TIME TO DIAGNOSIS OF CANCER IN CHILDREN AND YOUNG ADULTS IN SECONDARY CARE SERVICES

Christopher David Lethaby

MB ChB, MRCPCH(UK)

Submitted in accordance with the requirements for the degree of

Doctor of Medicine

The University of Leeds

School of Medicine,

Division for Epidemiology and Biostatistics, LIGHT

August 2013

"The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others."

Published Thesis Work

1. Lethaby CD. Picton S. Kinsey SE. Phillips R. van Laar M. Feltbower RG. A systematic review of time to diagnosis in children and young adults with cancer. [Review] Archives of Disease in Childhood. 98(5):349-55, 2013 May.

   Thesis chapters: Systematic review Chapter 3

   My contribution: the review design, the literature search, data extraction and analysis, writing the first draft and subsequent redrafting.

   Contribution of other authors: RGF, SP and SEK helped with the concept and significantly with the redrafting of the article for submission. MvL reviewed a sample of the literature and helped refine the list of included articles and was also involved in redrafting the article. RP provided a significant level of guidance on the review design, including data analysis, and also helped redraft the article.

ACKNOWLEDGEMENTS

I wish to thank my supervisors Dr Richard Feltbower, Miss Marlous van Laar, Professor Sally Kinsey and Dr Susan Picton for all their help throughout this work. Without their support and guidance the completion of this thesis would not have been possible and I greatly value all the time they gave me over the two years. Dr Feltbower and Miss van Laar were always available to help me overcome the technical and statistical hurdles I faced and gave me valuable advice on the direction of the project. The clinical expertise and critical appraisal provided by Professor Kinsey and Dr Picton were vital to the development of this work into what is presented here.

The funding for this study and my position as a clinical research fellow was provided by the Jeremy Neil Allen Foundation through the Leeds Teaching Hospitals NHS Trust Charitable committee.

I would also like to thank Dr Robert Phillips for his help and guidance especially relating to the systematic review of literature on early diagnosis research in children and young adults.

It was a real privilege to work with and learn from the teams at the Centre for Epidemiology and Biostatistics and the Yorkshire Regional Centre for Paediatric Oncology and Haematology. I hope I have the opportunity to carry on being involved with these two expert units as my career moves forward.

I would like to thank my parents John and Marie for all their support and always providing me with support and love. They have always been there for me and encouraged me to work hard and give the best of myself. I would also like to thank my parents-in-law Chris and Keith for all their support.

Finally I would like to thank my wife Tracey who has been with me through the highs and the lows. She is an inspiration to me and I am a lucky man to be able to share my life with her.

ABSTRACT

Purpose: A prolonged time to diagnosis (TTD) for cancer patients has been highlighted as a factor potentially contributing to worse outcomes. The majority of early diagnosis research in childhood and young adult (CYA) cancer has focused on primary care. This population-based study aimed to investigate TTD in secondary care services for CYAs diagnosed with cancer and its effect on survival in Yorkshire, UK.

Method: 1098 cases of cancer aged between 0-24 years were identified from the Yorkshire Specialist Registry of Cancer in Children and Young People over a 6 year period. ICD-10 codes contained within the in-patient HES episodes were reviewed against accepted UK CYA cancer awareness campaigns in order to identify alert signs and symptoms preceding the date of definitive diagnosis. The cohort was analysed in terms of the time spent in hospital care, number of alert and non-alert code containing events and 1 and 3-year survival, the latter modelled using Cox regression.

Results: 457 (41.6%) cases had no identifiable alert code containing episodes preceding their date of diagnosis. In two thirds of the remaining cases (437/641) the alert codes only occurred within the month preceding diagnosis. Cases with alert codes present within the month prior to diagnosis had a significantly poorer survival compared to patients with no alert code containing episodes (hazard ratio=1.67, p=0.003). For cases with a more prolonged TTD, there was a significantly poorer survival for 15-24 year olds (hazard ratio = 2.48, p=0.001) but not for 0-14 year olds (hazard ratio=0.98, p=0.964).

Conclusions: In Yorkshire, secondary care services appear to be organised effectively to deal with timely diagnosis of CYA cancers. This research supports the current focus of early diagnosis research at primary care for childhood cancer but indicates the need for further investigation of TTD in secondary care particularly in the TYA population.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

List of Terms	Abbreviation
Accident & Emergency	A&E
Centre for Epidemiology and Biostatistics	CEB
Central Nervous System	CNS
Children and Young Adults	CYA
Continuous Inpatient Spells	CIPS
Department of Health	DoH
Diagnostic Interval	DI
Finished Consultant Episodes	FCE
Former Yorkshire Regional Health Authority	FYRHA
General Practice	GP
Hospital Episodes Statistics	HES
International Classification of Childhood Cancer	ICCC
International Classification of Disease	ICD
International Society of Paediatric Oncology	SIOP
National Awareness & Early Diagnosis Initiative	NAEDI
National Cancer Intelligence Network	NCIN
National Health Service	NHS
National Institute of Clinical Excellence	NICE
Non-Rhabdomyosarcoma Soft Tissue Sarcoma	NRSTS
Patient Interval	PI
Patient Identifiable Data	PID
Soft Tissue Sarcoma	STS
Strategic Health Authority	SHA
Teenage(rs) and Young Adult(s)	TYA
Time to Diagnosis	TTD
United Kingdom	UK
World Health Organisation	WHO
Yorkshire & Humber Children & Young People's Cancer Network	YHCYPCN
Yorkshire Specialist Registry of Cancer in Children & Young People	YSRCCYP
United Kingdom Association of Cancer Registries	UKACR

Chapter 1 Introduction

1.1 Motivation

1.1.1 Why childhood and young adult cancer?

Although cancer is a rare occurrence in children and young adults (CYA), it has profound implications for patients and the people close to them. The diagnosis of cancer is devastating at any age, however life years lost in those who die and the disability experienced due to the disease and its treatment by those who survive is of particular concern in the CYA population.

The incidence of CYA cancers in the UK is increasing and so is the number and proportion surviving their cancer (1). Survival rates for cancer in CYA vary by diagnosis, age and between countries (2, 3). The improvements in the survival rates for teenagers and young adults have been more modest than those seen in younger children (2, 4). Advances in cancer treatments, improvements in supportive care, increased awareness of CYA cancers, advances in diagnostic investigations, centralisation of cancer specific services and the recruitment of more patients into national and international trials are all potentially contributing to improving survival.

Despite the reduction in mortality in the UK, survival in CYA patients treated in England lags behind our European counterparts (5, 6). This worrying observation was highlighted in the EUROCARE-4 study published in 2009, a study that analysed survival and survival time trends in young Europeans covering 83 cancer registries in 23 countries from 1995-2002 (5). The 5-year survival for CYA in England was below the European average for both age groups (5).

CYA cancers need special consideration for a number of reasons: they are the leading cause of natural death within the CYA age-range (7), and the pattern of increasing incidence and falling mortality is leading to an ever increasing cohort of survivors (4). In 2011 it was estimated that there were

40,000 survivors of childhood cancer alive in the UK of whom 60% would experience at least one adverse late effect of their treatment as well as being at an increased risk of a second malignancy (8). Those providing care for CYA cancer patients must aim to continue to improve survival while reducing the burden of the treatment. It has been suggested that reducing the time to diagnosis (TTD) for cancer in CYA in the UK is one way of reducing the burden of disease and treatment and improving survival and survivorship.

1.1.2 Why early diagnosis of cancer?

The NHS Cancer Plan (2000) gave cancer a high priority within the NHS (9). The aim was to reduce death rates, improve prospects for survival and quality of life through improvements in early detection and effective screening and address healthcare inequalities (9). The Cancer Reform Strategy (2007) built on this and set out a five year plan, which identified earlier diagnosis of cancer as a key area for improving cancer care within the UK (6). In 2005 the National Institute for Health and Clinical Excellence (NICE) published updated guidance that specifically aimed to improve clinical outcomes and the experience of cancer for CYA and their families (10). The Yorkshire and Humber Children and Young People's Cancer Network (YHCYPCN) emerged from the recommendations made within the 2005 NICE guidance, and through work with the National Cancer Action Team (NCAT), has implemented the cancer strategy reforms to provide the best care for young cancer users in the region.

NICE published referral guidance for suspected cancer for GPs in 2005 containing a childhood cancer section (see Appendix 1). The pre-diagnostic period has been further highlighted in the recent Department of Health (DH) Improving Outcomes: A Strategy for Cancer 2011 (11). Charities, non-NHS and non-government organisations are also working to improve the TTD for CYA. For example, the Headsmart campaign launched in 2011 aims to improve patient and professional awareness of CYA brain tumours. Headsmart is a partnership between the Children's Brain Tumour Research Centre in Nottingham, Royal College of Paediatrics and Child Health and

Brain Tumour Research (previously the Samantha Dickson Brain Tumour Trust) (12).

Since 2009 research into early diagnosis in the UK has been coordinated and supported by the National Awareness and Early Diagnosis Initiative (NAEDI), which primarily focuses on adult cancers in the primary care setting. To date the majority of guidance and research on early diagnosis has been primarily focused on patient awareness, improving access to primary care and referral pathways. This project will therefore provide a novel approach to an under-researched area of healthcare in which the importance of early diagnosis is poorly understood.

1.1.3 Why secondary care and childhood and young adult cancer?

Healthcare within the UK is based around a hierarchical structure consisting of primary, secondary, tertiary and quaternary levels of healthcare. The initial point of healthcare engagement can vary across the age spectrum, but is predominantly focused at the primary healthcare setting with primary care physicians often referred to as having a “gate-keeper” role. However, focusing on childhood healthcare engagement reveals an increased involvement of emergency care and outpatient care situated in the secondary care environment. This point was highlighted in an article by Gill *et al* published in 2013, who showed a year-on-year increase in emergency admission rates in under 15s between 2003 to 2010 in a population based study of Hospital Episodes Statistics (HES) data (13). This publication contained data on a number of common illnesses of childhood and did not include data specifically on cancer-related admissions. Nonetheless it demonstrates the important role secondary care services play at the point of contact for childhood illness.

The prominence of emergency admission at diagnosis for cancer patients under 25 years of age is highlighted in the National Cancer Intelligence Network (NCIN) Routes to Diagnosis publication in 2010 (14, 15). This work used HES data linked to cancer registration information to map the routes to diagnosis for cancer across all ages within England in 2007 (14). The

authors identified 269 distinct routes of admission and highlighted that patients aged less than 25 are most likely to present within the routes defined as an emergency (15). This work mainly focused on adult cancers and didn't provide any insight into the specific challenges to diagnosis faced by CYA with cancer.

In a similar manner to the NCIN routes to diagnosis work, the National Audit of Cancer Diagnosis in Primary Care 2011 provided an insight into current practice in primary care cancer diagnosis, which encompasses the entire cancer population (16). This work focused partly on those aged under 25: 25% of cases diagnosed with cancer in the under 25s were referred on from primary care via the two week referral for suspected cancer pathway; 40% of males and 45% of females were referred on from primary care via emergency routes; 65-70% of cases had only 1 or 2 GP attendances prior to referral for suspected cancer (16).

All this information gives an indication of the important role of secondary care services in the initial contact and diagnosis of cancer in those aged under 25 years. However, to date there is no body of work that focuses specifically on this area of healthcare in this age group. This thesis aims to provide a unique insight into the role of secondary care services in the diagnosis of cancer among those aged 0-24 years through the use of a population-based study which links registry to electronic health-records data.

1.2 Aims & Objectives

1.2.1.1 Aims

- Describe the pattern of variation of TTD for CYA within secondary care services in the Yorkshire and Humber region and highlight the major factors that influence TTD for this population.
- Assess how variations in TTD within secondary care services affect the outcome for CYA patients and construct recommendations for healthcare providers to aid early diagnosis of CYA cancers.

1.2.1.2 Objectives

- To assess variations in the time interval from the point at which a case presents to secondary care with a sign or symptom that potentially identifies their cancer to the definitive diagnosis, using population based cancer registry data linked to HES data.
- To examine whether correlation exists between the TTD and outcomes such as stage at diagnosis and survival. If correlation is found, the data will be analysed further to determine whether these effects are attributable to specific cancer subtypes, ages or other population characteristics.
- Assess how reliable and accurate HES data are in the analysis of a patient's TTD and survival.

1.3 Outline of thesis

The purpose of this project is to develop our understanding of the time spent in hospital preceding a diagnosis of cancer in CYA and to investigate how variations in this time period affect outcomes for CYA in Yorkshire. The current knowledge base will be explored and described through in-depth analysis of cancer registry data linked to HES data. This project will aim to provide new information and develop recommendations for secondary care services to aid early diagnosis of cancer in CYA in secondary care.

Chapter 2 will provide background that further supports the motivation for this research as outlined in Chapter 1 as well as providing the reader with important information regarding the complex nature of cancer among 0-24 year olds in the UK. The chapter will cover two main areas: the first introduces some of the main considerations for early diagnosis research; the second discusses major challenges for early diagnosis research in CYAs, highlighting the unique and complex nature of the disease that affects the TTD.

Chapter 3 will review the latest literature on early diagnosis in CYA and will pull out key factors that influence TTD as well as highlighting what is known of the association between TTD and specific outcomes for CYA. This section will inform the development of the methodology for this study.

Chapter 4 will provide an overview of the research methods, including information regarding the Yorkshire Specialist Registry of Cancer in Children and Young People (YSRCCYP), HES data and cross-validation with clinical notes. The chapter will outline common awareness campaigns and early diagnosis guidance for CYA cancer used to define alert symptoms and signs. The method of defining and flagging the alert episodes within the HES linked patient records and the subsequent survival analysis of patients with and without alert signs will be presented. A brief description of the process of validating the HES record with a small but representative sample of medical case notes will conclude this chapter.

Chapter 5 will present the results describing TTD within secondary care services for a population based cohort of CYA diagnosed with cancer between 2004 and 2009 in Yorkshire.

Chapters 6 will discuss the findings and outline recommendations for secondary care services and early diagnosis research in CYA, with the aim of improving TTD. These sections will also discuss the unique perspective on TTD provided by this study into secondary care and outline future work. At the end of this chapter the conclusions will bring together the main points from the preceding chapters to give an overview of the project.

Chapter 2 Time to Diagnosis

This chapter will provide the background that underpins and supports the work throughout this thesis. General considerations needed to construct a robust method for early diagnosis research are discussed in the initial section of this chapter; including the important time periods that contribute to the overall TTD for cancer. The second section discusses factors that influence the TTD in CYA cancers. The final section of this chapter will bring all the key issues raised together in a conclusive summary.

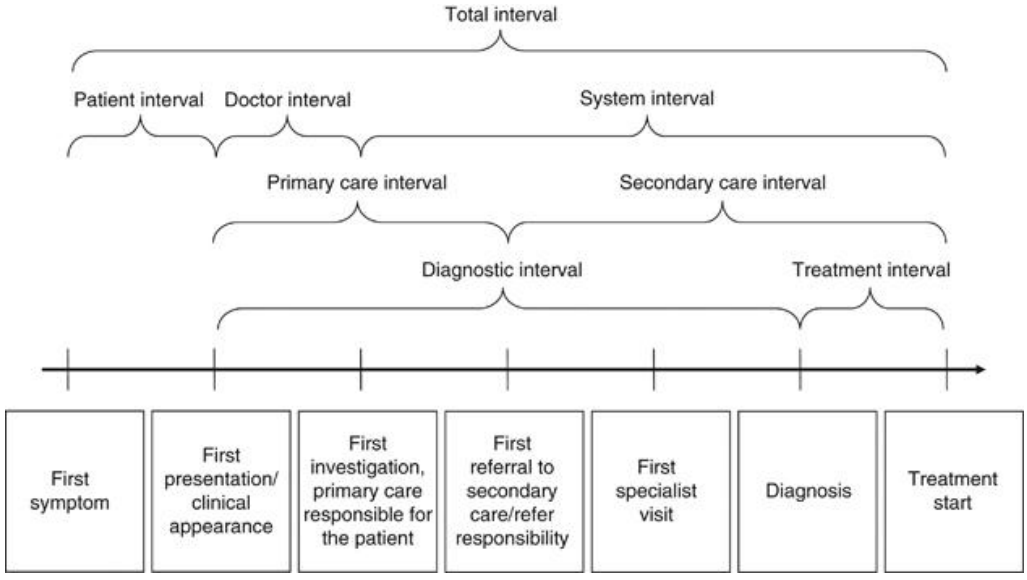
2.1 Early diagnosis research

Survival for cancer in the UK has been shown to be inferior to many of our European counterparts (5). A number of areas within UK cancer care have been highlighted as potentially contributing to poorer survival. The Cancer Reform Strategy highlighted the early diagnosis of cancer as a key area for improvement in the UK cancer care and subsequently the TTD has come under increasing scrutiny (6). The theory that prolonging the time spent prior to receiving the diagnosis of cancer results in more advanced disease at diagnosis and therefore a worse outcome has been widely accepted within UK healthcare research. Significance has been placed upon delays within the diagnostic process and how these may negatively impact on survival. National research initiatives, awareness campaigns and healthcare policy documents have set out plans to improve the TTD for cancer across the age-spectrum, some of which are specifically aimed at childhood and young adult cancer (12, 17).

