also existed between patients attending different hospitals in the Network.

Table 5:4. Baseline clinical and socio-demographic characteristics and diagnostic route distributed by travel distance in kilometres (km) and time in minutes (mins) by car and public transport to diagnostic hospital.

	Patients with travel to hospital by car data available	Distance to hospital by private car (km)	Time to hospital by private car (mins)	Patients with travel to hospital by public transport data available	Time to hospital by public transport (mins)
	N (%)	Median (IQR)	Median (IQR)	N (%)	Median (IQR)
Overall	*4099 (100.0)*	*9.5 (5.7 - 19.1)*	*10.8 (7.4 - 18.2)*	*3916 (100.0)*	*43.1 (30.7 - 64.8)*
Sex:					
Males	2177 (53.1)	9.7 (5.8 - 20.0)	11.0 (7.5 - 18.6)	2076 (53.0)	43.8 (31.4 - 66.0)
Females	1922 (46.9)	9.2 (5.5 - 18.3)	10.6 (7.4 - 17.2)	1840 (47.0)	42.5 (29.7 - 63.5)
Wilcoxon rank-sum test (p-value)	-	2.30 (0.02)	1.74 (0.08)	-	1.78 (0.08)
Age at diagnosis (years):					
<18	26 (0.6)	18.6 (9.7 - 27.4)	15.7 (10.8 - 21.9)	25 (0.6)	54.3 (44.7 - 71.8)
18-39	221 (5.4)	9.4 (5.6 - 17.5)	10.7 (7.3 - 17.2)	213 (5.4)	43.1 (28.2 - 63.6)
40-59	784 (19.1)	8.9 (5.5 - 18.4)	10.2 (7.3 - 16.7)	748 (19.1)	41.5 (29.4 - 60.4)
60-69	973 (23.7)	9.9 (5.9 - 20.5)	11.1 (7.5 - 19.1)	917 (23.4)	44.3 (31.8 - 67.3)
70-79	1244 (30.3)	9.6 (5.6 - 19.3)	10.9 (7.5 - 18.6)	1187 (30.3)	42.7 (30.5 - 65.3)
≥80	851 (20.8)	9.6 (5.7 - 18.6)	10.7 (7.4 - 17.4)	826 (21.1)	43.8 (31.4 - 66.4)
Kruskal-Wallis test (p-value)	-	14.41 (0.013)	13.14 (0.022)	-	14.09 (0.015)
Primary CNS lymphoma:					
Yes	141 (3.4)	25.2 (14.3 - 41.4)	20.7 (12.9 - 33.4)	128 (3.3)	64.2 (47.5 - 90.1)
No	3958 (96.6)	9.2 (5.6 - 18.5)	10.6 (7.4 - 17.6)	3788 (96.7)	42.5 (30.4 - 64.0)
Wilcoxon rank-sum test (p-value)	-	10.30 (<0.001)	9.24 (<0.001)	-	8.05 (<0.001)
Ann-Arbor stage:					
I	525 (12.8)	9.7 (5.7 - 17.5)	10.8 (7.4 - 17.0)	496 (12.7)	42.4 (30.3 - 64.1)
II	691 (16.9)	9.5 (5.7 - 18.4)	10.6 (7.5 - 17.4)	663 (16.9)	42.2 (31.0 - 64.8)
III	654 (16.0)	9.2 (5.8 - 18.5)	10.6 (7.4 - 17.6)	620 (15.8)	42.2 (29.5 - 62.2)

	Patients with travel to hospital by car data available	Distance to hospital by private car (km)	Time to hospital by private car (mins)	Patients with travel to hospital by public transport data available	Time to hospital by public transport (mins)
	N (%)	Median (IQR)	Median (IQR)	N (%)	Median (IQR)
IV	1734 (42.3)	9.7 (5.7 - 20.5)	11.1 (7.5 - 19.1)	1656 (42.3)	44.7 (31.1 - 66.5)
Not fully staged	495 (12.1)	9.2 (5.4 - 18.5)	10.2 (7.2 - 17.3)	481 (12.3)	43.0 (31.6 - 64.2)
Kruskal-Wallis test (p-value)	-	4.24 (0.38)	4.26 (0.37)	-	4.86 (0.30)
B Symptoms:					
Yes	1811 (44.2)	9.0 (5.4 - 18.9)	10.4 (7.2 - 17.9)	1725 (44.1)	41.9 (29.3 - 63.5)
No	2285 (55.7)	10.0 (5.9 - 19.4)	11.1 (7.6 - 18.3)	2189 (55.9)	44.4 (31.7 - 65.9)
Not known	3 (0.1)	21.9 (3.2 - 22.4)	17.6 (4.9 - 21.5)	2 (0.1)	44.9 (27.3 - 62.5)
Wilcoxon rank-sum test (p-value)[a]	-	-3.12 (0.002)	-2.61 (0.010)	-	-3.00 (0.003)
ECOG score:					
0	1191 (29.1)	10.0 (5.7 - 19.0)	11.0 (7.5 - 18.4)	1133 (28.9)	44.0 (31.1 - 66.5)
1	1538 (37.5)	9.1 (5.7 - 18.5)	10.6 (7.4 - 17.7)	1463 (37.4)	42.0 (30.3 - 62.2)
2	821 (20.0)	9.2 (5.6 - 19.2)	10.7 (7.4 - 17.6)	799 (20.4)	43.4 (30.8 - 66.7)
3	363 (8.9)	9.8 (5.7 - 20.7)	11.0 (7.5 - 18.8)	349 (8.9)	45.0 (31.9 - 63.9)
4	121 (3.0)	10.7 (5.5 - 25.0)	11.9 (7.6 - 21.9)	115 (2.9)	47.9 (28.4 - 74.4)
Not known	65 (1.6)	8.7 (5.3 - 22.0)	10.0 (7.4 - 18.6)	57 (1.5)	37.6 (28.1 - 55.6)
Kruskal-Wallis test (p-value)	-	3.23 (0.67)	4.27 (0.51)	-	6.78 (0.24)
Route to Diagnosis:					
Emergency presentation	1285 (31.4)	9.0 (5.6 - 19.0)	10.3 (7.3 - 17.5)	1234 (31.5)	42.0 (29.7 - 62.0)
Elective referrals	2624 (64.0)	9.8 (5.7 - 19.0)	11.0 (7.5 - 18.2)	2505 (64.0)	43.9 (31.3 - 66.2)
Unknown	190 (4.6)	9.8 (5.1 - 26.5)	10.9 (6.8 - 23.6)	177 (4.5)	44.0 (27.5 - 66.8)
Kruskal-Wallis test (p-value)	-	3.46 (0.18)	4.04 (0.13)	-	5.65 (0.06)
IMD income quintile:					
1 (least deprived)	824 (20.1)	11.3 (6.7 - 19.5)	12.3 (8.7 - 18.7)	771 (19.7)	46.4 (35.6 - 66.4)
2	886 (21.6)	10.7 (6.2 - 20.7)	12.0 (8.0 - 19.5)	834 (21.3)	45.5 (33.0 - 70.0)

	Patients with travel to hospital by car data available	Distance to hospital by private car (km)	Time to hospital by private car (mins)	Patients with travel to hospital by public transport data available	Time to hospital by public transport (mins)
	N (%)	Median (IQR)	Median (IQR)	N (%)	Median (IQR)
3	816 (19.9)	11.3 (6.2 - 24.1)	11.8 (7.8 - 21.6)	769 (19.6)	45.9 (31.6 - 65.9)
4	710 (17.3)	9.1 (5.3 - 19.0)	10.2 (7.1 - 18.0)	694 (17.7)	42.4 (28.3 - 66.1)
5 (most deprived)	863 (21.0)	6.7 (4.4 - 13.1)	8.5 (6.2 - 13.7)	848 (21.7)	36.4 (23.7 - 53.8)
Kruskal-Wallis test (p-value)	-	172.64 (<0.001)	182.93 (<0.001)	-	123.89 (<0.001)
Rural-urban classification:					
Rural	848 (20.7)	18.7 (10.2 - 29.6)	18.5 (11.6 - 27.0)	745 (19.0)	60.2 (40.1 - 86.5)
Urban	3251 (79.3)	8.3 (5.3 - 15.3)	9.7 (7.0 - 15.2)	3171 (81.0)	40.7 (29.2 - 59.4)
Wilcoxon rank-sum test (p-value)	-	19.52 (<0.001)	20.62 (<0.001)	-	15.02 (<0.001)
Cancer Alliance:					
West Yorkshire & Harrogate	2462 (60.1)	9.0 (5.5 - 16.2)	10.1 (7.0 - 15.6)	2427 (62.0)	42.1 (29.5 - 60.6)
Humber, Coast & Vale	1637 (39.9)	11.0 (5.9 - 28.8)	12.4 (8.1 - 26.0)	1489 (38.0)	45.6 (32.3 - 75.2)
Wilcoxon rank-sum test (p-value)	-	-9.07 (<0.001)	-12.76 (<0.001)	-	-7.02 (<0.001)
Diagnostic hospital:					
Airedale	144 (3.5)	11.5 (7.6 - 18.0)	14.4 (10.9 - 19.6)	134 (3.4)	44.5 (31.9 - 60.6)
Bradford Royal	372 (9.1)	6.3 (4.8 - 9.1)	8.0 (6.5 - 9.9)	370 (9.4)	37.0 (25.7 - 50.4)
Calderdale	322 (7.9)	9.4 (6.4 - 14.1)	10.7 (8.0 - 15.0)	321 (8.2)	47.1 (31.1 - 69.1)
Castle Hill, Humberside	316 (7.7)	12.7 (8.2 - 26.3)	14.3 (9.6 - 24.9)	281 (7.2)	63.1 (33.3 - 79.8)
Dewsbury & District	243 (5.9)	9.6 (4.9 - 21.2)	9.8 (5.6 - 18.0)	243 (6.2)	44.5 (28.5 - 77.5)
Grimsby	182 (4.4)	6.4 (4.9 - 11.7)	9.7 (8.4 - 13.8)	167 (4.3)	43.3 (32.8 - 49.2)
Harrogate	206 (5.0)	5.7 (2.9 - 17.6)	7.7 (4.8 - 16.9)	191 (4.9)	29.4 (20.4 - 53.2)
Huddersfield Royal Infirmary	69 (1.7)	7.3 (5.0 - 13.7)	9.7 (7.5 - 13.3)	69 (1.8)	39.3 (29.2 - 54.5)
Hull Royal Infirmary	362 (8.8)	13.0 (7.3 - 34.9)	12.8 (8.3 - 30.9)	323 (8.2)	45.2 (32.1 - 73.9)

	Patients with travel to hospital by car data available	Distance to hospital by private car (km)	Time to hospital by private car (mins)	Patients with travel to hospital by public transport data available	Time to hospital by public transport (mins)
	N (%)	Median (IQR)	Median (IQR)	N (%)	Median (IQR)
Leeds General Infirmary	334 (8.1)	10.5 (6.4 - 19.3)	10.2 (7.0 - 16.9)	332 (8.5)	37.9 (30.4 - 56.9)
Pinderfields General	199 (4.9)	11.5 (7.4 - 17.3)	11.9 (9.0 - 15.6)	199 (5.1)	46.0 (35.7 - 62.5)
Pontefract General	36 (0.9)	6.8 (5.1 - 12.7)	7.7 (6.0 - 12.8)	36 (0.9)	29.7 (19.4 - 40.4)
Scarborough	122 (3.0)	13.8 (3.2 - 30.0)	12.2 (4.5 - 25.7)	111 (2.8)	39.4 (16.7 - 71.9)
Scunthorpe General	181 (4.4)	11.4 (5.5 - 24.8)	12.4 (7.1 - 23.7)	168 (4.3)	33.1 (23.0 - 66.9)
St. James's, Leeds	537 (13.1)	9.9 (6.1 - 17.4)	10.8 (7.5 - 16.6)	532 (13.6)	44.9 (35.4 - 60.2)
York	474 (11.6)	11.0 (5.5 - 29.9)	14.7 (7.8 - 28.4)	439 (11.2)	45.8 (36.4 - 82.2)
Kruskal-Wallis test (p-value)	-	293.10 (<0.001)	351.99 (<0.001)	-	241.96 (<0.001)

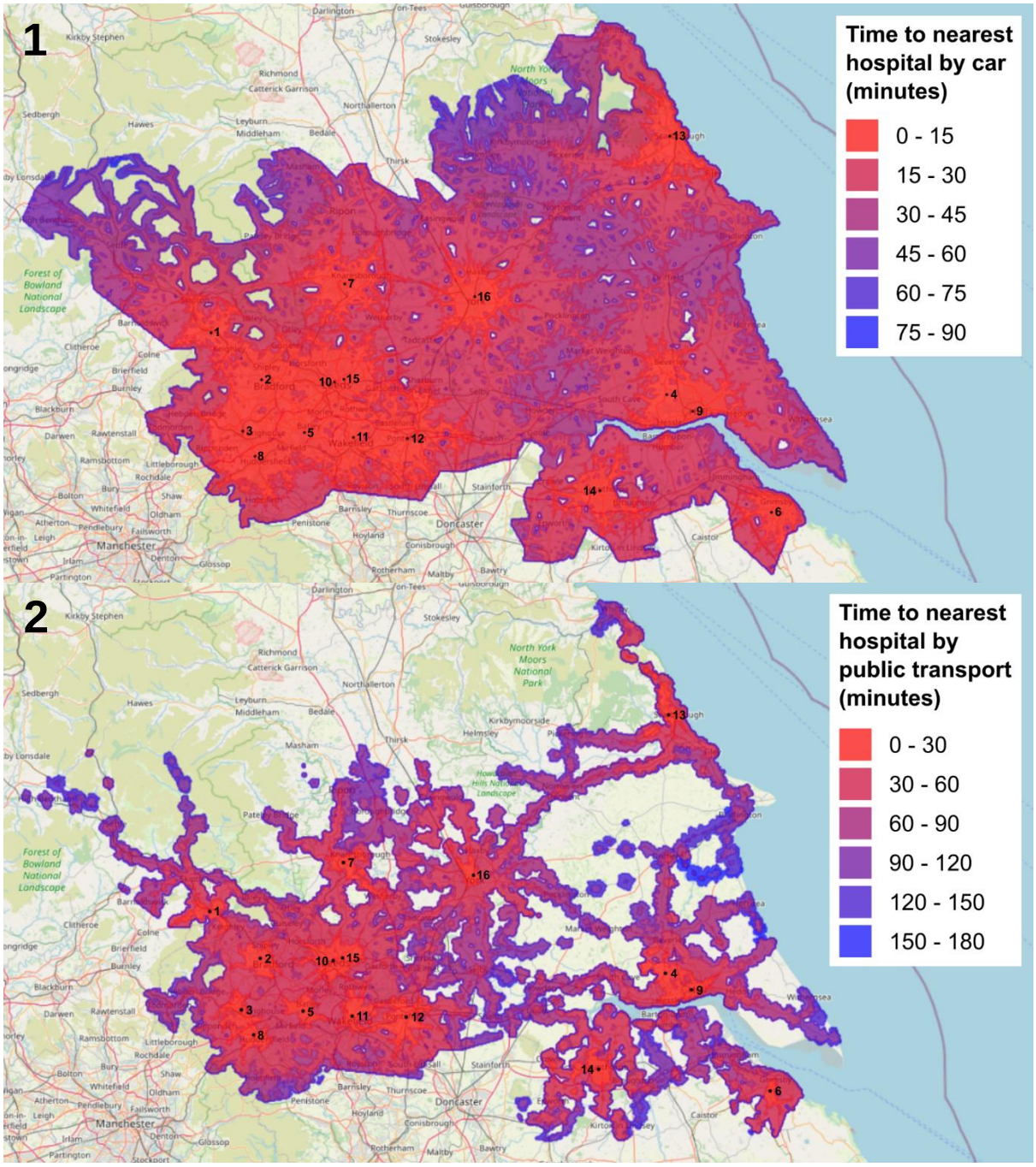
[a]Excluding those with ‘Not known’ values

IQR=Interquartile range, ECOG= Eastern Cooperative Oncology Group, IMD=Index of Multiple Deprivation

5.2.4 Accessibility to hospitals across the Network

As expected, there are considerably more GP practices in the region than hospitals, with each hospital serving a much larger catchment of patients' than a single GP practice. Consequently, the variation in patients' journey time to hospital is greater than the variation in journey time to GP as shown in Table 5:3 and Table 5:4. To understand where patients with longer journeys to the nearest hospital are located in the Network, heat maps showing journey time by car and public transport to the nearest hospital were produced (Figure 5:3). As detailed in section 4.5.8, these maps are produced by creating a 250m grid of points that cover the entire HMRN region and calculating the shortest time to travel to the nearest hospital from each point on the grid.

As shown by the legends, red indicates areas with the shortest journeys to nearest hospital and areas with longer journeys are indicated by the use of progressively more blue-toned colours. The different colours displayed on the heat maps in Figure 5:3 illustrate that access to hospitals by car and public transport is not equal across the Network. However, access is more uniform for journeys by car compared to journeys by public transport, as shown by the more homogenous use of red-toned colours and fewer 'blank' areas in Map 1 compared to Map 2. Access to hospital by car is particularly good in the areas surrounding urban centres such as York, Leeds and Bradford and less good to the east of York and around the North York Moors national park, excluding the area surrounding Scarborough, but still largely overlaid by colours indicating journey times of less than 45 minutes. Similarly, time to hospital by public transport is lower for those living in the urban and suburban areas surrounding hospitals and in the west of the Network and greater for those living in the less built-up areas in the east and some areas on the coast. There are also a considerable number and size of 'blank' areas on the map, indicating that the calculation was not possible and suggesting that access to hospitals by public transport is poor in these areas.



HMRN Hospitals			
1. Airedale	5. Dewsbury	9. Hull Royal Infirmary	13. Scarborough
2. Bradford	6. Grimsby	10. Leeds General Infirmary	14. Scunthorpe
3. Calderdale	7. Harrogate	11. Pinderfields	15. St James's, Leeds
4. Castle Hill	8. Huddersfield	12. Pontefract	16. York

Figure 5:3. Heat maps showing journey time to nearest hospital by 1) car 2) public transport across the HMRN region.

As noted in section 4.8.3, the hospitals in the Network can be organised into two Cancer Alliances; the West Yorkshire and Harrogate Cancer Alliance and the Humber, Coast and Vale Cancer Alliance. There is variation in the journey time by both car and public transport depending on the Cancer Alliance responsible for the patients care, with those under the care of the Humber, Coast and Vale travelling further to access healthcare than those under the care of West Yorkshire and Harrogate (Figure 5:4). This difference is least pronounced for journeys by car to the GP and most pronounced for journeys by public transport to hospital.

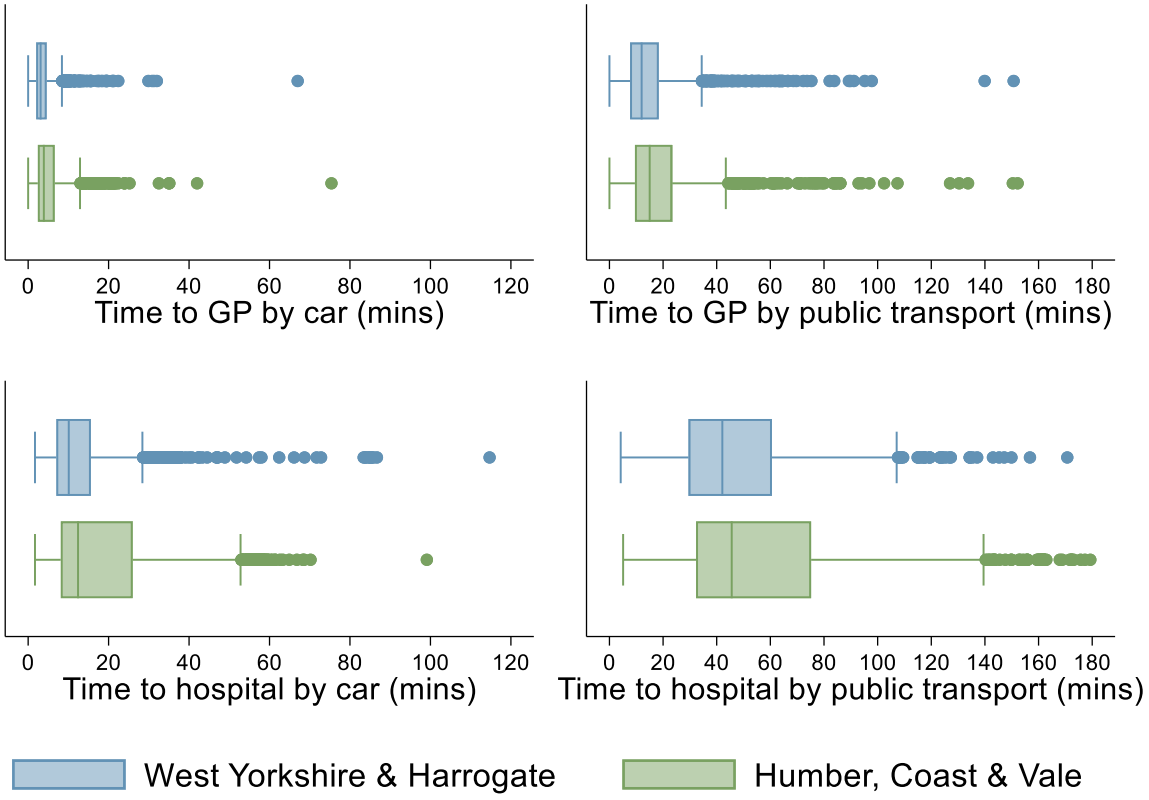


Figure 5:4. Box plots showing journey time in minutes (mins) to GP and Hospital by car and public transport stratified by Cancer Alliance.

When stratified by diagnostic hospital, there is little variation in time to GP between patients diagnosed at different hospitals for both journeys made by car and public transport (Figure 5:5). There is more variation in travel time to hospital, for example, most journey times for patients diagnosed at Bradford are clustered around the median indicating that there is more homogeneity of travel times for these patients compared to patients diagnosed at York and Hull Royal Infirmary where journey times are more dispersed around the median.

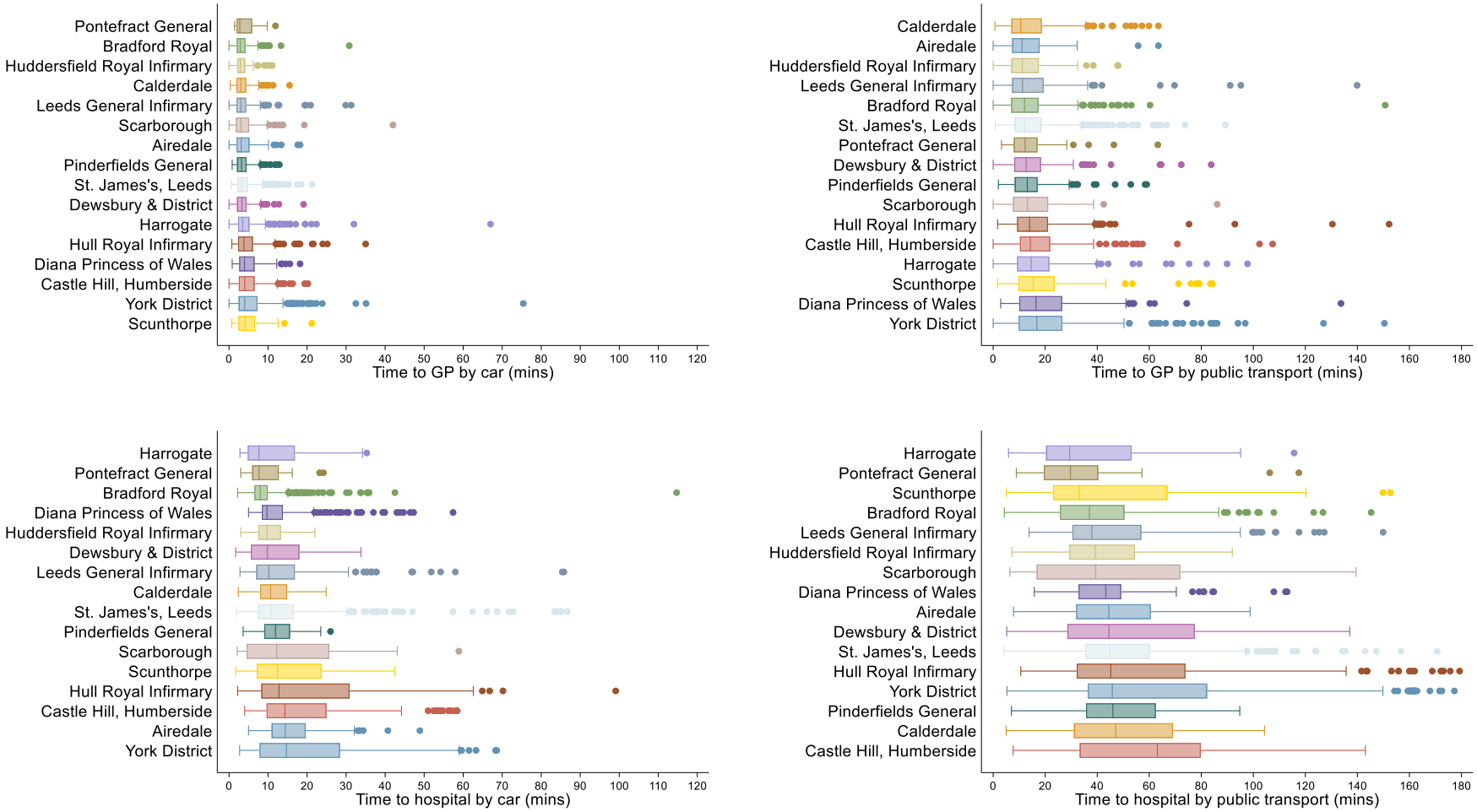


Figure 5:5. Box plots showing journey time in minutes (mins) to GP and Hospital by car and public transport stratified by diagnostic hospital.

To understand if the distribution of travel times varies between hospitals in the Network, for each travel time variable the overall percentage of patients diagnosed at each hospital was compared to the percentage belonging to each category of that variable (Table 5:5, Table 5:6, Table 5:7, Table 5:8). This revealed some differences, for example, 11.7% of all patients are diagnosed at York Hospital but of patients with the 'highest' journey times to GP by car or to hospital by public transport, 27.9% and 35.3% respectively are York patients. Similarly, 8.6% of patients overall are diagnosed at Hull Royal Infirmary but these patients make up 30.9% of the 'highest' travel to hospital by car group, but interestingly, only 11.4% of the 'highest' travel time by car to GP group, suggesting that access to GP practices is better than to hospitals for these patients. Harrogate patients make up 5% of total diagnoses, however, they account for 10.5% of the 'low' travel to hospital by public transport group. These differences in travel time distribution by diagnostic hospital further illustrate that travel to healthcare is not evenly distributed across the Network. However, these results should be interpreted with caution because there are considerable differences in sample size between these groups, which could be driving some of the differences observed.

Table 5:5. Travel time to GP by car distributed by diagnostic hospital.

		Time to GP by Car			
Diagnostic Hospital	**Overall**	**Low**	**Medium**	**High**	**Highest**
	N (%)	**N (%)**	**N (%)**	**N (%)**	**N (%)**
Total	*3946 (100.0)*	*1057 (100.0)*	*1866 (100.0)*	*822 (100.0)*	*201 (100.0)*
Airedale	143 (3.6)	44 (4.2)	63 (3.4)	31 (3.8)	5 (2.5)
Bradford Royal	362 (9.2)	111 (10.5)	196 (10.5)	53 (6.4)	2 (1.0)
Calderdale	311 (7.9)	110 (10.4)	145 (7.8)	54 (6.6)	2 (1.0)
Castle Hill, Humberside	302 (7.7)	63 (6.0)	131 (7.0)	85 (10.3)	23 (11.4)
Dewsbury & District	239 (6.1)	66 (6.2)	131 (7.0)	38 (4.6)	4 (2.0)
Diana Princess of Wales	146 (3.7)	26 (2.5)	64 (3.4)	45 (5.5)	11 (5.5)
Harrogate	203 (5.1)	50 (4.7)	102 (5.5)	28 (3.4)	23 (11.4)
Huddersfield Royal Infirmary	69 (1.7)	20 (1.9)	38 (2.0)	9 (1.1)	2 (1.0)
Hull Royal Infirmary	338 (8.6)	72 (6.8)	158 (8.5)	84 (10.2)	24 (11.9)
Leeds General Infirmary	325 (8.2)	113 (10.7)	156 (8.4)	49 (6.0)	7 (3.5)
Pinderfields General	197 (5.0)	54 (5.1)	109 (5.8)	29 (3.5)	5 (2.5)
Pontefract General	36 (0.9)	12 (1.1)	12 (0.6)	11 (1.3)	1 (0.5)
Scarborough	120 (3.0)	41 (3.9)	49 (2.6)	23 (2.8)	7 (3.5)
Scunthorpe	166 (4.2)	37 (3.5)	60 (3.2)	57 (6.9)	12 (6.0)
St. James's, Leeds	528 (13.4)	145 (13.7)	270 (14.5)	96 (11.7)	17 (8.5)
York	461 (11.7)	93 (8.8)	182 (9.8)	130 (15.8)	56 (27.9)
χ^2 (p-value)[a]	-	-	-	-	260.71 (<0.001)

[a]χ^2 comparing the distribution of diagnostic hospitals for all patients compared to those in each of the travel time categories.
Low: 0-2.2 mins, Medium: 2.3-5.1 mins, High: 5.2-10.4 mins, Highest: 10.5-75.4 mins.

Table 5:6. Travel time to hospital by car distributed by diagnostic hospital.

		Time to Hospital by Car			
Diagnostic Hospital	**Overall**	**Low**	**Medium**	**High**	**Highest**
	N (%)	**N (%)**	**N (%)**	**N (%)**	**N (%)**
Total	*4099 (100.0)*	*1029 (100.0)*	*2051 (100.0)*	*812 (100.0)*	*207 (100.0)*
Airedale	144 (3.5)	7 (0.7)	95 (4.6)	40 (4.9)	2 (1.0)
Bradford Royal	372 (9.1)	153 (14.9)	190 (9.3)	27 (3.3)	2 (1.0)
Calderdale	322 (7.9)	69 (6.7)	214 (10.4)	39 (4.8)	0 (0.0)
Castle Hill, Humberside	316 (7.7)	22 (2.1)	184 (9.0)	83 (10.2)	27 (13.0)
Dewsbury & District	243 (5.9)	100 (9.7)	86 (4.2)	57 (7.0)	0 (0.0)
Diana Princess of Wales	182 (4.4)	27 (2.6)	118 (5.8)	28 (3.4)	9 (4.3)
Harrogate	206 (5.0)	100 (9.7)	66 (3.2)	40 (4.9)	0 (0.0)
Huddersfield Royal Infirmary	69 (1.7)	17 (1.7)	49 (2.4)	3 (0.4)	0 (0.0)
Hull Royal Infirmary	362 (8.8)	69 (6.7)	153 (7.5)	76 (9.4)	64 (30.9)
Leeds General Infirmary	334 (8.1)	96 (9.3)	166 (8.1)	65 (8.0)	7 (3.4)
Pinderfields General	199 (4.9)	27 (2.6)	160 (7.8)	12 (1.5)	0 (0.0)
Pontefract General	36 (0.9)	16 (1.6)	18 (0.9)	2 (0.2)	0 (0.0)
Scarborough	122 (3.0)	46 (4.5)	23 (1.1)	50 (6.2)	3 (1.4)
Scunthorpe	181 (4.4)	51 (5.0)	72 (3.5)	57 (7.0)	1 (0.5)
St. James's, Leeds	537 (13.1)	130 (12.6)	295 (14.4)	93 (11.5)	19 (9.2)
York	474 (11.6)	99 (9.6)	162 (7.9)	140 (17.2)	73 (35.3)
χ^2 (p-value)[a]	-	-	-	-	286.09 (<0.001)

[a] χ^2 comparing the distribution of diagnostic hospitals for all patients compared to those in each of the travel time categories.
Low: 0-7.4 mins, Medium: 7.5-18.2 mins, High: 18.3-38.7 mins, Highest: 38.8-114.7 mins

.

Table 5:7. Travel time to GP by public transport distributed by diagnostic hospital.

		Time to GP by Public Transport				
Diagnostic Hospital	**Overall**	**Low**	**Medium**	**High**	**Highest**	**Timed out**
	N (%)	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**
Total	3946 (100.0)	979 (100.0)	1921 (100.0)	764 (100.0)	194 (100.0)	88 (100.0)
Airedale	143 (3.6)	45 (4.6)	69 (3.6)	20 (2.6)	2 (1.0)	7 (8.0)
Bradford Royal	362 (9.2)	113 (11.5)	186 (9.7)	51 (6.7)	12 (6.2)	0 (0.0)
Calderdale	311 (7.9)	107 (10.9)	138 (7.2)	54 (7.1)	11 (5.7)	1 (1.1)
Castle Hill, Humberside	302 (7.7)	48 (4.9)	162 (8.4)	67 (8.8)	15 (7.7)	10 (11.4)
Dewsbury & District	239 (6.1)	63 (6.4)	128 (6.7)	42 (5.5)	6 (3.1)	0 (0.0)
Diana Princess of Wales	146 (3.7)	27 (2.8)	63 (3.3)	41 (5.4)	15 (7.7)	0 (0.0)
Harrogate	203 (5.1)	41 (4.2)	99 (5.2)	36 (4.7)	14 (7.2)	13 (14.8)
Huddersfield Royal Infirmary	69 (1.7)	24 (2.5)	31 (1.6)	11 (1.4)	3 (1.5)	0 (0.0)
Hull Royal Infirmary	338 (8.6)	66 (6.7)	172 (9.0)	66 (8.6)	19 (9.8)	15 (17.0)
Leeds General Infirmary	325 (8.2)	103 (10.5)	151 (7.9)	59 (7.7)	10 (5.2)	2 (2.3)
Pinderfields General	197 (5.0)	50 (5.1)	116 (6.0)	25 (3.3)	6 (3.1)	0 (0.0)
Pontefract General	36 (0.9)	11 (1.1)	18 (0.9)	5 (0.7)	2 (1.0)	0 (0.0)
Scarborough	120 (3.0)	34 (3.5)	49 (2.6)	26 (3.4)	3 (1.5)	8 (9.1)
Scunthorpe	166 (4.2)	35 (3.6)	71 (3.7)	45 (5.9)	11 (5.7)	4 (4.5)
St. James's, Leeds	528 (13.4)	139 (14.2)	280 (14.6)	83 (10.9)	25 (12.9)	1 (1.1)
York	461 (11.7)	73 (7.5)	188 (9.8)	133 (17.4)	40 (20.6)	27 (30.7)
χ^2 (p-value)[a]	-	-	-	-	-	841.42 (<0.001)

[a]χ^2 comparing the distribution of diagnostic hospitals for all patients compared to those in each of the travel time categories.
Low: 0-8.3 mins, Medium: 8.4-20.6 mins, High: 20.7-37.7 mins, Highest: 38.0-152.2 mins.

Table 5:8. Travel time to hospital by public transport distributed by diagnostic hospital.

		Time to Hospital by Public Transport				
Diagnostic Hospital	**Overall**	**Low**	**Medium**	**High**	**Highest**	**Timed out**
	N (%)	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**
Total	4099 (100.0)	982 (100.0)	1955 (100.0)	784 (100.0)	195 (100.0)	183 (100.0)
Airedale	144 (3.5)	30 (3.1)	80 (4.1)	24 (3.1)	0 (0.0)	10 (5.5)
Bradford Royal	372 (9.1)	130 (13.2)	193 (9.9)	41 (5.2)	6 (3.1)	2 (1.1)
Calderdale	322 (7.9)	78 (7.9)	143 (7.3)	98 (12.5)	2 (1.0)	1 (0.5)
Castle Hill, Humberside	316 (7.7)	55 (5.6)	95 (4.9)	102 (13.0)	29 (14.9)	35 (19.1)
Dewsbury & District	243 (5.9)	67 (6.8)	79 (4.0)	82 (10.5)	15 (7.7)	0 (0.0)
Diana Princess of Wales	182 (4.4)	38 (3.9)	117 (6.0)	9 (1.1)	3 (1.5)	15 (8.2)
Harrogate	206 (5.0)	103 (10.5)	64 (3.3)	23 (2.9)	1 (0.5)	15 (8.2)
Huddersfield Royal Infirmary	69 (1.7)	22 (2.2)	38 (1.9)	9 (1.1)	0 (0.0)	0 (0.0)
Hull Royal Infirmary	362 (8.8)	67 (6.8)	161 (8.2)	51 (6.5)	44 (22.6)	39 (21.3)
Leeds General Infirmary	334 (8.1)	86 (8.8)	192 (9.8)	44 (5.6)	10 (5.1)	2 (1.1)
Pinderfields General	199 (4.9)	35 (3.6)	126 (6.4)	38 (4.8)	0 (0.0)	0 (0.0)
Pontefract General	36 (0.9)	18 (1.8)	16 (0.8)	0 (0.0)	2 (1.0)	0 (0.0)
Scarborough	122 (3.0)	44 (4.5)	33 (1.7)	31 (4.0)	3 (1.5)	11 (6.0)
Scunthorpe	181 (4.4)	75 (7.6)	48 (2.5)	31 (4.0)	14 (7.2)	13 (7.1)
St. James's, Leeds	537 (13.1)	86 (8.8)	338 (17.3)	86 (11.0)	22 (11.3)	5 (2.7)
York	474 (11.6)	48 (4.9)	232 (11.9)	115 (14.7)	44 (22.6)	35 (19.1)
χ^2 (p-value)[a]	-	-	-	-	-	769.88 (<0.001)

[a]χ^2 comparing the distribution of diagnostic hospitals for all patients compared to those in each of the travel time categories.
Low: 0-30.7 mins, Medium: 30.8-64.8 mins, High: 64.9-100.9 mins, Highest: 101.0-179.3 mins.

5.3 Hospital activity and route to diagnosis

To examine how long before diagnosis hospital activity increases for patients with DLBCL and the pattern of this activity surrounding time of diagnosis, the rate of activity of patients (cases) was compared to their matched controls for the two-years before- and year following- diagnosis. The control population is matched on age and sex and therefore has similar characteristics to the study population; 53.9% were male and the median age at pseudodiagnosis was 70.2 years compared to 53.0% male and a median age of 70.4 years at diagnosis for the cases.

Patients diagnosed with DLBCL experienced elevated hospital activity compared to their matched controls for the 24 months before they are diagnosed, for example, at 24 months before diagnosis the case-control rate ratio (RR) is 1.12 (95% confidence interval (CI): 1.03-1.22) (Figure 5:6). This activity steadily increases between 24- and 6- months before diagnosis (-6 months RR: 1.33 [95% CI: 1.24-1.43]) and then increases considerably from approximately six months before their diagnosis date until it peaks at one-month post-diagnosis (RR: 17.27 [95% CI: 16.68-17.89]). From two-months post-diagnosis the rate begins to decrease but remains elevated compared to the controls for the year following diagnosis (12 months RR: 4.58 [95% CI: 4.37-4.81]).

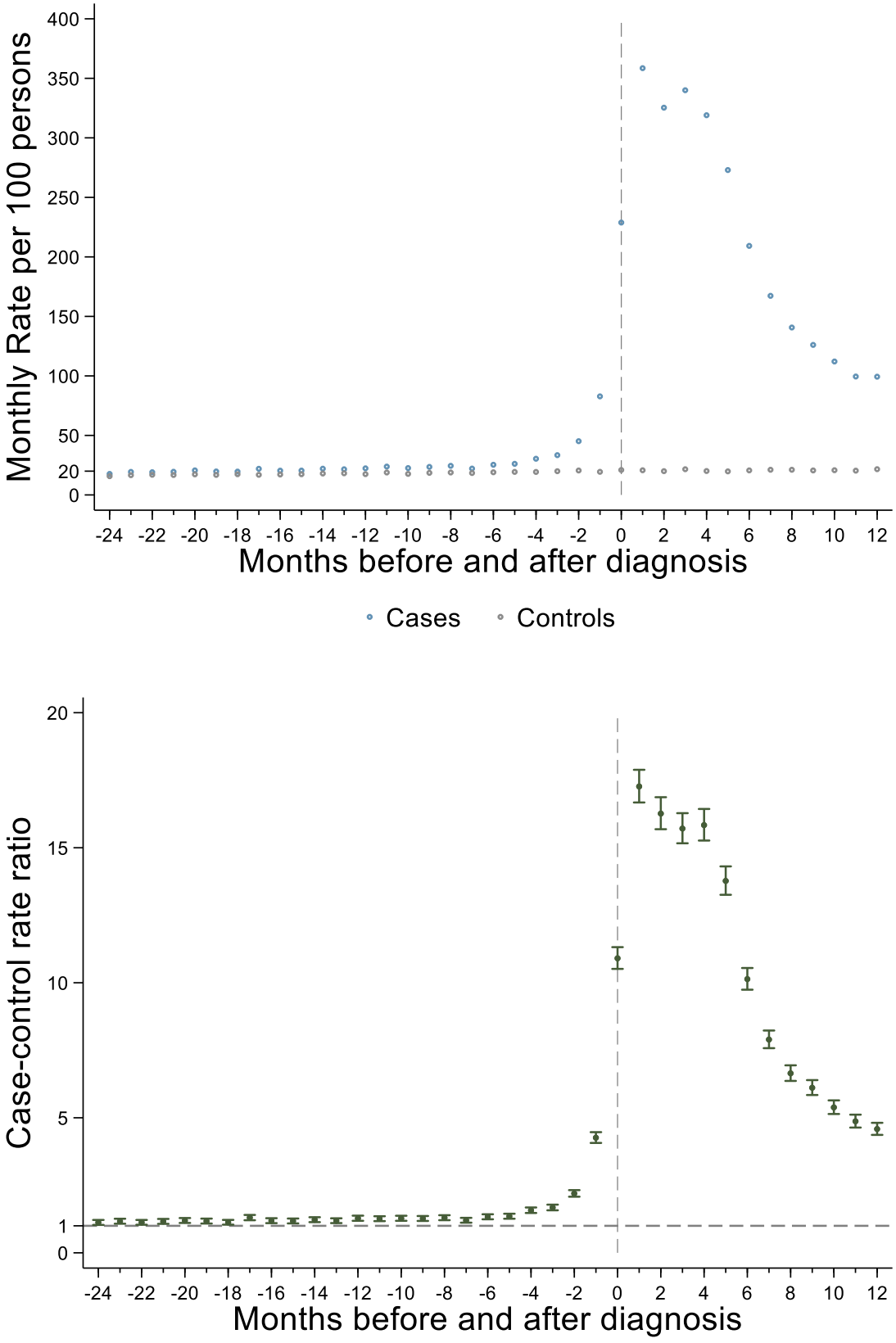


Figure 5:6. Rate ratio of monthly hospital visit activity 24 months pre-diagnosis and 12 months post-diagnosis (cases)/pseudodiagnosis (controls).

The most common RTD for patients in this cohort was the GP route (32.4%), followed by Emergency Presentation (31.4%), Two-Week Wait (TWW) (18.1%), Other Outpatient (11.3%) and Inpatient Elective (2.1%) routes (Table 5:9). In total, 192 (4.7%) patients were assigned to the Unknown category because: they could not be linked to the HES data, they did not have any inpatient or outpatient HES episodes in the 6 months preceding their date of sample or the diagnostic sample was a post-mortem sample taken after the patient had died. These patients tended to: be older, not be fully staged and have higher ECOG scores.

When examining the distribution of RTD by patient characteristics it was noted that the proportion of patients diagnosed via the TWW route varied considerably by diagnostic hospital, for example, less than 1% of York patients were assigned to the TWW route which is much lower than the total value for this route (18.1%) (Table 5:9). This raised concern about how uniformly across the hospitals in the Network TWW referrals are recorded in HES and as a result of this concern, RTD was collapsed into two groups for subsequent analysis; Emergency Presentations and Elective Referrals (which includes all other routes excluding the Unknown route). Approximately two-thirds of patients were diagnosed via Elective Referrals (n=2631, 67.1%) and one-third following an Emergency Presentation (n=1292, 32.9%) (Table 5:10).

Table 5:9. Baseline clinical and socio-demographic characteristics distributed by diagnostic route.

		Route to Diagnosis						
		Emergency Presentation		Elective Referrals				Unknown
	All Patients	Total	Total	Two-Week Wait	GP	Other Outpatient	Inpatient Elective	Total
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Total	*4115 (100.0)*	*1292 (31.4)*	*2631 (63.9))*	*745 (18.1)*	*1334 (32.4)*	*466 (11.3)*	*86 (2.1)*	*192 (4.7)*
Sex:								
Males	2179 (53.0)	671 (51.9)	1413 (53.7)	387 (51.9)	715 (53.6)	256 (54.9)	55 (64.0)	95 (49.5)
Females	1936 (47.0)	621 (48.1)	1218 (46.3)	358 (48.1)	619 (46.4)	210 (45.1)	31 (36.0)	97 (50.5)
χ^2 between individual RTDs (p-value)[a]	-	-	-	-	-	-	5.93 (0.20)	-
Age at diagnosis (years):								
Median (IQR)	70.4 (59.9 - 78.6)	70.6 (58.9 – 79.1)	70.4 (60.3-78.3)	70.5 (61.4 - 79.0)	70.5 (60.8 - 78.1)	69.7 (57.5 - 78.2)	69.0 (60.3 - 75.7)	70.9 (59.8 – 79.3)
Kruskal-Wallis test between individual RTDs (p-value)[a]	-	-	-	-	-	-	3.02 (0.56)	-
<18	27 (0.7)	11 (0.9)	14 (0.5)	3 (0.4)	3 (0.2)	7 (1.5)	1 (1.2)	2 (1.0)
18-39	225 (5.5)	73 (5.7)	141 (5.4)	40 (5.4)	69 (5.2)	28 (6.0)	4 (4.7)	11 (5.7)
40-59	786 (19.1)	263 (20.4)	488 (18.5)	124 (16.6)	245 (18.4)	104 (22.3)	15 (17.4)	35 (18.2)
60-69	976 (23.7)	285 (22.1)	650 (24.7)	194 (26.0)	332 (24.9)	99 (21.2)	25 (29.1)	41 (21.4)
70-79	1245 (30.3)	367 (28.4)	820 (31.2)	227 (30.5)	432 (32.4)	131 (28.1)	30 (34.9)	58 (30.2)
≥80	856 (20.8)	293 (22.7)	518 (19.7)	157 (21.1)	253 (19.0)	97 (20.8)	11 (12.8)	45 (23.4)

		Route to Diagnosis						
		Emergency Presentation		Elective Referrals				Unknown
	All Patients	Total	Total	Two-Week Wait	GP	Other Outpatient	Inpatient Elective	Total
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
χ^2 between individual RTDs (p-value)[a]	-	-	-	-	-	-	36.04 (0.02)	-
Primary CNS lymphoma:								
Yes	141 (3.4)	87 (6.7)	47 (1.8)	3 (0.4)	17 (1.3)	25 (5.4)	2 (2.3)	7 (3.6)
No	3974 (96.6)	1205 (93.3)	2584 (98.2)	742 (99.6)	1317 (98.7)	441 (94.6)	84 (97.7)	185 (96.4)
χ^2 between individual RTDs (p-value)[a]	-	-	-	-	-	-	87.83 (<0.001)	-
Ann-Arbor stage:								
I	526 (12.8)	105 (8.1)	404 (15.4)	91 (12.2)	242 (18.1)	59 (12.7)	12 (14.0)	17 (8.9)
II	696 (16.9)	147 (11.4)	519 (19.7)	151 (20.3)	269 (20.2)	81 (17.4)	18 (20.9)	30 (15.6)
III	655 (15.9)	129 (10.0)	506 (19.2)	182 (24.4)	226 (16.9)	80 (17.2)	18 (20.9)	20 (10.4)
IV	1740 (42.3)	698 (54.0)	962 (36.6)	250 (33.6)	484 (36.3)	198 (42.5)	30 (34.9)	80 (41.7)
Not fully staged	498 (12.1)	213 (16.5)	240 (9.1)	71 (9.5)	113 (8.5)	48 (10.3)	8 (9.3)	45 (23.4)
χ^2 between individual RTDs (p-value)[a]	-	-	-	-	-	-	219.28 (<0.001)	-
B Symptoms:								
Yes	1818 (44.2)	649 (50.2)	1098 (41.7)	338 (45.4)	541 (40.6)	173 (37.1)	46 (53.5)	71 (37.0)
No	2294 (55.7)	642 (49.7)	1533 (58.3)	407 (54.6)	793 (59.4)	293 (62.9)	40 (46.5)	119 (62.0)
Not known	3 (0.1)	1 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.0)

		Route to Diagnosis						
		Emergency Presentation	Elective Referrals					Unknown
	All Patients	Total	Total	Two-Week Wait	GP	Other Outpatient	Inpatient Elective	Total
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
χ^2 between individual RTDs (p-value)[a]	-	-	-	-	-	-	39.11 (<0.001)	-
								--
ECOG score:								
0	1196 (29.1)	160 (12.4)	984 (37.4)	320 (43.0)	490 (36.7)	144 (30.9)	30 (34.9)	52 (27.1)
1	1543 (37.5)	449 (34.8)	1044 (39.7)	277 (37.2)	535 (40.1)	197 (42.3)	35 (40.7)	50 (26.0)
2	823 (20.0)	370 (28.6)	413 (15.7)	102 (13.7)	213 (16.0)	82 (17.6)	16 (18.6)	40 (20.8)
3	365 (8.9)	218 (16.9)	124 (4.7)	30 (4.0)	63 (4.7)	28 (6.0)	3 (3.5)	23 (12.0)
4	122 (3.0)	72 (5.6)	37 (1.4)	14 (1.9)	14 (1.0)	8 (1.7)	1 (1.2)	13 (6.8)
Not known	66 (1.6)	23 (1.8)	29 (1.1)	2 (0.3)	19 (1.4)	7 (1.5)	1 (1.2)	14 (7.3)
χ^2 between individual RTDs (p-value)[a]	-	-	-	-	-	-	488.56 (<0.001)	-
IMD income quintile:								
1 (least deprived)	827 (20.1)	227 (17.6)	551 (20.9)	131 (17.6)	298 (22.3)	98 (21.0)	24 (27.9)	49 (25.5)
2	889 (21.6)	295 (22.8)	565 (21.5)	168 (22.6)	287 (21.5)	89 (19.1)	21 (24.4)	29 (15.1)
3	819 (19.9)	230 (17.8)	548 (20.8)	163 (21.9)	258 (19.3)	108 (23.2)	19 (22.1)	41 (21.4)
4	710 (17.3)	230 (17.8)	444 (16.9)	126 (16.9)	221 (16.6)	89 (19.1)	8 (9.3)	36 (18.8)
5 (most deprived)	865 (21.0)	307 (23.8)	521 (19.8)	156 (20.9)	269 (20.2)	82 (17.6)	14 (16.3)	37 (19.3)
Not known	5 (0.1)	3 (0.2)	2 (0.1)	1 (0.1)	1 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)

		Route to Diagnosis						
		Emergency Presentation		Elective Referrals				Unknown
	All Patients	Total	Total	Two-Week Wait	GP	Other Outpatient	Inpatient Elective	Total
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
χ^2 between RTDs (p-value)[a]	-	-	-	-	-	-	35.82 (0.003)	-
Rural-urban classification:								
Rural	850 (20.7)	261 (20.2)	554 (21.1)	169 (22.7)	286 (21.4)	80 (17.2)	19 (22.1)	35 (18.2)
Urban	3260 (79.2)	1028 (79.6)	2075 (78.9)	575 (77.2)	1047 (78.5)	386 (82.8)	67 (77.9)	157 (81.8)
Not known	5 (0.1)	3 (0.2)	2 (0.1)	1 (0.1)	1 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)
χ^2 between RTDs (p-value)[a]		-	-	-	-	-	6.06 (0.20)	-
Cancer Alliance:								
West Yorkshire & Harrogate	2469 (60.0)	790 (61.1)	1563 (59.4)	434 (58.3)	749 (56.1)	329 (70.6)	51 (59.3)	116 (60.4)
Humber, Coast & Vale	1646 (40.0)	502 (38.9)	1068 (40.6)	311 (41.7)	585 (43.9)	137 (29.4)	35 (40.7)	76 (39.6)
χ^2 between individual RTDs (p-value)	-	-	-	-	-	-	31.74 (<0.001)	-
Diagnostic hospital:								
Airedale	144 (3.5)	48 (3.7)	91 (3.5)	35 (4.7)	32 (2.4)	15 (3.2)	9 (10.5)	5 (2.6)
Bradford Royal	373 (9.1)	91 (7.0)	262 (10.0)	107 (14.4)	67 (5.0)	73 (15.7)	15 (17.4)	20 (10.4)
Calderdale	324 (7.9)	79 (6.1)	231 (8.8)	125 (16.8)	73 (5.5)	30 (6.4)	3 (3.5)	14 (7.3)
Castle Hill, Humberside	317 (7.7)	80 (6.2)	228 (8.7)	120 (16.1)	77 (5.8)	26 (5.6)	5 (5.8)	9 (4.7)
Dewsbury & District	243 (5.9)	88 (6.8)	142 (5.4)	21 (2.8)	92 (6.9)	23 (4.9)	6 (7.0)	13 (6.8)
Grimsby	184 (4.5)	52 (4.0)	125 (4.8)	50 (6.7)	61 (4.6)	8 (1.7)	6 (7.0)	7 (3.6)
Harrogate	208 (5.1)	41 (3.2)	150 (5.7)	73 (9.8)	61 (4.6)	12 (2.6)	4 (4.7)	17 (8.9)
Huddersfield Royal Infirmary	69 (1.7)	26 (2.0)	41 (1.6)	25 (3.4)	8 (0.6)	7 (1.5)	1 (1.2)	2 (1.0)

		Route to Diagnosis						
		Emergency Presentation	Elective Referrals					Unknown
	All Patients	Total	Total	Two-Week Wait	GP	Other Outpatient	Inpatient Elective	Total
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Hull Royal Infirmary	363 (8.8)	120 (9.3)	221 (8.4)	87 (11.7)	94 (7.0)	31 (6.7)	9 (10.5)	22 (11.5)
Leeds General Infirmary	334 (8.1)	120 (9.3)	198 (7.5)	9 (1.2)	134 (10.0)	54 (11.6)	1 (1.2)	16 (8.3)
Pinderfields General	199 (4.8)	66 (5.1)	131 (5.0)	24 (3.2)	78 (5.8)	26 (5.6)	3 (3.5)	2 (1.0)
Pontefract General	36 (0.9)	8 (0.6)	28 (1.1)	3 (0.4)	15 (1.1)	9 (1.9)	1 (1.2)	0 (0.0)
Scarborough	122 (3.0)	35 (2.7)	81 (3.1)	4 (0.5)	61 (4.6)	12 (2.6)	4 (4.7)	6 (3.1)
Scunthorpe General	183 (4.4)	54 (4.2)	121 (4.6)	48 (6.4)	54 (4.0)	14 (3.0)	5 (5.8)	8 (4.2)
St. James's, Leeds	539 (13.1)	223 (17.3)	289 (11.0)	12 (1.6)	189 (14.2)	80 (17.2)	8 (9.3)	27 (14.1)
York	477 (11.6)	161 (12.5)	292 (11.1)	2 (0.3)	238 (17.8)	46 (9.9)	6 (7.0)	24 (12.5)
χ^2 between individual RTDs (p-value)	-	-	-	-	-	-	748.01 (<0.001)	-
Distance to GP:								
Low	988 (24.0)	324 (25.1)	626 (23.8)	187 (25.1)	321 (24.1)	95 (20.4)	23 (26.7)	38 (19.8)
Medium	1974 (48.0)	612 (47.4)	1273 (48.4)	346 (46.4)	655 (49.1)	231 (49.6)	41 (47.7)	89 (46.4)
High	787 (19.1)	241 (18.7)	505 (19.2)	131 (17.6)	258 (19.3)	102 (21.9)	14 (16.3)	41 (21.4)
Highest	197 (4.8)	59 (4.6)	123 (4.7)	40 (5.4)	57 (4.3)	21 (4.5)	5 (5.8)	15 (7.8)
Not known	169 (4.1)	56 (4.3)	104 (4.0)	41 (5.5)	43 (3.2)	17 (3.6)	3 (3.5)	9 (4.7)
χ^2 between individual RTDs (p-value)[a]	-	-	-	-	-	-	9.33 (0.67)	-
Time to GP by car:								
Low	1057 (25.7)	336 (26.0)	677 (25.7)	201 (27.0)	346 (25.9)	108 (23.2)	22 (25.6)	44 (22.9)

		Route to Diagnosis						
		Emergency Presentation	Elective Referrals					Unknown
	All Patients	Total	Total	Two-Week Wait	GP	Other Outpatient	Inpatient Elective	Total
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Medium	1866 (45.3)	584 (45.2)	1198 (45.5)	333 (44.7)	615 (46.1)	212 (45.5)	38 (44.2)	84 (43.8)
High	822 (20.0)	256 (19.8)	528 (20.1)	136 (18.3)	269 (20.2)	105 (22.5)	18 (20.9)	38 (19.8)
Highest	201 (4.9)	60 (4.6)	124 (4.7)	34 (4.6)	61 (4.6)	24 (5.2)	5 (5.8)	17 (8.9)
Not known	169 (4.1)	56 (4.3)	104 (4.0)	41 (5.5)	43 (3.2)	17 (3.6)	3 (3.5)	9 (4.7)
χ^2 between individual RTDs (p-value)[a]	-	-	-	-	-	-	4.89 (0.96)	-
Time to GP by public transport:								
Low	979 (23.8)	312 (24.1)	627 (23.8)	189 (25.4)	319 (23.9)	97 (20.8)	22 (25.6)	40 (20.8)
Medium	1921 (46.7)	587 (45.4)	1253 (47.6)	343 (46.0)	648 (48.6)	224 (48.1)	38 (44.2)	81 (42.2)
High	764 (18.6)	252 (19.5)	471 (17.9)	124 (16.6)	235 (17.6)	96 (20.6)	16 (18.6)	41 (21.4)
Highest	194 (4.7)	62 (4.8)	117 (4.4)	32 (4.3)	56 (4.2)	25 (5.4)	4 (4.7)	15 (7.8)
Timed out	88 (2.1)	23 (1.8)	59 (2.2)	16 (2.1)	33 (2.5)	7 (1.5)	3 (3.5)	6 (3.1)
Not known	169 (4.1)	56 (4.3)	104 (4.0)	41 (5.5)	43 (3.2)	17 (3.6)	3 (3.5)	9 (4.7)
χ^2 between individual RTDs (p-value)[a]	-	-	-	-	-	-	12.36 (0.72)	-
Distance to hospital:								
Low	1025 (24.9)	336 (26.0)	635 (24.1)	169 (22.7)	343 (25.7)	100 (21.5)	23 (26.7)	54 (28.1)
Medium	2050 (49.8)	637 (49.3)	1338 (50.9)	400 (53.7)	653 (49.0)	241 (51.7)	44 (51.2)	75 (39.1)

		Route to Diagnosis						
		Emergency Presentation	Elective Referrals					Unknown
	All Patients	Total	Total	Two-Week Wait	GP	Other Outpatient	Inpatient Elective	Total
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
High	820 (19.9)	249 (19.3)	520 (19.8)	148 (19.9)	260 (19.5)	96 (20.6)	16 (18.6)	51 (26.6)
Highest	205 (5.0)	64 (5.0)	131 (5.0)	26 (3.5)	73 (5.5)	29 (6.2)	3 (3.5)	10 (5.2)
Not known	15 (0.4)	6 (0.5)	7 (0.3)	2 (0.3)	5 (0.4)	0 (0.0)	0 (0.0)	2 (1.0)
χ^2 between individual RTDs (p-value)[a]	-	-	-	-	-	-	13.60 (0.33)	-
Time to hospital by car:								
Low	1029 (25.0)	337 (26.1)	634 (24.1)	156 (20.9)	352 (26.4)	103 (22.1)	23 (26.7)	58 (30.2)
Medium	2051 (49.8)	642 (49.7)	1337 (50.8)	405 (54.4)	646 (48.4)	241 (51.7)	45 (52.3)	72 (37.5)
High	813 (19.8)	239 (18.5)	526 (20.0)	157 (21.1)	258 (19.3)	96 (20.6)	15 (17.4)	48 (25.0)
Highest	207 (5.0)	68 (5.3)	127 (4.8)	25 (3.4)	73 (5.5)	26 (5.6)	3 (3.5)	12 (6.2)
Not known	15 (0.4)	6 (0.5)	7 (0.3)	2 (0.3)	5 (0.4)	0 (0.0)	0 (0.0)	2 (1.0)
χ^2 between RTDs (p-value)[a]	-	-	-	-	-	-	19.52 (0.08)	-
Time to hospital by public transport:								
Low	982 (23.9)	327 (25.3)	602 (22.9)	192 (25.8)	298 (22.3)	88 (18.9)	24 (27.9)	53 (27.6)
Medium	1955 (47.5)	627 (48.5)	1252 (47.6)	317 (42.6)	653 (49.0)	241 (51.7)	41 (47.7)	76 (39.6)
High	784 (19.1)	220 (17.0)	525 (20.0)	161 (21.6)	255 (19.1)	95 (20.4)	14 (16.3)	39 (20.3)
Highest	195 (4.7)	60 (4.6)	126 (4.8)	37 (5.0)	63 (4.7)	23 (4.9)	3 (3.5)	9 (4.7)
Timed out	183 (4.4)	51 (3.9)	119 (4.5)	36 (4.8)	60 (4.5)	19 (4.1)	4 (4.7)	13 (6.8)
Not known	16 (0.4)	7 (0.5)	7 (0.3)	2 (0.3)	5 (0.4)	0 (0.0)	0 (0.0)	2 (1.0)

					Route to Diagnosis			
		Emergency Presentation			Elective Referrals			Unknown
	All Patients	Total	Total	Two-Week Wait	GP	Other Outpatient	Inpatient Elective	Total
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
χ^2 between individual RTDs (p-value)[a]	-	-	-	-	-	-	22.87 (0.12)	-

[a]Excluding those with 'Not known' values.
ECOG= Eastern Cooperative Oncology Group, IMD=Index of Multiple Deprivation

The Emergency Presentation route is often used as a proxy for diagnostic delay. The rates of hospital activity were stratified by diagnostic route to compare the pattern of activity of those diagnosed as Emergencies with those diagnosed following Elective Referrals. The overall shape of the plots for hospital activity and corresponding rate ratios (RR) are broadly similar for patients diagnosed following an Emergency Presentation and those diagnosed following an Elective Referral (Figure 5:7). However, patients diagnosed as emergencies have elevated hospital activity for a greater period of time compared to patients diagnosed following an Elective Referral, whose increase begins closer to the date of diagnosis (month 0). For example, the RR at 12 months before diagnosis for patients diagnosed as an Emergency is 1.49 (95% CI: 1.32-1.68) compared to 1.16 (95% CI: 1.05-1.27) for those diagnosed following an Elective Referral. Another interesting difference is when the peak in the rate of hospital activity occurs, which is one-month post-diagnosis for Elective Referrals, but not until three-months post-diagnosis for the Emergency group.

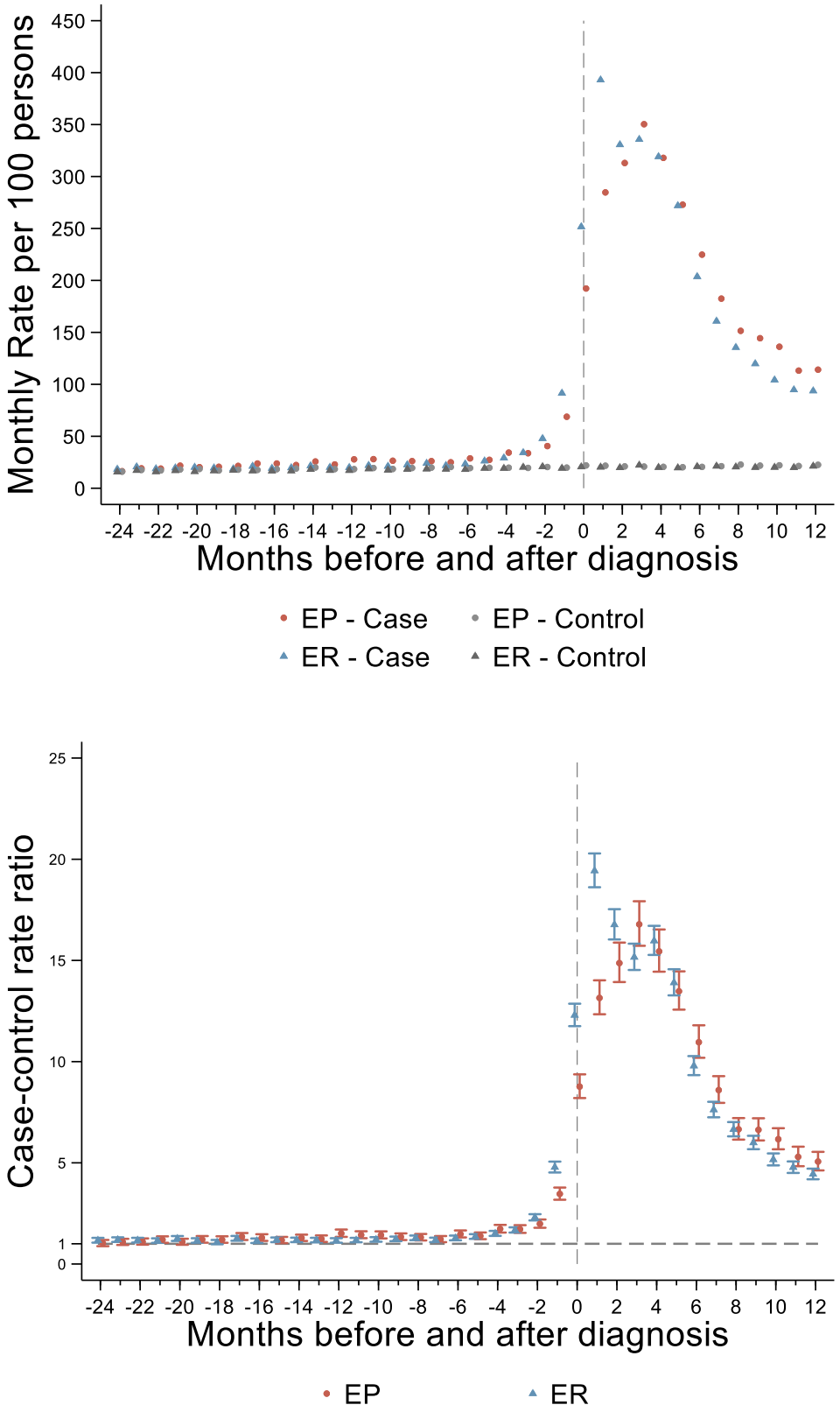


Figure 5:7. Rate ratio of monthly hospital visit activity 24 months pre- and 12 months post- diagnosis (cases)/pseudodiagnosis (controls) by diagnostic route.
EP=Emergency Presentation, ER=Elective Referral

The risk of Emergency Presentation, stratified by patient characteristics, was estimated using logistic regression (Table 5:10). No difference in the odds of being diagnosed as an Emergency by sex was detected. Patients aged 80 years and older (COR: 1.26 [95% CI: 1.05-1.53]) had increased odds of Emergency Presentation as did Children (crude odds ratio (COR): 1.76 [95% CI: 0.79-3.90]), although this result did not reach statistical significance, likely due to small numbers in this age group. However, this association was not seen for the over 80s when baseline characteristics were taken into account while a stronger association with younger age was observed (AOR: 2.66 [95% CI: 1.07-6.62]).

Those with PCNSL had over double the odds of Emergency Presentation compared to other DLBCL patients (AOR: 2.26 [95% CI: 1.48-3.46]). Patients with more advanced disease and higher ECOG scores were also diagnosed following an Emergency Presentation more frequently. For example, patients assigned an ECOG score of 0 had over nine times lower odds of Emergency Presentation compared to patients with ECOG scores of 4 (AOR: 9.71 [95% CI: 6.15-15.32]) and this was attenuated but remained in the adjusted model at over nine times the odds. Deprived patients were also more likely to be diagnosed as an Emergency, but this association was not observed when the analyses were adjusted for other characteristics.

No differences were observed between patients living in rural versus urban areas or by Cancer Alliance, however, differences did exist by hospital. Patients diagnosed at approximately two-thirds of the Network's hospitals had lower odds of Emergency Presentation compared to patients diagnosed at St. James's, Leeds. For example, patients diagnosed at Castle Hill hospital had less than half the odds of being diagnosed as an Emergency (COR: 0.45 [95% CI: 0.33-0.62]). However, when adjusted for other baseline characteristics this association only remained for patients diagnosed at Bradford Royal, Calderdale, Castle Hill and Harrogate hospitals and the increased risk of Emergency Presentation for patients diagnosed at St James's can be explained by other variables in the model such as age and PCNSL subtype.

Table 5:10. Crude and adjusted odds ratios of Emergency Presentation by patients baseline clinical and socio-demographic characteristics.

	Total	Emergency presentation	Elective referrals	Crude odds ratio	Adjusted odds ratio
	N	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
Total	3923	1292 (32.9)	2631 (67.1)	-	-
Sex:					
Males	2084	671 (32.2)	1413 (67.8)	Ref	Ref
Females	1839	621 (33.8)	1218 (66.2)	1.07 (0.94-1.23)	1.00 (0.87 - 1.17)
Age at diagnosis (years):					
<18	25	11 (44.0)	14 (56.0)	1.76 (0.79-3.90)	2.66 (1.07 - 6.62)
18-39	214	73 (34.1)	141 (65.9)	1.16 (0.85-1.57)	2.01 (1.42 - 2.83)
40-59	751	263 (35.0)	488 (65.0)	1.20 (0.99-1.46)	1.73 (1.39 - 2.15)
60-69	935	285 (30.5)	650 (69.5)	0.98 (0.81-1.18)	1.15 (0.94 - 1.42)
70-79	1187	367 (30.9)	820 (69.1)	Ref	Ref
≥80	811	293 (36.1)	518 (63.9)	1.26 (1.05-1.53)	0.98 (0.79 - 1.22)
Trend χ^2 (p-value)	-	-	-	11.40 (0.02)	-
Primary CNS lymphoma:					
Yes	134	87 (64.9)	47 (35.1)	3.97 (2.77-5.70)	2.26 (1.48 - 3.46)
No	3789	1205 (31.8)	2584 (68.2)	Ref	Ref
Ann-Arbor stage:					
I	509	105 (20.6)	404 (79.4)	Ref	Ref
II	666	147 (22.1)	519 (77.9)	1.09 (0.82-1.45)	1.05 (0.78 - 1.42)
III	635	129 (20.3)	506 (79.7)	0.98 (0.73-1.31)	0.89 (0.65 - 1.21)
IV	1660	698 (42.0)	962 (58.0)	2.79 (2.21-3.53)	1.77 (1.37 - 2.29)
Not fully staged	453	213 (47.0)	240 (53.0)	3.41 (2.57-4.53)	2.13 (1.56 - 2.91)
Trend χ^2 (p-value)[c]	-	-	-	43.89 (<0.001)	-
B Symptoms:					
Yes	1747	649 (37.1)	1098 (62.9)	1.41 (1.23-1.61)	1.20 (1.03 - 1.40)
No	2175	642 (29.5)	1533 (70.5)	Ref	Ref

	Total	Emergency presentation	Elective referrals	Crude odds ratio	Adjusted odds ratio
	N	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
Not known	1	1 (100.0)	0 (0.0)	-	-
ECOG score:					
0	1144	160 (14.0)	984 (86.0)	Ref	Ref
1	1493	449 (30.1)	1044 (69.9)	2.64 (2.16-3.23)	2.50 (2.02 - 3.08)
2	783	370 (47.3)	413 (52.7)	5.51 (4.43-6.85)	4.92 (3.87 - 6.26)
3	342	218 (63.7)	124 (36.3)	10.81 (8.20-14.26)	9.22 (6.81 - 12.48)
4	109	72 (66.1)	37 (33.9)	11.97 (7.78-18.40)	9.71 (6.15 - 15.32)
Not known	52	23 (44.2)	29 (55.8)	4.88 (2.75-8.64)	4.03 (2.19 - 7.41)
Trend χ^2 (p-value)	-	-	-	15.31 (0.001)	-
IMD income quintile:					
1 (least deprived)	778	227 (29.2)	551 (70.8)	Ref	Ref
2	860	295 (34.3)	566 (65.7)	1.27 (1.02-1.56)	1.19 (0.94 - 1.51)
3	778	230 (29.6)	548 (70.4)	1.02 (0.82-1.27)	1.03 (0.81 - 1.33)
4	674	230 (34.1)	444 (65.9)	1.26 (1.01-1.60)	1.22 (0.95 - 1.59)
5 (most deprived)	828	307 (37.1)	521 (62.9)	1.43 (1.16-1.76)	1.28 (0.99 - 1.65)
Not known	5	3 (75.0)	1 (25.0)	-	-
Trend χ^2 (p-value)[c]	-	-	-	7.96 (0.05)	-
Rural-urban classification:					
Rural	815	261 (32.0)	555 (68.0)	0.95 (0.80-1.12)	1.21 (0.98 - 1.48)
Urban	3103	1028 (33.1)	2075 (66.9)	Ref	Ref
Not known	5	3 (75.0)	1 (25.0)	-	-
Cancer Alliance:					
West Yorkshire & Harrogate	2353	790 (33.6)	1563 (66.4)	Ref	-
Humber, Coast & Vale	1570	502 (32.0)	1068 (68.0)	0.93 (0.81-1.07)	-
Diagnostic hospital:					
Airedale	139	48 (34.5)	91 (65.5)	0.68 (0.46-1.01)	0.87 (0.57 - 1.35)

	Total	Emergency presentation	Elective referrals	Crude odds ratio	Adjusted odds ratio
	N	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
Bradford Royal	353	91 (25.8)	262 (74.2)	0.45 (0.33-0.60)	0.55 (0.40 - 0.76)
Calderdale	310	79 (25.5)	231 (74.5)	0.44 0.33-0.60)	0.48 (0.34 - 0.68)
Castle Hill, Humberside	308	80 (26.0)	228 (74.0)	0.45 (0.33-0.62)	0.70 (0.49 - 0.99)
Dewsbury & District	230	88 (38.3)	142 (61.7)	0.80 (0.58-1.10)	1.10 (0.77 - 1.58)
Grimsby	177	52 (29.4)	125 (70.6)	0.54 (0.37-0.78)	0.68 (0.45 - 1.03)
Harrogate	191	41 (21.5)	150 (78.5)	0.35 (0.24-0.52)	0.50 (0.32 - 0.77)
Huddersfield Royal Infirmary	67	26 (38.8)	41 (61.2)	0.82 (0.48-1.38)	0.83 (0.48 - 1.46)
Hull Royal Infirmary	341	120 (35.2)	221 (64.8)	0.70 (0.53-0.93)	0.74 (0.54 - 1.02)
Leeds General Infirmary	318	120 (37.7)	198 (62.3)	0.79 (0.59-1.05)	0.78 (0.57 - 1.08)
Pinderfields General	197	66 (33.5)	131 (66.5)	0.65 (0.46-0.92)	0.80 (0.54 - 1.16)
Pontefract General	36	8 (22.2)	28 (77.8)	0.37 (0.17-0.83)	0.35 (0.15 - 0.83)
Scarborough	116	35 (30.2)	81 (69.8)	0.56 (0.36-0.86)	0.68 (0.42 - 1.11)
Scunthorpe General	175	54 (30.9)	121 (69.1)	0.58 (0.40-0.83)	0.76 (0.50 - 1.15)
St. James's, Leeds	512	223 (43.6)	289 (56.4)	Ref	Ref
York	453	161 (35.5)	292 (64.5)	0.71 (0.55-0.93)	0.86 (0.64 - 1.16)

[a]Crude odds ratio for Emergency Presentation vs Elective Referrals with 95% confidence interval.

[b]Odds ratio for Emergency Presentation vs Elective Referrals with 95% confidence interval adjusted for all other variables in the table excluding diagnostic Cancer Alliance because of collinearity with other variables and including time to GP and hospital by car and public transport. Excluding those with 'Not known' B-symptoms, IMD quintile or rural-urban classification (n=6).

[c]Excluding those 'Not fully staged' and with 'Not known' values.