A field of research is developing that focuses on the time patients spend in the pre-diagnosis and treatment period of disease. Early diagnosis research focuses on the pre-diagnosis symptomatic period of the patient pathway, this period is complex and can be divided in a number of different ways depending on the focus of the research. In order for the methods applied in early diagnosis studies to be robust and reproducible it is important that

clear definitions for the time periods being studied are present at the outset of any early diagnosis research project. Figure 2.1 illustrates the time period from first symptom until start of treatment, indicating specific events that define the boundaries for various time intervals and potential variable routes to diagnosis (18). This section will discuss the time interval that make up the pre-diagnosis period up to and including the date of diagnosis. The focus of this study is the period of time spent in secondary care by CYA's with cancer, which is part of the diagnostic interval identified in Figure 2.1. The date of definitive diagnosis is taken as the end point for the period being studied and the point of entry into secondary care as the initiating time point for the interval being studied. The rationale for the focus of this study will be discussed further in Chapter 4 (Methods).

Figure 2.1 Outline of milestones and time intervals in the route to diagnosis and treatment (Olesen, 2009)



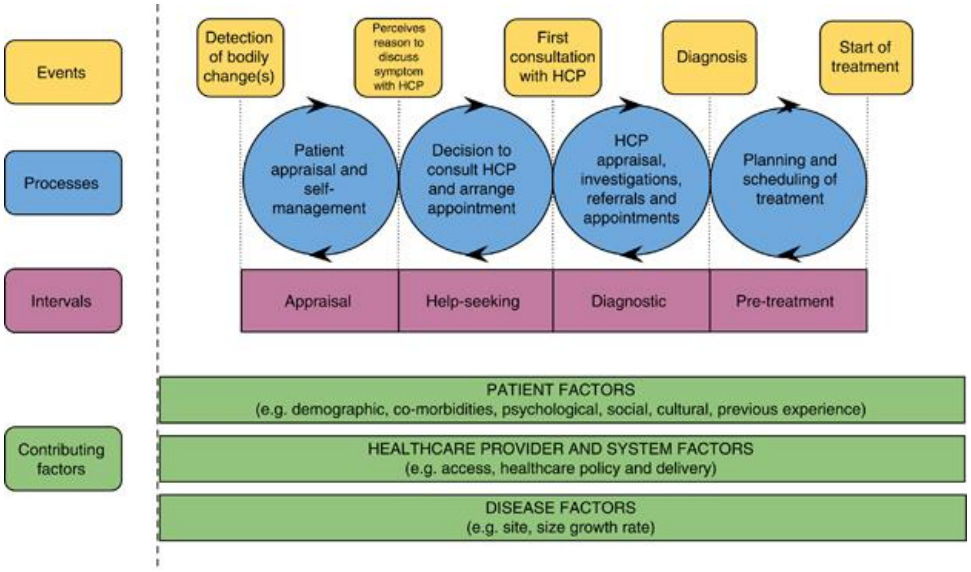
The majority of this research has, so far, lacked clear structure and the methodology and terminology applied along with the definitions for intervals being studied often vary widely as do the outcomes measured (19). This has resulted in debatable conclusions being drawn, especially regarding the association of a prolonged TTD of cancer with overall outcome. In response to the absence of a clear direction in early diagnosis research an

international consensus group published guidance for the early diagnosis researcher in 2012, under the title - “The Aarhus statement: improving design and reporting of studies on early cancer diagnosis” (19). The systematic review included within the Aarhus statement identified all studies of symptomatic cancer patients presenting to primary care, the group concluded:

> *“There is little consistency in the definitions and measurement of key time points and intervals;*
>
> - *There is little guidance for researchers in designing studies that require the measurement of diagnostic time points and intervals;*
> - *Little work in this field explicitly uses a theoretical framework to underpin definitions and measurement of diagnostic intervals;*
> - *There is a lack of transparency and precision over the methods and instruments in early diagnosis research….”(19)*

The paper identifies Walters model of pathways to treatment as a clear theoretical framework that should underpin future early diagnosis research, see Figure 2.2 (20). They also discuss definitions of key time points and highlights the Olesen *et al* 2009 illustration of milestones and time intervals as a good guide to outline of terminology, see Figure 2.1 (18, 19). Defining the time-points that divide the intervals being studied is not straight forward. There may be a number of different time points that potentially define a particular event in the pathway to diagnosis. The exact timing of an event may also vary depending on the perspective from which it is being viewed, either patient or healthcare professional.

Figure 2.2 Model of pathways to treatment (Walter, 2012)



The Aarhus statement was the first of its kind and is an indispensable tool for early diagnosis research, whilst the publication doesn't specifically deal with cancer in children and young adults the issues raised are common across all age groups. Due to the rare nature of cancer in children and young adults the difficulties faced by early diagnosis researchers within these age-groups are heightened. Walter's model presents the "Events", "Intervals", "Contributing factors" and "Processes" involved in a person receiving a diagnosis and subsequent treatment for cancer. It is important to recognise that the "Processes" don't run in a linear fashion, as suggested in Figure 2.1, at any point a "Contributing factor" can result in reappraisal and rescheduling of the whole process (20). This model applies to all types of cancer and across the spectrum of ages; as such it will be referred to throughout this thesis as the theoretical framework from which the study methods have developed. The literature review chapter will in part explore the extent to which such theoretical models have been applied in early diagnosis research in children and young people to date.

2.1.1 Symptom recognition

Figure 2.1 highlights the first symptom as the initiating event for the TTD and treatment. As a cancer develops there is a point in time where the person or carer realises a change in the function or feeling within their body, which

subsequently triggers concern and eventually manifests a desire to engage with a healthcare professional. This event and the subsequent trigger to contact a healthcare professional may take time, the length of which will be determined by such factors as a person's awareness of their own body, knowledge of health and desire to act upon their concerns. The first theoretical model dealing with the process a patient goes through to receive a diagnosis of cancer was developed by Safer et al 1979, this separates the time between symptom recognition and healthcare engagement into 3 intervals (21). Anderson adapted Safer's model in 1995 and conceptualised a number of delay intervals punctuated by decision-making processes(22). The Anderson model of total patient delay was further refined by Walter in 2012 to produce a model of pathways to treatment, Figure 2.2, this model is now widely accepted amongst early diagnosis researchers (20).

A prospective study of the interval between symptom recognition and healthcare engagement would be ethically challenging. Therefore, identification of the point of first symptom recognition has to rely on patient recall, either at the point of data collection or as recorded by a healthcare professional in the medical records at the point of contact or diagnosis. How a patient recalls a symptom and what they confer to the health professional may be influenced by a number of factors, three groups of contributing factors are identified in Figure 2.2; patient factors, healthcare factors and disease factors. The factors that affect CYA's with cancer and the challenges faced by early diagnosis researchers will be discussed in later in this chapter and the present literature explored in more depth in Chapter 3.

2.1.2 Healthcare engagement

All patients require some degree of healthcare engagement prior to their cancer diagnosis. The route to diagnosis through the healthcare system will be influenced by the structure of the healthcare system, the initial point of contact and the patients attitudes towards healthcare. Defining the point of initial engagement can vary depending whether it is being viewed from the patient or healthcare professionals perspective. A professional may not view a symptom as indicative of cancer at the point of contact and therefore not

act to investigate or refer, an alternative view maybe taken by the patient when recalling the events leading to their diagnosis. It is therefore important that the early diagnosis researcher set out a clear definition for the time of initial healthcare engagement and Weller *et al* considers this as *'the point at which, given the presenting signs, symptoms, history and other risk factors, it would be at least possible for the clinician seeing the patient to have started investigation or referral for possible important pathology, including cancer'* (19). Defining the point of healthcare engagement is also influenced by the collection method and data sources used by the early diagnosis researcher, for example, retrospective healthcare records review for the date of initial engagement will be influenced by the completeness of information at the point of recording and interpretation of the point a significant event occurs, such as presentation with a symptom potentially indicating cancer. The early diagnosis researcher should therefore identify criteria for defining the point of engagement taking into account the limitations of the data being used.

It is also important to consider when and where a patient is referred for investigation or opinion as there may be several referrals made prior to the eventual successful diagnosis. The patterns of referral and complexity of the diagnostic route can influence the time spent prior to diagnosis and understanding how routes to diagnosis vary within a population is vital to the development of effective interventions for improving the TTD. Healthcare factors that influence the time taken to receive a diagnosis of cancer for CYA's in the UK will be discussed further within this chapter.

2.1.3 Date of diagnosis

The date of diagnosis of cancer is a key time point for this project, however there are several events that potentially identify this point, for examples:

- The date the doctors tells the patient they have cancer.
- The date a scan defines a mass likely to be cancer.
- The date a biopsy is taken which confirms a cancer.

- The date the pathologist signs and communicates the report to the treating physician.

In order for early diagnosis research to be robust and reproducible there must be a clear hierarchy of events accepted as defining the date of diagnosis. A hierarchy is important as the process of receiving the diagnosis of cancer can involve a number of steps and not all cancers are diagnosed using the same investigations. A point highlighted in the varied approaches to the diagnosis of central nervous system tumours, for many tumour types a biopsy is required and a histological diagnosis is made, however in cases of diffuse pontine glioma a radiological diagnosis is often made upon MRI scan findings and biopsies are rarely attempted due to the perceived dangers associated with surgery in the pons (23).

Weller *et al* suggests the use of the European Network of Cancer Registries: Hierarchy for Defining the Date of Diagnosis (24), and this is outlined below:

> *In the order of declining priority:*
>
> *1. Date of first histological or cytological confirmation of this malignancy (with the exception of histology or cytology at autopsy). This date should be, in the following order:*
>
> *(a) date when the specimen was taken (biopsy)*
>
> *(b) date of receipt by the pathologist*
>
> *(c) date of the pathology report*
>
> *2. Date of admission to the hospital because of this malignancy.*
>
> *3. When evaluated at an outpatient clinic only: date of first consultation at the outpatient clinic because of this malignancy.*
>
> *4. Date of diagnosis, other than 1, 2 or 3.*
>
> *5. Date of death, if no information is available other than the fact that the patient has died because of a malignancy.*
>
> *6. Date of death, if the malignancy is discovered at autopsy.*

2.2 Challenges facing early diagnosis research in childhood and young adults with cancer

CYA cancers are rare and only contributes 1% of new cancer cases in the UK each year, the incidence of cancer increases with age into adulthood, another reason for the predominant focus of early diagnosis research on adult cancers (25). The rare nature of cancer in childhood and young adults often leads to the grouping of these populations together into the 0 to 24 age boundary for research purposes, as seen in the NCIN routes to diagnosis work and the National Audit of Cancer Diagnosis in Primary Care (14, 16). Our American counterparts consider young adults as up to 30 years of age (26, 27). However the overall age-range for childhood and young adult cancers is generally agreed as 0 to 24 years inclusive in the UK and across Europe (28, 29).

Within this complex population it is difficult to establish clear subdivisions within UK health services and health services research, this is most evident in the varied boundaries used to define the teenage and young adult population. The age boundaries used to define childhood and young adult populations often varies between population-based cancer registries depending on the country or individual researcher. The UK defines childhood cancer as occurring from birth up to the age of 14-years (inclusive) (3). Teenage and young adult cancer services include patients from 13 to 24 years inclusive (25), however within population-based registries TYA age boundaries are also defined as 15 to 24 years inclusive (30). The divisions made between cancer care services for children, teenagers, young adults and older adults in the NHS are founded in the variations in disease profiles with age as well as the recognition of the spectrum of physical and psychological development that occurs across the 0-24 year's age range. Companionship of peers and age appropriate surroundings can have a profound influence on the overall experience of disease and treatment (25).

The processes that define the onset of adulthood such as physical growth, maturation of personality and development of a sense of independence occur over variable periods in the lives of young people. All these factors will influence awareness of illness and healthcare engagement in CYA's and

potentially the TTD. For the purposes of health research the splitting of the CYA population by smaller age boundaries results in smaller populations for analysis and often reduces the researchers ability to conduct timely studies and may limit conclusions drawn. The issue of grouping or splitting the CYA cancer population for ease of analysis will be further highlighted throughout the subsequent sections of this chapter.

2.2.1 Disease related factors

The type of cancer has a fundamental influence on the time taken to receive a diagnosis; the primary site, the growth rate and the way a tumour spreads can influence. The population of cancers that affect CYA's are multiple and varied, they behave differently between and within diagnostic groups. The majority of tumours in children originate from embryological cell lines unlike the epithelial origins of carcinomatous cancers which predominate in adults (31). The cancers seen in teenagers and young adults are a mixture of those seen in childhood and adulthood, but they often have unique patterns of behaviour. This section will initially explore some of the variations in incidence and survival between CYA tumours, and will go on to discuss disease factors implicated within the literature as potentially influencing the TTD such as the form of the tumour, presenting symptoms, site of the tumour development, rates of growth, size of tumour and how these affect survival outcome.

An established classification system should be applied in order to describe cancers across a population. A number of international classification systems for cancer have been developed to allow international comparison of incidence and survival (32). The main focus of any classification system should reflect the predominant nature of the disease within the study population. It should represent the numerically important groups as well as rarer population specific tumours (33).

The International Classification of Childhood Cancers (ICCC) is an adaptation of the International Classification of Diseases for Oncology and is designed to allow comparison of cancer in the paediatric population. ICCC is in its 3rd edition, published in 2005 and classifies cancer primarily by

morphology. This type of classification is most appropriate in the childhood population due to the often disseminated nature of these cancers' at presentation (33). The ICCC-3 system has a hierarchical structure based around 3 levels:

- 12 broad diagnostic groups
- 47 diagnostic sub-groups
- Optional "Extended Classification" which comprises 2-11 divisions for selected diagnostic subgroups

Within the teenage and young adult population the occurrence of both adult and childhood cancer's alongside unique tumour types has lead to the production of a specific classification system; the Birch *et al* system. This is similar to the ICCC system in that it is primarily based on morphology but is designed to reflect the predominant cancers of teenagers and young adults (34). Within this project a number of different classification systems are encountered. However, to ensure consistency throughout this project the ICCC system will be applied within the 0-24 year age range.

It is important that early diagnosis researchers understand variations in cancer incidence rates and survival for a study population, in order for them to identify the predominant and significant population within the study cohort. The most reliable and up to date incidence figures for CYA cancers are commonly presented for two subpopulations; childhood cancers and teenage and young adult cancers. The next paragraphs will explore the variations in incidence of childhood cancer and then variations in teenage and young adult cancers by diagnosis.

The average number of new cases of cancer in children from 2005 to 2007 within the UK was 1490 per year, accounting for around 0.5% of all new cases of cancer each year across all ages (3). For all childhood cancers leukaemia's constitute around 30%, CNS tumours 25%, lymphoma 10%, soft-tissue sarcoma, sympathetic nervous system tumours and renal tumours between 6-7% each, carcinoma, germ cell tumours and retinoblastoma 3% each, hepatic and other or unspecified tumours around 1% each (3). The UK has a below average incidence rate compared to the rest of Europe, 134 versus 141 age-standardised incidence rates (per

million) (35). There is a general trend to an increasing incidence of childhood cancer in recent years that can only partly be attributed to changes in diagnosis and registration of cancer in childhood. Other factors considered as contributing to this increasing incidence include life-style choices or changes in exposure to carcinogenic agents both to the child and the parents (35).