ECOG= Eastern Cooperative Oncology Group, IMD=Index of Multiple Deprivation

The association between travel to healthcare and diagnostic route was examined using box and whisker plots and logistic regression. As shown in Figure 5:8, the distribution of the travel times to GP varies little by RTD and no differences between the median values were detected (median distance: Emergency Presentation=1.94 km, Elective Referral=1.99 km. Median time: Emergency Presentation and Elective Referral=3.3 minutes). However, patients diagnosed as Emergency Presentations had shorter median travel times (car: 10.3 minutes, public transport: 42 minutes) to their hospital compared to patients diagnosed via the Elective Referrals route (car: 11 minutes, public transport: 43.9 minutes) (car: Kruskal-Wallis test=3.84, p=0.05; public transport: Kruskal-Wallis test=5.65, p=0.02). Furthermore, no differences in the odds of Emergency Presentation by travel time or distance to GP or hospital were observed (Table 5:11). This suggests that travel to healthcare does not act as a barrier to timely diagnosis for DLBCL patients and care is equitable across the HMRN region regardless of distance travelled to access it.

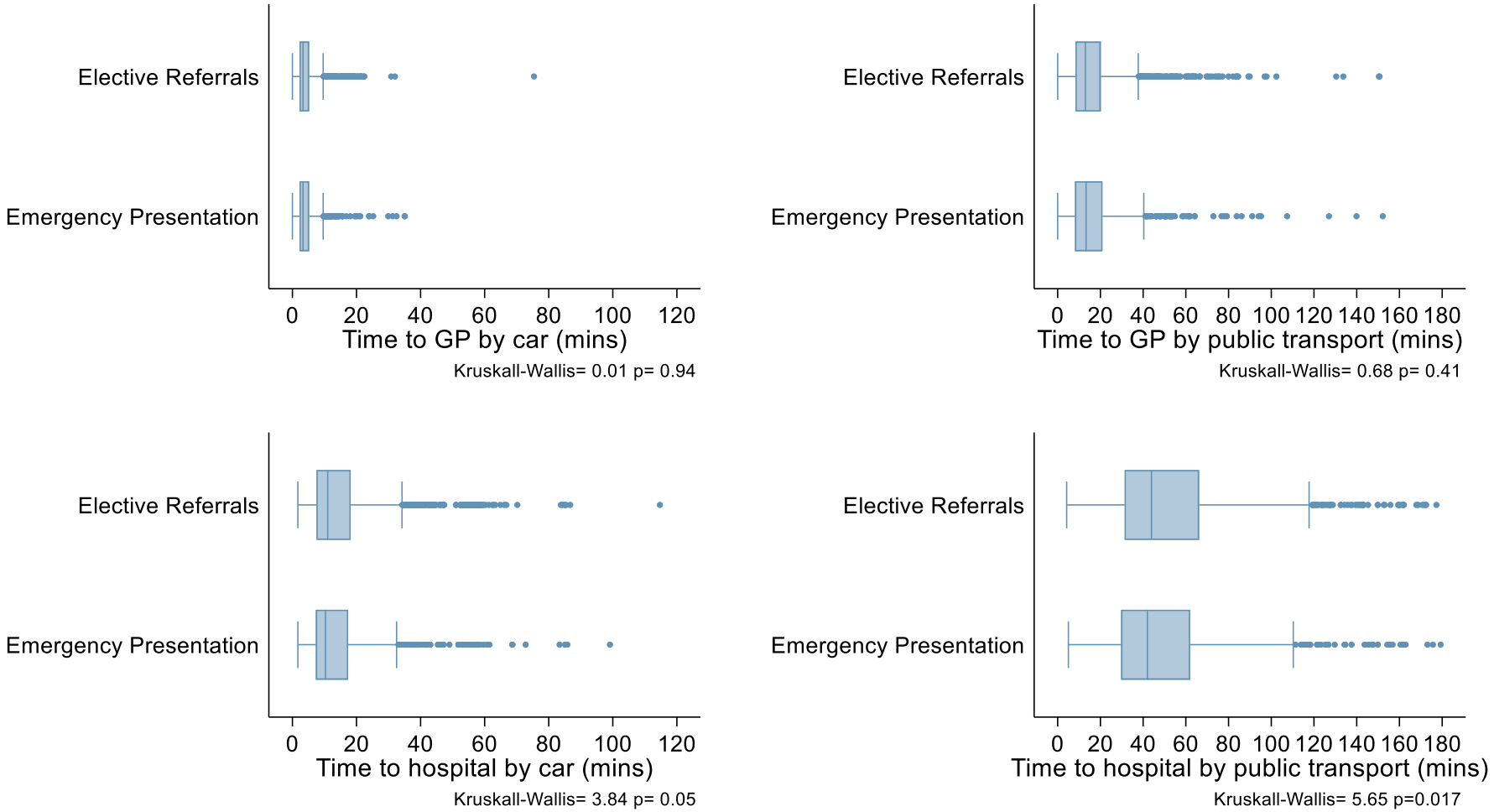


Figure 5:8. Box plots showing the distribution of travel time variables in minutes (mins) by patients' route to diagnosis.

Table 5:11. Crude and adjusted odds ratios of Emergency Presentation by travel to healthcare.

	Total	Emergency presentation	Elective referrals	Crude odds ratio	Adjusted odds ratio
	N	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
Total	3923	1292 (32.9)	2631 (67.1)	-	-
Distance to GP:					
Low	950	626 (65.9)	324 (34.1)	Ref	-
Medium	1885	1273 (67.5)	612 (32.5)	0.93 (0.79 - 1.10)	-
High	746	505 (67.7)	241 (32.3)	0.92 (0.75 - 1.13)	-
Highest	182	123 (67.6)	59 (32.4)	0.93 (0.66 - 1.30)	-
Not known	160	104 (65.0)	56 (35.0)	1.04 (0.73 - 1.48)	-
χ^2 (p-value)[c]	-	-	-	0.53 (0.77)	-
Time to GP by car:					
Low	1013	677 (66.8)	336 (33.2)	Ref	Ref
Medium	1782	1198 (67.2)	584 (32.8)	0.98 (0.83 - 1.16)	0.97 (0.78 - 1.22)
High	784	528 (67.3)	256 (32.7)	0.98 (0.80 - 1.19)	0.88 (0.64 - 1.22)
Highest	184	124 (67.4)	60 (32.6)	0.97 (0.70 - 1.36)	0.96 (0.54 - 1.69)
Not known	160	104 (65.0)	56 (35.0)	1.08 (0.76 - 1.54)	1.07 (0.70 - 1.63)
χ^2 (p-value)[c]	-	-	-	0.01 (0.99)	
Time to GP by public transport:					
Low	939	627 (66.8)	312 (33.2)	Ref	Ref
Medium	1840	1253 (68.1)	587 (31.9)	0.94 (0.80 - 1.11)	0.93 (0.74 - 1.17)
High	723	471 (65.1)	252 (34.9)	1.08 (0.88 - 1.32)	1.22 (0.88 - 1.69)
Highest	179	117 (65.4)	62 (34.6)	1.06 (0.76 - 1.49)	1.14 (0.67 - 1.95)
Timed out	82	59 (72.0)	23 (28.0)	0.78 (0.47 - 1.29)	0.86 (0.39 - 1.93)
Not known	160	104 (65.0)	56 (35.0)	1.08 (0.76 - 1.54)	-
χ^2 (p-value)[c]	-	-	-	1.79 (0.41)	-
Distance to hospital:					
Low	971	635 (65.4)	336 (34.6)	Ref	-
Medium	1975	1338 (67.7)	637 (32.3)	0.90 (0.76 - 1.06)	-
High	769	520 (67.6)	249 (32.4)	0.90 (0.74 - 1.11)	-
Highest	195	131 (67.2)	64 (32.8)	0.92 (0.67 - 1.28)	-

	Total	Emergency presentation	Elective referrals	Crude odds ratio	Adjusted odds ratio
	N	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
Not known	13	7 (53.8)	6 (46.2)	1.62 (0.54 - 4.86)	-
χ^2 (p-value)[c]	-	-	-	0.95 (0.62)	-
Time to hospital by car:					
Low	971	634 (65.3)	337 (34.7)	Ref	Ref
Medium	1979	1337 (67.6)	642 (32.4)	0.90 (0.77 - 1.06)	1.04 (0.81 - 1.33)
High	765	526 (68.8)	239 (31.2)	0.85 (0.70 - 1.05)	0.99 (0.69 - 1.43)
Highest	195	127 (65.1)	68 (34.9)	1.01 (0.73 - 1.39)	1.09 (0.63 - 1.87)
Not known	13	7 (53.8)	6 (46.2)	1.61 (0.54 - 4.84)	-
χ^2 (p-value)[c]	-	-	-	2.00 (0.37)	-
Time to hospital by public transport:					
Low	929	602 (64.8)	327 (35.2)	Ref	Ref
Medium	1879	1252 (66.6)	627 (33.4)	0.92 (0.78 - 1.09)	0.85 (0.66 - 1.08)
High	745	525 (70.5)	220 (29.5)	0.77 (0.63 - 0.95)	0.65 (0.46 - 0.92)
Highest	186	126 (67.7)	60 (32.3)	0.88 (0.63 - 1.23)	0.59 (0.35 - 1.00)
Timed out	170	119 (70.0)	51 (30.0)	0.79 (0.55 - 1.12)	0.79 (0.43 - 1.47)
Not known	14	7 (50.0)	7 (50.0)	1.84 (0.64 - 5.29)	-
χ^2 (p-value)[c]	-	-	-	1.77 (0.41)	-

[a]Crude odds ratio for Emergency Presentation vs Elective Referrals with 95% confidence interval.

[b]Odds ratio for Emergency Presentation vs Elective Referrals with 95% confidence interval adjusted for sex, age, subtype, stage, B symptoms, performance score, deprivation, rurality and diagnostic hospital. Excluding those with 'Not known' B-symptoms, deprivation or rural-urban classification (n=6).

[c]Excluding those 'with 'Not known' values.

5.4 Treatment

Information relating to treatment could be obtained for almost all patients with complete follow-up (n=4109, 99.9%) as shown in Figure 5:9. Most patients received treatment with curative intent (n=3425, 83.2%), typically with multi-agent chemotherapy (n=3401, 82.6%) as described in section 4.4.1. Those not treated curatively include those who: died shortly after diagnosis (died pre-treatment) (n=141, 3.4%), received single-agent chemotherapy or radiotherapy for symptom management only (n=127, 3.1%), received palliative/supportive care only (n=366, 8.9%) and a small number of patients who refused treatment (n=8, 0.2%). Some patients were not placed under the care of the haematology team following their diagnosis because they were already under the care of another speciality and the management of this condition took precedence, and these patients were excluded from any treatment analysis (n=42, 1.0%).

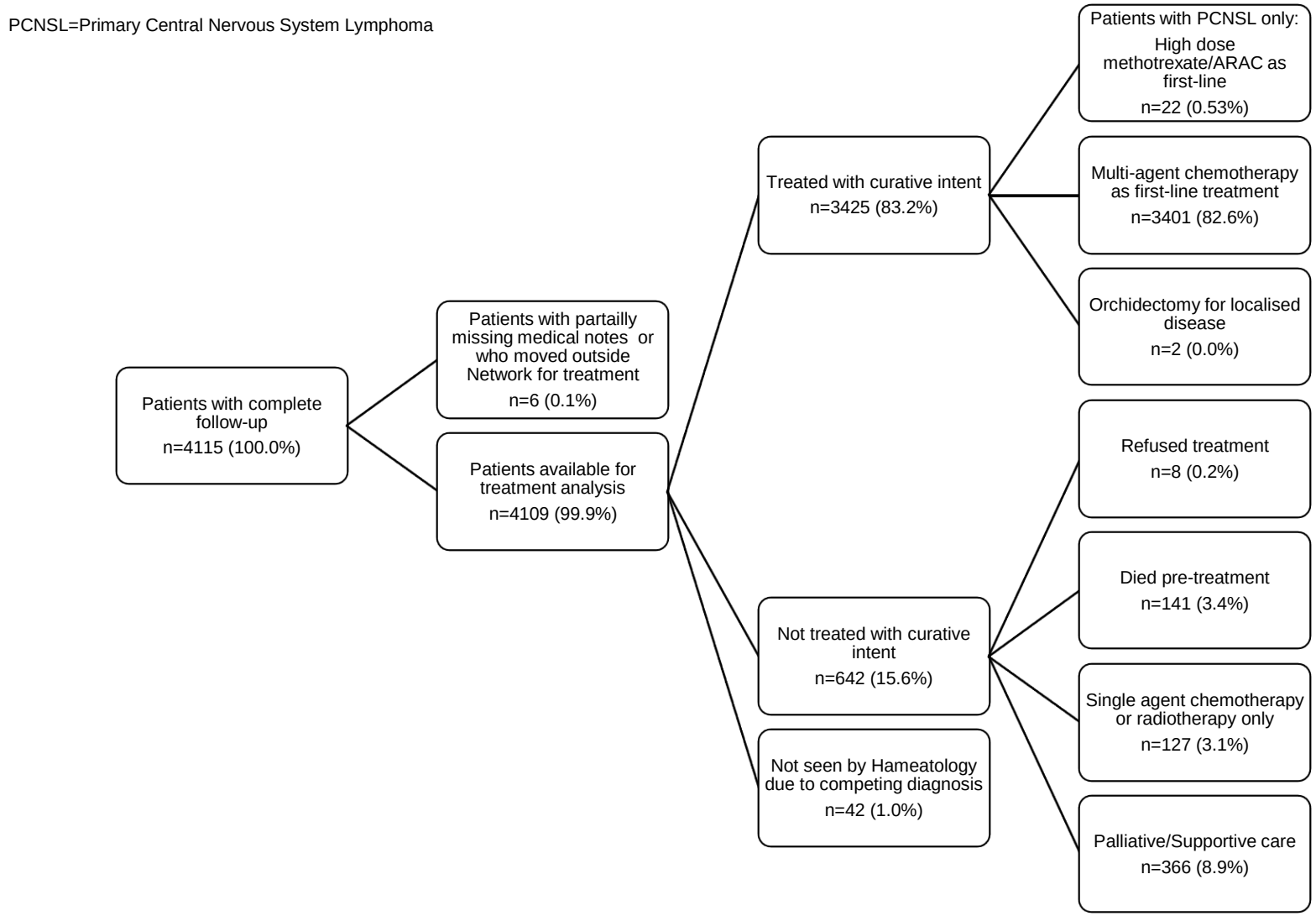


Figure 5:9. Completeness of treatment data, treatment intent and regimen type received.

Most patients received the standard treatment of R-CHOP as first-line treatment (n=2831, 82.7%). The next most common regimens given as first-line were R-CVP at 3.9% (n=132) and CHOP at 2.1% (n=72) (see full list of chemotherapy regimens in Table 4:2).

The characteristics of patients treated with curative intent differs from those not treated with curative intent, which includes those treated with supportive/palliative care or single-agent chemotherapy or radiotherapy and those who refused treatment (Table 5:12). This includes differences by sex with more patients treated with curative intent being male. However, in the adjusted model the association between sex and treatment with curative intent disappears, suggesting that the crude association is confounded by other factors, such as age. Differences were seen by age with a greater proportion of those not treated curatively belonging to the older age groups, particularly the aged 80 years and older group which makes up 60.1% of patients not treated curatively compared to 14.0% of patients treated curatively. Patients aged 80 years and older were 80% less likely to be treated with curative intent compared to those aged 70-79, whereas patients aged 18-39 were over seven times as likely to receive curative treatment.

PCNSL patients were half as likely to receive treatment with curative intent compared to other DLBCL patients (adjusted odds ratio (AOR): 0.50 [95% CI: 0.28-0.89]) (Table 5:12). Those with later stage disease or who were not fully staged also made up a higher percentage of the non-curative intent group than their overall percentage, for example, only 10.6% of patients were not fully staged but 34.5% of patients not treated curatively belong to this group. In the unadjusted model, advanced stage disease was negatively associated with receiving treatment with curative intent (COR: 0.46 [95% CI: 0.32-0.66]), however, this association remained but was reduced in the adjusted model (AOR: 0.76 [95% CI: 0.49-1.16]). Similarly, patients treated with curative intent had lower ECOG scores than those not treated curatively and patients with all ECOG scores greater than zero had lower odds of being treated with curative intent compared to those with a score of zero in both the unadjusted and adjusted models.

There were also differences by RTD, with those diagnosed as an Emergency making up almost half of the ‘not curative’ group compared to less than a third of the ‘treated curatively’ group and having lower odds of being treated curatively compared to those diagnosed following an Elective Referral, even after adjusting for other variables in the model including age and stage at diagnosis. No difference in the composition of patients treated curatively compared to non-curatively by IMD, rurality or Cancer Alliance were detected.

Little variation in the percentage of patients treated curatively or not by diagnostic hospital was observed (p=0.003) and in the adjusted model the confidence intervals for all the hospitals except three (Hull Royal Infirmary, Scarborough, York) overlapped the null value of 1. There was no difference in the composition of patients treated curatively or not or the odds of receiving curative treatment by any of the travel variables.

Table 5:12. Number of patients and odds of receiving treatment with curative intent by patients baseline clinical and socio-demographic characteristics.

	Total	Treated with curative intent	Not treated with curative intent	Crude odds ratio	Adjusted odds ratio
	N (%)	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
Total	*3926 (100.0)*	*3425 (87.2)*	*501 (12.8)*	-	-
Sex:					
Males	2079 (53.0)	1863 (54.4)	216 (43.1)	Ref	Ref
Females	1847 (47.0)	1562 (45.6)	285 (56.9)	0.64 (0.53 - 0.77)	0.89 (0.70 - 1.12)
χ^2 (p-value)[a]	-	-	22.32 (<0.001)	-	
Age at diagnosis (years):					
Median (IQR)	70.0 (59.3 - 78.2)	68.3 (57.9 - 76.2)	82.0 (75.1 - 87.2)	-	-
Kruskal-Wallis test (p-value)[a]	-	-	501.76 (<0.01)	-	-
<18	27 (0.7)	27 (0.8)	0 (0.0)	-	-
18-39	225 (5.7)	221 (6.5)	4 (0.8)	6.77 (2.48 - 18.50)	7.20 (2.18 - 23.84)
40-59	774 (19.7)	749 (21.9)	25 (5.0)	3.67 (2.37 - 5.69)	2.81 (1.76 - 4.51)
60-69	938 (23.9)	896 (26.2)	42 (8.4)	2.61 (1.82 - 3.74)	2.47 (1.67 - 3.67)
70-79	1182 (30.1)	1053 (30.7)	129 (25.7)	Ref	Ref
≥80	780 (19.9)	479 (14.0)	301 (60.1)	0.19 (0.15 - 0.25)	0.20 (0.15 - 0.27)
χ^2 (p-value)	-	-	620.34 (<0.001)	-	
Primary CNS lymphoma:					
Yes	135 (3.4)	107 (3.1)	28 (5.6)	0.54 (0.36 - 0.83)	0.50 (0.28 - 0.89)
No	3791 (96.6)	3318 (96.9)	473 (94.4)	Ref	Ref
χ^2 (p-value)	-	-	8.00 (0.005)	-	-
Ann-Arbor stage:					
I	521 (13.3)	484 (14.1)	37 (7.4)	Ref	Ref
II	691 (17.6)	667 (19.5)	24 (4.8)	2.12 (1.25 - 3.60)	2.29 (1.27 - 4.11)
III	650 (16.6)	618 (18.0)	32 (6.4)	1.48 (0.91 - 2.40)	1.56 (0.90 - 2.69)
IV	1647 (42.0)	1412 (41.2)	235 (46.9)	0.46 (0.32 - 0.66)	0.76 (0.49 - 1.16)
Not fully staged	417 (10.6)	244 (7.1)	173 (34.5)	0.11 (0.07 - 0.16)	0.28 (0.18 - 0.45)
χ^2 (p-value)[c]	-	-	93.32 (<0.001)	-	-
B Symptoms:					

	Total	Treated with curative intent	Not treated with curative intent	Crude odds ratio	Adjusted odds ratio
	N (%)	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
Yes	1730 (44.1)	1494 (43.6)	236 (47.1)	0.87 (0.72 - 1.05)	0.96 (0.75 - 1.22)
No	2196 (55.9)	1931 (56.4)	265 (52.9)	Ref	Ref
χ^2 (p-value)	-	-	2.15 (0.14)		
ECOG score:					
0	1194 (30.4)	1163 (34.0)	31 (6.2)	Ref	Ref
1	1510 (38.5)	1394 (40.7)	116 (23.2)	0.32 (0.21 - 0.48)	0.48 (0.31 - 0.74)
2	765 (19.5)	610 (17.8)	155 (30.9)	0.10 (0.07 - 0.16)	0.20 (0.13 - 0.31)
3	317 (8.1)	184 (5.4)	133 (26.5)	0.04 (0.02 - 0.06)	0.08 (0.05 - 0.14)
4	92 (2.3)	40 (1.2)	52 (10.4)	0.02 (0.01 - 0.04)	0.03 (0.02 - 0.07)
Not known	48 (1.2)	34 (1.0)	14 (2.8)	0.06 (0.03 - 0.13)	0.11 (0.05 - 0.27)
χ^2 (p-value)[c]	-	-	593.36 (<0.001)	-	-
Route to Diagnosis:					
Emergency	1205 (30.7)	965 (28.2)	240 (47.9)	0.42 (0.34 - 0.51)	0.77 (0.60 – 0.99)
Elective referrals	2583 (65.8)	2340 (68.3)	243 (48.5)	Ref	Ref
Unknown	138 (3.5)	120 (3.5)	18 (3.6)	0.69 (0.41 - 1.16)	0.79 (0.41 - 1.52)
χ^2 (p-value)[c]	-	-	81.57 (<0.001)	-	-
IMD income quintile:					
1 (least deprived)	792 (20.2)	710 (20.7)	82 (16.4)	Ref	Ref
2	857 (21.8)	746 (21.8)	111 (22.2)	0.78 (0.57 - 1.05)	0.89 (0.61 - 1.30)
3	784 (20.0)	675 (19.7)	109 (21.8)	0.72 (0.53 - 0.97)	0.65 (0.44 - 0.97)
4	666 (17.0)	582 (17.0)	84 (16.8)	0.80 (0.58 - 1.11)	0.90 (0.59 - 1.38)
5 (most deprived)	823 (21.0)	711 (20.8)	112 (22.4)	0.73 (0.54 - 0.99)	0.67 (0.44 - 1.01)
Not known	4 (0.1)	1 (0.0)	3 (0.6)	-	-
χ^2 (p-value)[c]	-	-	5.62 (0.23)	-	-
Rural-urban classification:					
Rural	817 (20.8)	716 (20.9)	101 (20.2)	1.04 (0.82 - 1.31)	0.82 (0.59 - 1.14)
Urban	3105 (79.1)	2708 (79.1)	397 (79.2)	Ref	Ref
χ^2 (p-value)	-	-	0.10 (0.75)	-	-
Cancer Alliance:					

	Total	Treated with curative intent	Not treated with curative intent	Crude odds ratio	Adjusted odds ratio
	N (%)	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
West Yorkshire & Harrogate	2347 (59.8)	2062 (60.2)	285 (56.9)	Ref	-
Humber, Coast & Vale	1579 (40.2)	1363 (39.8)	216 (43.1)	0.87 (0.72 - 1.05)	-
χ^2 (p-value)	-	-	2.00 (0.16)	-	-
Diagnostic hospital:					
Airedale	143 (3.6)	125 (3.6)	18 (3.6)	1.06 (0.61 - 1.85)	0.55 (0.27 - 1.11)
Bradford Royal	351 (8.9)	309 (9.0)	42 (8.4)	1.12 (0.74 - 1.70)	0.66 (0.39 - 1.10)
Calderdale	297 (7.6)	257 (7.5)	40 (8.0)	0.98 (0.64 - 1.49)	0.76 (0.44 - 1.30)
Castle Hill, Humberside	312 (7.9)	284 (8.3)	28 (5.6)	1.55 (0.97 - 2.47)	0.85 (0.48 - 1.53)
Dewsbury & District	226 (5.8)	201 (5.9)	25 (5.0)	1.23 (0.75 - 2.00)	1.23 (0.67 - 2.28)
Grimsby	175 (4.5)	154 (4.5)	21 (4.2)	1.12 (0.66 - 1.89)	1.05 (0.53 - 2.08)
Harrogate	200 (5.1)	177 (5.2)	23 (4.6)	1.17 (0.71 - 1.95)	0.92 (0.48 - 1.76)
Huddersfield Royal Infirmary	65 (1.7)	54 (1.6)	11 (2.2)	0.75 (0.37 - 1.50)	0.50 (0.21 - 1.19)
Hull Royal Infirmary	340 (8.7)	277 (8.1)	63 (12.6)	0.67 (0.46 - 0.98)	0.51 (0.31 - 0.83)
Leeds General Infirmary	328 (8.4)	290 (8.5)	38 (7.6)	1.16 (0.76 - 1.78)	0.98 (0.58 - 1.65)
Pinderfields General	195 (5.0)	176 (5.1)	19 (3.8)	1.41 (0.83 - 2.42)	1.21 (0.62 - 2.35)
Pontefract General	36 (0.9)	34 (1.0)	2 (0.4)	2.59 (0.61 - 11.05)	3.20 (0.58 - 17.69)
Scarborough	118 (3.0)	90 (2.6)	28 (5.6)	0.49 (0.30 - 0.81)	0.44 (0.22 - 0.87)
Scunthorpe General	174 (4.4)	156 (4.6)	18 (3.6)	1.32 (0.76 - 2.30)	1.03 (0.51 - 2.11)
St. James's, Leeds	506 (12.9)	439 (12.8)	67 (13.4)	Ref	Ref
York	460 (11.7)	402 (11.7)	58 (11.6)	1.06 (0.73 - 1.54)	1.71 (1.03 - 2.83)
χ^2 (p-value)	-	-	33.95 (0.003)	-	-
Distance to GP:					
Low	947 (24.1)	822 (24.0)	125 (25.0)	Ref	-
Medium	1880 (47.9)	1646 (48.1)	234 (46.7)	1.07 (0.85 - 1.35)	-
High	748 (19.1)	654 (19.1)	94 (18.8)	1.06 (0.79 - 1.41)	-
Highest	187 (4.8)	162 (4.7)	25 (5.0)	0.99 (0.62 - 1.56)	-
Not known	164 (4.2)	141 (4.1)	23 (4.6)	0.93 (0.58 - 1.51)	-
χ^2 (p-value)[c]	-	-	0.41 (0.94)	-	-
Time to GP by car:					
Low	1004 (25.6)	876 (25.6)	128 (25.5)	Ref	Ref
Medium	1783 (45.4)	1562 (45.6)	221 (44.1)	1.03 (0.82 - 1.30)	0.92 (0.64 - 1.32)

	Total	Treated with curative intent	Not treated with curative intent	Crude odds ratio	Adjusted odds ratio
	N (%)	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
High	784 (20.0)	681 (19.9)	103 (20.6)	0.97 (0.73 - 1.28)	0.76 (0.46 - 1.27)
Highest	191 (4.9)	165 (4.8)	26 (5.2)	0.93 (0.59 - 1.46)	0.60 (0.25 - 1.45)
Not known	164 (4.2)	141 (4.1)	23 (4.6)	0.90 (0.56 - 1.45)	1.06 (0.54 - 2.07)
χ^2 (p-value)[c]	-	-	0.43 (0.93)	-	-
Time to GP by public transport:					
Low	931 (23.7)	808 (23.6)	123 (24.6)	Ref	Ref
Medium	1842 (46.9)	1611 (47.0)	231 (46.1)	1.06 (0.84 - 1.34)	1.30 (0.90 - 1.88)
High	722 (18.4)	633 (18.5)	89 (17.8)	1.08 (0.81 - 1.45)	1.28 (0.76 - 2.14)
Highest	182 (4.6)	155 (4.5)	27 (5.4)	0.87 (0.56 - 1.37)	0.87 (0.38 - 1.98)
Timed out	85 (2.2)	77 (2.2)	8 (1.6)	1.47 (0.69 - 3.11)	1.37 (0.38 - 4.91)
Not known	164 (4.2)	141 (4.1)	23 (4.6)	0.93 (0.58 - 1.51)	-
χ^2 (p-value)[c]	-	-	1.93 (0.75)	-	-
Distance to hospital:					
Low	960 (24.5)	821 (24.0)	139 (27.7)	Ref	-
Medium	1971 (50.2)	1732 (50.6)	239 (47.7)	1.23 (0.98 - 1.54)	-
High	785 (20.0)	692 (20.2)	93 (18.6)	1.26 (0.95 - 1.67)	-
Highest	196 (5.0)	171 (5.0)	25 (5.0)	1.16 (0.73 - 1.83)	-
Not known	14 (0.4)	9 (0.3)	5 (1.0)	0.30 (0.10 - 0.92)	-
χ^2 (p-value)[c]	-	-	3.85 (0.28)	-	-
Time to hospital by car:					
Low	966 (24.6)	826 (24.1)	140 (27.9)	Ref	Ref
Medium	1969 (50.2)	1729 (50.5)	240 (47.9)	1.22 (0.98 - 1.53)	1.19 (0.79 - 1.79)
High	779 (19.8)	690 (20.1)	89 (17.8)	1.31 (0.99 - 1.75)	1.18 (0.67 - 2.10)
Highest	198 (5.0)	171 (5.0)	27 (5.4)	1.07 (0.69 - 1.67)	0.68 (0.28 - 1.66)
Not known	14 (0.4)	9 (0.3)	5 (1.0)	0.31 (0.10 - 0.92)	0.64 (0.07 - 6.19)
χ^2 (p-value)[c]	-	-	4.57 (0.21)	-	-
Time to hospital by public transport:					
Low	925 (23.6)	789 (23.0)	136 (27.1)	Ref	Ref

	Total	Treated with curative intent	Not treated with curative intent	Crude odds ratio	Adjusted odds ratio
	N (%)	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
Medium	1878 (47.8)	1639 (47.9)	239 (47.7)	1.18 (0.94 - 1.48)	1.15 (0.77 - 1.72)
High	747 (19.0)	664 (19.4)	83 (16.6)	1.38 (1.03 - 1.85)	1.59 (0.91 - 2.77)
Highest	185 (4.7)	163 (4.8)	22 (4.4)	1.28 (0.79 - 2.07)	1.84 (0.76 - 4.43)
Timed out	176 (4.5)	160 (4.7)	16 (3.2)	1.72 (1.00 - 2.97)	2.65 (0.96 - 7.28)
Not known	15 (0.4)	10 (0.3)	5 (1.0)	0.34 (0.12 - 1.02)	-
χ^2 (p-value)[c]	-	-	7.23 (0.12)	-	-

[a]Crude odds ratio for treated with curative intent vs not treated with curative intent, with 95% confidence interval.

[b]Odds ratio for treated with curative intent vs not treated with curative intent, with 95% confidence interval adjusted for all other variables in the table excluding diagnostic Cancer Alliance, distance to GP and distance to hospital because of collinearity with other variables. Excluding those with 'Not known' IMD (n=4).

[c]Excluding those with 'Unknown' and 'Not known' values.

5.5 Survival

Five-year overall survival (OS) was 51.7% and the majority of patients who did not survive five-years post-diagnosis died within the first year after diagnosis (one-year OS: 67.4%) (Figure 5:10 and Table 5:13). As expected, net survival (NS) was higher than overall survival; one-year 69.2% and five-year 59.5%.

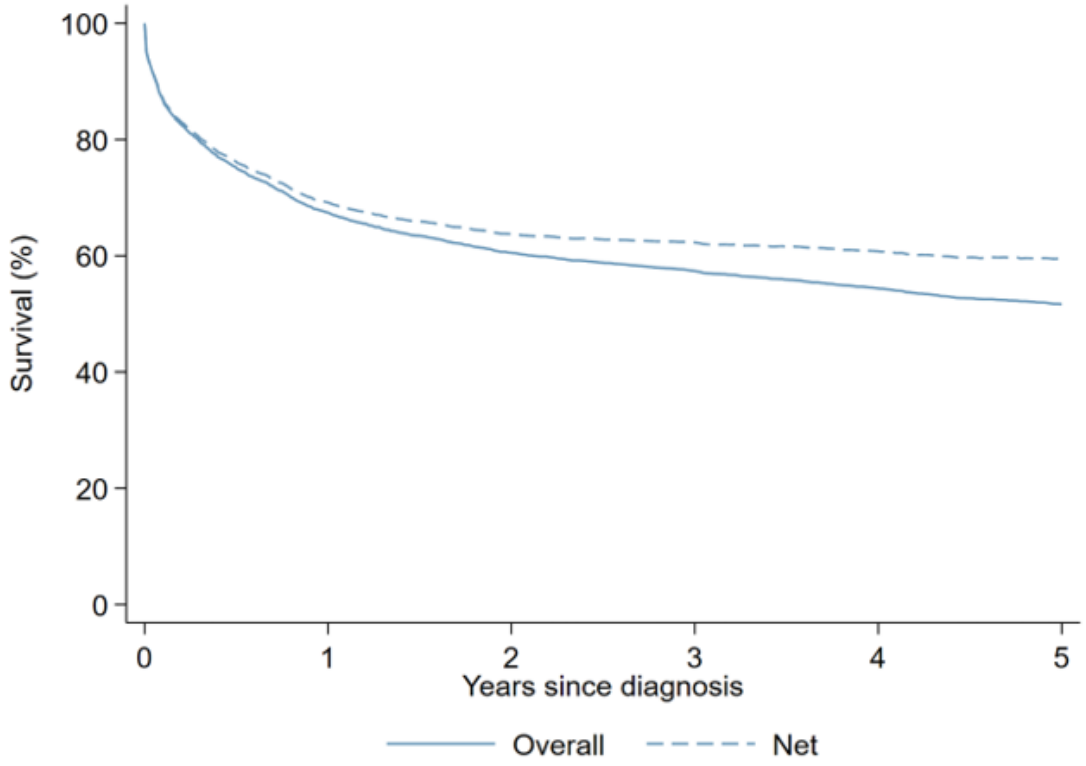


Figure 5:10. Overall and net survival curves.

Table 5:13. Survival at one to five years from diagnosis.

Time since diagnosis (years)	Patients alive at start of time period (n)	Deaths (n)	Overall survival (%)	Net survival (%)
0-1	4110	1339	67.4	69.2
1-2	2764	282	60.5	63.7
2-3	2471	128	57.4	62.3
3-4	2342	122	54.4	60.7
4-5	2194	106	51.7	59.5

5.5.1 Survival by demographic and baseline clinical characteristics

Overall and net survival are presented in Figure 5:11 and Table 5:14 stratified by patient characteristics and hazard ratios are estimated using Cox regression (Table 5:15). One- and five- year OS and NS differed by all patient characteristics, except for sex. However, an association between female sex and better five-year survival was revealed in the adjusted analyses. Most of this survival difference by sex is revealed with increasing magnitude when treatment intent, ECOG score and age are added to the model respectively (female-to-male hazard ratio (HR) adjusted for treatment intent, ECOG score and age: 0.81 [95% CI: 0.74-0.89]). Excluding those under 18 years old, increasing age is associated with worse survival at both one- and five- years, for example, those aged 18-39 have less than a third of the probability of death at five-years compared to patients aged 70-79. Furthermore, the difference between OS and NS widened with increasing age from less than 1% for the 18-39 group to 17.6% for the 80 years and older age group at five years, indicating that deaths in the younger age categories are almost all attributable to the patients' DLBCL whereas in the older age categories some patients will have died from other causes. Interestingly, the upper confidence interval for one-year NS of patients aged under 18 years old is 102.3%, indicating that a DLBCL diagnosis has little effect on survival for patients in this age group and the upper bound of greater than 100% is the explained by small counts in this age category. Those with PCNSL had markedly worse OS and NS compared to all other DLBCL patients, for example, five-year NS was 20.4% for PCNSL patients compared to 60.9% for all other patients. Advanced disease, B symptoms and higher ECOG scores were also associated with worse one- and five-year survival, for example, those with ECOG scores of four were over 16 times as likely to die within five-years compared to patients with a score of zero and this relationship remained when adjusted for baseline characteristics, although it was considerably weakened to six times as likely.

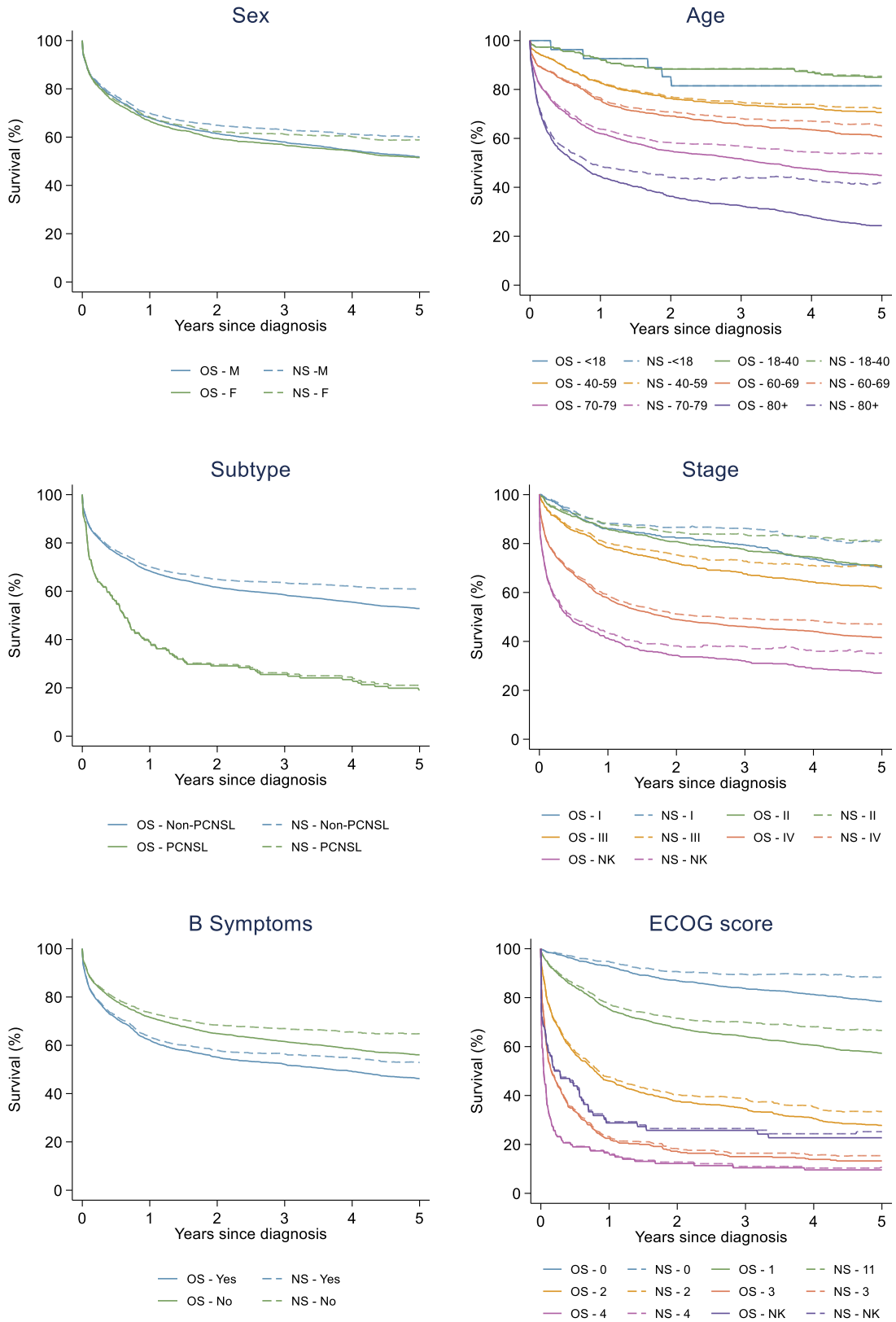


Figure 5:11. Overall (OS) and net survival (NS) curves stratified by patient characteristics.