Cancer in the 15 to 24 year age-range accounts for around 0.5% of all new cancers each year in the UK. The distribution of cancers within teenage and young adults in the UK is different to that in children. Nonetheless, the increasing incidence of cancer in children in recent decades is mirrored in the older population. Leukaemia contributes to a lesser extent in 15-24 year olds and there is a higher proportion of lymphoma, germ-cell tumours, carcinoma and bone sarcoma (36). CNS tumours are still common but contribute a smaller proportion of overall cases of cancer compared to the childhood population, neuroblastoma, retinoblastoma and renal tumours are very rare in the 15 to 24 year age group (4, 36, 37).

The outcome for a population with cancer is an important consideration for the early diagnosis researcher, not only when comparing results but also for targeting of effective interventions. Survival is the most commonly investigated outcome in cancer research as discussed in the Methods (Chapter 4). The next two paragraphs will consider survival for CYA cancers.

The five-year survival rates for all childhood cancers have increased over the last 40 years from under 30% in the late 1960's to under 80% by 2005 (3). Survival rates are not equal across all types of childhood cancer (3, 5) Retinoblastoma has almost 100% survival at five-years, such favourable survival has been related to a number of factors, some of which are listed:

- Screening of new-borns and infants identified at high-risk of retinoblastoma due to the high proportion of heritable disease with known autosomal dominant inheritance pattern.
- Neonatal screening for sporadic cases.
- Improved treatment in specialist retinoblastoma centres within the UK.

Acute Lymphoblastic Leukaemia is the most common cancer in childhood and the survival rate in the UK is around 88%, this has improved greatly over

the last 40 years, the success relates to identification of effective anti-neoplastic agents and progressive refinement of treatment protocols through ongoing clinical trials. The most recent trial, UKALL2011, has focused on reducing the burden of treatment whilst maintaining survival through risk stratification and modifications of therapeutic regimes. The success of ALL has not been realised in all childhood tumour types; osteosarcoma has seen little improvement in five-year survival since the early 1980's and survival rates in the UK remain static in children at <55% (38). Survival figures for CNS tumours (ICCC group 3) have improved from 40% five-year survival in the late 1960's to 70% by 2005, however there are gaps in survival within the population of CNS tumours (39). For some tumours the outcome remains dismal, for instance in cases of diffuse intrinsic pontine glioma more than 90% of children will not survive beyond two years post diagnosis (23). Alternatively the five-year survival for astrocytoma in childhood is around 80% (3). Shifting patterns in survival are not confined to inter-diagnosis variation, survival rates can vary within a specific diagnostic group, for example, neuroblastoma has a highly favourable outcome if diagnosed below the age of 18 months compared to all other ages (40). This pattern of variable survival with age is seen in Wilms' tumours where five-year survival in one to four year-olds is up to 86% compared to 70% five-year survival if diagnosed between 10-14 years (41).

The improving trends in survival over time seen in childhood cancer are reflected to a lesser degree in teenagers and young adults. In the England five-year overall survival between 1979-1984 stood at 63% this rose to 74% between 1996-2001(29). Survival patterns vary within this population, the younger end of this age range have better survival rates for Leukaemia and CNS, whereas older patients with GCTs' fare better (29, 30). Unfortunately survival has remained relatively static for certain tumour types such as high-grade glioma, bone sarcoma or soft-tissue sarcoma (29). As with most paediatric cancers females do better than males, the exception being germ cell tumours.

The age boundaries used to define the teenage and young adult cancer within health research are not as well defined as those for childhood cancer. The publications by Birch and Alston on survival and incidence of cancer in

teenagers and young adults in England published in 2008 defined the age boundaries for the population as 13-24 years inclusive (29, 36). Croucher published a paper on trends in incidence and survival in teenagers and young adults in the south-east of England using different age boundaries 15-24 inclusive (30). Arora published a paper looking at contrasting incidence rates for cancer in young people across England and India and identified the study age boundaries at 15 to 29, Birch and Alston were also authors on this paper (42). There are differing age groups within the boundaries set by these authors. Birch and Alston define three sub-groups (13-16, 17-20, 21-24) (29, 36) compared to two groups in Croucher (15-19, 20-24) (30). The narrow age-ranges used by Birch and Alston in the 2008 paper allow clear representation of changing patterns in incidence and survival by diagnosis and sexes across the teenage and young adult population. However, the five-year population boundaries used by Croucher reflect the predominant divisions used across childhood, teenage and young adult research. The use of five year age bands has been advocated in a recent publication from the child and young people's health outcomes forum in response to the challenges set out in 'Getting it right for children and young people' published in 2010 (43). The independent forum indicates the use of five year bands will improve local, national and international of data on incidence and outcomes of general health in children and young people (44). Within this publication they also recommend several outcome indicators for child health, one of which is "***time from NHS presentation to diagnosis or start of treatment"*** (44).

The previously highlighted variations in incidence and survival allude to the heterogeneity of this population of tumours. Figure 2.2 identifies several disease related factors that may influence the TTD. This section will discuss some disease related factors that have been implicated within the literature as potentially influencing the TTD for CYA cancers, these include:

- The form of the tumour
- The presenting symptoms
- The site tumour development
- The growth rate of a tumour

- The size of the tumour

CYA cancers can be described by their form, as either solid tumours, such as renal or CNS tumours or alternatively as liquid cancers such as leukaemia. The form of the tumour can influence the presentation of the tumour and the TTD. The most common liquid tumour seen in CYA's in the UK is acute lymphoblastic leukaemia (ALL), this is a neoplasm of lymphoblasts that occurs following dysregulation of normal haemopoetic stem cells (45). ALL presents most frequently with disseminated disease including signs of bone marrow infiltration often manifesting as clinical signs of anaemia, thrombocytopenia or neutropenia (45). There are multiple organs that can manifest signs of ALL, for instance the central nervous system, testicles, mediastinum, bones, joints and eyes, but no primary site of disease. ALL can therefore present in a variety of guises, mimicking a number of more common presentations such as infections or inflammatory disease and may not necessarily be the primary differential diagnosis. There are no clear staging criteria for ALL as patients present with widespread disease. Therefore a risk stratification system for the disease has developed based upon other factors such as age at diagnosis, cytogenetics and white cell count at presentation.

In contrast to ALL, Nephroblastoma is a solid tumour also known as Wilms' Tumour and is the predominant renal tumour in childhood, it accounts for 6-8% of all childhood malignancies and less that 1% of TYA malignancies (46). Wilms' tumour commonly presents with a painless mass in the abdomen that may be found incidentally on abdominal examination or brought to a healthcare professional's attention due to increasing abdominal distension. Haematuria, pain, weight loss, persistent urinary tract infections and constipation can all be the presenting symptom in this tumour as well as other more unusual complaints such as hypertension, coagulopathy due to acquired von Wilebrands disease or as an emergency presentation with an intra-abdominal haemorrhage (45). Only around 10% of patients have metastatic disease at presentation and the lungs are the most common site of metastatic spread (45). Overall survival is high, reaching 80-90% five-year survival (41). Survival for Wilms' tumours varies widely by the extent of the disease which is defined by the stage at diagnosis; stage IV, metastatic

disease at presentation is associated with a poor survival just above 50% (31). Earlier diagnosis in Wilms' tumours before stage IV disease has developed would theoretically improve the outcome.

The site of tumour development in solid tumours can influence the TTD. Certain tumour sites in the body such as the abdomen may predispose to a more insidious onset of disease meaning more extensive disease at presentation, as discussed in Wilms' tumour above. Neuroblastoma also primarily presents in the abdomen, 60% of cases have a primary tumour in the abdomen with 30 to 50% located in the adrenal gland, other common places are the pelvis and cervical region (45). This is the most common solid tumour of infancy and the second most common extra-cranial malignant tumour of childhood with the peak incidence at 18-23 months, 80% of neuroblastoma cases occurs below the age of four years-old (31). Approximately 50% of patients will present with metastatic disease, the common sites are liver, lymph nodes, bone, bone marrow and skin (31, 33).

Stage of disease at diagnosis is a factor that contributes significantly to the risk stratification of this neuroblastoma with higher stage disease indicating higher risk of a poor outcome, however other factors such as age and cytogenetic markers also influence risk stratification. Within the UK evidence has suggested neuroblastoma is generally diagnosed at a later stage than Germany and France and the UK appears to have a worse outcome (47). From this evidence it has been suggested that reducing the TTD and diagnosing the disease at an early stage could impact on outcome (47).

Nationwide mass screening programmes of infants for neuroblastoma were introduced in Japan in 1985, this was done by measuring urine catecholamines at 6 months of age (48, 49). The programme identified tumours with mostly favourable biological markers, however in the few tumours with less favourable biology there appeared to be a benefit from an early diagnosis (48, 49). Trials have also been conducted in Austria, Germany and Canada, which all identified an increase in the incidence of tumours but no impact on overall survival (50-52). This finding is felt to be due to the identification of asymptomatic tumours that naturally regress (51, 52). However, in light of the failed screening programmes and significant

influence of tumour biology in the risk stratification of neuroblastoma the link between TTD and outcome remains unclear.

In rhabdomyosarcoma (RMS) the primary site of disease has a major influence on how the disease presents. This is the most common soft-tissue sarcoma of childhood, this solid tumour accounts for around 60% of STS in the 0-14 year age range, the median age is five years with a peak incidence of two years of age (53). RMS can present in a number of anatomical sites, with head and neck RMS making up 40% of childhood RMS followed by 25% genitourinary, 20% in the extremities, 10% trunk walls and 10% other sites. The outcome for a patient with RMS is influenced by a number of factors, including; the site of primary disease, as well as age at diagnosis, size of primary tumour, extent of disease at diagnosis, histological sub-type and prediction of treatment sequalae. The site of the disease will alter the symptoms with which it presents, certain symptoms may cause more alarm, prompting the patient to seek help quickly and triggering rapid investigation and referral by healthcare professionals.

The rate a tumour progresses varies between and within diagnostic groups. Variations in the rate of tumour progression can be examined across the lymphoma population. The doubling time of a tumour indicates the time taken for a group of cells to double in size and indicates the speed of proliferation. The clinical behaviour of lymphomas varies widely and there are multiple discrete entities within this broad diagnostic group, the doubling time has been reported for a number of types of lymphoma affecting all ages: Burkitt's lymphoma predominantly seen in children and young adults has a very short doubling time, 24 to 48 hours, compared to follicular lymphoma, nearly always seen in older adults, which can have a doubling time of up to 1 year (54, 55). The speed at which the clinical features of lymphoma progress can indicate the specific diagnosis. It is generally accepted within CYA lymphoma's that Non Hodgkin's lymphoma has a more rapid progression and presentation than Hodgkin's lymphoma (45). Morley-Jacob published an update on lymphoma in children and young adults in 2011 and highlighted early referral of suspected cases for biopsy as a key learning point, presumably with the aim of improving stage at diagnosis and reducing treatment burden(56). Theoretically improving the TTD within the

lymphoma population has the potential to impact on disease and treatment related morbidity rather than improving overall high rates of survival.

The rate of tumour development can indicated the aggressiveness of a tumour. Within CNS tumour group the highly varied biological behaviour has necessitated the assignment of a “malignancy scale” to accompany the morphological diagnosis. This scale predicts the biological behaviour of the tumour, giving an indication of how aggressive a tumour is and how rapidly it will progress, which in turn influences the choice of treatment (57). The WHO grading system is applied internationally and are four grades applied within this system; grade one reflects lesions with a low proliferation index with most tumours amenable to surgical resection and more favourable cure rates, with the exception of some brainstem and optic pathways tumours; grade two tumours have a more infiltrative nature and a greater tendency to recur and progress to a higher malignant grade; grade three tumours have histological evidence of malignancy and require adjuvant radiotherapy and/or chemotherapy to give the best chance of cure; grade four tumours are highly malignant, rapidly progressing tumours that are generally associated with less favourable outcomes (57). The grading of tumour is important and guides treatment, prognosis and late-effects of treatment.

CNS tumours are the second most common tumour of childhood accounting for around 25% of tumours in this age group, their incidence reduces in teenagers and young adults to a nadir between 15-20 years of age, overall CNS tumours account for 9% of tumours in teenagers and young adults (4, 58). These tumours are the leading cause of cancer related deaths under the age of 25, and for those who survive around 60% will have a significant neurocognitive deficit.

Astrocytoma is the predominant CNS tumour across the CYA population, these tumours fall within the histopathological diagnosis of glioma tumours which arise from glial cells (59). Astrocytomas account for around 40% of childhood CNS tumours and a higher proportion of CNS tumours in the 15-24 years age group (3, 4). The morphological diagnosis of astrocytoma, ICCC diagnostic sub-group IIIb, identifies a group of tumours with highly varied patterns of behaviour that range from pilocytic astrocytomas that are

benign in nature with excellent survival rates to anaplastic astrocytomas an aggressive neoplasm with a poor outcome.

Low-grade gliomas generally have a very good prognosis, 95% five-years overall survival and comparatively low treatment burden compared to high grade gliomas, with most successfully treated by surgery alone. High-grade gliomas are highly malignant aggressive tumours that are defined by site and histological phenotype they usually occur between the ages of 5 to 10 years-old in otherwise healthy children (60). Important predictors of outcome in high-grade glioma are the degree of surgical clearance, the use of radiotherapy to the resected tumour bed and more recently the addition of high-dose chemotherapy have led to improvements in survival for high-grade tumours (61). Despite these precautions the often highly aggressive and invasive treatment required in brain tumours results in significant neurocognitive deficits and disability in long-term survivors (62).

The size of the tumour at presentation is an important prognostic factor for certain bone sarcomas and soft-tissue sarcomas, a point previously discussed in relation to rhabdomyosarcoma. Malignant bone tumours account for around 4% of childhood cancers and 7-8% of teenage and young adult cancers in the UK (1, 3, 38). Osteosarcoma accounts for 50% of bone sarcomas in the CYA populations, Ewing's sarcomas account for 40% of bone sarcomas in childhood and 30% of bone sarcomas in teenagers and young adults. The survival from bone sarcoma is poor compared to most other tumours within this population; they are the fourth most common tumour of teenagers and young adult but the second leading cause of mortality. Five-year survival for osteosarcoma is similar in childhood cancer (~55-60%) and young adult cancer (~50%), there is a wider survival gap seen in Ewing's sarcoma by age with children (~60%) having a more favourable outcome than young adults (~30 to 40%) (2, 38). In osteosarcoma a poor prognosis is associated with a large tumour volume, incomplete tumour resection, a poor response to chemotherapy, the presence of metastases at diagnosis, axial site of tumour, age over 40 years and higher grades of tumour (63). Factors associated with a poor outcome for Ewing's sarcoma are primary tumour volume greater than 200ml, age greater than 14 years, bone metastases, bone marrow involvement and lung

metastases, these factors have been combined to produce a scoring system for Ewing's sarcoma (64).

The size of tumour at diagnosis is associated with a worse outcome in Ewing's sarcoma with a specific cut off of 200ml volume however such a threshold has not been established for osteosarcoma. In adult studies of bone sarcoma links have been shown between the time taken to receive a diagnosis and outcome for Ewing's sarcoma but not osteosarcoma, this may reflect the influence of variable tumour grade in osteosarcoma. The higher the grade of tumour the more aggressive the malignancy and the faster tumour growth, therefore rapid changes in size of mass will be noted by patients and highlighted to healthcare professionals. Due to the rarity of these tumours the numbers of patients available to study in the childhood and young adult cancer population are small.

2.2.2 Patient related factors

Walter identifies several patient factors that contribute to determine the TTD (Figure 2.2). The type of cancers affecting CYAs varies by the patient demographics such as the age and the sex. Across Europe the incidence of childhood cancer is highest in the 0-4 year age-range in both males and females, males have a higher incidence of cancer most notably in the 5-9 age range (65). There are variations in incidence between the 15-19 and 20-24 year age ranges with the latter having a higher incidence of cancer (30). Cancer incidence varies between the sexes and different incidence profiles are seen for cancers between males and females in the UK in the 15-24 year-old group. Testicular cancer is the most common cancer type in the male group 15-24 making up 27% of the total cancers for this age group (1), however in females of the same age malignant melanoma and lymphomas both provide around 17% of the overall cancers for this age group (1). The approach to healthcare is different between the sexes, in a study by Fern *et al* of young people with potential cancer symptoms the females attended more frequently than males and older females attended more frequently than younger females (66).