ECOG= Eastern Cooperative Oncology Group

Table 5:14. One- and five- year survival stratified by patient characteristics.

Patient Characteristic	Patients alive at start of time period (N)	Deaths at 1 year (N)	% 1-year Overall Survival (95% CI)	% 1-year Net Survival (95% CI)	Deaths at 5 years (N)	% 5-year Overall Survival (95% CI)	% 5-year Net Survival (95% CI)
Sex:							
Male	2179	696	68.1 (66.1-70.0)	69.9 (67.9-71.9)	1045	51.8 (49.7-53.9)	60.1 (57.4-62.7)
Female	1931	643	66.7 (64.6-68.7)	68.3 (66.2-70.5)	932	51.4 (49.1-53.6)	58.8 (56.1-61.6)
Age at diagnosis (years):							
<18	27	2	92.6 (73.5-98.1)	92.6 (82.9-102.3)	5	81.5 (61.1-91.8)	81.5 (67.2-95.9)
18-39	224	17	92.4 (88.1-95.2)	92.5 (89.0-96.0)	33	84.6 (79.1-88.7)	85.0 (80.7-90.2)
40-59	785	136	82.7 (79.8-85.2)	83.0 (80.4-85.7)	230	70.6 (67.3-73.7)	72.3 (69.0-75.6)
60-69	976	237	75.7 (72.9-78.3)	76.6 (73.9-79.3)	383	60.6 (57.4-63.6)	65.1 (61.8-68.4)
70-79	1244	473	62.0 (59.2-64.6)	63.8 (61.0-66.6)	683	44.9 (42.1-47.6)	54.0 (50.4-57.1)
≥80	854	474	44.4 (41.1-47.7)	48.8 (45.1-52.5)	643	24.3 (21.5-27.2)	41.9 (36.4-47.4)
Primary CNS lymphoma:							
Yes	141	86	39.0 (31.0-47.0)	39.4 (31.3-47.5)	114	19.1 (13.1-26.0)	20.4 (13.5-27.2)
No	3969	1253	68.4 (67.0-69.8)	70.2 (68.7-71.7)	1863	52.8 (51.2-54.3)	60.9 (59.0-62.8)
Stage:							
I	526	73	86.1 (82.9-88.8)	88.2 (85.2-91.3)	156	70.2 (66.0-73.4)	80.6 (75.6-85.6)
II	694	98	85.9 (83.1-88.3)	88.0 (85.3-90.7)	200	70.9 (67.3-74.1)	81.4 (77.2-85.7)
III	655	142	78.3 (75.0-81.3)	80.0 (76.8-83.3)	249	61.8 (57.9-65.4)	70.6 (66.2-75.0)
IV	1737	734	57.7 (55.4-60.0)	59.0 (56.7-61.4)	1010	41.5 (39.2-43.8)	47.1 (44.4-49.8)
Not fully staged	498	292	41.4 (37.0-45.6)	43.6 (39.0-48.2)	362	27.1 (23.2-31.0)	35.2 (29.1-41.4)
B Symptoms:							
Yes	1815	685	62.3 (60.0-64.4)	63.8 (61.5-66.1)	972	46.2 (43.8-48.4)	52.9 (50.1-55.7)
No	2292	651	71.6 (69.7-73.4)	73.5 (71.6-75.4)	1002	56.0 (54.0-58.0)	64.8 (62.3-67.4)
Not known	3	3	-	-	3	-	-

Patient Characteristic	Patients alive at start of time period (N)	Deaths at 1 year (N)	% 1-year Overall Survival (95% CI)	% 1-year Net Survival (95% CI)	Deaths at 5 years (N)	% 5-year Overall Survival (95% CI)	% 5-year Net Survival (95% CI)
ECOG score:							
0	1195	85	92.9 (91.3-94.2)	94.8 (93.3-96.3)	255	78.4 (76.0-80.7)	88.4 (85.4-91.4)
1	1542	380	75.4 (73.1-77.4)	77.4 (75.2-79.6)	656	57.2 (54.7-59.7)	66.6 (63.5-69.7)
2	822	444	45.9 (42.5-41.0)	47.6 (44.0-51.1)	591	27.6 (24.6-30.8)	33.4 (29.5-37.3)
3	364	282	22.5 (18.4-26.9)	23.3 (18.8-27.7)	315	13.2 (10.0-16.9)	15.4 (11.2-19.6)
4	121	101	16.5 (10.6-23.7)	16.9 (10.3-23.5)	109	9.5 (5.1-15.6)	10.7 (4.7-16.6)
Not known	66	47	28.8 (18.5-39.9)	29.3 (18.3-40.3)	51	22.7 (13.5-33.4)	25.2 (14.2-36.2)

ECOG= Eastern Cooperative Oncology Group

Table 5:15. Five-year survival: number of deaths, unadjusted and adjusted hazard ratios by patient characteristics.

	Total	Deaths	Unadjusted hazard ratio	Adjusted hazard ratio
	N (%)	N (%)	(95% CI) [a]	(95% CI) [b]
Total	4115 (100)	1981 (48.1)	-	-
Sex:				
Males	2179 (53.0)	1045 (52.8)	Ref	Ref
Females	1936 (47.0)	936 (47.2)	1.02 (0.94 - 1.12)	0.80 (0.73 - 0.87)
Age at diagnosis (years):				
<18	27 (0.7)	5 (0.3)	0.25 (0.10 - 0.59)	0.26 (0.11 - 0.64)
18-39	225 (5.5)	34 (1.7)	0.20 (0.14 - 0.29)	0.27 (0.19 - 0.38)
40-59	786 (19.1)	230 (11.6)	0.43 (0.37 - 0.49)	0.50 (0.43 - 0.58)
60-69	976 (23.7)	383 (19.3)	0.61 (0.54 - 0.70)	0.68 (0.60 - 0.77)
70-79	1245 (30.3)	684 (34.5)	Ref	Ref
≥80	856 (20.8)	645 (32.6)	1.74 (1.56 - 1.94)	1.01 (0.89 - 1.14)
Primary CNS lymphoma:				
Yes	141 (3.4)	114 (5.8)	2.44 (2.02 - 2.95)	1.69 (1.37 - 2.09)
No	3974 (96.6)	1867 (94.2)	Ref	Ref
Ann-Arbor stage:				
I	526 (12.8)	156 (7.9)	Ref	Ref
II	696 (16.9)	201 (10.1)	0.99 (0.80 - 1.22)	1.24 (1.00 - 1.54)
III	655 (15.9)	249 (12.6)	1.39 (1.14 - 1.70)	1.71 (1.39 - 2.10)
IV	1740 (42.3)	1013 (51.1)	2.69 (2.28 - 3.19)	2.21 (1.86 - 2.64)
Not fully staged	498 (12.1)	362 (18.3)	4.33 (3.59 - 5.23)	1.83 (1.51 - 2.23)
B Symptoms:				
Yes	1818 (44.2)	974 (49.2)	1.35 (1.24 - 1.47)	1.18 (1.08 - 1.29)
No	2294 (55.7)	1004 (50.7)	Ref	Ref
Not known	3 (0.1)	3 (0.2)	-	-
ECOG score:				
0	1196 (29.1)	255 (12.9)	Ref	Ref
1	1543 (37.5)	657 (33.2)	2.33 (2.02 - 2.69)	1.76 (1.52 - 2.04)
2	823 (20.0)	592 (29.9)	5.70 (4.92 - 6.61)	3.21 (2.74 - 3.75)

	Total	Deaths	Unadjusted hazard ratio	Adjusted hazard ratio
	N (%)	N (%)	(95% CI) [a]	(95% CI) [b]
3	365 (8.9)	316 (16.0)	10.78 (9.12 - 12.74)	3.92 (3.26 - 4.72)
4	122 (3.0)	110 (5.6)	16.26 (12.98 - 20.38)	6.14 (4.81 - 7.84)
Not known	66 (1.6)	51 (2.6)	8.09 (5.98 - 10.93)	3.99 (2.91 - 5.49)

[a]Unadjusted hazard ratios for risk of death within five years of diagnosis with 95% confidence intervals.

[b]Hazard ratios for risk of death within five years of diagnosis with 95% confidence intervals adjusted for all other variables in the table in addition to route to diagnosis, IMD quintile, rurality, Cancer Alliance, travel time by car and public transport to GP and hospital, and treatment intent.

ECOG= Eastern Cooperative Oncology Group

5.5.2 Survival by deprivation and rurality

No difference was observed in the one- and five- year OS and NS of patients when stratified by IMD quintile (Figure 5:12, Table 5:16, and Table 5:17). There was a small difference in one-year OS between patients living in rural and urban areas, however, no difference was detected in the five-year OS or the one- and five- year NS and there was no difference in the hazard of death at five-years.

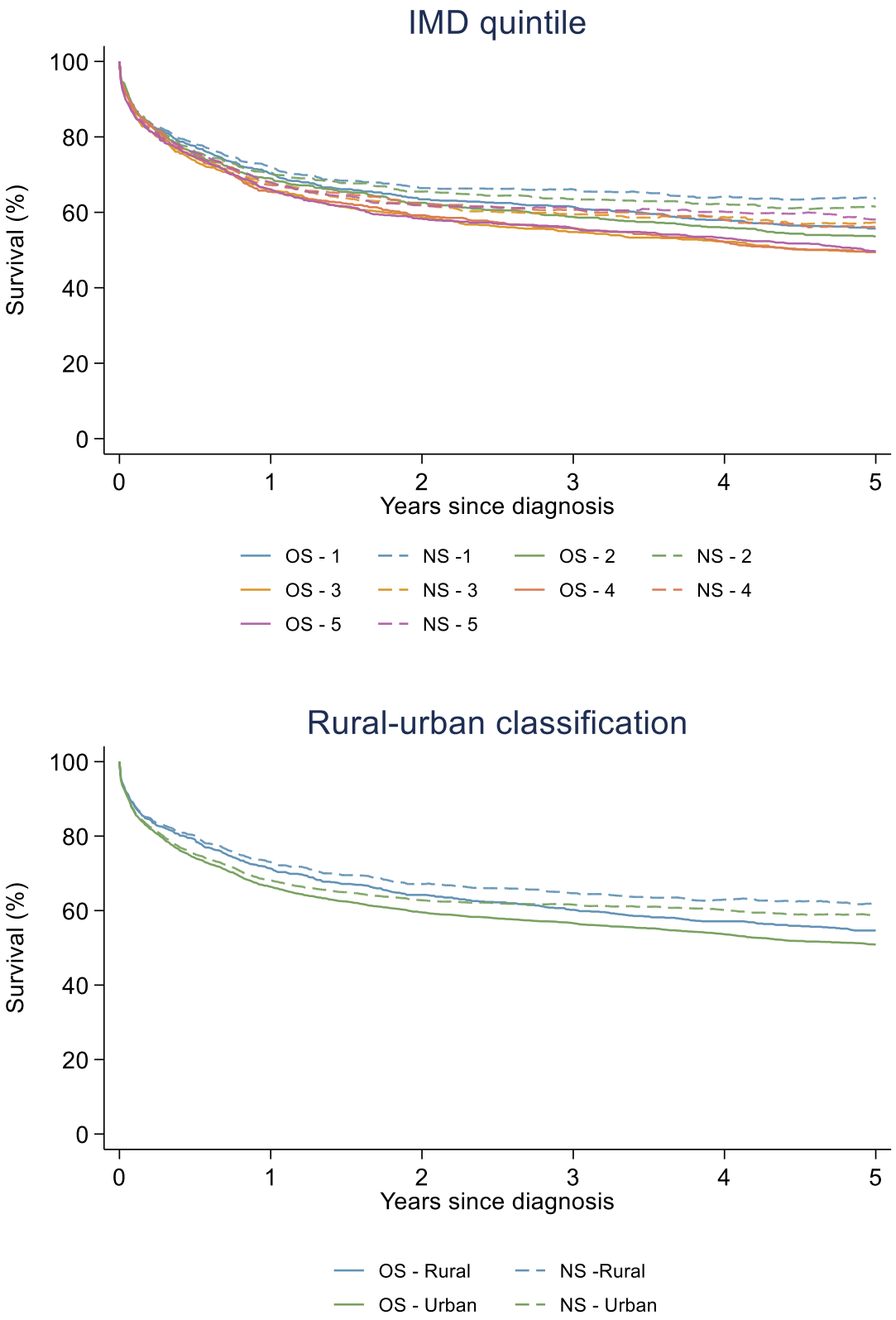


Figure 5:12. Overall survival (OS) and net survival (NS) curves stratified by Index of Multiple Deprivation (IMD) quintile and rurality.

Table 5:16. One- and five- year survival stratified by Index of Multiple Deprivation (IMD) quintile and rurality: HMRN patients newly diagnosed with diffuse large B-cell lymphoma between January 2005-August 2019.

Patient Characteristic	Patients alive at start of time period (N)	Deaths at 1 year (N)	% 1-year Overall Survival (95% CI)	% 1-year Net Survival (95% CI)	Deaths at 5 years (N)	% 5-year Overall Survival (95% CI)	% 5-year Net Survival (95% CI)
IMD income quintile:							
1 (least deprived)	827	246	70.3 (67.0-73.2)	71.9 (68.7-75.1)	365	55.7 (52.2-59.0)	63.8 (59.7-67.8)
2	889	276	69.0 (65.8-71.9)	70.5 (67.4-73.7)	411	53.6 (50.3-56.9)	61.5 (57.6-65.5)
3	819	279	65.9 (62.5-69.0)	67.6 (64.2-70.9)	411	49.5 (46.0-52.9)	57.3 (53.1-61.6)
4	710	245	65.5 (61.9-68.9)	67.3 (63.6-70.9)	357	49.5 (45.7-53.1)	56.2 (51.5-60.9)
5 (most deprived)	865	293	66.1 (62.9-69.2)	68.1 (64.8-71.3)	433	49.6 (46.2-52.9)	58.1 (53.8-62.3)
Rural-urban classification:							
Rural	850	244	71.3 (68.1-74.2)	73.0 (68.9-76.2)	384	54.7 (51.2-57.9)	61.9 (57.8-65.9)
Urban	3260	1095	66.4 (64.8-68.0)	68.1 (66.5-69.8)	1593	50.9 (49.1-52.6)	58.9 (56.7-61.0)

IMD= Index of Multiple Deprivation

Table 5:17. Five-year survival: number of deaths, unadjusted and adjusted hazard ratios by Index of Multiple Deprivation (IMD) quintile and rurality.

	Total	Deaths	Unadjusted hazard ratio	Adjusted hazard ratio
	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
IMD income quintile:				
1 (least deprived)	827 (20.1)	365 (18.4)	Ref	Ref
2	889 (21.6)	411 (20.7)	1.06 (0.92 - 1.22)	0.99 (0.86 - 1.15)
3	819 (19.9)	411 (20.7)	1.19 (1.04 - 1.37)	1.14 (0.99 - 1.32)
4	710 (17.3)	357 (18.0)	1.19 (1.03 - 1.37)	1.07 (0.92 - 1.24)
5 (most deprived)	865 (21.0)	433 (21.9)	1.19 (1.03 - 1.36)	1.09 (0.94 - 1.26)
Not known	5 (0.1)	4 (0.2)	-	-
Rural-urban classification:				
Rural	850 (20.7)	384 (19.4)	0.88 (0.79 - 0.99)	0.90 (0.79 - 1.02)
Urban	3260 (79.2)	1593 (80.4)	Ref	Ref
Not known	5 (0.1)	4 (0.2)	-	-

[a]Unadjusted hazard ratios for risk of death within five years of diagnosis with 95% confidence intervals.

[b]Hazard ratios for risk of death within five years of diagnosis with 95% confidence intervals adjusted for all other variables in the table in addition to sex, age, subtype, stage, B symptoms, performance status, route to diagnosis, Cancer Alliance, travel time by car and public transport to GP and hospital, and treatment intent.

IMD= Index of Multiple Deprivation

5.5.3 Survival by travel to healthcare

Interestingly, those in the ‘low’ travel categories have the poorest survival while patients in the ‘timed out’ and ‘not known’ categories typically had the best (Figure 5:13, Table 5:18 and Table 5:19). For example, one-year OS for patients in the ‘low’ travel time to hospital by public transport group was 63.9% (95% CI: 60.8-66.8) compared to 72.7% (95% CI: 65.6-78.6) for the ‘timed out’ group and 81.8% (95% CI: 44.7-95.1) for the ‘not known’ group. Similarly, the risk of death for the five-years following diagnosis for those in the ‘timed out’ category for time to hospital by public transport was three quarters (HR: 0.73 [95% CI: 0.57-0.93]) that of the ‘low’ category, however, this association was not significant when adjusted for baseline characteristics (HR: 0.85 [95% CI: 0.59-1.22]). None of the differences detected in one- and five- year OS or NS by travel reached statistical significance, which may in part be the result of small sample sizes for some of the categories.

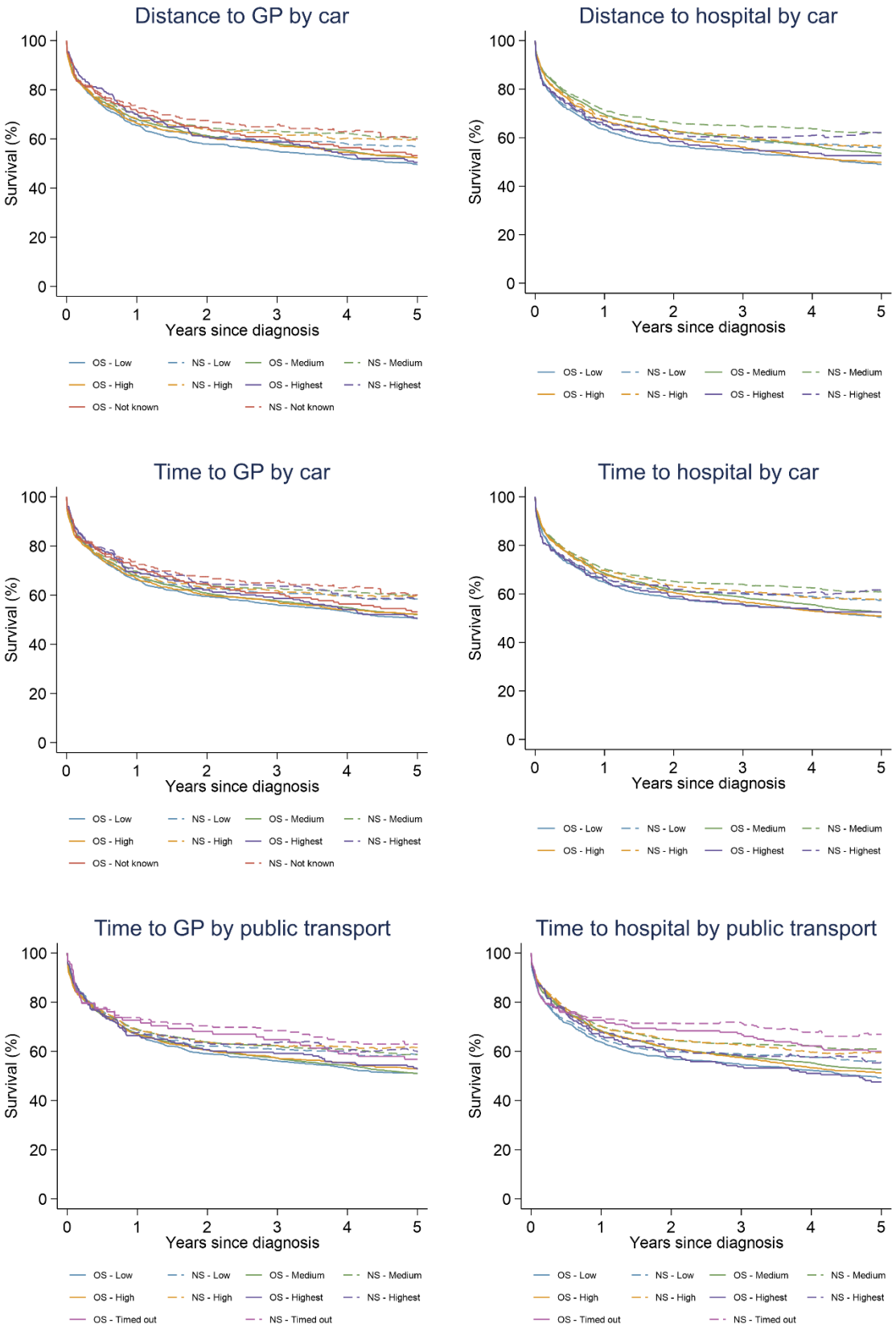


Figure 5:13. Overall survival (OS) and net survival (NS) curves stratified by travel to healthcare.

Table 5:18. One- and five- year survival stratified by travel to healthcare.

Patient Characteristic	Patients alive at start of time period (N)	Deaths at 1 year (N)	% 1-year Overall Survival (95% CI)	% 1-year Net Survival (95% CI)	Deaths at 5 years (N)	% 5-year Overall Survival (95% CI)	% 5-year Net Survival (95% CI)
Distance to GP:							
Low	988	341	65.5 (62.4-68.4)	67.1 (64.1-70.2)	495	49.6 (46.4-52.7)	56.9 (53.1-60.7)
Medium	1974	627	68.2 (66.1-70.2)	70.0 (67.9-72.2)	936	52.4 (50.2-54.6)	60.8 (58.1-63.6)
High	787	265	66.3 (62.9-69.5)	67.9 (64.5-71.3)	373	52.3 (48.7-55.7)	59.5 (55.3-63.8)
Highest	197	59	70.0 (63.1-75.9)	71.7 (65.1-78.3)	97	50.5 (43.2-57.2)	58.3 (49.8-66.8)
Not known	164	47	71.3 (63.8-77.6)	73.1 (66.0-80.2)	76	53.2 (45.2-60.5)	60.1 (50.2-69.9)
Time to GP by car:							
Low	1057	358	66.1 (63.2-68.9)	67.8 (64.9-70.8)	521	50.5 (47.4-53.5)	58.4 (54.8-62.1)
Medium	1866	598	67.9 (65.8-70.0)	69.7 (67.6-71.9)	890	52.1 (49.8-54.4)	60.1 (57.2-62.9)
High	822	274	66.7 (63.3-69.8)	68.3 (65.0-71.6)	391	52.2 (48.7-55.5)	59.6 (55.5-63.7)
Highest	201	62	69.2 (62.3-75.0)	70.7 (64.1-77.3)	99	50.6 (43.5-57.3)	58.5 (50.1-66.9)
Not known	164	47	71.3 (63.8-77.6)	73.1 (66.0-80.2)	76	53.2 (45.2-60.5)	60.1 (50.2-69.9)
Time to GP by public transport:							
Low	979	323	67.0 (64.0-69.9)	68.7 (65.7-71.8)	477	51.1 (47.9-54.2)	58.8 (55.0-62.7)
Medium	1921	628	67.3 (65.2-69.4)	69.1 (66.9-71.2)	937	51.0 (48.8-53.2)	58.7 (55.9-61.5)
High	764	252	67.0 (63.5-70.2)	68.8 (65.4-72.2)	358	52.9 (49.3-56.4)	61.6 (57.3-66.0)
Highest	194	65	66.5 (59.4-72.6)	67.6 (60.8-74.4)	91	52.6 (45.3-59.4)	59.7 (51.2-68.2)
Timed out	88	24	72.7 (62.1-80.8)	73.7 (64.3-83.2)	38	56.8 (45.8-66.4)	62.9 (51.3-74.6)
Not known	164	47	71.3 (63.8-77.6)	73.1 (66.0-80.2)	76	53.2 (45.2-60.5)	60.1 (50.2-69.9)
Distance to hospital:							
Low	1025	375	63.4 (60.4-66.3)	65.0 (62.0-68.1)	522	48.9 (45.8-51.9)	55.9 (52.1-59.7)
Medium	2050	623	69.6 (67.6-71.5)	71.4 (69.4-73.5)	947	53.6 (51.4-55.7)	62.0 (59.3-64.7)
High	820	268	67.3 (64.0-70.4)	69.0 (65.6-72.3)	408	49.9 (46.4-53.3)	56.8 (53.3-61.6)
Highest	205	71	65.4 (58.4-71.4)	66.9 (60.3-73.6)	97	52.6 (45.5-59.1)	62.1 (53.9-70.3)
Not known	10	2	80.0 (40.9-94.6)	80.2 (56.7-103.8)	3	70.0 (32.9-89.2)	76.6 (47.5-105.6)

Patient Characteristic	Patients alive at start of time period (N)	Deaths at 1 year (N)	% 1-year Overall Survival (95% CI)	% 1-year Net Survival (95% CI)	Deaths at 5 years (N)	% 5-year Overall Survival (95% CI)	% 5-year Net Survival (95% CI)
Time to hospital by car:							
Low	1029	361	64.9 (61.9-67.8)	66.6 (63.5-69.6)	508	50.4 (47.3-53.5)	57.4 (53.6-61.2)
Medium	2051	646	68.5 (66.4-70.5)	70.3 (68.2-72.4)	972	52.4 (50.2-54.5)	60.9 (58.1-63.6)
High	813	259	68.1 (64.8-71.2)	69.8 (66.5-73.0)	396	51.0 (47.5-54.4)	57.9 (53.8-62.0)
Highest	207	71	65.7 (58.8-71.7)	67.3 (60.7-73.9)	98	52.5 (45.5-59.1)	61.6 (53.5-69.8)
Not known	10	2	80.0 (40.9-94.6)	80.2 (56.7-103.8)	3	70.0 (32.9-89.2)	76.6 (47.5-105.6)
Time to hospital by public transport:							
Low	982	355	63.9 (60.8-66.8)	65.4 (62.3-68.5)	497	49.2 (46.0-52.3)	55.6 (51.8-59.4)
Medium	1955	618	68.4 (66.3-70.4)	70.2 (68.1-72.3)	923	52.6 (50.4-54.8)	61.1 (58.3-63.8)
High	784	248	68.4 (65.0-71.5)	70.2 (66.9-73.6)	380	51.2 (47.6-54.6)	59.4 (55.1-63.8)
Highest	195	66	66.1 (59.0-72.3)	67.7 (60.9-74.5)	101	47.5 (40.3-54.4)	55.4 (46.8-63.9)
Timed out	183	50	72.7 (65.6-78.6)	73.8 (67.3-80.3)	73	60.0 (52.5-66.7)	66.9 (58.7-75.2)
Not known	11	2	81.8 (44.7-95.1)	82.0 (60.3-103.8)	3	72.7 (37.1-90.3)	78.7 (51.8-105.5)

Table 5:19. Five-year survival: number of deaths, unadjusted and adjusted hazard ratios by travel to healthcare.

	Total	Deaths	Unadjusted hazard ratio	Adjusted hazard ratio
	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
Total	4115 (100)	1981 (48.1)	-	-
Distance to GP:				
Low	988 (24.0)	495 (25.0)	Ref	-
Medium	1974 (48.0)	936 (47.2)	0.92 (0.83 - 1.03)	-
High	787 (19.1)	373 (18.8)	0.93 (0.82 - 1.07)	-
Highest	197 (4.8)	97 (4.9)	0.94 (0.75 - 1.17)	-
Not known	169 (4.1)	80 (4.0)	0.92 (0.72 - 1.16)	-
Time to GP by car:				
Low	1057 (25.7)	521 (26.3)	Ref	Ref
Medium	1866 (45.3)	890 (44.9)	0.95 (0.86 - 1.06)	0.95 (0.83 - 1.09)
High	822 (20.0)	391 (19.7)	0.97 (0.85 - 1.10)	1.01 (0.83 - 1.22)
Highest	201 (4.9)	99 (5.0)	0.96 (0.78 - 1.19)	1.10 (0.79 - 1.53)
Not known	169 (4.1)	80 (4.0)	0.94 (0.74 - 1.19)	1.00 (0.77 - 1.29)
Time to GP by public transport:				
Low	979 (23.8)	477 (24.1)	Ref	Ref
Medium	1921 (46.7)	937 (47.3)	1.00 (0.89 - 1.11)	1.03 (0.90 - 1.19)
High	764 (18.6)	358 (18.1)	0.96 (0.84 - 1.10)	1.00 (0.82 - 1.22)
Highest	194 (4.7)	91 (4.6)	0.96 (0.76 - 1.20)	0.87 (0.63 - 1.20)
Timed out	88 (2.1)	38 (1.9)	0.84 (0.60 - 1.16)	0.89 (0.56 - 1.41)
Not known	169 (4.1)	80 (4.0)	0.96 (0.75 - 1.21)	-
Distance to hospital:				
Low	1025 (24.9)	522 (26.4)	Ref	-
Medium	2050 (49.8)	947 (47.8)	0.85 (0.76 - 0.94)	-
High	820 (19.9)	408 (20.6)	0.94 (0.83 - 1.07)	-
Highest	205 (5.0)	97 (4.9)	0.92 (0.74 - 1.14)	-
Not known	15 (0.4)	7 (0.4)	-	-
Time to hospital by car:				
Low	1029 (25.0)	508 (25.6)	Ref	Ref
Medium	2051 (49.8)	972 (49.1)	0.92 (0.83 - 1.03)	1.03 (0.89 - 1.19)

	Total	Deaths	Unadjusted hazard ratio	Adjusted hazard ratio
	N (%)	N (%)	(95% CI)[a]	(95% CI)[b]
High	813 (19.8)	396 (20.0)	0.96 (0.84 - 1.09)	1.00 (0.81 - 1.23)
Highest	207 (5.0)	98 (4.9)	0.96 (0.77 - 1.19)	0.90 (0.65 - 1.24)
Not known	15 (0.4)	7 (0.4)	-	-
Time to hospital by public transport:				
Low	982 (23.9)	497 (25.1)	Ref	Ref
Medium	1955 (47.5)	923 (46.6)	0.89 (0.80 - 0.99)	0.86 (0.75 - 1.00)
High	784 (19.1)	380 (19.2)	0.92 (0.80 - 1.05)	0.89 (0.73 - 1.08)
Highest	195 (4.7)	101 (5.1)	1.00 (0.81 - 1.24)	0.99 (0.73 - 1.33)
Timed out	183 (4.4)	73 (3.7)	0.73 (0.57 - 0.93)	0.85 (0.59 - 1.22)
Not known	16 (0.4)	7 (0.4)	-	-

[a]Unadjusted hazard ratios for risk of death within five years of diagnosis with 95% confidence intervals.

[b]Hazard ratios for risk of death within five years of diagnosis with 95% confidence intervals adjusted for all other variables in the table in addition to sex, age, subtype, stage, B symptoms, performance status, route to diagnosis, IMD quintile, rurality, Cancer Alliance and treatment intent. Excluding distance to GP and distance to hospital due to collinearity with time to GP and time to hospital.

5.5.4 Survival by treatment intent

As expected, survival differs dramatically depending on treatment intent as illustrated in Figure 5:14 and Table 5:20. Most patients not treated curatively are unfit for aggressive treatment and are treated palliatively or for symptom management only. Therefore, it is unsurprising that over 85% of patients who did not receive treatment with curative intent died within one-year of their diagnosis and only 5% were alive five-years after diagnosis, compared to 61% of those treated with curative intent. After adjusting for baseline characteristics patients not treated curatively were still five times as likely to die within five-years of diagnosis compared to those treated curatively (Table 5:21).

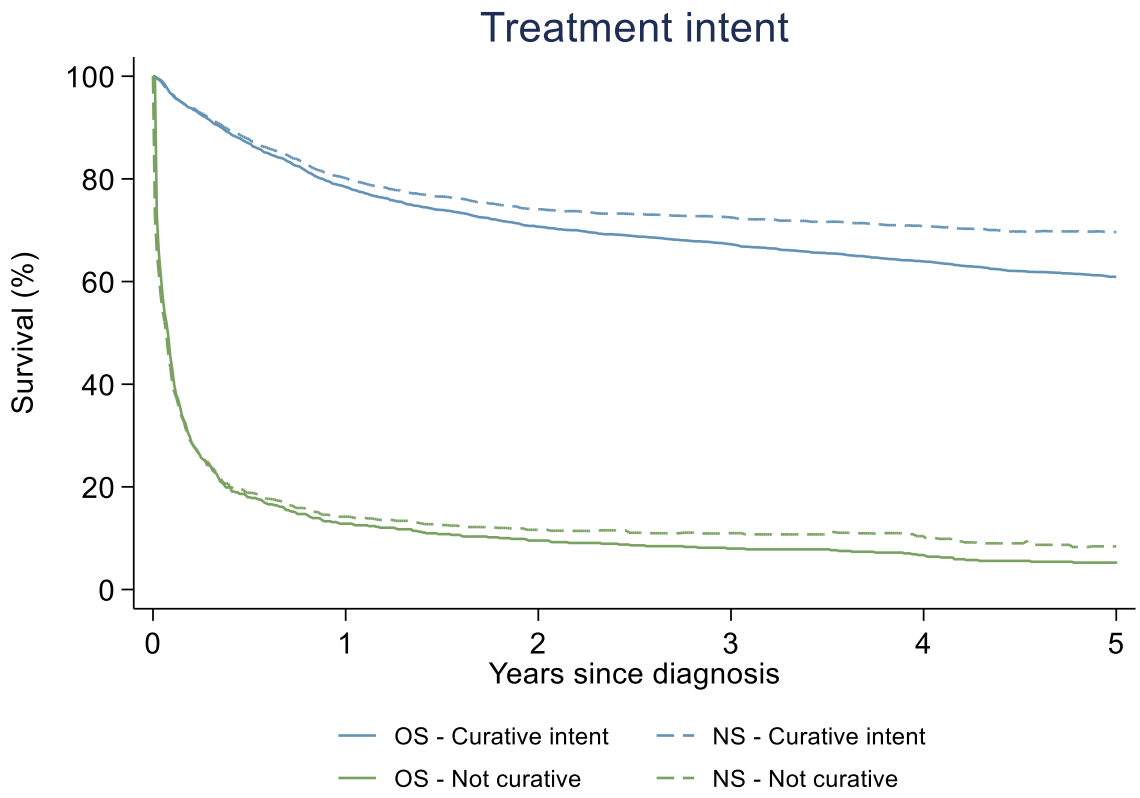


Figure 5:14. Overall survival (OS) and net survival (NS) curves stratified by treatment intent.

Table 5:20. One- and five- year survival stratified by treatment intent.

Patient Characteristic	Patients alive at start of time period (N)	Deaths at 1 year (N)	% 1-year Overall Survival (95% CI)	% 1-year Net Survival (95% CI)	Deaths at 5 years (N)	% 5-year Overall Survival (95% CI)	% 5-year Net Survival (95% CI)
Treatment intent:							
Curative	3424	742	78.3 (76.9-79.7)	80.1 (78.7-81.6)	1330	60.9 (59.2-62.5)	69.6 (67.7-71.6)
Not curative	639	557	12.8 (10.4-15.6)	14.2 (11.3-17.0)	606	5.1 (3.5-7.0)	8.4 (4.9-11.8)
Not known	47	40	9.5 (3.0-20.6)	9.8 (1.4-18.2)	41	7.1 (1.9-17.5)	7.5 (0.0-0.15)

Table 5:21. Five-year survival: number of deaths, unadjusted and adjusted hazard ratios by treatment intent.

	Total N (%)	Deaths N (%)	Unadjusted hazard ratio (95% CI)[a]	Adjusted hazard ratio (95% CI)[b]
Treatment intent:				
Curative intent	3425 (83.2)	1330 (67.1)	Ref	Ref
Non-curative	642 (15.6)	609 (30.7)	8.78 (7.94 - 9.72)	5.06 (4.44 - 5.76)
Not known	48 (1.2)	42 (2.1)	-	-

[a]Unadjusted hazard ratios for risk of death within five years of diagnosis with 95% confidence intervals.

[b]Hazard ratios for risk of death within five years of diagnosis with 95% confidence intervals adjusted for all other variables in the table in addition to sex, age, subtype, stage, B symptoms, performance status, route to diagnosis, IMD quintile, rurality, Cancer Alliance, travel time by car and public transport to GP and hospital.

5.5.5 Survival by diagnostic route

One- and five- year OS and NS is considerably higher for patients diagnosed following an Elective Referral compared to an Emergency Presentation (Figure 5:15, Table 5:22 and Table 5:23). For example, at one-year following diagnosis only half of patients diagnosed as an Emergency Presentation are alive, compared to over three-quarters diagnosed following an Elective Referral. The shape of the two survival curves also differs, with the curve representing survival of patients diagnosed as an Emergency having a steeper gradient at 0-1 years post-diagnosis compared to that for patients diagnosed following an Elective Referral. The gradient for the Emergency Presentation curve then starts to plateau and the two curves are almost parallel between 1-5 years post-diagnosis. This is reflected in the difference in both OS and NS between patients diagnosed as Elective Referrals compared to Emergencies which is greater at one-year (OS: 26.1%, NS: 27.1%) than at five-years (OS: 21.2%, NS: 25.8%).