The age and sex of a patient can also have an effect on healthcare engagement. The age of a patient has a significant influence their pre-diagnosis experience, young children cannot access healthcare independently and rely on their parents or carers to recognise illness and act for them. Therefore the experiences and beliefs of the parents can dictate the time spent prior to diagnosis. TYAs face a number of potential hurdles when engaging with healthcare services; reduced parental surveillance combined with developing an understanding of their bodies and their health may result in a prolonged patient interval as they struggle to understand changes within their body that are manifestations of disease. This age group also face a number of major life events as they transition into adult life, including leaving home, starting higher education, getting a job or starting a family all of which can be stressful and impact on health and well-being.

Where a person is born and lives may impact on the chance of developing cancer, the incidence of cancer varies across Europe with the highest rates in North Europe (67). The incidences rates for CYA cancer vary across the UK, most notably in teenage and young adult cancers where there appear to be a clear north south divide (28).

2.2.3 Healthcare related factors

The point of access and route taken through a healthcare system to reach a definitive diagnosis of cancer will depend on the availability, structure and delivery of healthcare within a country or region. As discussed in the introduction the routes to diagnosis for cancer in England has been studied at the primary care level (Primary care audit) (16) and throughout secondary care (NCIN routes to diagnosis) (14, 15). The NCIN routes to diagnosis document identifies 269 distinct routes to diagnosis for cancer and divided these into 8 groups of routes to diagnosis, these are listed below (14):

- Screen detected
- Two week wait – urgent GP referrals with suspected cancer
- GP/outpatient – Routine or urgent referrals but not part of the two week wait referral route

- Other Outpatient
- Inpatient Elective
- Emergency presentations
- Diagnosed on death certificate
- Unknown

Both the NCIN and Primary care documents suggest that the 0-24 year age group of patients diagnosed with cancer access healthcare in a different manner to the majority of their adult counterparts, with larger proportion accessing specialist care through emergency routes. These two documents provide an insight into the routes to diagnosis for 0 to 24 year olds with cancer but they do not focus on the specific challenges faced by this age group. There are suggestions in the literature that the referral pathways for CYA cancers may vary from adults, the role of the 2 week urgent referral pathway in childhood cancer has been questioned. Bragonier & Kenyon 2012 published a retrospective review of 312 two week referrals, highlighting the extremely low pick up rate of the 2 week referral pathway in childhood cancer (68). The vast majority of cancers in this population being diagnosed via alternative referral routes (69). The profile of admissions within secondary care services will be explored within this study. Access to health care is influenced by geography across the globe and even within the UK there are variations in healthcare provision that impact on outcome for CYA's in the general and not just those with cancer (70). The socio-economic status of a person can impact on their access to healthcare, even in the UK with a free to all at point of access healthcare system. The NCIN routes to cancer diagnosis publication in 2010 identified more affluent patients as less likely to present with cancer through an emergency admission (71).

The diagnosis of cancer in CYAs is a rare event and the symptoms they present with are often common to other more prevalent diagnoses, for this reason many healthcare professionals don't consider cancer in young patients (72). This can result in a protracted and negative experience within the healthcare system prior to diagnosis.

2.3 Conclusions

Based on the evidence presented within this chapter it is clear that there are many challenges facing all early diagnosis researchers, particularly in relation to CYA cancer. Low incidence rates of cancer within 0-24 year age range results in a limited sample for study. Highly variable patterns of incidence by age, sex and diagnosis coupled with variable focus points for cancer classification systems and different age boundaries for study populations act as impediments to clear comparisons between studies in 0 to 24 year age group.

The site, size, pattern of spread and speed of tumour proliferation can all influence the presentation and progression of the disease and therefore the TTD. Variations in behaviour of tumours between and within diagnostic groups makes intra-population comparisons difficult and leads to uncertainty regarding conclusions drawn in relation to the time taken to reach a definitive diagnosis across CYA cancer as a whole.

Every advance in the fields of paediatric, teenage and young adult oncology results in a constant shift in the screening, diagnosis, treatments and follow-up for cancer patients. The resultant changes in incidence and survival must be carefully interpreted by epidemiologists and health providers in order to inform future advances in the field.

Early diagnosis research in any age group should be underpinned by a theoretical framework and accompanied by clear definitions of significant milestones that outline the study intervals being scrutinised. The Aarhus statement provides such guidance. The systematic review that makes up the next chapter will add to the information gathered relating to early diagnosis thus far and explore the literature relevant to TTD in CYA cancers, assessing the strengths and limitations of CYA early diagnosis research to date.

Chapter 3 Systematic Review

3.1 Introduction

Improving the early diagnosis of cancer in all ages has been identified as a key factor to address in cancer care within the NHS (6, 11). Delays to the diagnosis of cancer have been implicated as contributing to poorer survival outcomes for both adults and CYA with cancer in the UK compared to many of our European counterparts. This systematic review will identify early diagnosis research that focuses on delayed diagnosis and TTD for cancer in CYA populations across the globe. The quality of methods used and the terminology applied to the field of early diagnosis research in CYA will be explored. A particular focus will be placed upon the application of the term “delay”, a term that carries negative connotations and is synonymous with early diagnosis research to date. The review will seek to investigate how TTD varies within the CYA population, identify factors that contribute to a prolonged TTD and explore the association of TTD with outcome in CYA cancers. Finally pulling all this together to inform the development of a robust reproducible method for the study of TTD in secondary care services within Yorkshire.

3.1.1 Early diagnosis research within the adult cancer population

Chapter 2 identifies several reasons for the predominant focus on adult cancers in early diagnosis research to date, especially in the UK. It is therefore important to consider the findings of major adult systematic reviews. Improving TTD for adult cancer patients has been repeatedly highlighted as a key area for improvement of cancer care within the NHS (6, 9, 11), however the impact of the time taken to achieve a diagnosis of cancer on patient outcomes is unclear (73).

Systematic reviews relating to delayed diagnosis in breast, upper gastro-intestinal tract, and colorectal cancer have been identified (74-77). These look at the association between delayed diagnosis and outcome, the factors that influence pre-hospital delay and delayed presentation amongst a number of other aims (74-77). The systematic review by Richards *et al* 1999 of the "…influence of delay on survival in breast cancer patients…" linked delays to diagnosis of 3-6 months with a lower survival in breast cancer patients (77). This review highlights the scope of adult early diagnosis to include large study populations, included are 87 studies published between 1907 and 1996 involving over 101 945 patients(77). The extensive study population and the author's attempts to deal with bias in the construct of the review add validity to the conclusions. However the methodology didn't discuss what constitutes a "delay" to diagnosis, with this term relating to the entire time a patient spends prior to their diagnosis. In such studies the term "delay" carries negative connotations, inferring undue prolongation of the time prior to the receiving of a diagnosis of cancer.

The results of these studies have been used by healthcare planners, government departments, cancer charities and those organisations involved in research planning as fuel to focus efforts on reducing delays in diagnosis within UK healthcare.

3.1.2 Previous systematic reviews of delayed diagnosis in childhood and young adult cancer

To date two reviews have focused on TTD for childhood and young adult cancer. In 2007 Dang-Tan *et al* published a review of 23 epidemiological papers, including study populations under the age of 30 (78). The authors identify the early diagnosis of CYA cancer as "…a fundamental goal of oncology.." citing the opportunity of timely treatment in early stage disease (78). This review provides an analysis of the factors associated with delay to diagnosis in childhood and young adult cancer grouping factors as patient, cancer or healthcare-related (78). The authors identify diagnostic delays as being longer than patient or parent delay. The main factors relating to delays in diagnosis are patient's age, parental education, presenting symptoms,

tumour site, stage at diagnosis and initial point of medical contact (78). The authors touch on the challenges of early diagnosis research, in particular they discuss the varied terminology applied within the field and outline a set of milestones within the cancer care pathway. Unfortunately they continue to use the term delay to refer to the entirety of the time-intervals being studied and the limitations in the methodology used within the reviewed studies are not explored (78).

The second review published by Brasme *et al* in 2012 is more extensive and identified 98 papers relating to the "distribution, determinants and consequences" of TTD of paediatric cancers (79). Papers containing adults were included if at least 70% of the study population was paediatric. This review again provides an extensive analysis of the factors associated with delay, identifying and discussing some of the medico-legal issues surrounding delayed diagnosis in childhood cancers (79). The authors identify several factors associated with prolonged delays to diagnosis including older age, level of qualification of the medical professional at initial contact, the tumour site and histology and presentation with non-specific symptoms (79). They also conclude that delayed diagnosis is associated with a poor outcome in retinoblastoma and possibly leukaemia, nephroblastoma and rhabdomyosarcoma but no association was shown for CNS tumours, osteosarcoma and Ewings sarcoma (79).

Caution must be taken in relation to the conclusions drawn from the Brasme paper. The statistical methods of weighted means of the median and mean TTD are of debatable efficacy, an issue that will be further explored later within this chapter. Brasme *et al* provide little insight into the intrinsic limitations of the applied methodologies within early diagnosis research in children and young adults, nor is there any proposed solution to the issue of how we interpret delay and TTD (79).

There are also a number of systematic reviews focusing on the presentation of cancer in CYA, such as Wilne *et al* 2007 that draws together CYA literature to identify patterns in clinical presentation of CNS tumours in childhood (80). The systematic review by Wilne *et al* developed a guideline to assist in the identification and referral for childhood CNS tumours.

Subsequently this has evolved into the national campaign, known as HeadSmart, aimed at raising awareness of the symptoms and signs of brain tumours children and teenagers, in both patients, carers and health-professionals (80).

3.2 Methods and materials

3.2.1 Data search

A literature search from 1948 to May 2012 was undertaken, using a predefined search protocol in: Medline, EMBASE, EMBASE Classic, CRD databases, Cochrane Library, Medline in-Process & Other Non-Indexed Citations. Citations searches, reference lists and colleague recommendations were also reviewed. One reviewer (CL) screened each title and abstract for inclusion, and a second independent reviewer (MvL) checked a random sample of potential citations (20%). Concordance between reviewers was high, and differences of opinion resolved by discussion.

3.2.1.1 Inclusion criteria

Primary research studies published in English were considered if they:

- Focused on children and young adults (0 to 30 years), the upper age limit was extended to 30 to include US studies involving young adults.
- Quantified the time between onset of symptoms and definitive diagnosis of cancer for at least 15 cases.
- Used diagnostic groups similar to the International Classification of Childhood Cancer (ICCC) (33).

Papers were excluded if they focused only on melanoma, due to the inconsistency of classification and registration of skin cancers within this study population.

3.2.1.2 Quality assessment

Methodological quality of the papers was assessed by the scoring systems outlined in Macdonald *et al*, 2006 (81) and against the Aarhus checklist (19).

A scoring system for descriptive studies as used to assess general methodological quality due to the majority of the studies having a cross-sectional retrospective design. This system assessed studies by:

- The presence of a hypothesis or research question
- Whether the source of the cases was identified
- Whether inclusion/exclusion criteria were defined
- Whether the sample size stated
- Whether there was a discussion of bias
- Whether the analytical method was described

Each of the 6 criteria was scored on a present (1) or not present basis (0) basis, and an overall score out of 6 generated. Studies with robust well constructed methods have higher scores and poor quality methods score lower.

As previously discussed in Chapter 2 the Aarhus statement published in 2012 provided specific guidance for the early diagnosis researcher (19). This publication included a checklist to aid the design of consistent and transparent methods in early diagnosis research. This checklist consists of 20 items; 7 relating to “definitions of time points and intervals”; 13 relate to “measurement” of which 3 address the context, relevance to definitions and acknowledgement of theoretical framework, 8 relate to questionnaires or interview and 2 to clinical notes review and databases (19). This scoring system was used to generate a score out of 20 to reflect the methodological quality of the studies specifically relating to early diagnosis.

3.2.1.3 Data extraction

Data extracted included: study type and period, published year, country, cancer type, explanatory factors, sample size and numbers excluded. Descriptive statistics extracted included: range, inter-quartile range, median, mean and standard deviation of TTD along with summary data for patient

and professional intervals. We also recorded authors' defined study time period.

3.2.1.4 Included studies

Of the initial 1665 abstracts, 65 full text articles were assessed, and 32 papers met the inclusion criteria and were reviewed. The studies varied; by country of publication, with papers predominantly published in Europe and North America; by diagnostic group, 11 papers studying multiple diagnostic groups, the rest focusing in on one single broad diagnostic group; and by age of study population though the majority focused upon children and young teenagers. The studies included sample sizes ranging from 29 to 2896, the median sample size was 139 cases and the mean 344, and these results highlight the predominantly small study populations. The limited number of large samples size studies may reflect the rare nature of CYA cancer and the limited availability of large regional or national CYA cancer registries across the globe. Table 3.1 outlines the summary information and data presented for each study.

Table 3.1 Summary data presented within the papers reviewed

Author/Year/Country	Cancer	Sample size	Age-Range (Years)	Median (days)			Mean (days)	Standard Deviation	Range (Days)
				TTD	PI	DI			
Brasme 2011 France	Medulloblastoma	166	<15	65					3-457
Klitbo 2011 Denmark	Brain Tumours	46	0-17	51	6	3			0-365
Shay 2011 Israel	Brain Tumours	330	0-18				234		3-3283
Wilne 2011 UK	Brain Tumours	139	<17	100					0-2520
Hayashi 2009 Japan	Brain Tumours	54	0-15	20.5					
Kukal 2009 Swiss	Brain Tumours	315	0-16	60	14	14			0-3480
Reulecke 2008 Germany	Brain Tumours	245	<20	24			59		0-795
Mehta 2002 Canada	Brain Tumours	104	≤17	91			222		
Halperin 2001 US	Medulloblastoma	122	0->17*				100	±149.8	
Edgeworth 1996 UK	Brain Tumours	74	0-16				140	203.7	0-910
Bai 2011 China	Retinoblastoma	572	0-14	61			125	±179.4	3-1094
Wallach 2006 Swiss	Retinoblastoma	139	Children*				114		
Rodrigues 2004 Brazil	Retinoblastoma	327	≤12	91			176	200.6	3-1459
Wirix 2002 Belgium	Retinoblastoma	33	0-7				97		61-365
Goddard 1999 UK	Retinoblastoma	100	<9*	56	18	14			7-672
Chotel 2008 UK	STS (Synovial)	35	3-16				686^		2-2548
Ferrari 2010 Italy	STS	575	≤21	61					7-1824
Goyal 2004 UK	Bone tumours	115	4-22	116					30-1398
Yang 2009 Hong Kong	Osteosarcoma	51	3-20	61	30	21			4-361
Crawford 2007 US	CNSGCT∞	30	6-17				255		
LaQuaglia 1992 US	Adenocarcinoma	29	≤21	61					12-547
Loh 2012 Singapore	Multiple	390	0-18	37	21	8			1-1982
Cecen 2011 Turkey	Multiple	329	0-19	53	3	28			0-2520
Stefan 2011 S.Africa	Multiple	194	0-15	34	5	20			2-1826
Haimi 2010 Israel	Multiple	315	0-20	49	7	28	110	188.9	0-1456
James 2010 Nigeria	Multiple	64	1-14	92	14	62	169	196	15-1098
Dang-Tan 2008 Canada	Multiple	2896	0-19	30	9	8			13-69ø
Martin 2007 US	Multiple	235	15-29				75	24.4	
Thulesius 2000 Sweden	Multiple	64	0-16	63+	35+	21+			1-1393+
Saha 1993 UK	Multiple	184	0-15	28~					
Pollock 1992 US	Multiple	2665	0-29				(50)*1		
Flores 1986 US	Multiple	79	<20				(77)*2		

** Age ranges not clearly defined, however all refer to the study of children*

^Chotel – Paper reports mean PI as 301 days (0-1092) and mean DI as 350 days (0-2534)

øDang-Tan – The interquartile range is represented

∞CNSGCT – Central Nervous System Germ Cell Tumours

+Thulesius – Reported values for brain tumour patients, paper also reports Leukaemia TTD 21, PI 1, DI 0

~Saha – Median TTD 28 days (7-364 days) for 101 males and 28 days (7-504 days) for 83 females

**1Pollock – This is a combined mean from the mean for each cancer group, included to display paper in review results*

3.2.1.5 Quality of studies, applied methods and summary statistics

All included papers were observational studies and data collection was predominantly retrospective using patient interviews or case note/cancer registry review. The majority of studies were cross-sectional focusing on an institution or clinic, the others population-based retrospective cohort studies.