Patients who present as an Emergency are twice as likely to die in the five-years following diagnosis, compared to patients diagnosed following an Elective Referral (HR: 2.05 [95% CI: 1.87-2.25]) (Table 5:23). All patients are vulnerable to this survival disadvantage, excluding those with PCNSL diagnosed as an Emergency whose risk of death is no greater than patients who are diagnosed following an Elective Referral (HR: 0.98 [95% CI: 0.66-1.45]) (Figure 5:16). As described earlier, PCNSL is an aggressive subtype and survival for these patients is much poorer than other DLBCL patients and therefore a lack of effect of diagnostic route on survival is unsurprising.

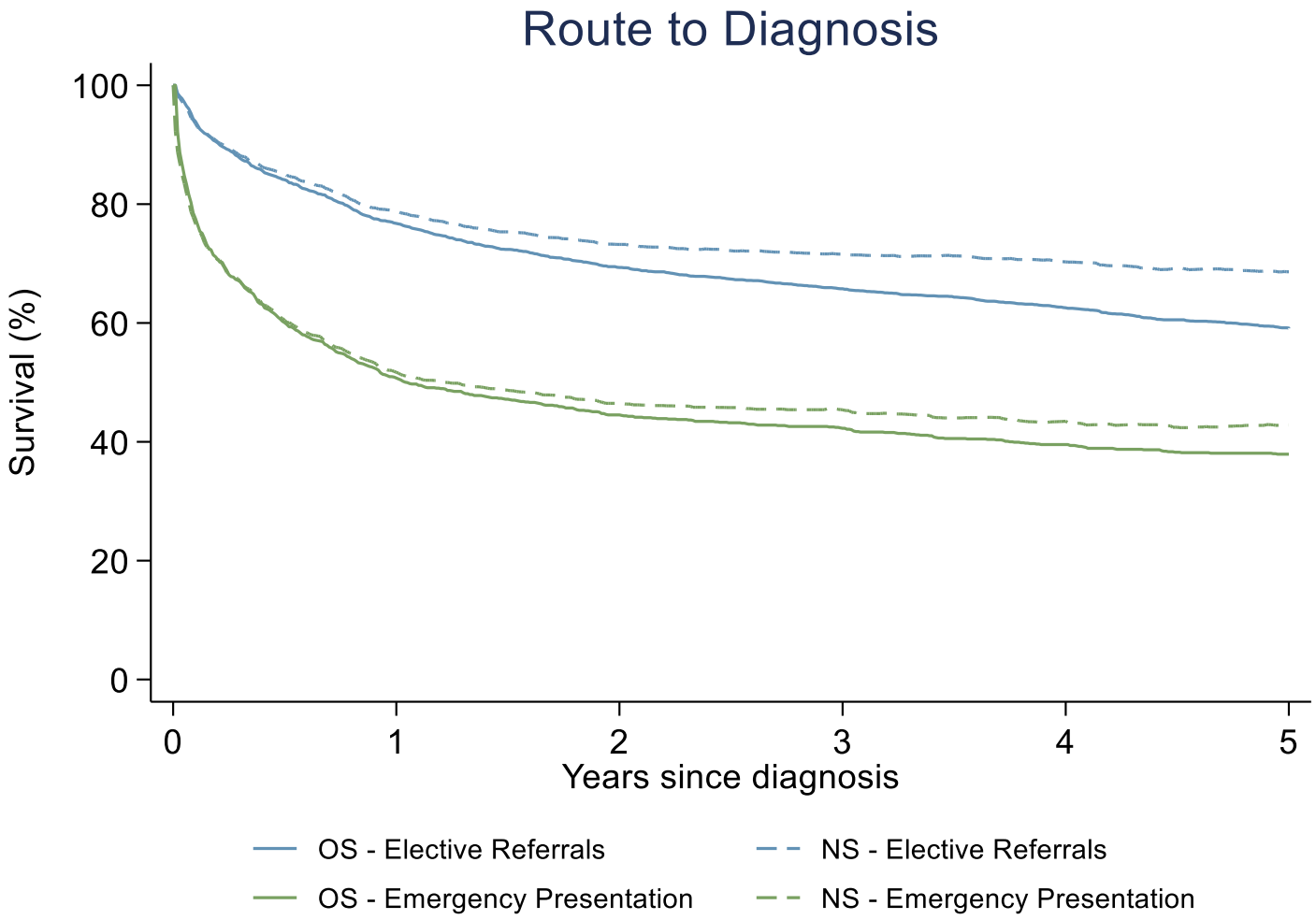


Figure 5:15. Overall survival (OS) and net survival (NS) curves stratified by route to diagnosis.

Table 5:22. One- and five- year survival stratified by route to diagnosis.

Patient Characteristic	Patients alive at start of time period (N)	Deaths at 1 year (N)	% 1-year Overall Survival (95% CI)	% 1-year Net Survival (95% CI)	Deaths at 5 years (N)	% 5-year Overall Survival (95% CI)	% 5-year Net Survival (95% CI)
Route to Diagnosis:							
Emergency Presentation	1289	640	50.6 (47.8-53.3)	51.7 (48.9-54.5)	802	37.9 (35.3-40.6)	42.8 (39.7-46.0)
Elective referrals	2629	614	76.7 (75.0-78.3)	78.8 (77.1-80.4)	1072	59.1 (57.2-61.0)	68.6 (66.2-71.0)
Unknown	192	89	53.5 (46.2-60.3)	54.6 (48.0-61.2)	107	42.4 (35.1-49.5)	46.6 (38.9-54.4)

Table 5:23. Five-year survival: number of deaths, unadjusted and adjusted hazard ratios by diagnostic route.

	Total N (%)	Deaths N (%)	Unadjusted hazard ratio (95% CI)[a]	Adjusted hazard ratio (95% CI)[b]
Route to Diagnosis:				
Emergency Presentation	1292 (31.4)	802 (40.5)	2.05 (1.87 - 2.24)	1.38 (1.25 - 1.53)
Elective referral	2631 (63.9)	1072 (54.1)	Ref	Ref
Unknown	192 (4.7)	107 (5.4)	2.01 (1.65 - 2.45)	1.92 (1.56 - 2.36)

[a]Unadjusted hazard ratios for risk of death within five years of diagnosis with 95% confidence intervals.

[b]Hazard ratios for risk of death within five years of diagnosis with 95% confidence intervals adjusted for sex, age, subtype, stage, B symptoms, performance status, route to diagnosis, IMD quintile, rurality, Cancer Alliance, travel time by car and public transport to GP and hospital and treatment intent.

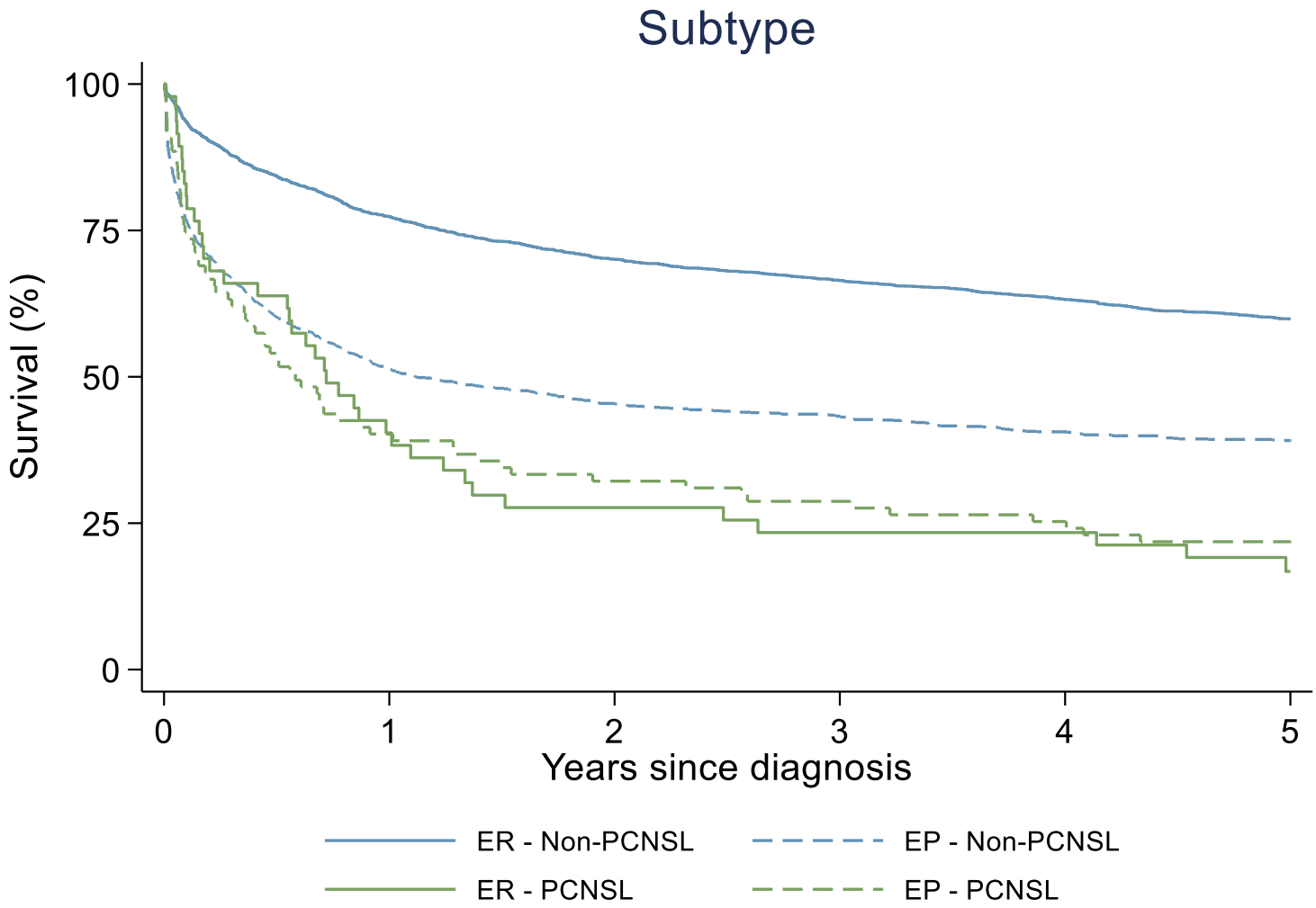


Figure 5:16. Overall survival (OS) stratified by subtype and route to diagnosis.
ER=Elective Referral, EP=Emergency Presentation, PCNSL=Primary central nervous system lymphoma.

The increased risk of death for patients diagnosed as an Emergency was more evident in the first year following diagnosis than for the subsequent years (Table 5:24). After this, survival remained worse than that of patients diagnosed following an Elective Referral, but to a lesser extent and in the fully adjusted model no difference between the diagnostic routes was detected (HR: 0.90 [95% CI: 0.75-1.10]). Patients not treated with curative intent do poorly regardless of their diagnostic route, as illustrated by the steep drop in the survival curves immediately after and up to a year following diagnosis as shown in Figure 5:17. Patients treated with curative intent do much better, although for those diagnosed following an Emergency Presentation the risk of death is still elevated in the first year. However, if patients survive past year one after receiving aggressive treatment, they typically have good outcomes regardless of diagnostic route as illustrated in Figure 5:18.

Table 5:24. Hazard ratios with 95% confidence intervals (95% CI) distributed by treatment intent and route to diagnosis.

		Route to Diagnosis						
		Emergency Presentation		Elective Referrals				
Treatment intent	**Time since diagnosis (years)**	**Total (N)**	**Deaths**	**Total (N)**	**Deaths**	**Unadjusted hazard ratio (95% CI)[a]**	**Adjusted hazard ratio (95% CI)[b]**	**Adjusted hazard ratio (95% CI)[c]**
All patients	0-5	1267	778	2613	1059	2.03 (1.85-2.23)	2.23 (2.02-2.44)	1.34 (1.22-1.48)
	0-1	1267	616	2613	602	1.36 (1.22-1.53)	1.41 (1.26-1.57)	1.20 (1.06-1.35)
	1-5	651	162	2011	457	1.13 (0.94-1.35)	1.27 (1.06-1.52)	0.90 (0.75-1.10)
Curative intent	0-5	965	480	2340	813	1.72 (1.54-1.93)	1.90 (1.70-2.13)	1.20 (1.06-1.35)
	0-1	965	329	2340	393	1.16 (1.00-1.34)	1.20 (1.03-1.39)	1.08 (0.92-1.26)
	1-5	636	151	1947	420	1.13 (0.94-1.37)	1.26 (1.05-1.52)	0.91 (0.75-1.11)

[a]Unadjusted hazard ratios for risk of death with 95% confidence intervals

[b]Hazard ratios for risk of death with 95% confidence intervals adjusted for sex and age.

[c]Hazard ratios for risk of death with 95% confidence intervals adjusted for sex, age, subtype, stage, B-symptoms, performance status and IMD quintile.

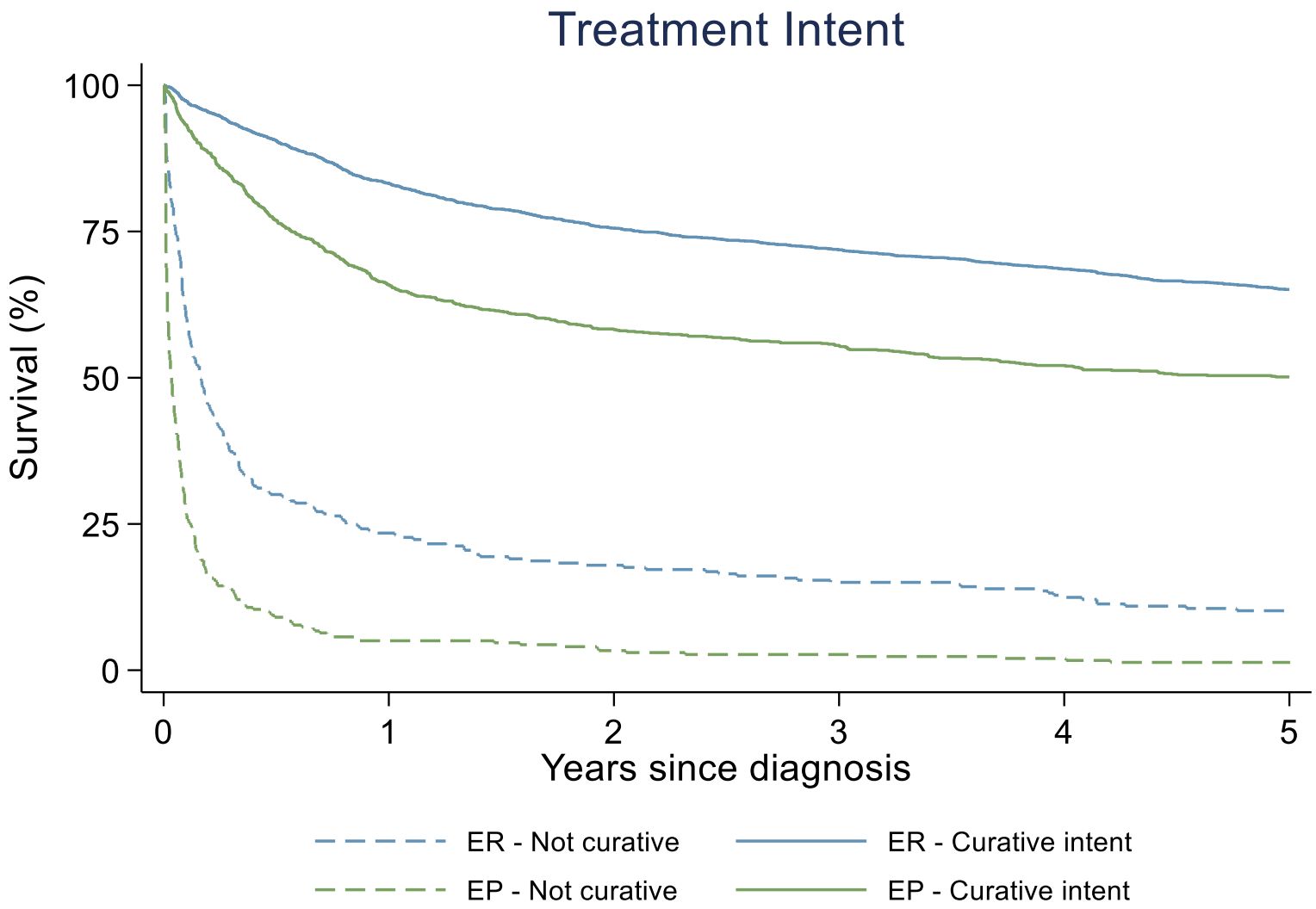


Figure 5:17. Overall survival (OS) stratified by treatment intent and route to diagnosis.

ER=Elective Referral, EP=Emergency Presentation.

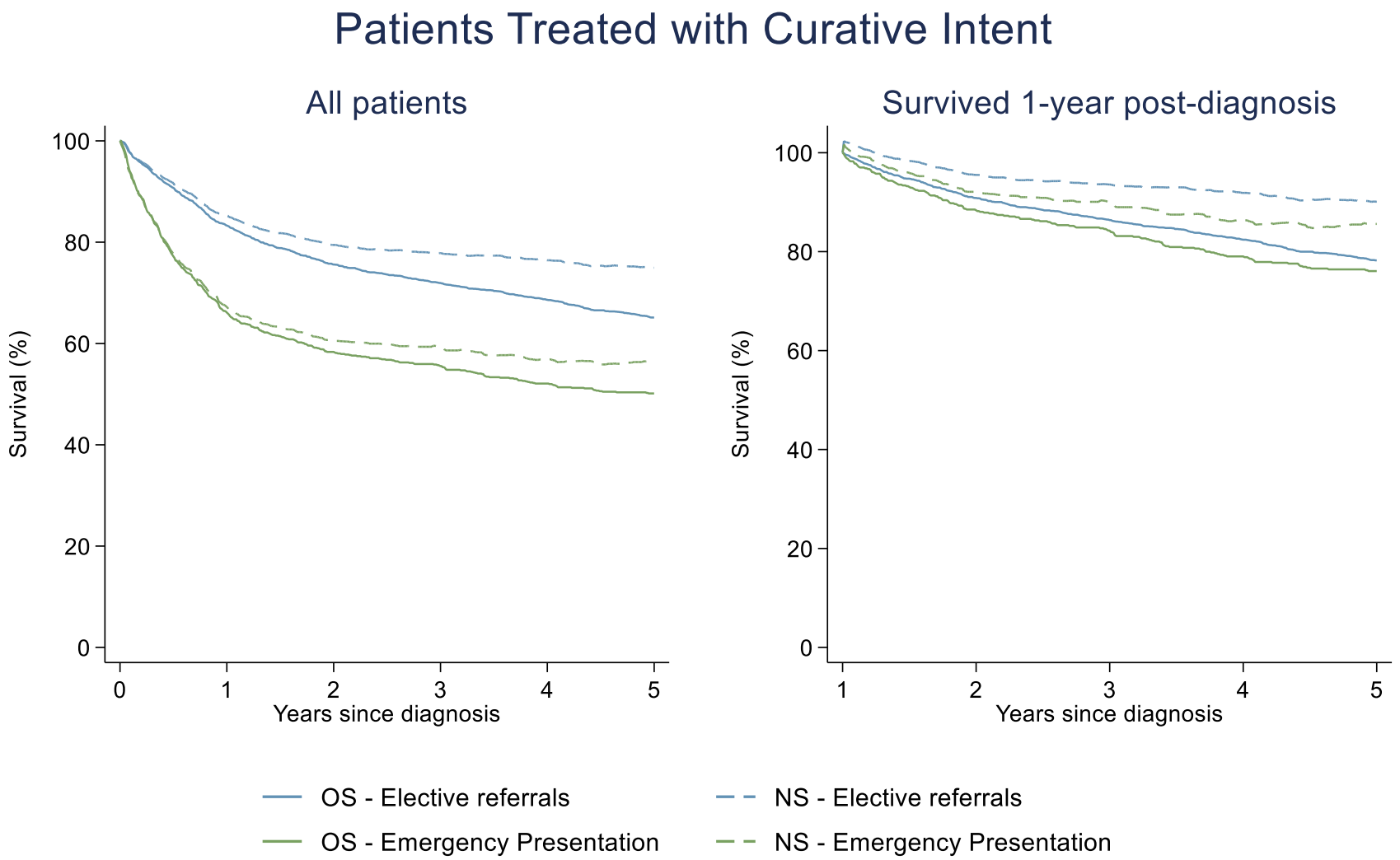


Figure 5:18. Overall survival (OS) and net survival (NS) for patients treated with curative intent stratified by route to diagnosis; all patients and those who survived at least 1-year following diagnosis.

6 Discussion

6.1 Key findings and conclusions

6.1.1 Summary

This thesis examined the impact of travel time and distance to healthcare on the diagnostic route, treatment and survival of patients with diffuse large B-cell lymphoma (DLBCL) in England for the first time. It was shown that journeys to healthcare took four times as long by public transport compared to journeys by car, and that the travel burden of accessing healthcare differed across the Network. Interestingly, the most deprived patients had the shortest journeys to both their General Practice (GP) and diagnostic hospital. The most common route to diagnosis was routine GP referral, closely followed by Emergency Presentation. No association between diagnostic route and travel time or distance was detected and similarly, no differences in treatment intent or survival by travel were revealed. However, those diagnosed as Emergencies were less likely to receive curative treatment and had significantly worse survival than patients diagnosed via Elective Referral routes, and this survival disadvantage was particularly pronounced during the first year following diagnosis. This survival disadvantage could not be explained fully by differences in treatment and was not driven by differences in the age of patients diagnosed as Emergencies.

6.1.2 Associations between travel to healthcare, patients' characteristics and diagnostic route

No differences in median travel distance or time to GP or hospital was detected by diagnostic route or stage at diagnosis (Table 5:3 and Table 5:4). The only other study identified that examined the impact of travel to healthcare and diagnostic delay for DLBCL patients reported a significant difference in the mean distance between home address and tertiary care for patients who experienced delayed diagnosis compared to those who did not (Zurko, Wade and Mehta, 2019). However, the study was small and conducted in the USA where the healthcare system differs to that of the UK's considerably, meaning the findings are not comparable.

Interesting associations between travel and patients' characteristics were observed. For example, median travel times to the GP and hospital were lower for patients living in deprived areas compared to less-deprived areas by both car and public transport. This could be explained by a greater number of GP practices, and the tendency for hospitals to be located in densely populated, inner-city areas where there is typically more deprivation. For example, St James's hospital in Leeds is located in a Lower Super Output Area (LSOA) that ranks in the 10% most deprived in the country (Ministry of Housing, Communities & Local

Government, 2015). However, it is important to consider that geographical proximity does not necessarily equate to shorter journeys, for example, if a patient relies on public or active transport (such as walking and cycling), they may have longer travel times to healthcare than patients who live further away. Median journey times for public transport are considerably longer than the equivalent by car, as illustrated in Table 5:2 which showed that a visit to either the GP or hospital took four times longer on public transport than the equivalent journey by car. Estimates of journey times by active transport such as by bicycle or on foot were not made, but it can be assumed that these would also, on average, take longer. The longer time by public transport could be explained by a number of factors, including: an indirect route, frequent stops and additional time for travel to- and between- stops or stations. Unfortunately, it is not possible to ascertain how patients travelled to access healthcare as this data is not collected by HMRN, which is discussed in more detail in section 6.2.2. There is evidence to suggest that most cancer patients travel by car to access healthcare, for example, a study of cancer patients in Northern England reported that 87% of patients travelled by car (including taxis and hospital cars) and only 5% travelled by bus (Haynes et al., 2006). A more recent report from Healthwatch found that 82% of cancer patients had access to a car to travel to their appointments (including lifts from friends and family) and 9% travelled using the bus or train (Healthwatch England, 2019). However, there is no breakdown of the characteristics of patients using the different transportation modes and so it is unclear if some groups of patients are less likely to have access to a car, such as those with low socioeconomic status.

Another point to consider is that living closer to their registered GP may not mean that patients living in deprived areas can access these services more easily, or get an appointment sooner than patients who live further away, if the availability of appointments is limited at their local surgery. This was demonstrated by findings reported by the Royal College of General Practitioners that showed significant variation in the number of patients per GP across England, and that GPs serving deprived areas have on average 300 more patients each than GPs in affluent areas (Royal College of General Practitioners, 2024). Furthermore, between 2015 and 2022, despite a 20% increase in the total FTE GP workforce, there was a 15% decrease in qualified FTE GPs per 1000 patients, which may make accessing appointments difficult for some patients (Pettigrew et al., 2024).

Notable differences in travel to healthcare also existed for patients with primary CNS lymphoma (PCNSL) who travelled twice as far to access their diagnostic hospital than other DLBCL patients. PCNSL is an aggressive subtype and it is likely that when it is suspected in clinic, patients are referred from their local hospital to specialised centres in the Network,

such as St James's in Leeds, for diagnosis and treatment, resulting in greater median travel distances and times. Interestingly, patients aged under 18 years lived closest to their GP practice, however, they also lived the furthest away from their diagnostic hospital. The longer journey time to hospital is likely explained by all paediatric patients in the Network being placed under the care of a specialist team in Leeds, meaning that many patients will have longer distances to travel to access this service than if they were under the care of their local hospital.

Predictably, differences in median travel times between patients living in urban and rural areas of the Network and by diagnostic hospital existed. This point was further emphasised by heat maps (Figure 5:3) which showed that journeys to nearest hospital were shorter by both car and public transport in urban areas, particularly to the west of the region, illustrating that access to healthcare is not uniform across the Network.

6.1.3 Examining hospital activity and diagnostic route

The hospital activity of DLBCL patients was examined and compared to that of their matched controls for the two-years before and year-after diagnosis to determine if and when activity was elevated compared to a DLBCL-free population (Figure 5:6 and Figure 5:7). Following the National Institute for Health and Care Excellence (NICE) guidelines for suspected cancer, most symptomatic patients should be urgently referred to secondary care by their GP. Therefore, in an ideal scenario where DLBCL patients present to their GP with symptoms and are then swiftly referred to, and diagnosed by, Haematology, we could expect to see most of the increase in hospital activity in the months immediately before diagnosis. This was not the case, and the rate ratio of hospital activity was elevated for two years before diagnosis meaning that some DLBCL patients were attending secondary care more frequently than expected for an extended period of time before diagnosis. If these attendances were related to symptoms of their DLBCL, this represents considerable diagnostic delay for these patients, for example, if patients with vague symptoms are referred to specialities such and Ear, Nose and Throat with swollen lymph nodes, instead of to Haematology. The specialities that patients with DLBCL are referred to in the months before diagnosis could be explored in future work.

When stratified by route to diagnosis (RTD), the Elective Referral group had elevated activity earlier than the Emergency Presentation group, suggesting that patients diagnosed as an Emergency do not present to secondary care earlier than those diagnosed following an Elective Referral, and may even present later. Despite this, for most of the months in the pre-diagnostic period the rate ratios were slightly higher for the Emergency Presentation

group, meaning that Emergency Presenters were attending hospital more frequently for inpatient stays, outpatient appointments, or a combination of these during this time. However, it is important to note that the case-control rate ratios during this time period are low and only reach 1.5 at 12- and four- months pre-diagnosis for those diagnosed following an Emergency and Elective Referral respectively, indicating that hospital activity was not elevated for all patients during this time. Another interesting difference between the pattern of hospital activity of patients diagnosed following an Emergency compared to an Elective Referral is the time when activity peaks. For Elective Referrals, there is a marked increase in activity from six months before diagnosis, which then peaks at diagnosis, before reducing during the post-diagnostic period. This pattern is what we might expect if we assumed that DLBCL patients attend hospital more frequently in the months leading up to and surrounding diagnosis for referrals, investigations and the commencement of treatment, followed by a reduction in attendances as the disease is managed. The starting point of increased hospital activity is not as well-defined for the Emergency Presentation group and does not peak until three months post-diagnosis.

It is important to remember that the monthly rates presented in the results are calculated using averages of activity across all patients and even when the rate is high, some patients may not contribute any activity while others may contribute a lot. Consequently, the pattern of activity displayed in the plots is not necessarily representative of each individual patient's experience. In addition, not all Hospital Episode Statistics (HES) episodes included in the calculation of the rates will be related to the patients' DLBCL, and some will have resulted from interactions with the hospital for other health conditions. The background rate of hospital activity is accounted for by inclusion of the controls (described in 4.3.4) which are matched on age, sex and residency in the study region. Therefore, we can be confident that the higher rates of activity in the cases compared to the controls is likely to be related to their DLBCL and conclude that the prolonged period of elevated case-control hospital activity rate ratios indicates that some DLBCL patients are symptomatic and seeking medical help for an extended period of time before diagnosis and may have experienced avoidable diagnostic delay. However, inclusion of the controls may not account for all background hospital activity as patients are not matched on other characteristics likely to influence their background rate of hospital activity, such as comorbidities.

No differences in RTD by travel were detected and these associations were confirmed when other factors such as patients age and deprivation were adjusted for in the analyses (Table 5:10). This suggests that travel does not act as a barrier to timely diagnosis of DLBCL in this region, which is in keeping with what has previously been found. For example, it has been

reported that no association between rurality and prolonged diagnostic pathway exists for patients with cancer in Scotland (Murchie et al., 2020; Maxwell et al., 2023) and a study of DLBCL patients in Canada reported that travel distance was not predictive of diagnostic delay (Nikonova et al., 2015). However, a study in the USA reported an increased risk of delayed diagnosis for DLBCL patients with increasing travel to tertiary care, although the association was weak (Zurko, Wade and Mehta, 2019). Interestingly, a study of solid tumour cancers in England reported an association with longer journey times to GP and increased risk of Emergency Presentation and Death Certificate Only diagnostic routes (Murage et al., 2019). This study used national data and it is possible that regional variations could explain the differences reported for HMRN patients. Furthermore, the presentation of patients with solid tumour cancers may differ to those with haematological malignancies which could impact associations with travel.

RTD varied by patient characteristics including age, stage at diagnosis and deprivation. For example, patients aged under 40 were more likely to present as an Emergency, with paediatric patients having over two and a half times the risk of Emergency Presentation compared to patients aged 70-79. However, it is important to remember that the Emergency Presentation route is not always a marker of diagnostic delay and for some patients it is the most appropriate diagnostic route (Murchie et al., 2017), which is likely to be the case for younger DLBCL patients. It is also possible that the symptoms of DLBCL are more alarming in younger people compared to older patients, resulting in a higher chance of them being referred to A&E by their GP, or presenting at A&E themselves, even before advanced-stage disease. However, studies have reported that younger adult cancer patients attend their GP more frequently before being referred to secondary care than older patients, suggesting that in many cases cancer is not suspected by GPs in younger patients (Lyratzopoulos et al., 2012). Advanced stage disease at diagnosis was also associated with the Emergency Presentation route, and indicates that this route does detect patients who are likely to have experienced delay during their diagnostic pathway.

The proportion of patients from each Index of Multiple Deprivation (IMD) quintile varied by diagnostic route and a trend for increasing risk of Emergency Presentation with increasing deprivation was observed (Table 5:10). However, this association can be explained by the confounding effect of other variables such as stage at diagnosis and performance status as this association disappeared when other factors were accounted for. This contradicts a report from a large study of DLBCL in England, which found that more deprived patients were more likely to be diagnosed as an Emergency in both univariate and multivariate analysis (Smith et al., 2021b). This study used national data, which includes a high

proportion of ‘not otherwise specified’ cases of lymphoma due to the challenges of data collection and does not include important covariables such as stage at diagnosis, which could explain the differences between the results of this study and the results of this thesis. This highlights the importance of having clinical data, which is not always available in national datasets.

6.1.4 Associations between travel to healthcare and treatment intent

Reassuringly, there was no evidence of an association between travel to healthcare and treatment intent, indicating that the delivery of care is equitable for DLBCL patients across HMRN’s Clinical Network, regardless of travel burden (Table 5:12). This aligns with results from a study of patients with acute leukaemia in Denmark, where longer distance to specialised cancer centres did not impact whether patients received intensive chemotherapy or consolidation with a stem cell transplant after adjusting for confounders (Tøstesen et al., 2019). Conversely, a study in the USA found that increasing distance from the transplant centre was associated with lower chance of receiving a stem cell transplant for patients with lymphoma and leukaemia (Mitchell and Conklin, 2015). However, the standard treatment for DLBCL, R-CHOP, is delivered at all hospitals across the Network, meaning that most patients have short travel times to access treatment, which could explain why an association was not detected.

Patients diagnosed following an Emergency were less likely to receive treatment with curative intent, even after adjusting for baseline characteristics. It is worth noting that the effect of RTD on receiving curative treatment is considerably reduced in the adjusted analysis, indicating that much of this relationship, but not all, can be explained by confounding from other variables, such as age and stage at diagnosis. It is probable that some patients diagnosed via the Emergency RTD will have experienced avoidable delays during their diagnostic pathway and the reduction in their odds of receiving curative treatment illustrates the long-lasting negative impact of diagnostic delay on these patients’ outcomes and the considerable benefit that earlier diagnosis could achieve. Interestingly, patients whose RTD was Unknown were not less likely to receive curative treatment. A possible explanation is that some of these patients were seen privately, although, a more likely explanation is inaccuracies in the HES data. For example, some patients could have been wrongly assigned to the Unknown RTD because their hospital activity in the six months leading up to diagnosis was not recorded in the HES data in error. This raises the issue of data quality and completeness in HES, which is a common concern when using routine datasets for research (Boyd et al., 2017).

No association between socioeconomic status and treatment with curative intent was detected, which is consistent with previous publications using HMRN data (Smith et al., 2015). However, differences were observed by hospital, for example, compared to patients diagnosed at St James' hospital in Leeds, those diagnosed at Hull Royal Infirmary or Scarborough hospital were less likely to be treated curatively, while York patients were more likely. Treatment for DLBCL is nationally standardised and therefore variations in treatment between individual hospitals or clinicians are unlikely to explain this. Instead, it may be explained by a number of other factors, such as differences in the prevalence of comorbidities between the populations that these hospitals serve. For example, it is possible that patients living in Hull or Scarborough and the surrounding areas are more likely to smoke and have higher rates of smoking related comorbidities, such as cardiovascular disease, which may make them unfit for aggressive treatment. Hull and Scarborough are both coastal; coastal areas have been highlighted as having particularly poor health outcomes, including in Professor Chris Whitty's annual report as Chief Medical Officer in 2021 (Department of Health and Social Care, 2021). The reasons for poorer health in coastal areas are complex and include; a higher proportion of retired people, a higher prevalence of smokers, excessive alcohol consumption and concentrated areas of deprivation (Department of Health and Social Care, 2021). For example, at the lower tier local authority level of the 2015 IMD, Scarborough was highlighted as the most deprived area in North Yorkshire and Hull was ranked 3rd most deprived nationally, while York ranked 191st (Ministry of Housing, Communities & Local Government, 2015). The relationship between coastal areas and outcomes of patients with DLBCL could be explored in future work, as the Network covers both coastal and non-coastal areas that can be used to draw meaningful comparisons.

As previously described, IMD is an area-based measure that is susceptible to ecological fallacy (section 1.5.2), which occurs when inferences about individuals are made from aggregate data under the assumption that what is true at the group level is also true for each individual belonging to that group. An example is the assumption that no patient belonging to quintile one is affected by income deprivation, which is unlikely to be true. This point is exemplified by a study that showed that the IMD had low concordance with individual socioeconomic status and in a cohort of cancer patients had a predictive value for individual deprivation only marginally above that of chance (Ingleby et al., 2020). It is therefore possible that the IMD does not fully capture the impact of living in deprivation, resulting in residual confounding and possibly leading to the null result reported between deprivation and treatment intent. Although proxies for low income are used in its calculation, no direct measure of personal income are used and instead the focus is on identifying those who

satisfy means tests to claim benefits and tax credits. However, there may be many people living in financial poverty who do not meet the threshold to successfully claim and therefore may not be included in the numerator of these calculations. Furthermore, for ease of drawing comparisons between groups and to prevent groups sizes becoming too small, IMD ranks were condensed into quintiles in this thesis. In doing so, much of the granularity of the ranks is lost which may dilute any associations with deprivation and explain the null result described in the paragraph above. Finally, the IMD is a measure of deprivation, not affluence, and is better at distinguishing between areas with higher levels of deprivation than those with less deprivation. For example, an area may be assigned a low rank because few people living in it have low incomes, but if there are also few people living in it that have high incomes, it can simultaneously be among the least deprived in the country while not being amongst the most affluent. Therefore, it cannot be assumed that areas with low ranks are affluent and in doing so the IMD fails to fully capture socioeconomic status by only focussing on deprivation. However, with a lack of wide-scale individual level data needed to accurately assess individual socioeconomic status, the IMD provides valuable insights into associations that may otherwise be overlooked.

6.1.5 Associations between travel to healthcare and survival

No statistically significant differences in the survival of patients by travel were detected, except between the low and timed-out categories for travel to hospital by public transport (Figure 5:13 and Table 5:18). However, some differences in the absolute values were observed, for example, patients in the ‘low’ travel category for distance to hospital had five-year net survival of 55.6% compared to 62.6% for the ‘highest’ category. This is interesting, as it suggests that those who live closest to the hospital have a survival disadvantage, which might be explained by a higher prevalence of comorbidities or lower compliance to treatment in these areas, which may be related to socioeconomic status.

There are a limited number of studies that have examined the impact of travel on survival of DLBCL patients, none of which are based in the UK; however, the results of this thesis align with these studies. In Canada, where healthcare is publicly funded and universally available, no effect of travel distance on survival has also been reported (Lee et al., 2014). Similarly, it was reported that distance to treatment facility was not independently associated with survival of patients in the USA (Vargo et al., 2015). Finally, a study of French DLBCL patients reported a U-shaped association between travel to nearest centre and survival, with those in the intermediate travel category having the worst survival (Guyader-Peyrou et al., 2017).

Although no differences were detected between survival of patients by travel, other interesting associations with survival were revealed. As expected, patient characteristics such as age and stage at diagnosis were strongly associated, with younger patients and those with early-stage disease having better survival. No association with deprivation and survival was detected when confounders such as age and subtype were accounted for. This echoes what has previously been reported for HMRN patients (Smith et al., 2015) but contradicts findings of other studies in England (Smith et al., 2021a). A German study that examined the impact of socioeconomic status of patients with DLBCL over three decades reported that survival inequalities disappeared with the introduction of rituximab, suggesting that the lack of association in more recent years can be attributed to all patients receiving standard chemo-immunotherapy regardless of socioeconomic status (Ghandili et al., 2024).