Very few studies offered a clear hypothesis and most cross-sectional studies lacked clearly defined inclusion criteria and did not deal with bias in their construct (82-86). All papers scored moderate to low scores when applying the descriptive study assessment tool, and a similar outcome was seen when applying the Aarhus checklist (19). Only 1 study set out a clear set of milestones (87). The circumstances in which a delay to diagnosis became unacceptable were clearly defined by Shay *et al* 2011, if the patient encountered any of the 6 defined circumstances prior to diagnosis the delay was considered unacceptable (figure 3.4) (88).

Summary statistics presented varied between studies, see Table 3.1. The median and range were often presented for TTD, patient-interval (PI) and the diagnostic-interval (DI) due to extreme outliers and positively skewed distributions. However, the inter-quartile ranges were rarely cited. Furthermore only 4 of 32 papers presented the mean, median, range and standard deviation in combination (89-92).

3.3 Results

3.3.1 Time to diagnosis, patient interval and diagnostic interval

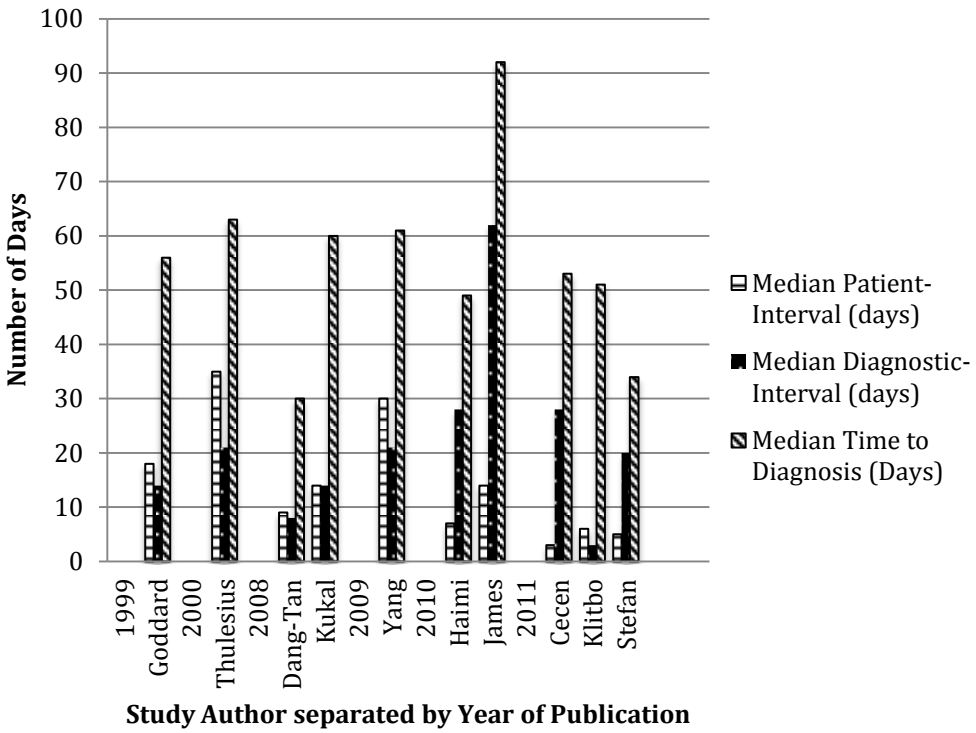
Time between symptom recognition and definitive diagnosis were described as: TTD (78, 88, 93), delay to diagnosis (89, 94-96), pre-diagnostic interval (85, 97-99), lag-time (26, 27, 86, 90, 91, 100, 101) and duration of symptoms (84, 92, 102-105).

“Patient-Interval” (PI) (19, 93) was used to describe the time between symptom onset and first clinical presentation. Other terms used were: patient delay (78, 83, 90, 91, 95, 101, 102), symptom-interval and onset of symptoms to presentation (103).

Diagnostic-Interval (DI) referred to the period from primary engagement with a healthcare professional to definitive diagnosis (also referred to as doctor, physician or healthcare delay) (78, 83, 90, 91, 95, 101, 102).

Initial comparison of median PI, DI and TTD, taken from the 10 studies that included these summary statistics, would suggest that the median TTD is not accounted for by the cumulative value of the median PI and median DI, see Figure 3.1. Further inspection of the summary statistics reveals that this discrepancy is due to the variable skew in each of the distribution for TTD, PI and DI within a study. It is therefore inappropriate to draw conclusions relating to the contribution of PI or DI to TTD based on the descriptive data provided.

Figure 3.1 Comparison of the median time to diagnosis, median patient-interval and median diagnostic interval by year of publication



3.3.2 Time to diagnosis and disease related factors

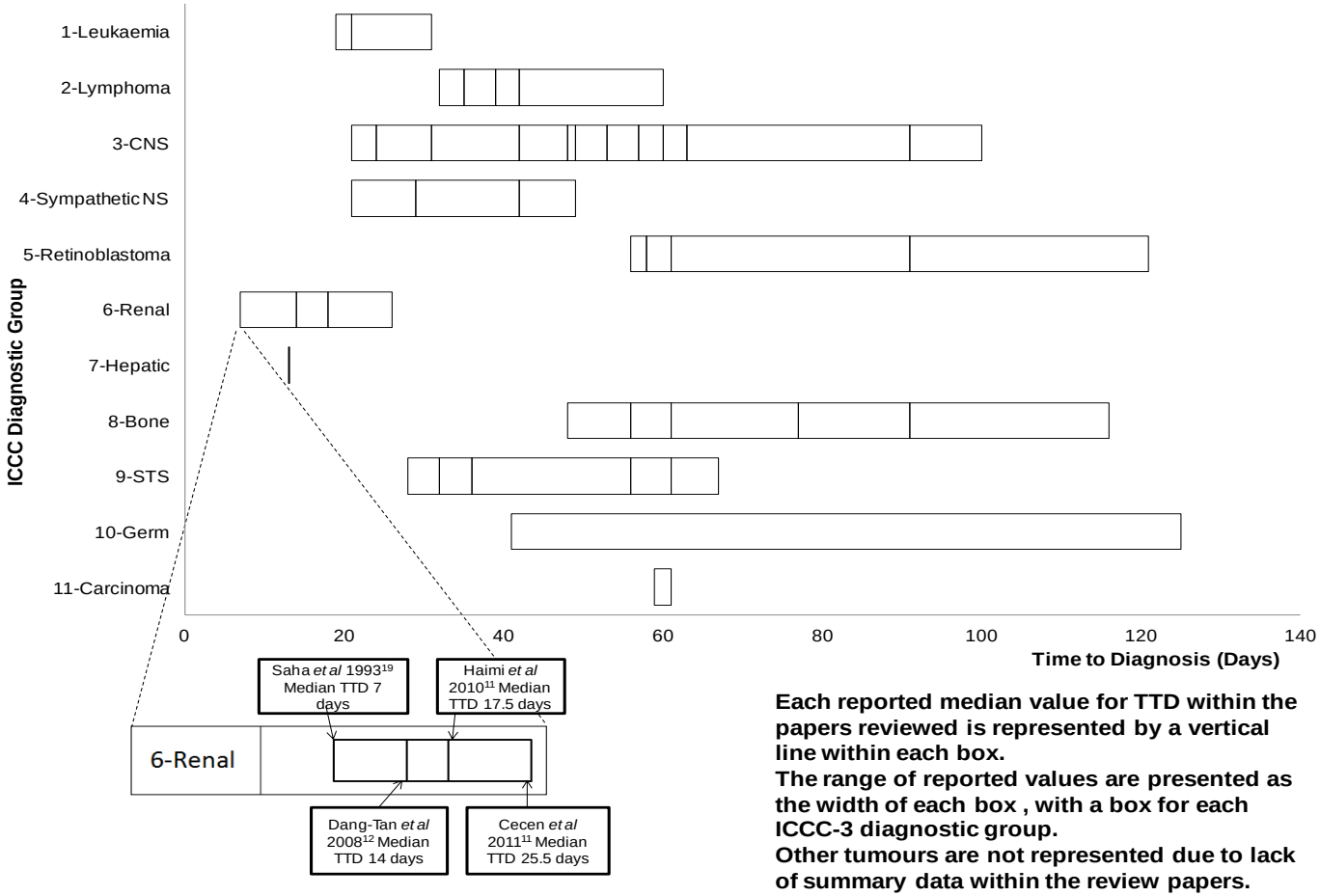
In Chapter 2 the challenges facing early diagnosis research in CYA's were explored and factors relating to the disease, the patient and the healthcare system explored. Specific factors influencing the TTD in CYA's identified in the reviewed literature will be divided into these three key areas and are explored in the next three sections, the final section 3.3.5 will include any identified associations between TTD and outcomes. Thus establishing the present knowledge base for early diagnosis research in CYA's

Eleven studies included multiple diagnostic groups (26, 27, 86, 87, 90, 91, 93, 95, 96, 101, 105), but most focused on individual groups including brain tumours (n=10) (84, 85, 88, 94, 97-99, 106-108), retinoblastoma (n=5) (89, 92, 100, 109, 110), bone tumours or soft-tissue sarcoma's (n=4) (82, 83, 102, 103), central nervous system germ-cell tumours (n=1) (104) and adenocarcinoma of the colon and rectum (n=1) (111).

Wide variation exists between diagnostic groups regarding the frequency of reported median values for TTD. Brain tumours had the most frequently reported median TTD (n=12) and hepatic tumours the least (n=1) (Figure 3.2). A wide range of median values for TTD were reported within each diagnostic group (Figure 3.2). TTD varied by diagnostic group, e.g. there was no overlap between the reported median TTD for leukaemia and lymphoma, whilst the renal tumours appeared to have the shortest TTD. Ten papers reported the median values for the TTD, PI and DI (Table 3.1).

Diagnosis of bone tumours (27, 87), brain tumours (86, 90, 105), germ cell tumours (93) and retinoblastoma (87, 93) were associated with longer TTD compared to leukaemia (26, 87, 105) and renal tumours (Figure 3.2) (86, 87, 90, 93, 105). There were significant variations by sub-type, e.g. medulloblastoma had a significantly shorter TTD than other brain tumours (85, 106), Ewing's sarcoma a longer TTD than osteosarcoma (83) and non-rhabdomyosarcoma soft-tissue sarcomas (NRSTS) a longer TTD compared to Ewing's-family Soft Tissue Sarcomas (82).

Figure 3.2 Graph of median values for time to diagnosis by diagnostic group



Tumour site was associated with variations in TTD, PI and DI in solid tumours. Soft tissue sarcomas (STS) in the bodies' extremities were associated with a prolonged TTD (82) as were peri-articular tumours in CYA patients (102), whereas axial bone sarcoma's were found to have a longer TTD compared to limb tumours (83). Tumours presenting with an abdominal mass or distension such as Wilms' tumours and neuroblastoma have been associated with a shorter TTD (93, 96). The location of a brain tumour was associated with a variable TTD with brain stem tumours exhibiting a longer TTD compared to all other locations (106) and supratentorial midline tumours having a longer TTD than supratentorial hemispheric and infratentorial tumours (85).

A prolonged TTD was associated with a larger tumour volume at presentation in STS in CYA (82, 83). The size of the tumour at diagnosis is associated with a poor prognosis in STS and osteosarcoma. However in CYA with osteosarcoma a prolonged TTD has not been associated with increase tumour size at diagnosis at the present time. This fact may reflect predominant influence on TTD of tumour grade, hence higher grade tumours grow faster and have a worse outcome than slower growing lower grade tumours (63).

The relationship between the stage of disease at presentation and TTD was variable in brain and retinoblastoma studies, with more advanced disease at presentation associated a longer TTD (89, 90, 92, 109). In other brain tumour studies more advanced stage and grade of tumour at presentation was associated with shorter TTD (84, 85, 107). As discussed in the tumour biology section, the grade of a brain tumour influences the outcome and aggressiveness of the treatment strategy, with higher grade tumours growing faster, therefore less time afforded for the body to adapt and presenting symptoms can be severe. Slow-growing lower grade tumours may progress at speeds that allow the body to compensate for changes to brain functioning and the symptoms and signs of resulting deficits more subtle.

In terms of delay, brain tumours were the most extensively investigated group of tumours with presenting symptoms repeatedly highlighted as

factors influencing TTD (85, 88, 97, 99, 106, 107). Prolonged TTD was associated with psychological symptoms, motor dysfunction, ataxia, head tilt, cranial nerve palsies, reduced visual acuity and endocrine or growth abnormalities at presentation (88, 94, 97), however, these findings were not consistently shown (107). The systematic review of childhood CNS tumour presentations by Wilne *et al* 2007 identified 56 symptoms or signs at presentation in 4171 patients in 74 studies. The more common signs and symptoms, identified as the 28 signs and symptoms occurring in more at 5% of cases, were included in further analysis. It is therefore difficult to adequately investigate TTD with respect to the presentation of CNS tumours due to the highly variable patterns of symptoms and signs at presentation occurring in relatively small groups of patients.

3.3.3 Time to diagnosis and patient related factors

The majority of reviewed articles included childhood and teenage populations, 7 papers included cases 20 years or over (26, 27, 82, 83, 90, 103, 111) Martin *et al* 2007 was the only study focusing solely on teenagers and young adults (15-29) (26). Authors often sub-divided the study population into age groups for comparison. However different age divisions were used with the studies and were dictated by the overall age-range, the peak incidence of a tumour or age thresholds for certain treatments.

A longer TTD was significantly associated with older age at diagnosis in a number of studies that involved childhood and young adult bone tumours, leukaemia, lymphoma, brain tumours, retinoblastoma and soft-tissue sarcoma (27, 85, 90, 96). Nevertheless, no correlation between age and TTD was shown in some brain tumour studies and the study focusing solely on teenagers and young adults (91, 95, 107, 108). Most authors hypothesise that as children become young adults they are subject to reduced parental surveillance, have limited knowledge of their own physical health and face hurdles when accessing healthcare (26, 46). Childhood and young adult patients would appear to have a longer TTD than older adults for certain cancers such as bone tumours, possibly due to the rarity of such tumours in

this population and a resulting lack of awareness of the diagnosis amongst general healthcare professionals (112).

3.3.4 Time to diagnosis and healthcare related factors

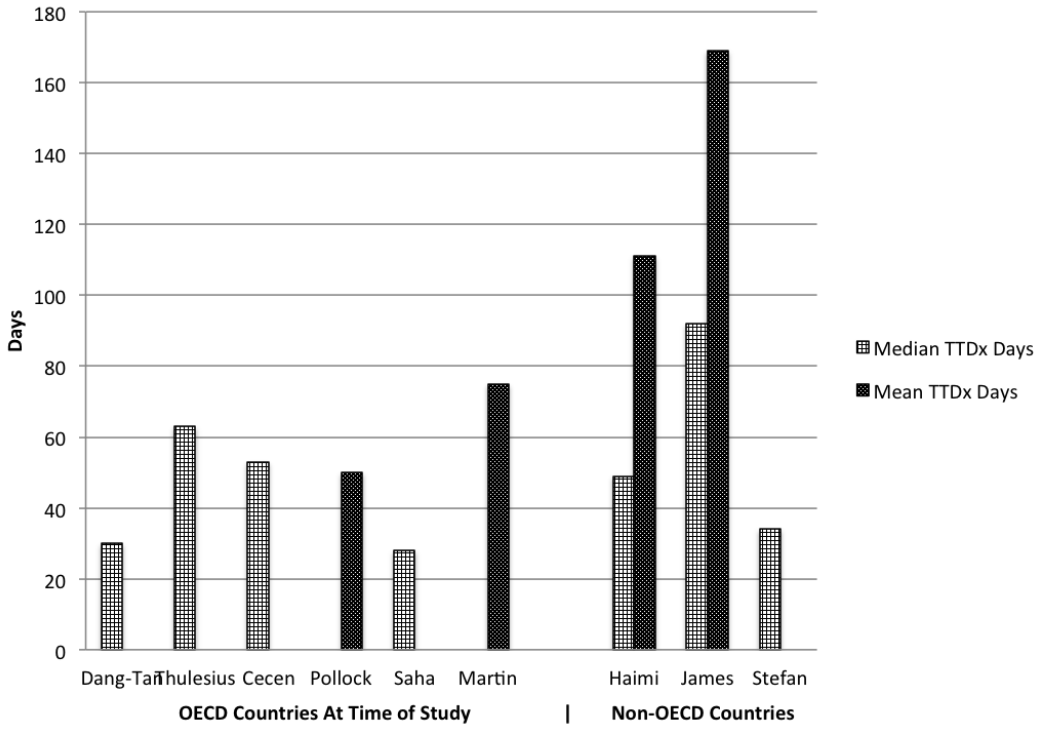
The country of publication varied: Europe (n=15) (82, 83, 85, 86, 93, 94, 97, 99-102, 107-110) with 6 from UK populations, 8 from North America (USA n=6, Canada n=2) (26, 27, 84, 87, 104-106, 111), 4 from Asia (China n=2, Japan n=1, Singapore n=1) (89, 96, 98, 103), 2 from Africa (South Africa and Nigeria) (91, 95), 2 from the Middle East (Israel) (88, 90) and 1 from South America (Brazil) (92). In order to consider TTD on a global scale the studies were divided according to their Organisation of Economic Co-operation and Development (OECD) status at the time of study. The median and mean TTD for studies that represented TTD across multiple ICCC tumour types are plotted in Figure 3.3. Haimi 2010 is included in the non-OECD group as Israel did not undergo accession to OECD until 2010 and the study was conducted from 1993 to 2001 (113).