Patients diagnosed following an Emergency Presentation were twice as likely to die within the five years following diagnosis and this difference could not be fully explained by differences in treatment (Table 5:23). Emergency RTD was associated with significantly worse survival for all patients except those with PCNSL. The lack of association between Emergency RTD and survival for PCNSL could be explained by this group already having poor survival and therefore there is little opportunity for diagnostic route to exert an effect, or it may be too small to be detected. However, what seems more likely is that the Emergency Presentation route is the most appropriate diagnostic route for patients with this aggressive subtype. Interestingly, the risk of death associated with the Emergency RTD is much more pronounced in the first year following diagnosis and has a much smaller impact on survival after a year (Table 5:24). DLBCL is potentially curable with intensive chemo-immunotherapy and these results suggest that if patients survive treatment, typically delivered over a six- to eight- month period following diagnosis, the long-term effect of diagnostic route on survival is negligible. This highlights the importance of earlier diagnosis for DLBCL patients, who are frequently diagnosed via the Emergency Presentation route, as this could confer a considerable survival benefit given that patients diagnosed with earlier stage disease typically respond better to treatment and have superior survival.

6.2 Strengths and Limitations

6.2.1 Generalisability

As described previously (section 1.2.2), the sociodemographic structure of the HMRN cohort is broadly representative of the national population with regards to age, sex, ethnicity, urban/rural distribution and levels of deprivation and patients are treated following national guidelines meaning that most findings using the HMRN cohort can be confidently extrapolated to the national population. Furthermore, under section 251 of the NHS Act 2006, all patients diagnosed with a haematological malignancy in the region are included in this population-based study, and consequently the results are not vulnerable to selection bias. In addition to this, healthcare in England is predominantly provided by a single provider, the NHS, meaning that differences in care between different regions should be minimal. Lastly, DLBCL represents a large proportion of haematological malignancy diagnoses, meaning the number of patients included in the analyses in this thesis was large, further adding to the robustness of the data. Therefore, the findings of this thesis most likely reflect the experience of DLBCL patients nationally and may even be applicable to other high-income countries, although care should be taken when extrapolating findings due to different healthcare systems and scales of travel based on the size of the country which limit their generalisability. In other countries such as the USA and Australia, some areas are particularly remote, much more so than in the UK, and patients living in these locations might have to travel hundreds of miles to access specialist cancer services, which is incomparable to remote patients in Yorkshire and the Humber. For example, a study of adolescent and young adult acute leukaemia patients in the USA estimated that almost a quarter would have to travel more than an hour (one-way) to access a specialist cancer centre (Parsons et al., 2024), whereas in HMRN over 95% of DLBCL patients could access their hospital by car in under 40 minutes.

A limitation to the generalisability of this thesis' results is that the analyses were not adjusted for ethnicity, which has been shown to be associated with Emergency Presentation for patients with colorectal cancer in England (Wallace et al., 2014). Attitudes and behaviours relating to help-seeking for cancer symptoms are influenced by ethnic and cultural factors, including stigma and health secrecy, which can contribute to delays during the diagnostic pathway (Scott and Hoskin, 2024). However, although the HMRN cohort is broadly representative of the national population in terms of ethnicity, numbers of non-white patients at the individual subtype level are small and therefore it was not possible to examine the influence of ethnicity on the diagnostic pathway or adjust for it in any of the analyses.

It is also important to note that because many variables are included in these analyses, a standard 5% significance level would provide a significant result by chance for one in every 20 variables tested, meaning results should be interpreted with caution. Furthermore, care must be taken when extrapolating these findings beyond 2019, the cut-off point for inclusion in this analysis, due to the impact of the coronavirus pandemic on patients' diagnostic experience and survival. For example, during the pandemic, routine services including diagnostic services were reduced, which contributed to an increase in diagnostic delay for cancer patients which will likely have an effect on long-term survival, although this may not be fully realised for some time (Maringe et al., 2020; Guven et al., 2024). There have also been considerable changes to how healthcare is delivered following the pandemic, for example, there has been an increase in the number of online GP consultations and a reduction in face-to-face appointments, which could also influence patient's diagnostic pathway, treatment and survival.

6.2.2 Travel calculations

A limitation of the study is that despite calculating travel times by both private car and public transport, we do not know which mode of transport patients actually used to access healthcare for their cancer diagnosis, as this data is not collected by HMRN. If this information was available, patients could be assigned a single travel time to their GP and hospital for the mode of transport actually used, instead of two separate times for car and public transport. Examining the two travel modes individually, an association between decreasing deprivation and increasing travel times was observed. However, as described previously, journeys by public transport take much longer than those made by car and therefore if deprived patients are more likely to use the latter to access healthcare, even if their journey time by the individual modes is lower, their actual journey times could be longer. It is reasonable to assume that some groups of patients, including those living in deprivation and the elderly, are more likely to use public transport and consequently have longer travel times than other groups of patients, even if they live close to healthcare facilities. For example, Age UK reported that about a quarter of bus journeys made by people aged over 65 were for medical appointments (Age UK, 2017). Similarly, approximately 40% of those in the lowest income households do not have access to a car, principally because of affordability (Government Office for Science and Social and Political Science Group, Institute for Transport Studies, University of Leeds, 2019). Therefore, by examining associations between patient outcomes and travel times by public transport and car separately, the true effect of travel on patients' outcomes may be masked. Furthermore, it is possible that by combining the two modes of transport to create a single journey time

variable that more accurately reflects the spectrum of travel times to healthcare, an interaction between travel and deprivation could be detected. This has previously been reported for receipt of active treatments of patients with cancer, whereby deprived patients are impacted more by increased travel than less deprived patients (Crawford et al., 2009). It is important to note that the NHS does offer support for some patients with transport difficulties, for example, through patient transport services for those with poor mobility. In addition to this, there is the Healthcare Travel Cost Scheme that reimburses low-income patients for expenses related to transport to hospital such as bus fares or parking (NHS, 2023). However, not all patients on low incomes are eligible for support and this support does not extend to assistance with the logistics of travel, which can be challenging. Additionally, expenses are claimed for costs that the patient initially must pay, which may not be possible for some.

As opposed to collecting data on patient's actual journeys, travel distances and times were calculated using TRACC (described in section 4.5.1) a travel-time analysis tool, which does not take into account specific travel conditions on a particular day, such as road works, traffic congestion or inclement weather conditions, that could increase journey time. In addition to this, public transport services can be unreliable and infrequent meaning that patients may add extra time to account for late or cancelled services when planning their journey which also is not reflected in the times calculated by TRACC. This could result in an underestimation of journey times, which if random will likely have minimal effect on the results, however if this is not the case and underestimation is more likely to occur for patients with certain characteristics, bias could be introduced. The distance variable represents the shortest route between the patient's address and their healthcare provider using the road network and is therefore less vulnerable to the effect of day-to-day fluctuations in travel conditions. However, despite this, travel time was the preferred measure of travel burden in this thesis, as distance is a simplistic measure and two journeys of the same distance can have considerably different times depending on the topography of the landscape, the road type and more.

The road network data used in this thesis is from 2016 and the public transport data is from 2017 and therefore may not accurately reflect the travel conditions of patients diagnosed before or after these dates. Changes to the road network could have a considerable impact on travel time and/or distance, such as pedestrianisation of city centres areas. However, in the absence of travel data directly collected from patients, these calculations provide good estimates of travel and facilitate examination of its impact on patient outcomes in the Network.

Another limitation of the travel analysis was missing data which can occur for a number of reasons, for example, the patient's home address or GP address was not recorded at diagnosis. Fortunately, this was only the case for a small number of patients. However, a greater number of patients had missing travel data because the calculation could not be completed in TRACC, for example, because it was not possible to complete the journey by public transport within three hours. Therefore, some patients with 'missing' travel data will be those with the longest journeys to healthcare and consequently a category was made for patients whose calculation 'timed-out' so that their outcomes could also be compared.

Finally, the travel burden associated with a diagnosis of, and treatment for, DLBCL cannot be fully captured by a single journey time or distance. For example, the standard treatment for most DLBCL patients is six cycles of chemo-immunotherapy, each cycle lasting 3 weeks, meaning that even if patients have relatively short journeys to healthcare, the cumulative travel burden will be considerable, particularly if complications arise and interim appointments are required. Therefore, to accurately capture the travel burden associated with multiple visits to healthcare over the course of the disease, a cumulative total of travel starting at diagnosis and ending when treatment is complete or the patient dies, may be more appropriate. This cumulative total may differ between patients with different characteristics, such as by age, and impact their outcomes, for example, if complications associated with their treatment are less likely to be detected.

6.2.3 Hospital activity rates

When using HES data to examine diagnostic delay, it is important to remember that HES only records hospital episodes and therefore we can only examine the patient's diagnostic pathway once they have entered secondary care. Using this data alone, delays that occur within the primary care setting, or between primary and secondary care, cannot be estimated. For example, it can be deduced from the HES data what type of referral led to an outpatient appointment, such as two-week wait (TWW), but there is no indication of when this referral was made, as the date attached to the HES episode is for the outpatient appointment, not the referral. There is also no information relating to how long the patient had attended primary care or the number of GP appointments they had before a referral was made. Furthermore, as described previously, all HES activity is included in the calculation of the rates as it is not possible to accurately determine what activity is relevant to the patients DLBCL diagnosis and what is not. Consequently, some irrelevant activity will be included in the estimates, however, the inclusion of the matched controls in this analysis should have accounted for most of this background rate of activity.

6.2.4 Routes to diagnosis

Most studies investigating diagnostic route of cancer patients in England use data from the National Cancer Registration and Analysis Service (NCRAS) which assigns RTD using a combination of routine datasets (Elliss-Brookes et al., 2012). Some of this data is not available to HMRN, including the Cancer Waiting Times (CWT) data, which is an obvious limitation of this thesis. Consequently, the algorithm was modified reflecting data routinely collected by HMRN (described in section 4.4.2) as it was not possible to complete some of the steps in the original algorithm without the National Screening data, CWT data and cancer registry data. For example, Emergency Presentations identified from the HES data that occurred within 28 days before a 'decision to treat date' were prioritised over TWW referrals identified from the CWT data in the original algorithm. A 'decision to treat date' is the formal date when a treatment plan was agreed between the patient and their clinicians and is recorded in the CWT dataset to monitor compliance with treatment time standards. In the original paper, 39% of patients were assigned to a diagnostic route using non-HES data (CWT and Screening data) (Elliss-Brookes et al., 2012). However, despite this limitation, the algorithm which uses HES data supplemented with minimal data from HMRN (including date of sample and date of death) produces broadly similar results to those reported using the original algorithm (Figure 6:1), suggesting that patients' diagnostic route can be accurately inferred from HES data alone (Elliss-Brookes et al., 2012; Smith et al., 2021b). For example, the proportion of patients diagnosed following a TWW referral was 18% in both the original analysis and this thesis, which is surprising, as the method for assigning TWW differed considerably between the two algorithms. It is possible that the absent data sources may be used less frequently to assign diagnostic route for patients with haematological malignancies compared to other cancer types, especially as there is no screening programme for these diseases, meaning that few, if any, DLBCL patients will begin their diagnostic journey this way.

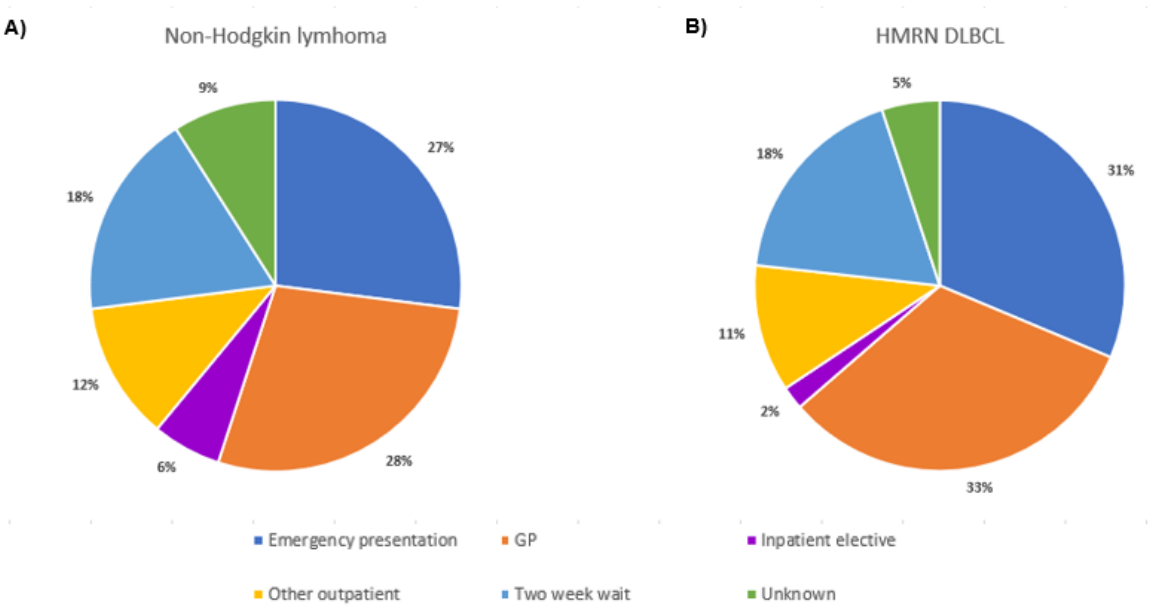


Figure 6:1. A) Percentage of non-Hodgkin lymphoma patients assigned to each route to diagnosis by Elliss-Brookes et al, 2012. B) Percentage of HMRN patients newly diagnosed with diffuse large B-cell lymphoma between January 2005-August 2019 assigned to each route to diagnosis.

It is important to consider that this thesis relies heavily on the accuracy of HES data which is created from data routinely submitted to NHS England from healthcare providers for the purpose of reimbursement for care and commissioning of services, not research. It is assumed that NHS providers accurately record all hospital activity, however as described previously, submissions may be incomplete or contain errors which could result in some patients being assigned to the wrong diagnostic route. For example, when examining patient characteristics by RTD, it was revealed that although the overall percentage of patients allocated to the TWW route was the same for both algorithms, some hospitals in the Network contributed very little to the TWW route (Table 5:9). For example, less than 0.5% of patients diagnosed at York Hospital were assigned to TWW, while almost half were assigned to the routine GP referral route. Some variation in the allocation of RTD can be expected between hospitals, reflecting differences in the characteristics of the patients they serve. However, it can be inferred from the discrepancies in the percentage of patients assigned to the TWW and routine GP referrals route at York compared to other hospitals, that some outpatient appointments at York hospital that occurred following a TWW referral were erroneously recorded as occurring after routine GP referrals in HES. As a result of this concern, diagnostic routes were collapsed into Emergency Presentations and Elective Referrals, which includes all other routes other than Unknown. However, this meant that any

differences in outcomes between patients diagnosed via different Elective Referral routes, such as Inpatient Elective and Other Outpatient, could not be detected. It is possible that such differences exist due to the different composition of patient characteristics by route, such as stage at diagnosis. However, given the small number of patients impacted, it is likely that there would be insufficient power to detect such differences and therefore combining the routes is favourable. The quality of HES data has improved over time (Herbert et al., 2017) and it is possible that the TWW code has been more widely adopted and consequently, if this analysis was repeated with more recent data, comparisons could be made across all diagnostic routes.

Although comparable, there are some differences in the proportions of patients assigned to each RTD, as shown in Figure 6:1. Notably, the Emergency Presentation group is 4% larger in this thesis compared to those reported by Elliss-Brookes et al., 2012 which could be explained by a number of factors. Firstly, the proportions presented in the original paper are for all non-Hodgkin lymphoma (NHL) patients, whereas this analysis was restricted to DLBCL patients. NHL is a heterogenous group of cancers, which includes DLBCL, an aggressive disease, as well as indolent diseases, such as follicular lymphoma. Patients with DLBCL may have increased risk of diagnosis following an Emergency Presentation compared to patients with other types of NHL because the opportunity for diagnosis at an early stage is shorter, explaining the higher proportion of Emergency RTD in this analysis. There is evidence to support this suggestion, as a recent study reported that 34.1% of DLBCL patients were diagnosed as an Emergency compared to 12.3% of those with follicular lymphoma (Smith et al., 2021b). Furthermore, the time period of diagnosis is different for the two samples of patients, 2006-2008 for the original analysis compared to 2005-2019 in this thesis, in which time there have been changes to cancer policy as discussed in section 1.4.2. It seems unlikely that the distribution of patients diagnosed via each route is static over time and it has been reported that across 35 cancer types, Emergency Presentations decreased by 4.7% between 2006-2017 (Herbert et al., 2019b). Considering this trend, we might expect there to be fewer Emergency Presentations, however, this may be counteracted by the more aggressive presentation of DLBCL compared to other NHLs as discussed above. Finally, the patients are from different areas, the original research using national data while patients included in this thesis all reside in the Yorkshire and the Humber region. Although the HMRN population is broadly representative of the national population (section 1.2.2), regional differences in health seeking behaviour could influence diagnostic route. Although there is no screening programme for DLBCL, differences in the uptake of cancer screening programmes between regions provides evidence that differences in health-seeking behaviour exist for other cancers and it is

possible that regional differences in how DLBCL patients interact with the healthcare system also exist (Massat et al., 2015; Hirst et al., 2018). It is also important to note that HMRN patients diagnosed with DLBCL between 2006-2008 (n=832) will be included in both samples, as the Elliss-Brooke et al., 2012 analysis included all patients diagnosed with NHL (n=25,413) in England during this time. This overlap should be considered when drawing comparisons, however, HMRN patients constitute a small percentage (~3.3%) of this national total, so the effect of this is likely to be minimal.

The purpose of the RTD algorithm is to examine how cancer patients enter secondary care. The proportion of patients presenting as Emergencies is frequently used to infer the extent of diagnostic delay, which is avoidable and often attributed to problems in primary care. However, in a sample of Scottish patients with either breast, colorectal, lung, melanoma, prostate or upper GI cancer, only 19% of Emergency Presenters were found to experience missed opportunities for earlier diagnosis in primary care, implying that for many patients the Emergency Presentation route may not signify diagnostic delay (Murchie et al., 2017). As noted, data relating to the primary care portion of the diagnostic route is not available in HMRN and therefore it was not possible to ascertain if DLBCL patients who presented as an emergency experienced delays before entering secondary care. However, most cancer patients initially present to primary care (Mitchell et al., 2015; Abel et al., 2017) and those with haematological malignancies may be particularly vulnerable to delays during this time. For example, a paper comparing the percentage of patients who presented to their GP three or more times before a hospital referral across 24 cancer types reported that this was highest for patients with multiple myeloma at 50% (Lyratzopoulos et al., 2012). Another study reported that over 30% of NHL patients had three or more pre-diagnostic GP consultations and over 10% had five or more (Mendonca, Abel and Lyratzopoulos, 2016). It is important to note that a greater number of GP appointments before diagnosis may not indicate diagnostic delay and could be explained by GPs carrying out appropriate investigations to rule out other causes for symptoms that are vague and common in self-limiting conditions, before referring patients to specialists in secondary care and in doing so fulfil their roles as ‘gatekeepers’. This explanation was proposed by authors of a study in Scotland that found that rural cancer patients have more GP consultations and investigations before they are referred to secondary care than their urban counterparts (Maxwell et al., 2023). It is also important to remember that for some acute cancers with sudden onset of symptoms, such as acute leukaemia, the Emergency Presentation route is likely to be the most appropriate diagnostic route (McPhail et al., 2013) and diagnosis via other routes may be more indicative of delay and result in extremely poor outcomes for these patients.

The limitations of the original algorithm should also be addressed, many of which also apply to the algorithm used in this thesis, such as the assumption that all hospital activity in the six months before diagnosis is related to the patient's cancer, which may not be the case. For example, during the six months leading up to diagnosis a patient may be routinely referred to secondary care by their GP for symptoms of their cancer, resulting in their diagnosis a few weeks later. However, if sometime during the 28 days before the patient is diagnosed, they were admitted to hospital following an attendance at A&E for something unrelated, such as an injury sustained in a road traffic accident, the patient would be assigned to the Emergency Presentation RTD despite the GP route being more appropriate for their cancer diagnosis. However, with the data available in HES it is not possible to create an algorithm that can be applied to large datasets that is capable of discerning what activity is related to the cancer diagnosis and filtering out irrelevant episodes. Primarily, most outpatient records do not have diagnostic coding available and even if they did it is unlikely that these could be used to identify episodes of interest because by definition pre-diagnostic hospital episodes occur before a formal diagnosis and therefore are unlikely to include the relevant code for DLBCL (Elliss-Brookes et al., 2012). It was estimated by the authors of the paper where the algorithm was published, that under the assumption that HES activity relating- and not relating- to the cancer diagnosis were equally likely to be detected by the algorithm, the resulting error rate in assigning RTD was not likely to exceed 10% (Elliss-Brookes et al., 2012). This means that if the algorithm estimated that 27% of patients were diagnosed as an Emergency Presentation, we can be reasonably confident that no more than 2.7% of these patients were wrongly assigned to this route. However, certain groups of patients with higher background hospital activity rates, such as older patients and those with comorbidities, may be at increased risk of being assigned to an 'incorrect' RTD. Furthermore, the RTD algorithm only uses activity in the six months leading up to diagnosis, with a particular focus on activity occurring 28 days before diagnosis, to assign a diagnostic route. There is evidence that some patients with DLBCL have elevated hospital activity for two years before diagnosis (section 5.3) and consequently, using the original algorithm may truncate the diagnostic pathway of many DLBCL patients and may not be appropriate for describing their diagnostic route.

Overall, the RTD algorithm provides valuable insights into DLBCL patients' interactions with the secondary care system before diagnosis and highlights patients who likely experienced delays in their diagnosis.

6.3 Implications and recommendations for future work

This thesis has contributed novel findings regarding the impact of travel on DLBCL patients' RTD and survival. DLBCL patients are vulnerable to delayed diagnosis which negatively impacts their survival; however, these results suggest that travel to healthcare does not act as a barrier to earlier diagnosis and better survival for these patients. This is in keeping with the small number of studies that examined the impact of travel on diagnostic route and survival of patients with DLBCL (Guyader-Peyrou et al., 2017; Lee et al., 2014; Nikonova et al., 2015; Vargo et al., 2015; Zurko, Wade and Mehta, 2019). These studies were based in Canada, the USA and France, which have different healthcare systems and geographical size to the UK. For example, Canada and the USA are much vaster than the UK and France, and the healthcare systems in the UK, Canada and France are publicly funded while in the USA this is not the case. This suggests that compared to other cancers, outcomes of patients with DLBCL may be less susceptible to differences in travel, regardless of setting and these results can be confidently extrapolated to DLBCL patients nationally, if not to other developed countries too.

These findings may also be applicable to patients with other haematological malignancies living in the UK, with journeys to access care likely to be comparable to those of DLBCL patients in HMRN. A notable exception to this is Scotland, where some patients reside in much more remote locations such as offshore islands including Orkney and Shetland and face lengthy journeys, including a ferry crossing or a flight, to access specialist cancer services on the mainland. It has been reported that patients with other cancers who live on Scottish islands have inferior survival compared to mainland patients, which can be explained by fewer outpatient appointments following diagnosis (Turner et al., 2023). For most cancer patients in the UK, journeys to access care, even if to a specialist centre, will be relatively short and likely similar to what has been reported in this thesis. DLBCL is an aggressive disease, which may limit the opportunity for travel to exert an effect on the diagnostic pathway and survival of these patients and therefore it is possible that an effect of travel on long-term survival might be more pronounced for patients with less aggressive, chronic diseases that are incurable and require long-term management, such as follicular lymphoma. Although, it is also likely that any association detected between increased travel and delayed diagnosis for more indolent subtypes may have less impact on survival. This is because these patients are often placed on active monitoring following diagnosis and are only treated once the disease advances and therefore prompt treatment is not as crucial as it is for patients with aggressive diseases like DLBCL (Dreyling et al., 2021).

The work in this thesis could also be expanded to include treatment decisions made by DLBCL patients that may be influenced by travel, such as whether those who are not cured by first-line treatment with R-CHOP (relapsed and refractory disease) decide to travel for stem cell transplants as recommended in the treatment guidelines (Chaganti et al., 2025). These patients may be very unwell and could face long journey times because these treatments are only delivered at highly specialised centres, and so decline the transplant because of this. Another decision that may be influenced by travel is whether patients recommended for consolidation radiotherapy, such as those with limited stage disease who have not received full-dose chemo-immunotherapy, agree to it. Consolidation radiotherapy is given after a patient has received R-CHOP to remove any remaining cancer cells and reduce the risk of recurrence. Refusing this treatment could negatively impact the chance of achieving long-term disease-free survival, although the benefits of consolidation radiotherapy are debated, with some studies reporting an improvement to survival while others report no benefit (Berger et al., 2020; Campbell, 2013; Grass et al., 2019). It has been reported that the likelihood of receiving radiotherapy decreases with increasing travel time to nearest radiotherapy centre for cancer patients in Northern England (Jones et al., 2008a) and a public consultation by NHS England revealed travel times as the most important concern for patients (NHS England, 2018). Furthermore, a study based in the USA reported that DLBCL patients who travelled over 30 miles to their treatment facility were 25% less likely to receive multiagent chemotherapy with radiotherapy compared to chemotherapy alone than patients who lived 30 miles or closer (Vargo et al., 2015). In the HMRN region, radiotherapy is offered at two centres, St James' in Leeds and Castle Hill in Hull, meaning that the travel time to this treatment will be considerable for many patients and is likely to exceed the 45-minute maximum target for radiotherapy services suggested by the National Radiotherapy Advisory Group (National Radiotherapy Advisory Group, 2007). It is plausible that patients may decline radiotherapy based on journey times, especially as it is delivered on consecutive days over several weeks, and there is uncertainty surrounding the benefits of the treatment.

The results of this thesis may not fully represent the current situation for patients with DLBCL in England, as there have been substantial changes to healthcare during the timeframe of this work and after 2019, the cut-off point for inclusion in this thesis. For example, more GP surgeries are merging as groups, such as York Medical Group which is made up seven surgeries across the city (York Medical Group, 2025). Patients may be offered appointments at any surgery in the group as opposed to their registered surgery and as a result, may need to travel further to access appointments. There are also the immediate and long-term impacts of the coronavirus (COVID) pandemic on the delivery of healthcare that should be

considered. The pandemic began in March 2020 and in response to this an already struggling NHS responded by cancelling all but the most urgent non-COVID care, including elective surgeries and cancer treatments, to cope with the increased demand to the service during the initial ‘waves’ of the pandemic (McCabe et al., 2020; British Medical Association, 2024). In addition to the impact of cancelled appointments, social distancing and travel restrictions made it difficult for some patients to arrange travel to access health services during this time (NHS England, 2021). These short-term effects of the pandemic on the healthcare system are likely to have caused considerable disruption to the diagnostic pathway, treatment, and consequently survival of cancer patients, including those with DLBCL. It is also important to note that patients with blood cancers are often immunocompromised due to the dysregulation of immune cells related to the disease and the effects of anti-cancer treatments that target these cells, making them particularly vulnerable to severe infection with COVID (Buske et al., 2022). Furthermore, as well as the initial impact of the pandemic, there have been long-term effects on the delivery of care as the NHS has struggled to recover from the disruption caused by COVID, for example, since the start of the pandemic there has been a 66% increase in elective treatment waiting lists (Nuffield Trust, 2024). Therefore, it is likely that the pandemic will have affected the way in which DLBCL patients enter the secondary care system, the proportion who experience diagnostic delay and their survival, and these effects may persist for years to come.

There have also been some positive changes to the healthcare system since 2019, including the introduction of the Faster Diagnosis Standard and Rapid Diagnostic Centres, as part of the NHS Long Term Plan (NHS, 2019). The Faster Diagnosis Standard applies to patients who: are referred by their GP on a suspected cancer pathway, are referred by their GP with breast symptoms where cancer is not initially suspected and those referred by the National Screening Service with an abnormal screening result. Under this standard, three-quarters of these patients should either have cancer diagnosed or eliminated within 28 days of being referred, meaning that patients with cancer can start treatment sooner and those without cancer are not put under prolonged, unnecessary stress while waiting for their result. This is intended to contribute towards achieving early diagnosis targets set out in the NHS Long Term Plan (2019), including that by 2028 75% of cancer patients will be diagnosed at stages I or II and 55,000 additional patients each year will survive five-years or more after their diagnosis. Rapid Diagnostic Centres are specialised centres designed to rapidly investigate and diagnose patients with serious conditions and include a novel Non-Specific Symptoms Pathway for patients at risk of being diagnosed with cancer, but who do not fulfil the current guidelines for an urgent referral. This offers a new diagnostic pathway for patients with haematological malignancies, many of whom present with non-specific symptoms such as

weight loss and fatigue and previously may not have been eligible for an urgent referral for suspected cancer. It is possible that these new centres will decrease the time between first-presentation and diagnosis for DLBCL patients and reduce their vulnerability to diagnosis following Emergency Presentation, which could also improve survival. In light of these changes to the healthcare system in recent years, it would be interesting to repeat these analyses in this thesis for patients diagnosed after 2019 to examine any changes in the presentation and survival of DLBCL patients over time.

It would also be interesting to examine the potential impact of increased centralisation of services in the future. The centralisation of services poses a dilemma in that the aim is to provide equitable high-quality care, but in doing so, inequalities in accessibility may be exacerbated by increasing travel times for some patients. Since the Calman-Hine report of 1995, which aimed to improve cancer outcomes through radical reform including reorganising services and creating specialist multidisciplinary teams, there has been a shift towards a highly specialised, centralised cancer service in the UK (Calman KC and Hine D, 1995). Further support for increased centralisation comes from the observation of many studies based in the USA that cancer patients treated at high-volume hospitals have better outcomes than those treated at low-volume hospitals, the results of which are likely to be confounded by factors such as the characteristics of patients who choose to travel further for care (Crawford, 2022). Finally, the centralisation of services is attractive to governments as healthcare resources are finite and economies of scale mean that reconfiguring services into fewer, specialised centres could generate cost savings (Crawford, 2021), although, robust evidence of cost savings from centralisation is limited (Bhattarai et al., 2016). Therefore, it is important to establish that the negative impact of increased travel associated with centralisation of services does not outweigh the benefit of improved quality of specialised services. No difference in outcomes of DLBCL patients in HMRN by travel to healthcare were detected in this thesis, however, the standard treatment during this time period, R-CHOP, is offered across the Network. The treatment guidelines have recently been updated, including R-CHOP being replaced with Pola-R-CHP as first-line treatment for patients with a prognostic score of 2 or more and the use of bispecifics such as glofitamab for patients who are unfit for stem cell transplants (Chaganti et al., 2025; Fox et al., 2024). There is also an increase in the use of new specialised treatments as second-line treatment that are only delivered at highly specialised cancer centres, for example, chimeric antigen receptor T-cell (CAR-T) therapy which is recommended for patients who have relapsed or refractory disease within 12 months of completing first-line treatment. With these changes to standard treatments and where they are delivered, any associations between travel, treatment and survival may change as a consequence.

6.4 Conclusions

The overall aim of this thesis was to examine the relationship between travel to healthcare and diagnostic route and survival of patients with DLBCL in the Yorkshire and the Humber region. This is the first time that the impact of travel on diagnostic route and survival of DLBCL patients has been estimated in England and the results suggest that there is no association between travel and patients' outcomes. These findings are reassuring, as it indicates that patients receive equitable care from the NHS regardless of how close they live to healthcare services.

Appendix 1: Route to diagnosis algorithms

Flow charts taken from: Elliss-Brookes, L. et al. (2012). Routes to diagnosis for cancer – determining the patient journey using multiple routine data sets. *British Journal of Cancer*, 107 (8), pp.1220–1226. [Online]. Available at: doi:10.1038/bjc.2012.408.

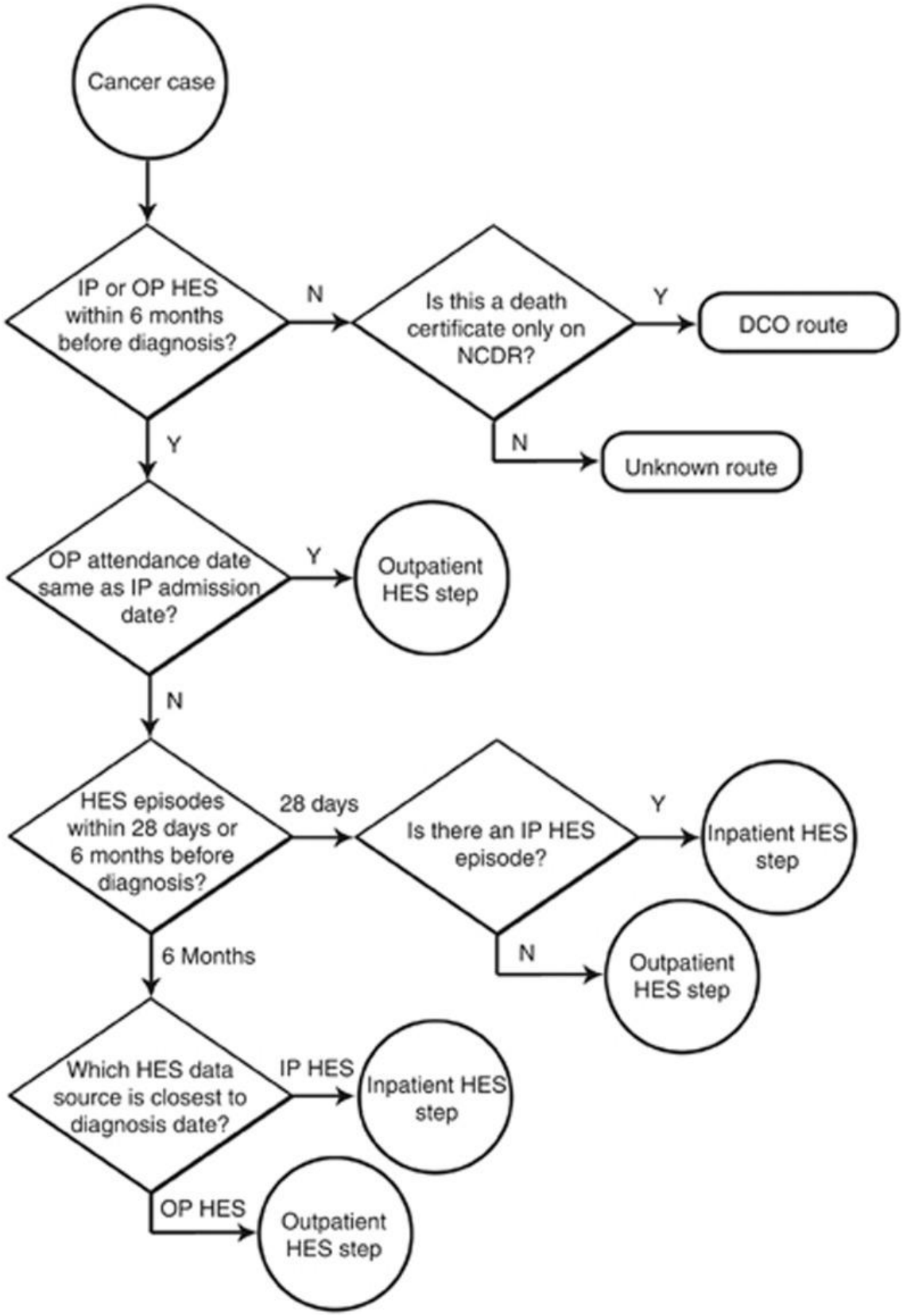


Figure A:1. Flow diagram for allocating the end point of the route using inpatient and outpatient data.

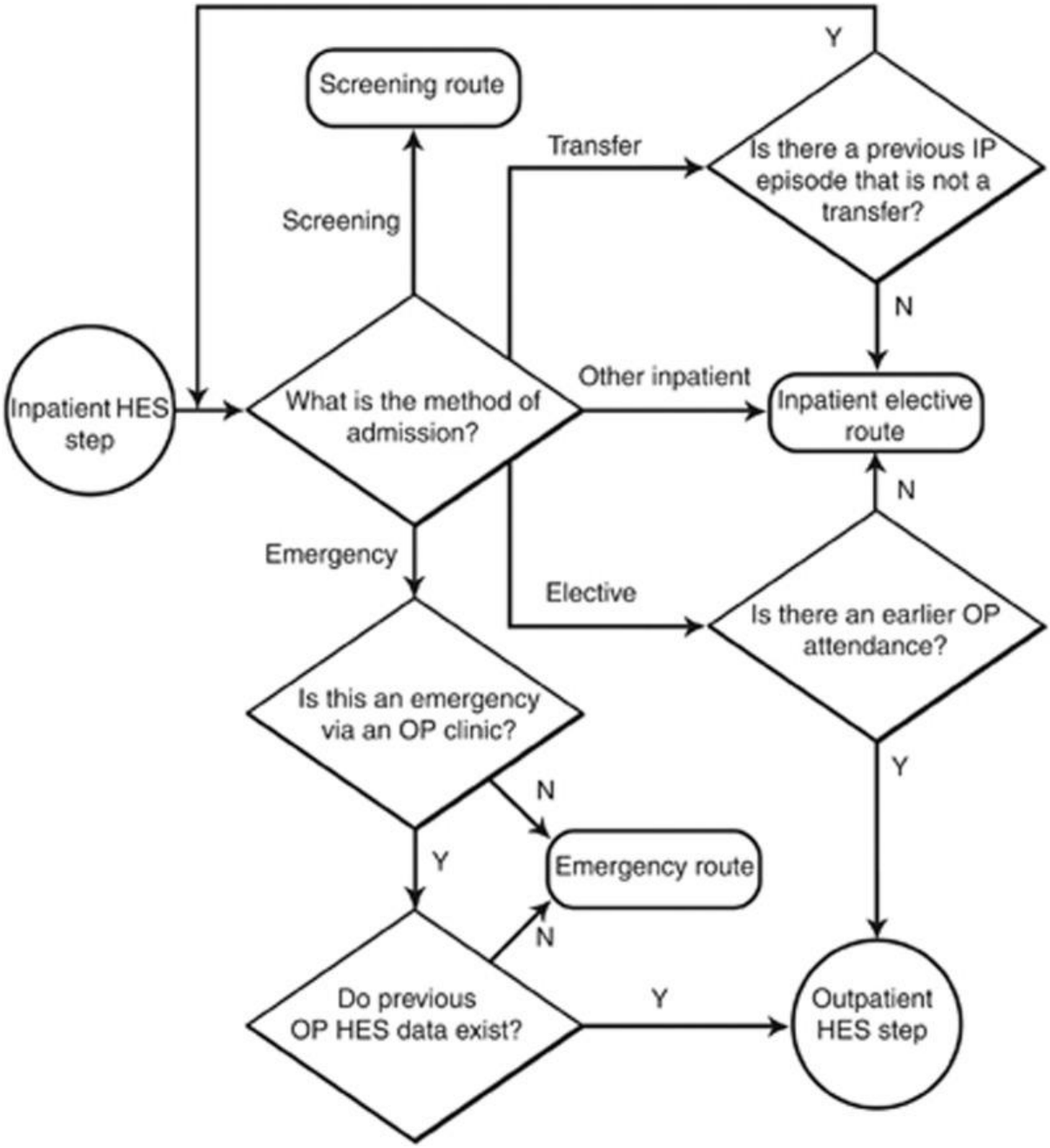


Figure A:2. Flow diagram for finding the start point or prior step for an inpatient step in a route.

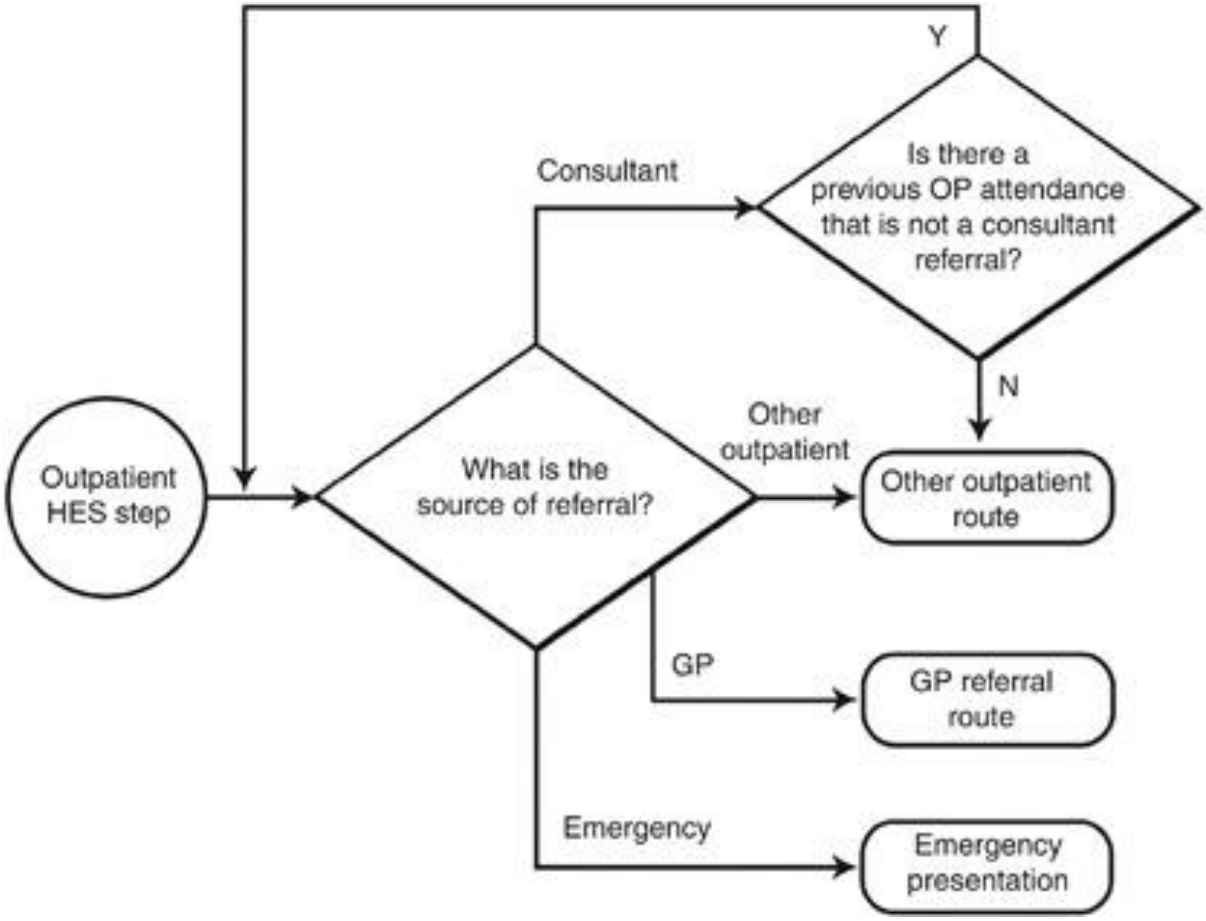


Figure A:3. Flow diagram for finding the start point or prior step for an outpatient step in a route.

Appendix 2: Stata code for assigning treatment ‘Treatline’

```
program define TreatLine2

syntax varname ,ORDERby(varname) [keep(string) drop(string) PREinterval(integer 30)
POSTinterval(integer -1) SEARCHagent(string) GENerate(string) SINgleagent COMbinedagent
ADDitive STRing]

capture mkdir "Data Files"

if "`searchagent'" != "" {

	if "`generate'" == "" {

		di as error`"Option 'generate' required when 'searchagent' is specified"'

		exit

	}
}

if "`additive'" != "" {

	use "Data Files\TreatmentFormatAll.dta", clear
}

local preinterval = 0 - `preinterval'
di `preinterval'

replace RegimenType    = RegimenTypeHILIS    if RegimenType == ""
replace RegimenGroup  = RegimenTypeHILIS    if RegimenGroup == ""
replace RegimenName  = RegimenTypeHILIS    if RegimenName == ""
replace RegimenAgents          = RegimenTypeHILIS    if RegimenAgents == ""
replace RegimenHILIS   = RegimenTypeHILIS    if RegimenHILIS == ""
save "Data Files\TreatmentFormatAll.dta", replace

di "filter on DateOfDiag"
replace start_date = end_date if (start_date == . | start_date == date("1/1/1000", "MDY")) &  end_date
!= .
drop if (start_date - DateOfDiag) < `preinterval' & start_date != date("1/1/1000", "MDY")
drop if (start_date > DateOfDiag2) & (DateOfDiag2 != .) & ((DateOfDiag2 - DateOfDiag) > 60)
encode hospital, gen(TreatHosp) label(TreatHosp)
encode response, gen(Response) label(Response)

if `postinterval' != -1 {

	drop if start_date - DateOfDiag > `postinterval'

}

di "Filtered"
```

```
noisily {
di "`varlist'"
local var = "`varlist'"
if "`keep'" != "" {
        tokenize "`keep'", parse(",")
        local step = 1
        local teststring = `"keep if `var' == "``step'""'
        di `"`teststring'"'
        local step = 2
        *macro list
        while "``step'" != "" {
                local teststring = `"`teststring' | `var' == "``step'""'
                di `"`teststring'"'
                local ++step
        }
}
else if "`drop'" != "" {
        tokenize "`drop'", parse(",")
        local step = 1
        local teststring = `"drop if `var' == "``step'""'
        di `"`teststring'"'
        local step = 2
        while "``step'" != "" {
                local teststring = `"`teststring' | `var' == "``step'""'
                di `"`teststring'"'
                local ++step
        }
}
`teststring'
```

```
pause

marksample touse

if "`additive'"=="additive" {

        do "Data Files\TreatmentLabels.do"

        *pause "adding now!"

}
else {
                preserve

                        keep EGU_ID

                        duplicates drop

                        sort EGU_ID

                        gen DeathLogged = 0

                        save "Data Files\TreatmentPathway.dta", replace


                        save "Data Files\TreatmentAgentUse.dta", replace


                restore
}

encode RegimenType, gen(AllRegimenType) label(RegimenType)
encode RegimenGroup, gen(AllRegimenGroup) label(RegimenGroup)
encode RegimenName, gen(AllRegimenName) label(RegimenName)
encode RegimenAgents, gen(AllRegimenAgents) label(RegimenAgents)
encode RegimenName, gen(AllTreatment) label(Treatment)
encode RegimenType if AllTreatment == . , gen(AllTreatType) label(Treatment)
encode RegimenTypeHILIS, gen(AllRegimenTypeHILIS) label(RegimenTypeHILIS)
encode RegimenHILIS, gen(AllRegimenHILIS) label(RegimenHILIS)
label save RegimenType RegimenGroup RegimenName RegimenAgents Treatment
RegimenTypeHILIS RegimenHILIS Response TreatHosp using "Data Files\TreatmentLabels.do",
replace

drop AllRegimenTypeHILIS AllRegimenType AllRegimenGroup AllRegimenName AllRegimenAgents
AllTreatment AllRegimenHILIS TreatHosp Response
di "check labels"

sort EGU_ID `orderby' tx_id
by EGU_ID (`orderby' tx_id):  gen tx_no = _n

tab tx_no
local x=r(r)
di `x'
quietly: sum tx_no
scalar define max_no = `r(max)'
```

```
pause
save "Data Files\TreatmentFormat.dta", replace

forvalues i=1 (1) `x'  {

	keep if tx_no == `i'

	di `i'

	local name=string(`i')
	di "name assigned"

	encode RegimenTypeHILIS, gen(RegTypeHILIS) label(RegimenTypeHILIS)
	di "type encoded"
	encode RegimenType, gen(RegType) label(RegimenType)
	di "RegimenType encoded"
	encode RegimenGroup, gen(RegGroup) label(RegimenGroup)
	di "RegimenGroup encoded"
	encode RegimenName, gen(RegName) label(RegimenName)
	di "RegimenName encoded"
	encode RegimenAgents, gen(RegAgents) label(RegimenAgents)
	di "RegimenAgents encoded"
	encode RegimenName, gen(Treatment) label(Treatment)
	di "RegimenName encoded to treatment"
	encode RegimenType if Treatment == . , gen(TreatType) label(Treatment)
	di "RegimenType encoded to treatment"
	replace Treatment = TreatType if Treatment == .
	di "Treatment completed"
	encode RegimenHILIS, gen(RegHILIS) label(RegimenHILIS)
	encode response, gen(Response) label(Response)
	encode hospital, gen(TreatHosp) label (TreatHosp)
	di "----------------------"
	di "All Variables Encoded"
	di ""
	di ""

	if "`searchagent'" != "" {

		local keepstring = `"EGU_ID `generate' `generate'Date"'

		tokenize `searchagent'

		local step = 1

		local teststring = `"RegimenType == "``step'"" | strmatch(RegimenAgents,"*``step'"*")"'

		di `"`teststring'"'

		local step = 2

		di "step -> 2"

		di `2'

		*macro list
```

```
while "``step"" != "" {

        di "teststring loop start"

        local teststring = `"`teststring' | RegimenType == "``step"" |
        strmatch(RegimenType,"*``step"*")"

        di `"`teststring""

        local ++step

}

di "teststring loop passed"

di `"`teststring""


gen `generate' = 1 if `teststring'

di "Search agent generated"


gen `generate'Date = start_date if `generate' == 1

di "Search agent date generated"


if "`singleagent"' != "" {

        local teststring = `"RegimenType == "``step"""

        di `"`teststring""

        local step = 2

        while "``step"" != "" {

                local teststring = `"`teststring' | RegimenType == "``step"""

                di `"`teststring""

                local ++step

        }

        gen "`generate'Single" = 1 if `"`teststring""
        gen "`generate'SingleDate" = start_date if "`generate'Single" == 1

        local keepstring = `" `generate'Single `generate'SingleDate"'

}

if "`combinedagent"' != "" {
```

```
                        local teststring = `"(RegimenType != "``step"" &
strmatch(RegimenType,"*``step"*"))"'

                        di `"`teststring""'

                        local step = 2

                        while "``step"" != "" {

                                local teststring = `"`teststring' | (RegimenType != "``step"" &
strmatch(RegimenType,"*``step"*"))"'

                                di `"`teststring""'

                                local ++step

                        }

                        di "combined teststring completed"

                        gen "`generate'Combined" = 1 if `"`teststring""'
                        gen "`generate'CombinedDate" = start_date if "`generate'Combined" == 1

                        local keepstring = `" `generate'Combined `generate'CombinedDate"'

                }

                preserve

                        keep `keepstring'

                        save "Data Files\TreatmentAgentElement.dta", replace

                        use "Data Files\TreatmentAgentUse.dta", clear

                        merge 1:1 EGU_ID using "Data Files\TreatmentAgentElement.dta", update

                        drop _merge

                        save "Data Files\TreatmentAgentUse.dta", replace

                restore

        }

        gen post = start_date - DateOfDiag

        gen trial = 1 if RegimenTypeHILIS == "clinical trial"

        keep EGU_ID RegType RegGroup RegName RegAgents Treatment RegTypeHILIS
        RegHILIS Response TreatHosp start_date end_date cycles post trial EventDate EventType

        renvars RegType, postfix(`name')
        renvars RegGroup, postfix(`name')
        renvars RegName, postfix(`name')
```

```
        renvars RegAgents, postfix(`name')
        renvars Treatment, postfix(`name')
        renvars RegTypeHILIS, postfix(`name')
        renvars RegHILIS, postfix(`name')
        renvars Response, postfix(`name')
        renvars TreatHosp, postfix(`name')
        renvars start_date, postfix(`name')
        renvars end_date, postfix(`name')
        renvars post, postfix(`name')
        renvars cycles, postfix(`name')
        renvars trial, postfix(`name')
        *renvars AML_Induction, postfix(`name')

        sort EGU_ID
        pause "check first diag"
        save "Data Files\TreatmentElement.dta", replace

        use "Data Files\TreatmentPathway.dta", clear
        qui{
        do "Data Files\TreatmentLabels.do"
        }

        sort EGU_ID
        merge 1:1 EGU_ID using "Data Files\TreatmentElement.dta", update
        tab RegType`i' EventType,miss
        replace RegType`i' = 999 if RegType`i' == . & DeathLogged == 0 & EventType == "D"
        replace DeathLogged = 1 if RegType`i' == 999
        replace RegGroup`i' = 999 if RegType`i' == 999
        replace RegName`i' = 999 if RegType`i' == 999
        replace RegAgents`i' = 999 if RegType`i' == 999
        replace Treatment`i' = 999 if RegType`i' == 999

        capture {
        label define RegimenType 999 "Died", add
        label define RegimenGroup 999 "Died", add
        label define RegimenName 999 "Died", add
        label define RegimenAgents 999 "Died", add
        label define Treatment 999 "Died", add
        }

        drop _merge
        *local p = `i' -1
        *replace RegType`p' = RegType`p' + " + radiotherapy" if RegType`i' == "chemoptherapy"
        sort EGU_ID
        quietly: compress
        save "Data Files\TreatmentPathway.dta", replace

        use "Data Files\TreatmentFormat.dta", clear
        qui{
        do "Data Files\TreatmentLabels.do"
        }
}

}

if "`string'" == "string"{
```

```
use "Data Files\TreatmentPathway.dta", clear
*Rename variables to make it clear these are the treatment variables
*Treatment#
forvalues i = 1 (1) `=scalar(max_no)'{
	decode Treatment`i', gen(temp)
	drop Treatment`i'
	rename temp Treatment`i'
}
*RegName#
forvalues i = 1 (1) `=scalar(max_no)'{
	decode RegName`i', gen(temp)
	drop RegName`i'
	rename temp RegName`i'
}
*RegGroup#
forvalues i = 1 (1) `=scalar(max_no)'{
	decode RegGroup`i', gen(temp)
	drop RegGroup`i'
	rename temp RegGroup`i'
}
*RegType#
forvalues i = 1 (1) `=scalar(max_no)'{
	decode RegType`i', gen(temp)
	drop RegType`i'
	rename temp RegType`i'
}
*RegAgents#
forvalues i = 1 (1) `=scalar(max_no)'{
	decode RegAgents`i', gen(temp)
	drop RegAgents`i'
	rename temp RegAgents`i'
}
*HILISRegimen#
forvalues i = 1 (1) `=scalar(max_no)'{
	decode RegHILIS`i', gen(temp)
	drop RegHILIS`i'
	rename temp RegHILIS`i'
}
*HILISType#
forvalues i = 1 (1) `=scalar(max_no)'{
	decode RegTypeHILIS`i', gen(temp)
	drop RegTypeHILIS`i'
	rename temp RegTypeHILIS`i'
}
*Response*
forvalues i = 1 (1) `=scalar(max_no)'{
	decode Response`i', gen(temp)
	drop Response`i'
	rename temp Response`i'
}
*TreatHosp*
forvalues i = 1 (1) `=scalar(max_no)'{
	decode TreatHosp`i', gen(temp)
	drop TreatHosp`i'
	rename temp TreatHosp`i'
}
```

```
	quietly: compress
	save "Data Files\TreatmentPathway.dta", replace

}

use "Data Files\TreatmentPathway.dta", clear
reshape long start_date end_date cycles post trial Treatment RegName RegGroup RegType
RegAgents RegHILIS RegTypeHILIS Response TreatHosp, i(EGU_ID) j(tx_no) favor(memory)
if "`string'" == "string"{
	drop if Treatment==""
}
else{
	drop if Treatment==.
}
quietly: compress
save "Data Files\TreatmentPathwayLong.dta", replace

end
```

References

Abdel-Rahman, M. et al. (2009). What if cancer survival in Britain were the same as in Europe: how many deaths are avoidable? *British Journal of Cancer*, 101 (2), pp.S115–S124. [Online]. Available at: doi:10.1038/sj.bjc.6605401.

Abdel-Rahman, M. E. et al. (2014). No socioeconomic inequalities in ovarian cancer survival within two randomised clinical trials. *British Journal of Cancer*, 111 (3), pp.589–597. [Online]. Available at: doi:10.1038/bjc.2014.303.

Abel, G. A. et al. (2015). Cancer-specific variation in emergency presentation by sex, age and deprivation across 27 common and rarer cancers. *British Journal of Cancer*, 112 (1), pp.S129–S136. [Online]. Available at: doi:10.1038/bjc.2015.52.

Abel, G. A. et al. (2017). Emergency diagnosis of cancer and previous general practice consultations: insights from linked patient survey data. *Br J Gen Pract*, p.bjgp17X690869. [Online]. Available at: doi:10.3399/bjgp17X690869.

Afshar, N. et al. (2020). Differences in cancer survival by remoteness of residence: an analysis of data from a population-based cancer registry. *Cancer causes & control: CCC*, 31 (7), pp.617–629. [Online]. Available at: doi:10.1007/s10552-020-01303-2.

Afshar, N., English, D. R. and Milne, R. L. (2019). Rural–urban residence and cancer survival in high-income countries: A systematic review. *Cancer*, 125 (13), pp.2172–2184. [Online]. Available at: doi:10.1002/cncr.32073.

Afshar, N., English, D. R. and Milne, R. L. (2021). Factors Explaining Socio-Economic Inequalities in Cancer Survival: A Systematic Review. *Cancer Control : Journal of the Moffitt Cancer Center*, 28, p.10732748211011956. [Online]. Available at: doi:10.1177/10732748211011956.

Age UK. (2017). *Painful journeys - Why getting to hospital appointments is a major issue for older people*. [Online]. Available at: https://www.ageuk.org.uk/siteassets/documents/reports-and-publications/reports-and-briefings/active-communities/rb_dec17_painful_journeys_indepth_report.pdf.

Ali, N. et al. (2024). A systematic review on the qualitative experiences of people living with lung cancer in rural areas. *Supportive Care in Cancer*, 32 (3), p.144. [Online]. Available at: doi:10.1007/s00520-024-08342-4.

Australian Bureau of Statistics. (2021). *Australian Statistical Geography Standard (ASGS) Edition 3*. [Online]. Australian Bureau of Statistics. Available at: https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/latest-release [Accessed 1 October 2023].

Barrett, J. and Hamilton, W. (2005). Pathways to the diagnosis of prostate cancer in a British city. A population-based study. *Scandinavian Journal of Urology and Nephrology*, 39 (4), pp.267–270. [Online]. Available at: doi:10.1080/00365590510031282.

Barrington, D. A. et al. (2016). Distance from a Comprehensive Cancer Center: A Proxy for Poor Cervical Cancer Outcomes? *Gynecologic oncology*, 143 (3), pp.617–621. [Online]. Available at: doi:10.1016/j.ygyno.2016.10.004.

Beal, E. W. et al. (2020). Travel to a high volume hospital to undergo resection of gallbladder cancer: does it impact quality of care and long-term outcomes? *HPB*, 22 (1), pp.41–49. [Online]. Available at: doi:10.1016/j.hpb.2019.05.004.

Berger, M. D. et al. (2020). Impact on survival through consolidation radiotherapy for diffuse large B-cell lymphoma: a comprehensive meta-analysis. *Haematologica*, 106 (7), pp.1923–1931. [Online]. Available at: doi:10.3324/haematol.2020.249680.

Bergin, R. J. et al. (2018). Rural-Urban Disparities in Time to Diagnosis and Treatment for Colorectal and Breast Cancer. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 27 (9), pp.1036–1046. [Online]. Available at: doi:10.1158/1055-9965.EPI-18-0210.

Berkman, A. M. et al. (2021). Impact of Race, Ethnicity, and Socioeconomic Status over Time on the Long-term Survival of Adolescent and Young Adult Hodgkin Lymphoma Survivors. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 30 (9), pp.1717–1725. [Online]. Available at: doi:10.1158/1055-9965.EPI-21-0103.

Bhattarai, N. et al. (2016). *Economic evaluations on centralisation of specialised healthcare services: a systematic review of methods*. [Online]. Available at: doi:10.1136/bmjopen-2016-011214 [Accessed 2 June 2025].

Blick, C. et al. (2010). The impact of the two-week wait rule on the diagnosis and management of bladder cancer in a single UK institution. *Annals of The Royal College of Surgeons of England*, 92 (1), pp.46–50. [Online]. Available at: doi:10.1308/003588410X12518836440207.

Boscoe, F. P., Henry, K. A. and Zdeb, M. S. (2012). A Nationwide Comparison of Driving Distance Versus Straight-Line Distance to Hospitals. *The Professional Geographer: The Journal of the Association of American Geographers*, 64 (2). [Online]. Available at: doi:10.1080/00330124.2011.583586.

Boslooper, K. et al. (2017). No outcome disparities in patients with diffuse large B-cell lymphoma and a low socioeconomic status. *Cancer Epidemiology*, 48, pp.110–116. [Online]. Available at: doi:10.1016/j.canep.2017.04.009.

Boyd, A. et al. (2017). *Understanding Hospital Episode Statistics (HES).* London, UK: CLoser. [Online]. Available at: : https://www.closer.ac.uk/wp-content/uploads/ CLOSER-resource-understanding-hospital-episode-statistics-2018.pdf.

Bright, C. J. et al. (2020). Does geodemographic segmentation explain differences in route of cancer diagnosis above and beyond person-level sociodemographic variables? *Journal of Public Health (Oxford, England)*, 43 (4), pp.797–805. [Online]. Available at: doi:10.1093/pubmed/fdaa111.

British Medical Association. (2024). *COVID-19: Impact of the pandemic on healthcare delivery*. [Online]. The British Medical Association. Available at: https://www.bma.org.uk/advice-and-support/covid-19/what-the-bma-is-doing/covid-19-impact-of-the-pandemic-on-healthcare-delivery [Accessed 20 May 2025].

Bryder, D., Rossi, D. J. and Weissman, I. L. (2006). Hematopoietic Stem Cells. *The American Journal of Pathology*, 169 (2), pp.338–346. [Online]. Available at: doi:10.2353/ajpath.2006.060312.

Burns, R. et al. (2016). Economic burden of malignant blood disorders across Europe: a population-based cost analysis. *The Lancet Haematology*, 3 (8), pp.e362–e370. [Online]. Available at: doi:10.1016/S2352-3026(16)30062-X.

Buske, C. et al. (2022). Managing hematological cancer patients during the COVID-19 pandemic: an ESMO-EHA Interdisciplinary Expert Consensus. *ESMO Open*, 7 (2), p.100403. [Online]. Available at: doi:10.1016/j.esmoop.2022.100403.

Calman KC and Hine D. (1995). *A policy framework for commissioning cancer services: a report by the Expert Advisory Group on Cancer to the Chief Medical Officers of England and Wales.* Department of Health. [Online]. Available at: http://www.surginet.org.uk/misc/interview/downloads/doh/cancerfr_CALMAN_HINE.pdf.

Campbell, B. A. (2013). The Role of Radiation Therapy in the Treatment of Stage I-II Diffuse Large B-Cell Lymphoma. *Current Hematologic Malignancy Reports*, 8 (3), pp.236–242. [Online]. Available at: doi:10.1007/s11899-013-0170-5.

Canale, T. D., Cho, H. and Cheung, W. Y. (2018). A population-based analysis of urban-rural disparities in advanced pancreatic cancer management and outcomes. *Medical Oncology (Northwood, London, England)*, 35 (8), p.116. [Online]. Available at: doi:10.1007/s12032-018-1173-9.

Cancer Research UK. (2022). *Cancer mortality for common cancers*. [Online]. Cancer Research UK. Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality/common-cancers-compared#heading-Zero [Accessed 30 April 2025].

Cancer Research UK. (2024). *Cancer incidence statistics*. [Online]. Cancer Research UK. Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence [Accessed 30 April 2025].

Carr-Hill, R. A. and Pritchard, C. W. (1992). The Registrar-General's Social Class Scheme: The Stevenson Version. In: Carr-Hill, R. A. and Pritchard, C. W. (Eds). *Women's Social Standing: The Empirical Problem of Female Social Class*. London: Palgrave Macmillan UK. pp.10–24. [Online]. Available at: doi:10.1007/978-1-349-22072-4_2 [Accessed 12 June 2025].

Carriere, R. et al. (2018). Rural dwellers are less likely to survive cancer – An international review and meta-analysis. *Health & Place*, 53, pp.219–227. [Online]. Available at: doi:10.1016/j.healthplace.2018.08.010.

Carstairs, V. and Morris, R. (1990). Deprivation and health in Scotland. *Health bulletin*, 48 (4), pp.162–175.

Chaganti, S. et al. (2025). Management of relapsed or refractory large B-cell lymphoma: A British Society for Haematology Guideline. *British Journal of Haematology*, 206 (6), pp.1593–1603. [Online]. Available at: doi:10.1111/bjh.20129.

Coiffier, B. et al. (2002). CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *The New England journal of medicine*, 346 (4), pp.235–242.

Coleman, M. P. et al. (2004). Trends and socioeconomic inequalities in cancer survival in England and Wales up to 2001. *British Journal of Cancer*, 90 (7), pp.1367–1373. [Online]. Available at: doi:10.1038/sj.bjc.6601696.

Comber, H. et al. (2016). Affluence and Private Health Insurance Influence Treatment and Survival in Non-Hodgkin's Lymphoma. *PloS One*, 11 (12), p.e0168684. [Online]. Available at: doi:10.1371/journal.pone.0168684.

Costa, L. J., Brill, I. K. and Brown, E. E. (2016). Impact of marital status, insurance status, income, and race/ethnicity on the survival of younger patients diagnosed with multiple myeloma in the United States. *Cancer*, 122 (20), pp.3183–3190. [Online]. Available at: doi:10.1002/cncr.30183.

Crawford, S. M. et al. (2009). Social and geographical factors affecting access to treatment of lung cancer. *British Journal of Cancer*, 101 (6), pp.897–901. [Online]. Available at: doi:10.1038/sj.bjc.6605257.

Crawford, S. M. (2021). How to provide specialist services: how do we know when centralisation is a good idea? *Postgraduate Medical Journal*, 97 (1144), pp.69–71. [Online]. Available at: doi:10.1136/postgradmedj-2020-138151.

Crawford, S. M. (2022). *Everyone should have ready access to a competent cancer service*. [Online]. Available at: doi:10.1136/bmj.o2786 [Accessed 2 June 2025].

Crump, M. et al. (2017). Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. *Blood*, 130 (16), pp.1800–1808. [Online]. Available at: doi:10.1182/blood-2017-03-769620.

Davies, A. J. et al. (2022). Five-Year Survival Results from the Remodl-B Trial (ISRCTN 51837425) Show Improved Outcomes in Diffuse Large B-Cell Lymphoma Molecular Subgroups from the Addition of Bortezomib to R-CHOP Chemoimmunotherapy. *Blood*, 140 (Supplement 1), pp.1770–1772. [Online]. Available at: doi:10.1182/blood-2022-159976.

Dejardin, O. et al. (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. [Online]. Available at: doi:10.1016/j.healthplace.2014.08.002.

Department for Environment Food & Rural Affairs. (2022). *Statistical Digest of Rural England*. [Online]. Available at: https://assets.publishing.service.gov.uk/media/630620558fa8f5536c077dcc/07_Statistical_Digest_of_Rural_England_2022_August_edition.pdf [Accessed 15 February 2025].

Department of Health. (2000). *The NHS Cancer Plan*. [Online]. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4014513.pdf [Accessed 18 June 2025].

Department of Health. (2006). *Our health, our care, our say: a new direction for community services*. Gov.uk. [Online]. Available at: https://assets.publishing.service.gov.uk/media/5a7c2baee5274a25a9140eab/6737.pdf [Accessed 18 June 2025].

Department of Health. (2007). *Cancer Reform Strategy*. London, UK. [Online]. Available at: https://www.nhs.uk/NHSEngland/NSF/Documents/Cancer%20Reform%20Strategy.pdf [Accessed 30 April 2025].

Department of Health and Social Care. (2021). *Chief Medical Officer's Annual Report 2021 Health in Coastal Communities – Summary and recommendations*. Gov.uk. [Online]. Available at: https://www.gov.uk/government/publications/chief-medical-officers-annual-report-2021-health-in-coastal-communities [Accessed 18 June 2025].

Dhakal, P. et al. (2020). Effects of distance from academic cancer center on overall survival of acute myeloid leukemia: A retrospective analysis of treated patients. *Clinical lymphoma, myeloma & leukemia*, 20 (10), pp.e685–e690. [Online]. Available at: doi:10.1016/j.clml.2020.05.016.

Dobson, C. et al. (2023). Barriers to Early Presentation amongst Rural Residents Experiencing Symptoms of Colorectal Cancer: A Qualitative Interview Study. *Cancers*, 15 (1), p.274. [Online]. Available at: doi:10.3390/cancers15010274.

Dreyling, M. et al. (2021). Newly diagnosed and relapsed follicular lymphoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 32 (3), pp.298–308. [Online]. Available at: doi:10.1016/j.annonc.2020.11.008.

Ellis-Brookes, L. (2015). Big data in action: The story behind Routes to Diagnosis. *UK Health Security Agency*. [Online]. Available at: https://ukhsa.blog.gov.uk/2015/11/10/big-data-in-action-the-story-behind-routes-to-diagnosis/.

Elliss-Brookes, L. et al. (2012). Routes to diagnosis for cancer – determining the patient journey using multiple routine data sets. *British Journal of Cancer*, 107 (8), pp.1220–1226. [Online]. Available at: doi:10.1038/bjc.2012.408.

Elting, L. S. et al. (2006). Generalizability of cancer clinical trial results: prognostic differences between participants and nonparticipants. *Cancer*, 106 (11), pp.2452–2458. [Online]. Available at: doi:10.1002/cncr.21907.

Ermann, D. A. et al. (2020). Academic Centers Compared With Nonacademic Centers for Patients With International Prognostic Index Risk-stratified Diffuse Large B-cell Lymphoma: A Survival Outcomes Analysis. *Clinical Lymphoma Myeloma and Leukemia*, 20 (4), pp.e174–e183. [Online]. Available at: doi:10.1016/j.clml.2019.11.012.

Exarchakou, A. et al. (2018). Impact of national cancer policies on cancer survival trends and socioeconomic inequalities in England, 1996-2013: population based study. *The BMJ*, 360, p.k764. [Online]. Available at: doi:10.1136/bmj.k764.

Forbes, L. J. L. et al. (2014). Risk factors for delay in symptomatic presentation: a survey of cancer patients. *British Journal of Cancer*, 111 (3), pp.581–588. [Online]. Available at: doi:10.1038/bjc.2014.304.

Forrest, L. F. et al. (2015). The role of receipt and timeliness of treatment in socioeconomic inequalities in lung cancer survival: population-based, data-linkage study. *Thorax*, 70 (2), pp.138–145. [Online]. Available at: doi:10.1136/thoraxjnl-2014-205517.

Fox, C. P. et al. (2019). Guidelines for the diagnosis and management of primary central nervous system diffuse large B-cell lymphoma. *British Journal of Haematology*, 184 (3), pp.348–363. [Online]. Available at: doi:10.1111/bjh.15661.

Fox, C. P. et al. (2024). The management of newly diagnosed large B-cell lymphoma: A British Society for Haematology Guideline. *British Journal of Haematology*, 204 (4), pp.1178–1192. [Online]. Available at: doi:10.1111/bjh.19273.

Gardy, J. et al. (2023). Cancer Survival and Travel Time to Nearest Reference Care Center for 10 Cancer Sites: An Analysis of 21 French Cancer Registries. *Cancers*, 15 (5), p.1516. [Online]. Available at: doi:10.3390/cancers15051516.

Gardy, J. et al. (2024). Access to primary care and mortality in excess for patients with cancer in France: Results from 21 French Cancer Registries. *Cancer*, 130 (23), pp.4096–4108. [Online]. Available at: doi:10.1002/cncr.35519.

Geographic Information Science & Analysis Team and Rural & Environment Science and Analytical Services Division. (2022). *Scottish Government Urban Rural Classification*. [Online]. Available at: https://www.gov.scot/publications/scottish-government-urban-rural-classification-2022/ [Accessed 12 June 2025].

Ghandili, S. et al. (2024). The changing influence of neighborhood socioeconomic status on long-term survival in diffuse large B-cell lymphoma patients: A German metropolitan case-control study spanning over three decades. *HemaSphere*, 8 (10), p.e70011. [Online]. Available at: doi:10.1002/hem3.70011.

Gibson, T. M. et al. (2014). Risk of diffuse large B-cell lymphoma after solid organ transplantation in the United States. *American journal of hematology*, 89 (7), pp.714–720. [Online]. Available at: doi:10.1002/ajh.23726.

Gotfrit, J. et al. (2020). The impact of driving time, distance, and socioeconomic factors on outcomes of patients with locally advanced rectal cancer. *Public Health in Practice*, 1, p.100012. [Online]. Available at: doi:10.1016/j.puhip.2020.100012.

Government Office for Science and Social and Political Science Group, Institute for Transport Studies, University of Leeds. (2019). *Inequalities in Mobility and Access in the UK Transport System*. [Online]. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/784685/future_of_mobility_access.pdf [Accessed 12 June 2025].

Graboyes, E. M. et al. (2018). Racial and Ethnic Disparities in Travel for Head and Neck Cancer Treatment and the Impact of Travel Distance on Survival. *Cancer*, 124 (15), pp.3181–3191. [Online]. Available at: doi:10.1002/cncr.31571.

Grass, G. D. et al. (2019). Radiotherapy for early stage diffuse large B-cell lymphoma with or without double or triple hit genetic alterations. *Leukemia & lymphoma*, 60 (4), pp.886–893. [Online]. Available at: doi:10.1080/10428194.2018.1506586.

Griffin, F. et al. (2023). Assessing the impact of rurality on oesophagogastric cancer survival in the North-East of Scotland- a prospective population cohort study. *The Surgeon*, 21 (3), pp.e97–e103. [Online]. Available at: doi:10.1016/j.surge.2022.04.011.

Gunnarsson, H., Ekholm, A. and Olsson, L. I. (2013). Emergency presentation and socioeconomic status in colon cancer. *European Journal of Surgical Oncology: The Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*, 39 (8), pp.831–836. [Online]. Available at: doi:10.1016/j.ejso.2013.04.004.

Guven, D. C. et al. (2024). Newly diagnosed cancer and the COVID-19 pandemic: tumour stage migration and higher early mortality. *BMJ Supportive & Palliative Care*, 14 (e1), pp.e456–e461. [Online]. Available at: doi:10.1136/bmjspcare-2021-003301.

Guyader-Peyrou, S. L. et al. (2017). Factors related to the relative survival of patients with diffuse large B-cell lymphoma in a population-based study in France: does socio-economic status have a role? *Haematologica*, 102 (3), p.584. [Online]. Available at: doi:10.3324/haematol.2016.152918.

Haddad, A. Q. et al. (2015). Association of Distance to Treatment Facility With Survival and Quality Outcomes After Radical Cystectomy: A Multi-Institutional Study. *Clinical Genitourinary Cancer*, 15 (6), pp.689-695.e2. [Online]. Available at: doi:10.1016/j.clgc.2017.05.006.

Haines, M. R. (2001). The urban mortality transition in the united states, 1800-1940. *Annales de démographie historique*, 101 (1), pp.33–64. [Online]. Available at: doi:10.3917/adh.101.0033.

Hamilton, W. (2012). Emergency admissions of cancer as a marker of diagnostic delay. *British Journal of Cancer*, 107 (8), pp.1205–1206. [Online]. Available at: doi:10.1038/bjc.2012.423.

Han, L. et al. (2024). The impact of transportation mode, socioeconomic deprivation and rurality on travel times to radiotherapy and surgical services for patients with prostate cancer: A national population-based evaluation. *Radiotherapy and Oncology*, 192, p.110092. [Online]. Available at: doi:10.1016/j.radonc.2024.110092.

Hanna, T. P. et al. (2020). *Mortality due to cancer treatment delay: systematic review and meta-analysis*. [Online]. Available at: doi:10.1136/bmj.m4087 [Accessed 9 April 2025].

Haynes, R. et al. (2006). Validation of travel times to hospital estimated by GIS. *International Journal of Health Geographics*, 5, p.40. [Online]. Available at: doi:10.1186/1476-072X-5-40.

Haynes, R., Pearce, J. and Barnett, R. (2008). Cancer survival in New Zealand: Ethnic, social and geographical inequalities. *Social Science & Medicine*, 67 (6), pp.928–937. [Online]. Available at: doi:10.1016/j.socscimed.2008.05.005.

Healthwatch England. (2019). *There and back - what people tell us about their experiences of travelling to and from NHS services*. [Online]. Available at: https://www.healthwatch.co.uk/sites/healthwatch.co.uk/files/20191016%20People%27s%20experiences%20of%20patient%20transport%20Formatted%20final.pdf [Accessed 12 June 2025].

Herbert, A. et al. (2017). Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *International Journal of Epidemiology*, 46 (4), pp.1093–1093i. [Online]. Available at: doi:10.1093/ije/dyx015.