On inspection the median values do not appear to show a difference between the 2 groups however the mean TTD are both shorter in the OECD inclusion group compared to the non-OECD countries. In the study by James *et al* 2010 of Nigerian children with cancer, the author concluded that a longer TTD compared to the developed world was due to significantly longer DI due to patients seeking alternative health-care such as witch doctors and the church as well as financial constraints (91). Alternatively, Haimi *et al* 2010 in their study of the Israeli healthcare system suggested that a small group of paediatric specialist get involved early in the care of children admitted with “persistent and progressive symptoms” leading to earlier diagnosis (90). The differences in these two studies in non-OECD or developing countries suggests a predominant influence of the healthcare system over the economic status of the country.

The first medical contact was found to influence TTD. For example, a person presenting to a non-paediatric specialist was found to have a longer TTD in Turkey (93) and Israel (88). In the UK a CYA patient presenting to a health visitor rather than a GP, or a GP rather than an A&E had a longer TTD (83,

98). A study of children and adolescents diagnosed with leukaemia or lymphoma in the Canadian healthcare system found that a longer PI was associated with a shorter time to treatment in a study of the time taken to receive treatment. The finding in this Canadian study by Dang-tan *et al* 2010 suggest that a longer PI results in more advanced disease stage at presentation to a healthcare professional and therefore more striking symptoms and signs that expedited the diagnostic process through to treatment. This finding is of particular interest as this was a prospective cohort study including over 1200 patients in the Canadian healthcare system, which is often seen as being most akin to that of the UK (114).

Figure 3.3 A comparison of the median and mean time to diagnosis by Organisation of Economic Co-operation and Development inclusion at the time of study



Dividing the studies by OECD status is crude and does not account for variations in the health-care structures within countries. The UK is seen as having a “gate-keeper” style healthcare service with GP’s acting as the primary point of contact and referral, where as in the US and many

European countries the patient can directly access a specialist. The rare nature of cancer in the young means most UK GP's are unlikely to encounter more than a single case of childhood or TYA cancer in 20 years experience (115). It is therefore unlikely that personal experience and reflection will improve the diagnosis of CYA cancers by GPs. The strategy of increasing the awareness and knowledge of GP's to the symptoms and signs of potential cancers within CYA may have an impact. Awareness campaigns in CYA cancers are already in place, as previously discussed the HeadSmart campaign sets out to raise awareness of the potential symptoms and signs of CNS tumours in patients and healthcare professional in order to improve on survival and neurocognitive outcomes for patients (12). The campaign has set a target to reduce the median TTD in the UK for brain tumours from 13 weeks to 5 weeks. The initial results of the campaign have suggested a reduction in the TTD (described as symptom-interval by the authors) and DI after the launch of HeadSmart, though they concede more data are need to identify a sustained reduction (116).

There are two sides to the story of presentation of CYA cancers to GPs, which are highlighted in three studies done in UK primary care services: The first is a study by Fern *et al* 2011 of Scottish GP practises that found around 70% of registered 15-24 year-olds attended their GP in a year and only 4% of consultations could be classified as relating to potential "alert" symptoms for cancer (66). This study suggests a small proportion of GP consultations maybe amenable to targeted interventions within the TYA population. Two studies by Dommett *et al* 2012 used a population-based, nested case-control method to investigate the features of childhood cancer (72) and teenager and young adult cancer (117) in primary care. These studies found a low positive predictive value of individual alert symptoms and consultations and suggested further studies of symptoms and consultation clustering to increase the positive predictive value. The three studies referenced the NICE guidelines for referral of suspected cancer in the identification of alert symptoms, the relevance of this document to this project will be discussed further within the methods section.

3.3.5 Time to diagnosis and outcome

Three distinct patterns of survival emerged from the review:

- An almost linear decrease in survival probability with increasing TTD. Observed in the Ewing's-family of Soft Tissue Sarcomas (82).
- A non-linear relationship showing an initial fall in survival probability to a nadir at a relatively short TTD, followed by an increasing survival probability with increasing TTD. Observed in children and young adults with brain (85) and non-rhabdomyosarcoma (82) tumours, a contrast to adult colorectal cancer populations (118).
- No significant difference in survival with increasing TTD was observed in the bone sarcoma population (83).

The relationship between TTD and outcome is also unclear. Differing patterns of survival with increasing TTD were reported across diagnostic groups, most notably in brain, bone and STS populations (82-84). Peaks and troughs in survival outlined earlier may signify the presence of other factors that have a stronger influence on outcome, such as tumour biology or response to treatment (85). Analysis comparing TTD in the low-grade glioma population and the medulloblastoma population concluded that tumour biology is "*dominant and overwhelms any opposing effect on survival of a delay in diagnosis*" (85).

Overall survival is not the only measure of outcome relevant to early diagnosis; other measures include recurrence, quality of life, treatment late-effects and survivorship (10). In the UK five-year survival for retinoblastoma between 2001-2005 was nearly 100%, yet the median TTD for this group of patients was amongst the longest (3). Furthermore, studies consistently showed a significant association between increasing TTD and other poor outcome measures including advanced localised or metastatic disease and increased treatment burden (89, 92, 100, 109). Improving TTD would therefore aim to impact on survivorship by reducing the treatment burden from chemotherapy, improving visual outcomes and long-term quality of life. CYA CNS tumour outcomes such as neurocognitive status at diagnosis and subsequent to treatment could be improved by reducing TTD, in a similar fashion to the aforementioned retinoblastoma population.

3.4 Conclusion

The overwhelming opinion amongst patients, clinicians and public health professionals is that improving TTD for childhood and young adult cancers will impact positively upon survival and survivorship. However the conclusions drawn from early diagnosis research in children and young adults to date are far from clear. This review highlights some limitations in this research area. It is clear that a theoretical framework is lacking from the vast majority of publications and subsequently the definitions used for key time-intervals and milestones are neither robust nor reproducible.

The Aarhus statement discussed in section 2.1 provides clear guidance for the early diagnosis researcher and the “Model of pathways to treatment” Walter *et al* 2011 provides a clear theoretical framework. The framework presented in Figure 2.2 will underpin the development of the method within this research project. Further considerations regarding the application of the term “delay” and a simple set of milestones will be discussed further in sections 3.4.1.

3.4.1 Qualifying the application of the term “delay”

The term “delay” is used in the majority of articles, most frequently describing an unqualified period of time between symptom onset and definitive diagnosis (85, 87, 89, 90, 92, 95, 96, 100, 101, 110). When authors attempted to qualify the point from which a delay becomes unacceptable or excessive they often used an arbitrary time point defined by their observations (104) or the median delay value for the study population (96, 103). Given the negative connotations of the term “delay” a set of defined criteria should be outlined at the start of the study to identify unacceptable circumstances during which a delay is experienced (78, 104). Shay *et al* 2011 and the National Patient Safety Agency (2010) defined criteria for delay in a thematic review of delayed diagnosis of cancer (figure 3.4) (88, 119). Such criteria should be accompanied by a clear model for time intervals to be studied and hierarchies for identification of major time points

(similar to the hierarchy of defining the date of diagnosis issued by the European Network of Cancer Registries) (24).

Figure 3.4 Table of Criteria for defining delays in diagnosis of cancer in children and young people

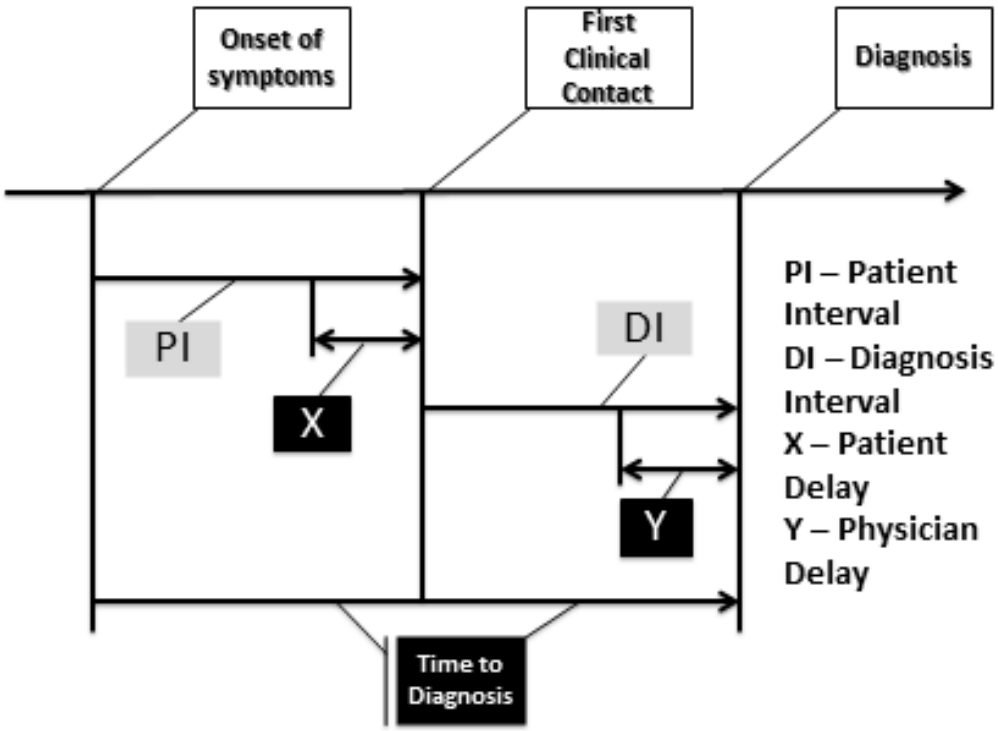
National Patient Safety Agency, Delayed diagnosis of cancer: Thematic review 2010	Diagnostic delay of pediatric brain tumors in Israel: a retrospective risk factor analysis, Shay *et al* 2012	Proposed criteria for defining delay; (1) defines delay within the patient interval, (2) & (3) defines delay within the diagnostic interval
(1) Delayed diagnosis in cancer is when someone who has cancer: (2) Is not investigated or referred for investigation; (3) *or* having been investigated, is not diagnosed at the time of the investigation; (4) *or* is diagnosed incorrectly; (5) *or* where a positive test result or diagnosis is not communicated effectively to a clinician with the ability to act on the information; (6) *or* where a positive test result or diagnosis is not acted upon and treatment commenced as appropriate.	(1) Delay was defined as unacceptable if any circumstances were encountered: (2) Neurological deterioration without proper medical attention (3) Repeated visits to the treating physician or to the emergency room which did not receive adequate attention (4) Presence of classical symptoms of raised intracranial pressure, such as morning headaches and vomiting, which did not lead to referral for brain imaging (5) Presence of administrative problems that led to a delay in performing imaging studies that were properly ordered (6) Presence of symptoms such as a strong headache that lasted for several months and was misdiagnosed as a psychological problem without excluding an organic origin by imaging (7) Presence of inadequately passive parents who did not seek medical help, in spite of several weeks of symptoms	(1) Delays occurs if any of the criteria are met: (2) Due to inactivity of the case or parent there is a prolonged time to engagement with healthcare professional (3) A case approaches a healthcare provider with accepted alert symptom/symptoms for cancer that according to guidance should prompt investigation or referral and no action or inappropriate action are taken by the healthcare provider (4) Investigations once conducted are not appropriately acted upon or incorrectly interpreted leading to neither referral nor treatment

3.4.2 Milestones in early diagnosis research

An outline of milestones to aid TTD research has been developed based on this review (figure 3.5). The terms PI, DI and TTD are purely descriptive and infer nothing about the significance or modifiability in duration of time lapse within each period. “Delay” describes a variable time within the PI, DI or TTD

that occurs if the outlined criteria for delay are met. More extensive models of pathways to treatment and milestones in early diagnosis research are outlined by Weller *et al* 2012. These should be used as the theoretical models from which early diagnosis research can develop (19).

Figure 3.5 Diagram of milestones in the time line to diagnosis of cancer with the presentation of delay intervals



3.4.3 Limitations of the Literature review

Potential language bias was caused through a lack of scope to access and translate foreign language papers. Search strategies did not identify papers that considered prognostic-factors or quantified TTD, PI or DI within the main text of the paper but not in the abstract. Publication bias has not been addressed within this study; however, we believe the effect of this is minimized due to the wide range of results obtained between and within diagnostic groups.

3.4.4 Summary points of Literature review

Summary points for the limitations found from the literature review were:

- Little consistency in the terminology of TTD and heterogeneity of research methods made comparisons between studies difficult.
- Summary data reported was inconsistent and incomplete, leading to difficulty interpreting comparative data. This is important as the data are skewed and contain extreme outliers.
- There was marked variability in reporting the association between TTD and outcome.

These points mirror the conclusions from the Aarhus statement, which identifies the common limitations in early diagnosis research to date across the age spectrum (19). There are distinctions to be made in early diagnosis research between the adult cohort and CYA cohort and this review and the Background (Chapter 2) highlights these. Cancer in CYA consists of a rare and heterogeneous population of tumours with distinct and varied inter and intra-diagnosis patterns of biology and behaviour that vary across the age range. The challenges faced by patients and their carers in receiving the diagnosis of cancer are compounded by the rarity of the diagnosis, however the fact that cancer is still a major cause of mortality and morbidity in this population justifies the intensity with which we strive to improve cancer care in this population. There are gaps in our knowledge regarding early diagnosis research and in the UK as with many other countries the focus has fallen on the primary care setting. Secondary care is incorporated in some research, but the majority of guidance is aimed at the primary care clinician, as such this project provides a unique opportunity to explore the role of secondary care services in the diagnostic pathway for CYA cancer patients.

Chapter 4 Methods

4.1 Overview

This study seeks to tackle the aims set out in Chapter 1 through the investigation of the time taken to diagnose cancer within secondary care services in a population of children and young people in the geographical region of the Former Yorkshire Regional Health Authority (FYRHA). The distribution of TTD within this population will be analysed against population demographics and disease profiles to identify if sub-groups exist within the population who are at risk of a prolonged TTD. There are two aims of this study; firstly the development of recommendations for healthcare providers that will aim to improve TTD, and secondly to assess and advance the use of HES data linked to population based cancer registries in the field of CYA cancers. This is a descriptive study that will employ survival analysis along with clinical knowledge of paediatric oncology.

The requirement for this study follows the evidence presented in chapters 1, 2 and 3. In the first two chapters the high profile of early diagnosis in cancer care service planning and research in the UK and across the globe was identified. The assumptions around the impact of prolonged TTD are explored along with the challenges facing researchers in the field of early diagnosis and cancer in CYA. Chapter 3 discussed specific considerations from the reviewed literature on early diagnosis research in CYA cancers as well as identifying general recommendations for early diagnosis research.

This chapter outlines the study design providing the justification for this approach, the data sources used, data sampling and analysis. Section 4.2 will provide a description of the overall study design and section 4.3 describes the ethical approval for this project. Following this 4.4 will identify and discuss the sources of data for the project; how they were developed, how the information for these sources was acquired and finally how this data was linked together to provide the dataset from which the study population was sampled. The sampling and identification of cases, hospital episodes

and cancer “alert” symptoms is discussed in section 4.8.4. Final sections 4.9 and 4.10 will outline the statistical analysis, data presentation methods and survival analysis for the subsequent results and discussion sections.

4.2 Study design

This study sits within the broad field of epidemiology, defined as *“The study of the distribution of health-related states or events in specified populations and the application of this study to control of health problems”* (*Eva Steliarova-Foucher*). This is an observational study design, as information is collected on factors associated with outcomes of interest within a population without any attempt to modify the exposure of the study population. This is also a health services research study involving retrospective analysis of a cohort identified using a population-based cancer register linked to hospital admissions data. The study cohort is used to assess the impact of an exposure (prolonged TTD of cancer) within a defined population (CYAs with cancer in Yorkshire) upon a certain outcome (survival). As a retrospective study the information will have already been recorded for the population, allowing the review of information over a long period of time.

4.2.1 Limitations of study design

There are a number of advantages and disadvantages to the design of this study. The observational nature means it is relatively cheap to conduct compared to other study designs such as Randomised Control Trials (RCT). However observational studies are not as robust as RCT in establishing causal links and adjusting for the effects of known and unknown confounding factors (120). The retrospective nature of the study allows access to information already collected and therefore the study needs only a short period of time for data collection compared to prospective studies that must collect information as they follow cases up. The fact that the information has already been collected is also a limitation as the study can only work from information recorded and is reliant on the completeness of data collected. If

data are missing or incomplete, this will impact on the validity and generalisability of the conclusions drawn.

The population-based design of this study is in contrast to the institution-based design that predominated in the CYA early diagnosis research studies identified in the systematic review section 3.3. Within the field of epidemiology based health services research a population-based study design is the gold standard, it allows the assessment of the impact of an exposure across a defined population and can be used to generate incidence rates, mortality rates and population-based survival rates. The use of this design will facilitate the development of recommendation more applicable to healthcare provision at local, national and international levels.

The study is also at risk from the influence of bias at a number of points, which is defined as a systematic error that could result in the inaccurate conclusions of the true effect of an exposure on the outcome. Specific areas for bias within this study will be discussed later within this chapter.

4.3 Ethical approval

This project forms part of the ongoing research programme of the Yorkshire Specialist Registry for Cancer in Children and Young People (YSRCCYP). National Information and Governance Board approval for the YSRCCYP comes under the United Kingdom Association of Cancer Registries NIGB application number 0001, available at http://www.nigb.nhs.uk/s251/registerapp/register311012.xls

The YSRCCYP protocol is currently in its 4th version protocol, where this project is outlined – Protocol Version 4 December 2011 REC: MREC/0/3/1.

Hospital Episodes Statistics data linked to the YSRCCYP will be used to analyse the TTD of cancer for children and young people (0 to 24 years) within secondary care services in the FYRHA. Approval to work with HES data was provided by Data Access Advisory Group – DAAG reference: OC/HES/015, as part of the work approved within the YSRCCYP.

The sole purpose for accessing the medical records as part of this project is to validate what is recorded within the HES records. Information is required from a small sample of less than 50 cases and will include:

- Corroboration of demographic details.
- The list of inpatient events preceding transfer to the tertiary care centre.
- Comparison of the diagnostic codes listed within the HES episodes with what is recorded within the medical records.
- The project will only publish data which is non-patient identifiable in line with the requirements of the registry protocol. Data protection will be discussed in more depth in section 4.7.

4.4 Data sources

This project will utilise data extracted from 2 main sources: the Yorkshire Specialist Register of Cancer in Children and Young People (YSRCCYP) and HES data. Clinical notes will be used as a third source for validation of HES data.

4.4.1 Yorkshire Specialist Register of Cancer in Children and Young People (YSRCCYP)

The YSRCCYP is a population-based register of young people with cancer who are diagnosed within the Yorkshire & Humber Strategic Health Authority (SHA) aged under 30-year at the time of diagnosis. The register has collected data on children aged under 15 at diagnosis since 1974 and young adults aged 15-29 at diagnosis since 1990. The register covers a population of around 5 million, nearly 2 million of whom are aged under 30 years. The geographical area covered includes the Yorkshire and Humber SHA and recently has been extended to incorporate the local authority areas of Barnsley, Doncaster, Rotherham and Sheffield.

The register contains demographic, diagnostic and clinical data relating to treatment. The diagnostic data are classified according to the ICCC classification system. The date of diagnosis is taken from the medical records and is recorded as the date of histopathological confirmation of the cancer diagnosis.

The use of the register to identify cases allows a population-based approach to this retrospective research. Each case is recorded within an Access database held on a secure server within the Centre of Epidemiology and Biostatistics, University of Leeds. Cases are ascertained using a number of different sources. The principal source for childhood cancer cases is the tertiary paediatric oncology centre based at the Leeds General Infirmary and there are also links to regional, extra-regional and national registries. Utilisation of multiple sources for case accrual allows the registry to present highly accurate data on all cancer cases diagnosed in the Yorkshire region.

4.4.2 International Statistical Classification of Disease

The International Statistical Classification of Disease (ICD) is a classification that has evolved over the last century to allow the counting of disease, symptoms, injuries, reasons for disease and health encounters as well as external causes of death (32). Developed by the WHO, this standard diagnostic tool is currently in its 10th edition (32). This system is designed to allow the classification of disease and health issues in many types of healthcare and health research as well as from vital health records such as death certificates. Globally it is used by a number of different organisations and disciplines such as:

- Epidemiologists to monitor incidence and prevalence of disease and other health issues, it is also used to compile mortality and morbidity data. The ICD codes are often used within cancer registries to generate comparable and meaningful summaries of data contained within patient records.
- Health management and policy makers to monitor quality, facilitate health economic analysis and inform provision of healthcare resources.

- Clinicians to monitor quality of care and perform healthcare research.

ICD-10 groups information about diseases and their causes into 22 chapters that cover:

- Communicable diseases (codes A00-B99).
- General diseases that affect the whole body (codes C00-F99), this section contains a section of codes relevant to the classification of neoplasms (codes C00-D48).
- Local diseases arranged by site (codes G00-N99).
- Developmental disease (codes O00-Q99).
- Injuries (codes S00-T98).
- External causes (codes V01-Y98).
- Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (codes R00-R99).

Each ICD-10 code starts with a letter followed by two digits which are further divided in sub-categories indicated by a point and a third digit to create a four unit character (32).

4.4.3 Hospital Episodes Statistics (HES)

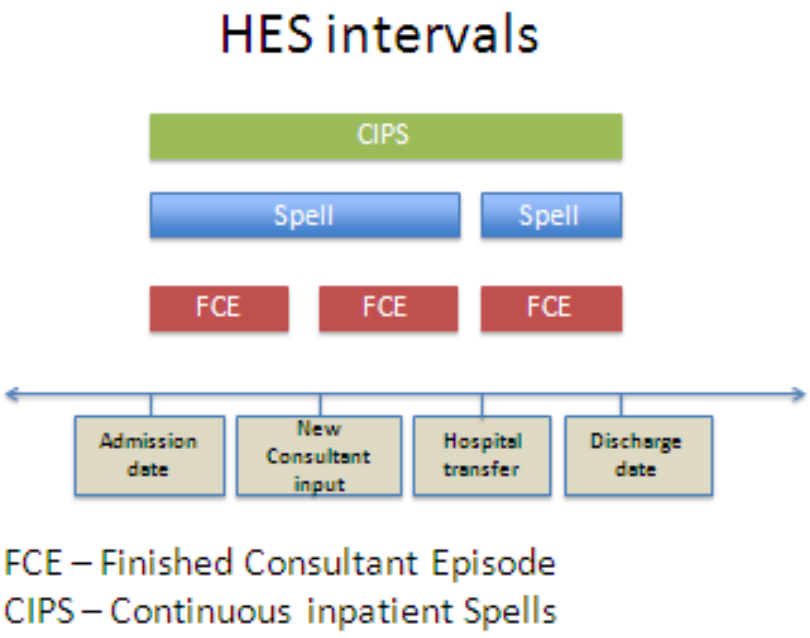
Hospital Episodes Statistics (HES) are a data warehouse for all inpatient admissions for NHS funded patients treated within English NHS Trusts (including the private sector) since 1989 (121). HES inpatient data provides information relevant to the patients diagnosis and treatment, demographics including ethnicity, administrative information such as waiting time and geographical information such as where a patient was treated (121). The process of generating HES data involves non-clinical staff collecting information recorded by medical staff in clinical notes. The data are collated centrally and quality assessed before being made available. A wide variety of organisations and individuals use this data including the government and the NHS, as well as researchers and institutions studying health economics, cancer intelligence and epidemiology. The outpatient data has been recorded in HES since 2003 and A&E attendances recorded since 2007.

For the purposes of this project, HES records for inpatient and outpatient admission have been linked to the diagnostic data in the YSRCCYP.

The multilevel data structure of HES is complex; containing episodes which lie within spells which lie within continuous inpatient spells (CIPS) as represented in Figure 4.1. This project will mainly focus at the level of patient episodes, rather than spells or CIPS, although CIPS will be considered in the analysis of admission routes. The use of episodes allows the analysis of diagnostic information included in HES for inpatient events in secondary care and focuses on individual consultant led teams. The focus on CIPS for admission routes allows for the analysis of the initial point of contact for an inpatient event.

Each Finished Consultant Episode contains a series of ICD-10 diagnostic codes which are used to describe the hospital event, and a sequence of up to 20 ICD-10 codes can be used to describe the episode.

Figure 4.1 Outline of HES inpatient time intervals



4.4.4 Data coding

In order to understand the process of coding of healthcare events I met with the lead coding trainer for Leeds Teaching Hospitals trust. This provided key information on how medical information is translated from the case notes to a sequence of diagnostic codes, the source from which HES episodes records are generated. Several key considerations are outlined for those working with HES data generated by medical coding:

- Coding is only as good as the information recorded within the clinical notes.
- There is no universal approach to coding within the NHS.
- The process of coding inpatient events has come under increasing scrutiny and is now subject to increasingly strict regulations.
- At the end of the month coders have 5 days to complete 95% of the coding and then a further month to complete the last 5%.
- Coders utilise a number of sources for coding information such as the case notes and electronic resources (e.g. results server).
- A symptom or sign is only recorded if it fails to relate to the diagnosis made during the episode. If a symptom appears and is found to relate to the diagnosis made at a later date within the episode this symptom is omitted from the list of diagnostic codes.
- A governing body exists for clinical coders that can help with queries or difficult cases.
- The quality of coding is scrutinised by a regular audit process.
- If a diagnosis, symptom or procedure cannot be coded, relevant consultants are contacted to discuss the issue.
- Back coding is not allowed, coders must only use information within case-notes, associated referral, discharge letters or other resource that relate to the episode being coded.
    - Therefore if a co-morbidity is not recorded it can’t be back filled from a previous entry.
- The clinical information within case notes can only be coded if an entry starts with the key words:
    - Diagnosis….

    - Treat as….
    - Probably….
    - Presumed….
- Some coders will code the Δ symbol, recorded by some medical professionals to indicate a differential diagnosis, however this is not a universally accepted approach.
- Coders cannot code if an entry in the medical notes starts with the key words:
    - Impression….
    - Likely….
    - Possible….

In 2002 the Department of Health consulted on its plans for NHS financial reforms and introduced the concept of Payment by Results (PbR) for the reimbursement of NHS hospitals in England for their activities (122). PbR works by assigning tariffs to patient treatments within a cluster of diagnosis and procedures which consume similar resources. This system was introduced in 2003-2004 and has been developed incrementally; in 2005-2006 the system included the majority of trusts but only elective cases; 2006-2007 saw the inclusion of non-elective, accident and emergency, outpatient and emergency admissions; 2008-2009 saw the final stage of financial implementation. The application of the financial currency that dictates the national tariffs was only applied to outpatient activity in 2009-2010 (122).

There are several key time points within the implementation of PbR that merit consideration within the methods of this study (122). The threshold year of 2006 saw a change in attitude within the NHS towards the accuracy of coding, which was driven by the desire to improve the recording of healthcare activity to facilitate more accurate reflection of utilisation of resources. This had the secondary effect of improving HES records for healthcare researchers and this transition will be explored as part of the validation process within this project. The later implementation of PbR to outpatient activity and the exclusion of diagnostic information from the mandatory requirements reduces the quality of information contained within outpatient HES records. The NCIN routes to diagnosis work did not consider

any of the diagnostic information in the process of sequencing secondary care events prior to diagnosis and therefore included outpatient HES data within the analysis (71).

4.5 Data linkage

The YSRCCYP cases were linked to the HES data using specified identifiers (NHS number, date of birth, sex, and postcode). The YSRCCYP therefore provided unique information on patient demographics and diagnosis, whilst the HES data provided unique information relevant to all inpatient hospital episodes and mutual data was included by the identifiers outlined (Figure 4.2). Within the study population only three cases identified from the YSRCCYP did not link to a single HES record.

Figure 4.2 Linked cancer register and HES dataset structure .

	Variables Merged from YSRCCYP					Variables Used for Linkage				Variables Merged from National Information Centre HES data							
	YCTR_ID	Date of Diagno	ICCC	ICCC Grou	Date of Death	SEX	Postcode	Date of Birth	NHS Number	Admission Meth	spellstart	spellend	ICD-10 Code 1	ICD-10 Code 2	ICD-10 code 3	ICD-10 Code 4	ICD-10 Code 5
Patient A	2006145	12/05/2006	61	6	.	F	LSX XXX	24/06/2003	4567000000	28	28/11/2004	29/11/2004	N180	Z911	N038	Z992	Z940
	2006145	12/05/2006	61	6	.	F	LSX XXX	24/06/2003	4567000000	21	01/02/2005	01/02/2005	T393	X600	F329	N180	
	2006145	12/05/2006	61	6	.	F	LSX XXX	24/06/2003	4567000000	11	02/05/2006	16/05/2006	R31X	N288			
Patient B	2007245	29/12/2007	61	6	.	M	WFX XXX	11/04/2005	4890000000	82	11/04/2005	15/04/2005	P070	Z380	P590	P742	E872
	2007245	29/12/2007	61	6	.	M	WFX XXX	11/04/2005	4890000000	21	18/01/2006	18/01/2006	K409	K429			
	2007245	29/12/2007	61	6	.	M	WFX XXX	11/04/2005	4890000000	12	19/01/2006	21/01/2006	K409				
	2007245	29/12/2007	61	6	.	M	WFX XXX	11/04/2005	4890000000	21	29/10/2006	29/10/2006	B349	R062			
	2007245	29/12/2007	61	6	.	M	WFX XXX	11/04/2005	4890000000	21	30/12/2006	30/12/2006	B349				
	2007245	29/12/2007	61	6	.	M	WFX XXX	11/04/2005	4890000000	21	01/11/2007	02/11/2007	R31X	R934			
Patient C	2008556	12/02/2008	62	6	24/06/2010	M	HGX XXX	05/10/1998	4230000000	21	02/02/2008	15/02/2008	R190				

4.6 Data quality and bias

Several areas of potential bias must be considered within the design of this study and the data sources used for the study. As previously stated the presence of bias within any research study is the presence of factors that influence the validity of the estimate of the true relationship between the study exposure and outcome measure (120). The main areas for consideration are in the ascertainment of cases for the study, the completeness of information within the retrospective records, classification of cases and assessment of TTD and the measurement of the outcome.

4.6.1 Ascertainment bias

The population based nature of this study ensures the study population includes all cases of cancer in CYA in the defined FYRHA and period of study, as long as they meet the inclusion criteria. As such the influence of selection bias upon the case profile of the population should not influence assessment of the exposure on the outcome. However, the fact that the cases are being sourced from a registry does raise the issues of completeness of the cohort, particularly raising the question; is the ascertainment of cases of cancer in CYA in Yorkshire complete within the Registry? This issue is discussed within the YSRCCYP protocol version 4, in which they identify a number of sources from which information is gathered and also sources used to cross-reference against those patients on the Registry, some are which listed below:

- Leeds Teaching Hospitals Trust Paediatric Oncology Department is the tertiary centre for CYA cancer within the Yorkshire region and provides the key source of notification for cases.
- The Northern and Yorkshire Cancer Registry and Information Service provides an annual list of CYA cancer patients within the region from which the YSRCCYP can cross-reference.
- The National Registry of Childhood Tumours provides another source for cross-checking annually.