Herbert, A. et al. (2019a). Are inequalities in cancer diagnosis through emergency presentation narrowing, widening or remaining unchanged? Longitudinal analysis of English population-based data 2006–2013. *Journal of Epidemiology and Community Health*, 73 (1), pp.3–10. [Online]. Available at: doi:10.1136/jech-2017-210371.

Herbert, A. et al. (2019b). Population trends in emergency cancer diagnoses: The role of changing patient case-mix. *Cancer Epidemiology*, 63, p.101574. [Online]. Available at: doi:10.1016/j.canep.2019.101574.

Hill, A. B. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, 58 (5), pp.295–300.

Hirst, Y. et al. (2018). Uptake of the English Bowel (Colorectal) Cancer Screening Programme: an update 5 years after the full roll-out. *European Journal of Cancer*, 103, pp.267–273. [Online]. Available at: doi:10.1016/j.ejca.2018.07.135.

HMRN. (2022a). *Factsheets: T-cell prolymphocytic leukaemia 2022.* [Online]. The Haematological Malignancy Research Network. Available at: www.hmrn.org/factsheets#factsheets#t-cell_prolymphocytic_leukaemia [Accessed 13 April 2022].

HMRN. (2022b). *Haematological Malignancy Research Network (HMRN) - Incidence statistics*. [Online]. Available at: https://hmrn.org/statistics/incidence [Accessed 3 March 2026].

HMRN. (2022c). *Survival statistics: Myeloma*. [Online]. Available at: https://hmrn.org/statistics/survival [Accessed 3 March 2026].

HMRN. (2025). *Survival statistics: Large B-cell lymphomas*. [Online]. The Haematological Malignancy Research Network. Available at: https://hmrn.org/statistics/survival [Accessed 30 April 2025].

Hong, S. et al. (2016). Association of socioeconomic status with autologous hematopoietic cell transplantation outcomes for lymphoma. *Bone Marrow Transplantation*, 51 (9), pp.1191–1196. [Online]. Available at: doi:10.1038/bmt.2016.107.

Hong, Y. D. et al. (2023). Association of individual low-income status and area deprivation with mortality in multiple myeloma. *Journal of Geriatric Oncology*, 14 (2), p.101415. [Online]. Available at: doi:10.1016/j.jgo.2022.12.003.

Howell, D. A. et al. (2013). Time-to-diagnosis and symptoms of myeloma, lymphomas and leukaemias: a report from the Haematological Malignancy Research Network. *BMC Blood Disorders*, 13 (1), p.9. [Online]. Available at: doi:10.1186/2052-1839-13-9.

Howell, D. A. et al. (2019). Disease-related factors affecting timely lymphoma diagnosis: a qualitative study exploring patient experiences. *Br J Gen Pract*, 69 (679), pp.e134–e145. [Online]. Available at: doi:10.3399/bjgp19X701009.

Howell, D. A., Smith, A. G. and Roman, E. (2008). Help-seeking behaviour in patients with lymphoma. *European Journal of Cancer Care*, 17 (4), pp.394–403. [Online]. Available at: doi:10.1111/j.1365-2354.2007.00897.x.

Ingleby, F. C. et al. (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), p.e041714. [Online]. Available at: doi:10.1136/bmjopen-2020-041714.

Ingleby, F. C. et al. (2022). An investigation of cancer survival inequalities associated with individual-level socio-economic status, area-level deprivation, and contextual effects, in a cancer patient cohort in England and Wales. *BMC Public Health*, 22 (1), p.90. [Online]. Available at: doi:10.1186/s12889-022-12525-1.

International Non-Hodgkin's Lymphoma Prognostic Factors Project. (1993). A predictive model for aggressive non-Hodgkin's lymphoma. *The New England Journal of Medicine*, 329 (14), pp.987–994. [Online]. Available at: doi:10.1056/NEJM199309303291402.

Isaac, K. M. et al. (2021). Epidemiology of acute myeloid leukemia in Virginia: Excellent survival outcomes for patients in rural Appalachia. *Cancer Reports (Hoboken, N.J.)*, 4 (4), p.e1354. [Online]. Available at: doi:10.1002/cnr2.1354.

Islami, F. et al. (2021). Annual Report to the Nation on the Status of Cancer, Part 1: National Cancer Statistics. *JNCI Journal of the National Cancer Institute*, 113 (12), pp.1648–1669. [Online]. Available at: doi:10.1093/jnci/djab131.

Jaffe, E. S. et al. (Eds). (2001). *World Health Organization classification of Tumours. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon, France: IARC Press.

Johnson, K. J. et al. (2021). Associations between geographic residence and US adolescent and young adult cancer stage and survival. *Cancer*, 127 (19), pp.3640–3650. [Online]. Available at: doi:10.1002/cncr.33667.

Jones, A. P. et al. (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. [Online]. Available at: doi:10.1016/j.ejca.2008.02.001.

Jones, A. P. et al. (2008b). Travel times to health care and survival from cancers in Northern England. *European Journal of Cancer*, 44 (2), pp.269–274. [Online]. Available at: doi:10.1016/j.ejca.2007.07.028.

Jones, A. P. et al. (2010). Geographical access to healthcare in Northern England and post-mortem diagnosis of cancer. *Journal of Public Health*, 32 (4), pp.532–537. [Online]. Available at: doi:10.1093/pubmed/fdq017.

Kane, E. et al. (2017). Emergency admission and survival from aggressive non-Hodgkin lymphoma: A report from the UK's population-based Haematological Malignancy Research Network. *European Journal of Cancer*, 78, pp.53–60. [Online]. Available at: doi:10.1016/j.ejca.2017.03.013.

Keegan, T. H. M. et al. (2016). Impact of Treatment and Insurance on Socioeconomic Disparities in Survival after Adolescent and Young Adult Hodgkin Lymphoma: A Population-Based Study. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 25 (2), pp.264–273. [Online]. Available at: doi:10.1158/1055-9965.EPI-15-0756.

Kenzik, K. M. et al. (2024). Estimating the Impact of Rurality in Disparities in Cancer Mortality. *JCO Oncology Practice*, 20 (7), pp.993–1002. [Online]. Available at: doi:10.1200/OP.23.00626.

Kogevinas, M. et al. (1991). Socioeconomic differences in cancer survival. *Journal of Epidemiology and Community Health*, 45 (3), pp.216–219. [Online]. Available at: doi:10.1136/jech.45.3.216.

Laing, K. A. et al. (2014). Prostate cancer in Scotland: does geography matter? An analysis of incidence, disease characteristics and survival between urban and rural areas. *Journal of Clinical Urology*, 7 (3), pp.176–184. [Online]. Available at: doi:10.1177/2051415813512303.

Lamb, M. et al. (2024). Lymphoid blood cancers, incidence and survival 2005-2023: A report from the UK's Haematological Malignancy Research Network. *Cancer Epidemiology*, 88, p.102513. [Online]. Available at: doi:10.1016/j.canep.2023.102513.

Larfors, G. et al. (2017). The impact of socio-economic factors on treatment choice and mortality in chronic myeloid leukaemia. *European Journal of Haematology*, 98 (4), pp.398–406. [Online]. Available at: doi:10.1111/ejh.12845.

Le Floch, A.-C. et al. (2020). Socioeconomic deprivation is associated with decreased survival in patients with acute myeloid leukemia. *Cancer Epidemiology*, 66, p.101699. [Online]. Available at: doi:10.1016/j.canep.2020.101699.

Lee, B. et al. (2014). Effect of Place of Residence and Treatment on Survival Outcomes in Patients With Diffuse Large B-Cell Lymphoma in British Columbia. *The Oncologist*, 19 (3), pp.283–290. [Online]. Available at: doi:10.1634/theoncologist.2013-0343.

Lee, S. F. et al. (2021). Socioeconomic inequalities in treatment and relative survival among patients with diffuse large B-cell lymphoma: a Hong Kong population-based study. *Scientific Reports*, 11 (1), p.17950. [Online]. Available at: doi:10.1038/s41598-021-97455-5.

Lidsky, M. E. et al. (2017). Going the Extra Mile: Improved Survival for Pancreatic Cancer Patients Traveling to High-volume Centers. *Annals of Surgery*, 266 (2), p.333. [Online]. Available at: doi:10.1097/SLA.0000000000001924.

Lim, J. et al. (2023). Impact of travel distance on outcomes for clinical trial patients: the Kinghorn Cancer Centre experience. *Internal Medicine Journal*, 53 (2), pp.242–249. [Online]. Available at: doi:10.1111/imj.15561.

Lister, T. A. et al. (1989). Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. *Journal of Clinical Oncology*, 7 (11), pp.1630–1636. [Online]. Available at: doi:10.1200/JCO.1989.7.11.1630.

Lyratzopoulos, G. et al. (2012). Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *The lancet oncology*, 13 (4), pp.353–365. [Online]. Available at: doi:10.1016/S1470-2045(12)70041-4.

Lyratzopoulos, G., Vedsted, P. and Singh, H. (2015). Understanding missed opportunities for more timely diagnosis of cancer in symptomatic patients after presentation. *British Journal of Cancer*, 112 (1), pp.S84–S91. [Online]. Available at: doi:10.1038/bjc.2015.47.

Mackenbach, J. P. et al. (2008). Socioeconomic Inequalities in Health in 22 European Countries. *New England Journal of Medicine*, 358 (23), pp.2468–2481. [Online]. Available at: doi:10.1056/NEJMsa0707519.

Macmillan Cancer Support. (2024a). *Cancer statistics in the UK*. [Online]. Macmillan Cancer Support. Available at: https://www.macmillan.org.uk/about-us/what-we-do/research/cancer-statistics-fact-sheet [Accessed 13 May 2025].

Macmillan Cancer Support. (2024b). *Tackling cancer inequality: Macmillan's partnership with the Farming Community Network*. [Online]. Macmillan Cancer Support. Available at: https://www.macmillan.org.uk/cancer-information-and-support/stories-and-media/blogs/rural-communities-project [Accessed 30 May 2025].

MacVicar, E. et al. (2020). Analysing the impact of living in a rural setting on the presentation and outcome of colorectal cancer. A prospective single centre observational study. *The Surgeon*, 18 (6), pp.354–359. [Online]. Available at: doi:10.1016/j.surge.2020.02.001.

Maheswaran, R. and Morley, N. (2018). Incidence, socioeconomic deprivation, volume-outcome and survival in adult patients with acute lymphoblastic leukaemia in England. *BMC cancer*, 18 (1), p.25. [Online]. Available at: doi:10.1186/s12885-017-3975-0.

Maishman, T. et al. (2013). A novel adaptive trial design: randomised evaluation of molecular guided therapy for diffuse large b-cell lymphoma with bortezomib (REMODL-B) with two interim analyses to explore safety and efficacy. *Trials*, 14 (1), p.O77. [Online]. Available at: doi:10.1186/1745-6215-14-S1-O77.

Maringe, C. et al. (2020). The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: a national, population-based, modelling study. *The Lancet Oncology*, 21 (8), pp.1023–1034. [Online]. Available at: doi:10.1016/S1470-2045(20)30388-0.

Marmot, M. and Bell, R. (2012). Fair society, healthy lives. *Public Health*, 126, pp.S4–S10. [Online]. Available at: doi:10.1016/j.puhe.2012.05.014.

Marmot, M. G. et al. (2013). The benefits and harms of breast cancer screening: an independent review. *British Journal of Cancer*, 108 (11), pp.2205–2240. [Online]. Available at: doi:10.1038/bjc.2013.177.

Massat, N. J. et al. (2015). *Variation in cervical and breast cancer screening coverage in England: a cross-sectional analysis to characterise districts with atypical behaviour*. [Online]. Available at: doi:10.1136/bmjopen-2015-007735 [Accessed 14 April 2025].

Maxwell, S. et al. (2023). The impact of rurality on patient experience and diagnostic pathway intervals in Scotland's cancer patients: Further results from a national cancer diagnosis audit. *Cancer Epidemiology*, 86, p.102414. [Online]. Available at: doi:10.1016/j.canep.2023.102414.

McCabe, R. et al. (2020). Adapting hospital capacity to meet changing demands during the COVID-19 pandemic. *BMC Medicine*, 18 (1), p.329. [Online]. Available at: doi:10.1186/s12916-020-01781-w.

McPhail, S. et al. (2013). Emergency presentation of cancer and short-term mortality. *British Journal of Cancer*, 109 (8), pp.2027–2034. [Online]. Available at: doi:10.1038/bjc.2013.569.

Mendonca, S. C., Abel, G. A. and Lyratzopoulos, G. (2016). Pre-referral GP consultations in patients subsequently diagnosed with rarer cancers: a study of patient-reported data. *Br J Gen Pract*, 66 (644), pp.e171–e181. [Online]. Available at: doi:10.3399/bjgp16X683977.

Ministry of Housing, Communities & Local Government. (2015). *The English indices of deprivation 2015*. [Online]. GOV.UK. Available at: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 [Accessed 13 May 2025].

Ministry of Housing, Communities & Local Government. (2019). *The English Indices of Deprivation 2019, Research Report*. [Online]. Available at: https://assets.publishing.service.gov.uk/media/5d8b364ced915d03709e3cf2/IoD2019_Research_Report.pdf [Accessed 1 May 2025].

Mitchell, E. D. et al. (2015). The role of primary care in cancer diagnosis via emergency presentation: qualitative synthesis of significant event reports. *British Journal of Cancer*, 112 (s1), pp.S50–S56. [Online]. Available at: doi:10.1038/bjc.2015.42.

Mitchell, J. M. and Conklin, E. A. (2015). Factors Affecting Receipt of Expensive Cancer Treatments and Mortality: Evidence from Stem Cell Transplantation for Leukemia and Lymphoma. *Health Services Research*, 50 (1), p.197. [Online]. Available at: doi:10.1111/1475-6773.12208.

Mizen, A. et al. (2015). Quantifying the Error Associated with Alternative GIS-based Techniques to Measure Access to Health Care Services. *AIMS public health*, 2 (4), pp.746–761. [Online]. Available at: doi:10.3934/publichealth.2015.4.746.

Murage, P. et al. (2016). Geographical disparities in access to cancer management and treatment services in England. *Health & Place*, 42, pp.11–18. [Online]. Available at: doi:10.1016/j.healthplace.2016.08.014.

Murage, P. et al. (2017). Impact of travel time and rurality on presentation and outcomes of symptomatic colorectal cancer: a cross-sectional cohort study in primary care. *The British Journal of General Practice*, 67 (660), pp.e460–e466. [Online]. Available at: doi:10.3399/bjgp17X691349.

Murage, P. et al. (2019). Geographical access to GPs and modes of cancer diagnosis in England: a cross-sectional study. *Family Practice*, 36 (3), pp.284–290. [Online]. Available at: doi:10.1093/fampra/cmy077.

Murchie, P. et al. (2017). Does emergency presentation of cancer represent poor performance in primary care? Insights from a novel analysis of linked primary and secondary care data. *British Journal of Cancer*, 116 (9), pp.1148–1158. [Online]. Available at: doi:10.1038/bjc.2017.71.

Murchie, P. et al. (2018). Impact of rurality on processes and outcomes in melanoma care: results from a whole-Scotland melanoma cohort in primary and secondary care. *The British Journal of General Practice*, 68 (673), p.e566. [Online]. Available at: doi:10.3399/bjgp18X697901.

Murchie, P. et al. (2020). Impact of geography on Scottish cancer diagnoses in primary care: Results from a national cancer diagnosis audit. *Cancer Epidemiology*, 66, p.101720. [Online]. Available at: doi:10.1016/j.canep.2020.101720.

Murchie, P. et al. (2021a). Geographic variation in diagnostic and treatment interval, cancer stage and mortality among colorectal patients – An international comparison between Denmark and Scotland using data-linked cohorts. *Cancer Epidemiology*, 74, p.102004. [Online]. Available at: doi:10.1016/j.canep.2021.102004.

Murchie, P. et al. (2021b). Is place or person more important in determining higher rural cancer mortality? A data-linkage study to compare individual versus area-based measures of deprivation. *International Journal of Population Data Science*, 6 (1), p.1403. [Online]. Available at: doi:10.23889/ijpds.v6i1.1403.

National Disease Registry Service. (2025). *Get Data Out*. [Online]. Get Data Out. Available at: https://nhsd-ndrs.shinyapps.io/get_data_out/ [Accessed 12 June 2025].

National Institute for Health and Care Excellence. (2023). *Polatuzumab vedotin in combination for untreated diffuse large B-cell lymphoma. Technology appraisal guidance [TA874].* [Online]. Available at: https://www.nice.org.uk/guidance/ta874/resources/polatuzumab-vedotin-in-combination-for-untreated-diffuse-large-bcell-lymphoma-pdf-82613673152197 [Accessed 30 April 2025].

National Radiotherapy Advisory Group. (2007). *Report to Ministers from National Radiotherapy Advisory Group*. [Online]. Available at: https://data.parliament.uk/DepositedPapers/Files/DEP2008-0036/DEP2008-0036.pdf [Accessed 12 June 2025].

Neal, R. D. et al. (2014). Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. *British Journal of Cancer*, 110 (3), pp.584–592. [Online]. Available at: doi:10.1038/bjc.2013.791.

Nennecke, A. et al. (2014). Survival of cancer patients in urban and rural areas of Germany--a comparison. *Cancer Epidemiology*, 38 (3), pp.259–265. [Online]. Available at: doi:10.1016/j.canep.2014.02.011.

Ng, A. P. and Alexander, W. S. (2017). Haematopoietic stem cells: past, present and future. *Cell Death Discovery*, 3, p.17002. [Online]. Available at: doi:10.1038/cddiscovery.2017.2.

NHS. (2019). *NHS Long Term Plan*. [Online]. Available at: https://www.longtermplan.nhs.uk/online-version/chapter-3-further-progress-on-care-quality-and-outcomes/better-care-for-major-health-conditions/cancer/ [Accessed 20 May 2025].

NHS. (2023). *How to organise transport to and from hospital*. [Online]. NHS. Available at: https://www.nhs.uk/nhs-services/hospitals/going-into-hospital/how-to-organise-transport-to-and-from-hospital/.

NHS England. (2014). *Five Year Forward View*. [Online]. Available at: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf [Accessed 20 May 2025].

NHS England. (2018). *NHS England Modernising Radiotherapy Services in England - Public Consultation Report*. [Online]. Available at: https://www.engage.england.nhs.uk/consultation/radiotherapy-service-specification-consultation/user_uploads/smart-consult-public-consultation-report.pdf [Accessed 13 May 2025].

NHS England. (2021). *Urgent cancer diagnostic services during COVID-19*. [Online]. Available at: https://www.england.nhs.uk/coronavirus/documents/c0789-urgent-cancer-diagnostic-services-during-covid-19/ [Accessed 20 May 2025].

NHS England. (2023a). *Cancer Registrations Statistics, England 2021 [online]*. [Online]. NHS England. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/cancer-registration-statistics/england-2021---summary-counts-only/cancer-incidence-by-stage#related-links [Accessed 7 April 2025].

NHS England. (2023b). *Cancer Survival in England, cancers diagnosed 2016 to 2020, followed up to 2021*. [Online]. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/cancer-survival-in-england/cancers-diagnosed-2016-to-2020-followed-up-to-2021#resources [Accessed 3 January 2024].

NICE. (2005). *Referral Guidelines for Suspected Cancer*. London: Department of Health.

NICE. (2015). *Suspected cancer: recognition and referral*. London, UK. [Online]. Available at: https://www.nice.org.uk/guidance/ng12/resources/suspected-cancer-recognition-and-referral-1837268071621 [Accessed 21 April 2016].

Nikonova, A. et al. (2015). Predictors of delay in diagnosis and treatment in diffuse large B-cell lymphoma and impact on survival. *British Journal of Haematology*, 168 (4), pp.492–500. [Online]. Available at: doi:10.1111/bjh.13150.

Nuffield Trust. (2024). *Still waiting: Is it just England that still has a backlog problem?* [Online]. Nuffield Trust. Available at: https://www.nuffieldtrust.org.uk/news-item/still-waiting-is-it-just-england-that-still-has-a-backlog-problem [Accessed 20 May 2025].

Office for National Statistics. (2016). *2011 rural/urban classification*. [Online]. ons.gov.uk. Available at: https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2011ruralurbanclassification [Accessed 30 April 2025].

Oh, D. L. et al. (2024). Disparities in cancer stage of diagnosis by rurality in California, 2015-2019. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 33 (11), pp.1523–1531. [Online]. Available at: doi:10.1158/1055-9965.EPI-24-0564.

Oken, M. M. et al. (1982). Toxicity and response criteria of the Eastern Cooperative Oncology Group. *American Journal of Clinical Oncology*, 5 (6), pp.649–655.

Orkin, S. H. (2000). Diversification of haematopoietic stem cells to specific lineages. *Nature Reviews Genetics*, 1 (1), pp.57–64. [Online]. Available at: doi:10.1038/35049577.

Parsons, H. M. et al. (2024). Travel-time barriers to specialized cancer care for adolescents and young adults with acute lymphoblastic leukemia. *JNCI Cancer Spectrum*, 8 (4), p.pkae046. [Online]. Available at: doi:10.1093/jncics/pkae046.

Pearson, C. et al. (2020). Cross-sectional study using primary care and cancer registration data to investigate patients with cancer presenting with non-specific symptoms. *BMJ Open*, 10 (1), p.e033008. [Online]. Available at: doi:10.1136/bmjopen-2019-033008.

Pennings, E. R. A. et al. (2024). Treatment and outcomes for patients with relapsed or refractory diffuse large B-cell lymphoma: a contemporary, nationwide, population-based study in the Netherlands. *Blood Cancer Journal*, 14 (1), pp.1–5. [Online]. Available at: doi:10.1038/s41408-023-00970-z.

Perry, N. J. et al. (2023). Impact of travel burden on the treatment of stage I and II breast cancer: A National Cancer Database analysis. *Surgery*, 174 (4), pp.794–800. [Online]. Available at: doi:10.1016/j.surg.2023.07.004.

Pettigrew, L. M. et al. (2024). *The changing shape of English general practice: a retrospective longitudinal study using national datasets describing trends in organisational structure, workforce and recorded appointments*. [Online]. Available at: doi:10.1136/bmjopen-2023-081535 [Accessed 19 May 2025].

Peyrade, F. et al. (2011). Attenuated immunochemotherapy regimen (R-miniCHOP) in elderly patients older than 80 years with diffuse large B-cell lymphoma: a multicentre, single-arm, phase 2 trial. *The Lancet Oncology*, 12 (5), pp.460–468. [Online]. Available at: doi:10.1016/S1470-2045(11)70069-9.

Pornet, C. et al. (2012). Construction of an adaptable European transnational ecological deprivation index: the French version. *Journal of Epidemiology and Community Health*, 66 (11), pp.982–989. [Online]. Available at: doi:10.1136/jech-2011-200311.

Price, S. et al. (2020). Trends in time to cancer diagnosis around the period of changing national guidance on referral of symptomatic patients: A serial cross-sectional study using UK electronic healthcare records from 2006–17. *Cancer Epidemiology*, 69, p.101805. [Online]. Available at: doi:10.1016/j.canep.2020.101805.

Rachet, B. et al. (2008). Survival from non-Hodgkin lymphoma in England and Wales up to 2001. *British Journal of Cancer*, 99 (1), pp.S104–S106. [Online]. Available at: doi:10.1038/sj.bjc.6604605.

Rachet, B. et al. (2009). Population-based cancer survival trends in England and Wales up to 2007: an assessment of the NHS cancer plan for England. *The Lancet Oncology*, 10 (4), pp.351–369. [Online]. Available at: doi:10.1016/S1470-2045(09)70028-2.

Rachet, B. et al. (2010). Socioeconomic inequalities in cancer survival in England after the NHS cancer plan. *British Journal of Cancer*, 103 (4), pp.446–453. [Online]. Available at: doi:10.1038/sj.bjc.6605752.

Raman, V. et al. (2019). Disparity of Colon Cancer Outcomes in Rural America: Making the Case to Travel the Extra Mile. *Journal of Gastrointestinal Surgery*, 23 (11), pp.2285–2293. [Online]. Available at: doi:10.1007/s11605-019-04270-5.

Ramsay, S. E. et al. (2014). Time trends in socioeconomic inequalities in cancer mortality: results from a 35 year prospective study in British men. *BMC cancer*, 14, p.474. [Online]. Available at: doi:10.1186/1471-2407-14-474.

Rauh, L. A. et al. (2018). Cervical cancer care in rural Virginia: The impact of distance from an academic medical center on outcomes & the role of non-specialized radiation centers. *Gynecologic Oncology*, 150 (2), pp.338–342. [Online]. Available at: doi:10.1016/j.ygyno.2018.06.019.

Renzi, C. et al. (2016). Do colorectal cancer patients diagnosed as an emergency differ from non-emergency patients in their consultation patterns and symptoms? A longitudinal data-linkage study in England. *British Journal of Cancer*, 115 (7), pp.866–875. [Online]. Available at: doi:10.1038/bjc.2016.250.

Riaz, I. B. et al. (2019). A Systematic Review on Predisposition to Lymphoid (B and T cell) Neoplasias in Patients With Primary Immunodeficiencies and Immune Dysregulatory Disorders (Inborn Errors of Immunity). *Frontiers in Immunology*, 10. [Online]. Available at: doi:10.3389/fimmu.2019.00777 [Accessed 7 May 2024].

Riaz, S. P. et al. (2011). Lung Cancer Incidence and Survival in England: An Analysis by Socioeconomic Deprivation and Urbanization. *Journal of Thoracic Oncology*, 6 (12), pp.2005–2010. [Online]. Available at: doi:10.1097/JTO.0b013e31822b02db.

Richards, M. A. (2009). The size of the prize for earlier diagnosis of cancer in England. *British Journal of Cancer*, 101 (2), pp.S125–S129. [Online]. Available at: doi:10.1038/sj.bjc.6605402.

Ringstrom, M. J. et al. (2018). Travel distance: Impact on stage of presentation and treatment choices in head and neck cancer. *American Journal of Otolaryngology*, 39 (5), pp.575–581. [Online]. Available at: doi:10.1016/j.amjoto.2018.06.020.

Ritter, A. J. et al. (2019). Rural and urban patients with diffuse large B-cell and follicular lymphoma experience reduced overall survival: a National Cancer DataBase study. *Leukemia & Lymphoma*, 60 (7), pp.1656–1667. [Online]. Available at: doi:10.1080/10428194.2018.1546855.

Roman, E. et al. (2022). Cohort Profile Update: The Haematological Malignancy Research Network (HMRN)'s UK population-based cohorts. *International Journal of Epidemiology*, p.dyab275. [Online]. Available at: doi:10.1093/ije/dyab275.

Rossi, D., Spina, V. and Gaidano, G. (2018). Biology and treatment of Richter syndrome. *Blood*, 131 (25), pp.2761–2772. [Online]. Available at: doi:10.1182/blood-2018-01-791376.

Royal College of General Practitioners. (2024). *GPs in deprived areas responsible for almost 2,500 patients per head*. [Online]. Royal College of General Practitioners. Available at: https://www.rcgp.org.uk/News/research-statement-conference-2024 [Accessed 22 May 2025].

Rubin, G. et al. (2015). The expanding role of primary care in cancer control. *The Lancet Oncology*, 16 (12), pp.1231–1272. [Online]. Available at: doi:10.1016/S1470-2045(15)00205-3.

Salonen, M. and Toivonen, T. (2013). Modelling travel time in urban networks: comparable measures for private car and public transport. *Journal of Transport Geography*, 31, pp.143–153. [Online]. Available at: doi:10.1016/j.jtrangeo.2013.06.011.

Sandlund, J. T. and Martin, M. G. (2016). Non-Hodgkin lymphoma across the pediatric and adolescent and young adult age spectrum. *Hematology*, 2016 (1), pp.589–597. [Online]. Available at: doi:10.1182/asheducation-2016.1.589.

Scott, E. C. S. and Hoskin, P. J. (2024). Health inequalities in cancer care: a literature review of pathways to diagnosis in the United Kingdom. *eClinicalMedicine*, 76, p.102864. [Online]. Available at: doi:10.1016/j.eclinm.2024.102864.

Siegel, J. et al. (2021). Travel distance and its interaction with patient and hospital factors in pancreas cancer care. *The American Journal of Surgery*, 221 (4), pp.819–825. [Online]. Available at: doi:10.1016/j.amjsurg.2020.08.023.

Singh, G. K. and Jemal, A. (2017). Socioeconomic and Racial/Ethnic Disparities in Cancer Mortality, Incidence, and Survival in the United States, 1950–2014: Over Six Decades of Changing Patterns and Widening Inequalities. *Journal of Environmental and Public Health*, 2017, p.2819372. [Online]. Available at: doi:10.1155/2017/2819372.

Sjöström, O. et al. (2020). Travel time to care does not affect survival for patients with colorectal cancer in northern Sweden: A data linkage study from the Risk North database. *PLoS ONE*, 15 (8), p.e0236799. [Online]. Available at: doi:10.1371/journal.pone.0236799.

Smith, A. et al. (2011). Incidence of haematological malignancy by sub-type: a report from the Haematological Malignancy Research Network. *British Journal of Cancer*, 105 (11), pp.1684–1692. [Online]. Available at: doi:10.1038/bjc.2011.450.

Smith, A. et al. (2015). Impact of age and socioeconomic status on treatment and survival from aggressive lymphoma: a UK population-based study of diffuse large B-cell lymphoma. *Cancer Epidemiology*, 39 (6), pp.1103–1112. [Online]. Available at: doi:10.1016/j.canep.2015.08.015.

Smith, A. et al. (2018). Cohort Profile: The Haematological Malignancy Research Network (HMRN): a UK population-based patient cohort. *International Journal of Epidemiology*, 47 (3), pp.700–700g. [Online]. Available at: doi:10.1093/ije/dyy044.

Smith, M. J. et al. (2021a). Association between multimorbidity and socioeconomic deprivation on short-term mortality among patients with diffuse large B-cell or follicular lymphoma in England: a nationwide cohort study. *BMJ Open*, 11 (11), p.e049087. [Online]. Available at: doi:10.1136/bmjopen-2021-049087.

Smith, M. J. et al. (2021b). Investigating the inequalities in route to diagnosis amongst patients with diffuse large B-cell or follicular lymphoma in England. *British Journal of Cancer*, 125 (9), pp.1299–1307. [Online]. Available at: doi:10.1038/s41416-021-01523-6.

Steliarova-Foucher, E. et al. (2017). International incidence of childhood cancer, 2001–10: a population-based registry study. *The Lancet Oncology*, 18 (6), pp.719–731. [Online]. Available at: doi:10.1016/S1470-2045(17)30186-9.

Stoyanov, D. S. et al. (2022). Impact of travel burden on clinical outcomes in lung cancer. *Supportive Care in Cancer*, 30 (6), pp.5381–5387. [Online]. Available at: doi:10.1007/s00520-022-06978-8.

Swerdlow, S. et al. (2008). *WHO Classification of tumours of haematopoietic and lymphoid tissues.*

Swerdlow, S. H. H. edt. (2016). *WHO classification of tumours of haematopoietic and lymphoid tissues*. p.592.

Tao, L. et al. (2014). Socioeconomic disparities in mortality after diffuse large B-cell lymphoma in the modern treatment era. *Blood*, 123 (23), pp.3553–3562. [Online]. Available at: doi:10.1182/blood-2013-07-517110.

Tataru, D. et al. (2018). Variation in geographical treatment intensity affects survival of non-small cell lung cancer patients in England. *Cancer Epidemiology*, 57, pp.13–23. [Online]. Available at: doi:10.1016/j.canep.2018.09.001.

The Independent Cancer Taskforce. (2015). *Achieving World-Class Cancer Outcomes a Strategy for England*. [Online]. Available at: https://www.cancerresearchuk.org/sites/default/files/achieving_world-class_cancer_outcomes_-_a_strategy_for_england_2015-2020.pdf.

Tilly, H. et al. (2022). Polatuzumab Vedotin in Previously Untreated Diffuse Large B-Cell Lymphoma. *New England Journal of Medicine*, 386 (4), pp.351–363. [Online]. Available at: doi:10.1056/NEJMoa2115304.

Tørring, M. L. et al. (2013). Evidence of increasing mortality with longer diagnostic intervals for five common cancers: A cohort study in primary care. *European Journal of Cancer*, 49 (9), pp.2187–2198. [Online]. Available at: doi:10.1016/j.ejca.2013.01.025.

Tøstesen, M. et al. (2019). Longer distance to specialized treatment centers does not adversely affect treatment intensity or outcomes in adult acute myeloid leukemia patients. A Danish national population-based cohort study. *Clinical Epidemiology*, 11, pp.769–780. [Online]. Available at: doi:10.2147/CLEP.S210456.

Townsend, P., Phillimore, P. and Alastair, B. (1988). *Health and deprivation: inequality and the North*. London: Croom Helm.

Turner, M. et al. (2017). A cancer geography paradox? Poorer cancer outcomes with longer travelling times to healthcare facilities despite prompter diagnosis and treatment: a data-linkage study. *British Journal of Cancer*, 117 (3), pp.439–449. [Online]. Available at: doi:10.1038/bjc.2017.180.

Turner, M. et al. (2023). The impact of travel time to cancer treatment centre on post-diagnosis care and mortality among cancer patients in Scotland. *Health & Place*, 84, p.103139. [Online]. Available at: doi:10.1016/j.healthplace.2023.103139.

Ulmer, K. K. et al. (2022). Disparity of ovarian cancer survival between urban and rural settings. *International Journal of Gynecological Cancer*, 32 (4), pp.540–546. [Online]. Available at: doi:10.1136/ijgc-2021-003096.

Unger, J. M. et al. (2018). Geographic Distribution and Survival Outcomes for Rural Patients With Cancer Treated in Clinical Trials. *JAMA network open*, 1 (4), p.e181235. [Online]. Available at: doi:10.1001/jamanetworkopen.2018.1235.

US Census Bureau. (2024). *Health Insurance Coverage in the United States: 2023*. [Online]. Census.gov. Available at: https://www.census.gov/library/publications/2024/demo/p60-284.html [Accessed 30 April 2025].

U.S. Department of Agriculture, Economic Research Service. (2023a). *Rural-Urban Commuting Area Codes- Documentation*. [Online]. Economic Research Service. Available at: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation [Accessed 12 June 2025].

U.S. Department of Agriculture, Economic Research Service. (2023b). *Rural-Urban Continuum Codes - Documentation*. [Online]. Economic Research Service. Available at: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation [Accessed 12 June 2025].

Vallance, A. E. et al. (2018). Socioeconomic differences in selection for liver resection in metastatic colorectal cancer and the impact on survival. *European Journal of Surgical Oncology: The Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*, 44 (10), pp.1588–1594. [Online]. Available at: doi:10.1016/j.ejso.2018.05.024.

Vardell, V. A. et al. (2023). Impact of academic medical center access on outcomes in multiple myeloma. *American Journal of Hematology*, 98 (1), pp.41–48. [Online]. Available at: doi:10.1002/ajh.26759.

Vargo, J. A. et al. (2015). Treatment Selection and Survival Outcomes in Early-Stage Diffuse Large B-Cell Lymphoma: Do We Still Need Consolidative Radiotherapy? *Journal of Clinical Oncology*, 33 (32), pp.3710–3717. [Online]. Available at: doi:10.1200/JCO.2015.61.7654.

Vetterlein, M. W. et al. (2017). Impact of travel distance to the treatment facility on overall mortality in US patients with prostate cancer. *Cancer*, 123 (17), pp.3241–3252. [Online]. Available at: doi:10.1002/cncr.30744.

Wallace, D. et al. (2014). Identifying patients at risk of emergency admission for colorectal cancer. *British Journal of Cancer*, 111 (3), pp.577–580. [Online]. Available at: doi:10.1038/bjc.2014.300.

Wang, L., Li, L. and Young, K. H. (2020). New agents and regimens for diffuse large B cell lymphoma. *Journal of Hematology & Oncology*, 13 (1), p.175. [Online]. Available at: doi:10.1186/s13045-020-01011-z.

Wasif, N. et al. (2016). Association of Distance Traveled for Surgery with Short- and Long-Term Cancer Outcomes. *Annals of Surgical Oncology*, 23 (11), pp.3444–3452. [Online]. Available at: doi:10.1245/s10434-016-5242-z.

WHO Classification of Tumours Editorial Board. (2024). *The WHO Classification of Tumours Haematolymphoid Tumours*. 5th ed. Lyon, France: International Agency for Research on Cancer. [Online]. Available at: Available at: https://publications.iarc.fr/Book-And-Report-Series/Who-Classification-Of-Tumours/Haematolymphoid-Tumours-2024 [Accessed 17 November 2024].

Woods, L. M. et al. (2022). Are deprivation-specific cancer survival patterns similar according to individual-based and area-based measures? A cohort study of patients diagnosed with five malignancies in England and Wales, 2008-2016. *BMJ open*, 12 (6), p.e058411. [Online]. Available at: doi:10.1136/bmjopen-2021-058411.

Xia, L. et al. (2018). Associations Between Travel Distance, Hospital Volume, and Outcomes Following Radical Cystectomy in Patients With Muscle-invasive Bladder Cancer. *Urology*, 114, pp.87–94. [Online]. Available at: doi:10.1016/j.urology.2017.12.014.

Yee, E. K. et al. (2021). Geographic impact on access to care and survival for non-curative esophagogastric cancer: a population-based study. *Gastric Cancer*, 24 (4), pp.790–799. [Online]. Available at: doi:10.1007/s10120-021-01157-w.

York Medical Group. (2025). *York Medical Group*. [Online]. York Medical Group. Available at: https://www.yorkmedicalgroup.co.uk/ [Accessed 18 June 2025].

Yost, K. et al. (2001). Socioeconomic Status and Breast Cancer Incidence in California for Different Race/Ethnic Groups. *Cancer Causes & Control*, 12 (8), pp.703–711.

Zhang, Y. et al. (2011). Risk Factors of Non-Hodgkin Lymphoma. *Expert opinion on medical diagnostics*, 5 (6), pp.539–550. [Online]. Available at: doi:10.1517/17530059.2011.618185.

Zurko, J. C., Wade, R. C. and Mehta, A. (2019). The impact of structural factors on diagnostic delay in diffuse large B-cell lymphoma. *Cancer Medicine*, 8 (4), pp.1416–1422. [Online]. Available at: doi:10.1002/cam4.2009.