Ascertainment bias is therefore likely to be minimal and as such the use of the YSRCCYP for identification of cases for inclusion within the study is robust. There is also a risk of ascertainment bias through the linkage of the two data sets, however as discussed within section 4.5 the data linkage within this project was high.

4.6.2 Information bias

The retrospective nature of this study means the completeness of data recording must be considered and is a key issue for this project. As stated within the coding section the process of coding medical records and the generation of HES records has changed dramatically with the introduction of PbR in 2003-2004, the implementation of which occurred gradually over the intervening years (122). The variable approach to coding of medical records between hospital trusts was also touched upon in section 0. These two factors brings into question the accuracy of the information recorded with the HES records, especially during the earlier period of the study. In order to assess the completeness and accuracy of HES records, a small population of hospital notes was reviewed. The process of HES record validation is discussed in section 4.11.2.

Unfortunately the use of case-notes review in retrospective studies also raises the question of missing information bias, especially an issue for studies of patient interval to healthcare engagement. The duration of symptoms as recorded within the medical notes is questionable and subject to recall bias and prone to incomplete recording. The alternative method of patient interview as outlined in the majority of CYA studies discussed in the systematic review (Chapter 3) is also prone to recall bias as the patient may have a skewed view of their disease given the outcome or their experiences during treatment. For these reasons this project will focus on the diagnostic interval; recorded as the duration of time in days between the presence of alert symptoms within a HES record and the date of definitive diagnosis.

4.6.3 Misclassification error

The method of analysing HES records against national CYA cancer awareness resources for the identification of alert codes is prone to bias given the processes of interpretation involved. There is a risk of misclassifying hospital events as not containing an alert code, given the presence of an established diagnosis. The ICD-10 classification system is based on diagnosis codes and generalised symptoms, symptoms which will not be recorded by coders unless they cannot be explained by the given diagnosis within each episode. For example, 'headache' will only appear in the HES record if no other explanation can be extracted from the hospital record, such as the diagnosis of 'migraine'.

There is also a risk of over subscribing significance to the codes of an episode due the fact the diagnosis of cancer is made after the occurrence of the episodes being studied. Many of the symptoms present within the CYA awareness campaigns occur within a number of paediatric conditions, an issue highlighted by Dommett *et al* 2012 in relation to the low positive predictive value of NICE referral guidance (72).

The issue of misclassification error is discussed within the NCIN routes to diagnosis work, where the authors concede they could not be sure the events occurring prior to diagnosis were related to the subsequent cancer diagnosis (71). Unlike this study, no attempts were made to relate the pre-diagnosis events to the subsequent diagnosis based on the coded content of the HES episodes. Therefore the method applied to this study adds another valuable and unique level of complexity to the analysis of routes to diagnosis.

4.7 Data security and confidentiality

Working within the ethical approval for the YSRCCYP this project is presented in a manner such that no patient identifiable data (PID) or potentially PID are divulged. Extreme care and attention was taken to ensure confidentiality and data security. All work with PID was carried out in

accordance with the guidance set out for those working in the YSRCCYP outlined in the data security policy of the Centre for Epidemiology and Biostatistics (CEB), the department in which the YSRCCYP Register is held.

The YSRCCYP outlines conditions for holding PID:

- No information will be published in which an individuals can be identified.
- No individuals whose information is present on the Registry will be approached directly.
- Data will only be released in accordance with the data security policy.

4.7.1 Working with Patient Identifiable Data (PID)

This project has worked with patient level data that was identifiable on many levels, however this work took place within the remit of the policy of data security. Work at the PID level took place within the CEB, which has restricted user access and can only be processed on a secure server where the YSRCCYP Register resides.

4.7.2 Presenting results that are non-Patient Identifiable

The results for this study will be non-PID, however the division of data into PID and non-PID is not straightforward. There are clear pieces of information that can identify patients when results are presented at an individual case level. Clear definitions of what constitutes PID have been taken from the guidelines issued by the UK Association of Cancer Registries (UKACR) "*Data will be regarded as identifiable if it includes any of the following data items: name, address, postcode, date of birth, date of death, NHS number, hospital number*" (123).

This is not the only constraint on presented data: there is still the potential for data to become identifiable despite the omission of the above identifiers. Patients can also be identified by non-classical indicators, such as a unique diagnosis (123). If a case falls within a defined geographical area or recognised population they may be identified by the amalgamation of

identifiers traditionally perceived as non-PID, such as age at diagnosis, route to diagnosis and diagnosis. The UKACR also provides guidance for potentially identifiable patient information:

> *"As a general rule, the following categories of data should be regarded as potentially identifiable:*
>
> *Individual records even if they do not include variables, such as names, full postcodes, and dates of birth which would make them obviously identifiable;*
>
> *Tabular data, based on small geographic areas, with cell counts of fewer than five cases/events (or where counts of less than five can be inferred by simple arithmetic) – hereafter referred to as "sparse cells"*
>
> *Tabular data containing cells that have underlying population denominators of less than approximately 1,000." (123)*

4.8 Population and sampling

A clear set of inclusion and exclusion criteria outlined in this section was applied to the linked dataset in order to define the population for this study.

4.8.1 Inclusion criteria

The study population was identified through a set of inclusion criteria:

1. Patients diagnosed with cancer up to their 25th birthday as recorded on the YSRCCYP.
2. Cases diagnosed between 1st January 2004 and 31st December 2009.
3. Cases treated in the FYRHA as recorded on the YSRCCYP.

The first criteria defines the study population as CYA, comprising a heterogeneous population of tumours as discussed in the background section (Chapter 2).

The second criteria defines a retrospective cohort of patients that covers a six-year time period, allowing for the analysis of variation in TTD across a changing healthcare system for a large geographical population. The second criteria also provided the follow-up time scale to allow 1 and 3 year survival analysis within the time frame of this research project.

The third inclusion criteria was imposed due to changes in the YSRCCYP in recent times with the introduction of cases from South Yorkshire. This data was still being retrospectively collected at the time of analysis and therefore these cases have not been included within the study.

4.8.2 Exclusion

Within the YSRCCYP a patient can potentially have an unlimited number of cancer diagnoses, which are identified by a diagnosis number. Most patients have one diagnosis, however, multiple diagnoses can occur and signify; relapsed disease, secondary malignancy or occasionally duplication of cases due to the revision of a primary diagnosis. Within this study all diagnoses other than the primary diagnoses were excluded from further analysis. This was an important step as the analysis of TTD in relation to cases of relapsed or secondary malignancies is more relevant to tertiary care services, due to the fact that the case will have had or still have ongoing contact with the oncology department prior to relapse or secondary diagnosis. The HES data doesn't allow differentiation between episodes relating to the different diagnoses, especially as some cases may have a recorded second diagnosis while still receiving treatment for their primary cancer.

4.8.3 Episodes (FCE)

Episode (FCE) level data is the base unit of representation for inpatient hospital events within HES records, as outlined in Figure 4.1, enabling the data to be investigated at a consultant event level. The analysis of inpatient events for the purposes of alert code identification was conducted at the

episode level due to the potential for multiple episodes with variable diagnostic sequencing.

4.8.3.1 Pre-diagnosis episodes

The linked HES inpatient data set contains the catalogue of inpatient events for the cases identified from the YSRCCYP. The pre-diagnosis episodes were separated from episodes subsequent to diagnosis for the process of data analysis. The date of diagnosis was taken from the date recorded within the YSRCCYP, and was defined as the date when the pathological specimen was taken. When a histopathological diagnosis was not made, the date of radiological confirmation was used as the date of diagnosis.

Using code written in Stata, episodes with a date of admission that preceded the date of diagnosis or coincided with the date of diagnosis were identified. Any episode in which a date of admission was subsequent to the date of diagnosis were removed from further analysis.

4.8.3.2 Cancer code containing episodes

The date a pathological specimen is taken may be preceded by the date a diagnosis made via another means such as clinical suspicion or radiology. In cases where the medical records identify a clinical or radiological diagnosis prior to the pathological diagnosis, there is a potential for an ICD-10 code for malignancy (C-codes) within a pre-diagnosis episode. The pre-diagnosis episodes will therefore be analysed for the presence of pre-diagnosis C codes and these codes can then be cross-referenced with the ICCC classification. The extent to which discrepancies in date of clinical diagnosis and pathological diagnosis occur will be assessed.

The NHS has also introduced guidance that outlines the duration of time deemed acceptable for referral, diagnosis and treatment of cases with cancer and suspected cancer. The guidance aims to promote timely care for cancer cases in order to improve outcomes. Cancer waiting times guidance applies to children and adults, and the main pathways are (124):

- Two-weeks from urgent GP referral for suspected cancer to first hospital assessment
- 31 days from decision to treat to first treatment
- 62 days from urgent GP referral for suspected cancer to first treatment, however this threshold is 31 days for suspected childhood cancer, testicular cancer and acute leukaemia.

The 31 day period from urgent referral to treatment for suspected childhood cancer will be adapted for this study to a 31 day period from first alert code occurrence to date of diagnosis. A 31 day threshold will be used within the analysis of alert code events in order to differentiate between cases where the first alert code event occurs within an appropriate time prior to diagnosis and cases where potentially there is a more prolonged TTD in secondary care.

4.8.4 Alert codes

In order to facilitate earlier diagnoses, patients must be aware of the potential importance of their symptoms, and the healthcare professional at the point of initial contact must regard the symptoms as concerning. A number of government, private and third sector organisations have been working to develop awareness and early diagnosis campaigns. For the purposes of this project, major awareness and early diagnosis guidance will be used to define ‘alert’ symptoms for an initial presentation of cancer in CYA patients. The identified ‘alert’ symptoms will be cross-referenced against the ICD-10 diagnostic codes and those codes identified as potentially relating to an ‘alert’ symptom will be highlighted and identified within HES episodes as they appear in the full data set.

The ‘alert’ symptoms were divided into one broad category relating to a general diagnosis of cancer, which do not vary across ICCC diagnostic groups. A second group of specific alert codes were defined within each ICCC group and these alert codes were group specific. The specific ‘alert’ symptoms for each diagnostic group were taken from the sources outlined in the next section. The awareness literature was selected as it aims to increase patient awareness, facilitate early diagnosis in professionals and is

relevant to CYA cancers combined with the fact that they are easy to access online for patients and professionals. Any code directly relating to the treatment of a neoplasm was also included within the pre-diagnosis alert codes.

4.8.4.1 Awareness campaigns

4.8.4.1.1 National Institute for Clinical Excellence guidance

The Department of Health first published guidance on referral to suspected cancer in 2000, this was revised in light of new evidence by NICE, in 2005 the document “Referral guideline for suspected cancer” were published and is still in use to date (125). The guidelines are aimed at the primary healthcare provider and offer best practice advice for the referral of suspected cancer in adults and children. The document identifies the importance of good communication and patient choice in delivering patient centred care, while at the same time acknowledging the difficulties of making a diagnosis of cancer, a point which is stressed in the childhood cancer population. Three categories are defined for acceptable referral times, which are as follows (125):

- Immediate – admission or referral within hours of suspecting the diagnosis.
- Urgent – within the 2 week national target.
- Non-urgent - All other referrals.

The bulk of the document relates to adult cancers categorised by primary site, however there is a chapter relevant to certain childhood cancers, this chapter categorises the tumour types in accordance with ICCC diagnostic groups. There is guidance of several of the major childhood cancers including: leukaemia, lymphoma, central nervous system tumours (children under two-years of age and children over two-years of age), neuroblastoma, Wilms’ tumour, bone tumours, soft tissue sarcoma and retinoblastoma, see appendix 1 (125). The guidelines are not exhaustive, however they provide a clear baseline for what should be expected from the healthcare provider at

the point of contact, for this reason the guidelines are pertinent to secondary care providers.

The terms ‘unexplained’ and ‘persistent’ appear repeatedly in relation to symptoms and signs and provide a challenge when cross-referencing signs and symptoms from the guidelines to HES diagnostic codes. The term ‘unexplained’ refers to a symptoms and/or sign that is present but lacks a clear diagnosis after history, exam and investigations by the primary physician, the term ‘persistent’ refers to a symptom and/or sign present beyond that expected of a self-limiting illness, an upper limit of 4-6 week is described (125). The method for identification of ICD-10 diagnostic codes relevant to ‘alert’ symptoms is described later within the methods section.

The NICE guidelines discussed provide a clear foundation from which to develop a portfolio of ‘alert’ symptoms and signs, but they only cover certain types of childhood cancers. TYA cancers are not covered and the adult guidance is aimed at specific cancer sites in the majority of cancer, therefore alternative sources were sort to provide a wider scope including young adult specific cancers.

4.8.4.1.2 Teenage Cancer Trust Awareness

The Teenage Cancer Trust is a UK charity dedicated to improving the lives of TYA’s between the ages 13 to 24 who are diagnosed with cancer (25). In April 2012 the charity launched an awareness week that aimed to ‘educate young people, parents, teachers and healthcare professionals about the signs and symptoms of cancer in young people’ (17). The campaign provides information packs for teachers and awareness posters for young people covering several tumour types: leukaemia, lymphoma, bone tumours, brain tumours, testicular cancer, ovarian cancer and skin cancer (17). A patient story relevant to the presentation accompanied the guidance for each tumour type outlined above, highlighting the reality as well as the rarity of cancer within this population.

4.8.4.1.3 HeadSmart

The HeadSmart campaign was launched in 2011 aims to ‘enhance awareness of symptoms of brain tumours in children and young adults’. This

is a partnership between the Royal College of Paediatric and Child Health, the Brain Tumour Charity, the Childrens Brain Tumour Research Centre and the Health Foundation Trust and the campaign is aimed at children, parents and healthcare professionals. The campaign introduction highlights the prolonged median TTD for brain tumours within the UK childhood population of 12 to 13 weeks as ‘unfavorable’ compared this to 5 weeks in other countries (12).

The guidance provided is relevant to the presenting symptoms of brain tumours in different age groups ranging from 0 to 18 years (12):

- Preschool or under 5-years
- Children or 5 to 11-years
- Young people 12-18- years
- Additional symptoms for all ages

As well as information leaflets, posters and symptom cards there are educational modules on-line for healthcare professional to access.

4.8.4.1.4 MacMillan Cancer

The Macmillan Cancer support is a registered charity that provides practical, medical and financial support and strives for better cancer care (126). The charity has an annual awareness week that takes place at the end of January each year, in 2013 the Cancertalk week encouraged schools to talk and teach students about cancer. The charity also provides referenced information on signs and symptoms of cancer in children and adults. The information provided is often only very brief but is relevant to a wide variety of childhood specific diagnoses, including: acute lymphoblastic leukaemia, acute myeloid leukaemia, brain tumours, Ewing's sarcoma, germ cell tumours, Hodgkin lymphoma, liver tumours, neuroblastoma, non-Hodgkin lymphoma, osteosarcoma, rare tumours, retinoblastoma, rhabdomyosarcoma and Wilms’ tumour (126).

4.8.4.1.5 Bone Cancer Research Trust

The Bone Cancer Research Trust (BCRT) is a charity dedicated to primary bone cancer within the UK and Ireland. The charity is committed to improving outcomes for patients with primary bone cancer through

awareness and information. The BCRT website contains awareness information for osteosarcoma and Ewing's sarcoma, which is aimed at patients, their families and friends and the general public (127).

The methodology outlined below will aim to identify if the information provided in the identified awareness literature, above, can be converted into ICD-10 diagnoses and used to identify ICD-10 codes indicative of a potential cancer within pre-diagnosis HES inpatient episodes. The awareness resources used within this study span a range of publication dates, 2005 to present day, these fall both within and following the time period of this study (2004 to 2009). The resources are therefore used to retrospectively analyse the HES data for potential signs and symptoms of cancer preceding a diagnosis and this study is not aiming to critique these resources against contemporary healthcare events.

4.8.4.2 Alert codes

The 'alert' symptoms identified from the awareness and early diagnosis sources outlined earlier within this chapter were collated into a single list of broad 'alert' symptoms and signs and lists of ICCC diagnostic group specific 'alerts' symptoms and signs.

The main HES dataset was separated into twelve broad diagnostic groups according to the ICCC code taken from the YSRCCYP. The ICD-10 codes contained within each patient spells were extracted, collated and a list of bespoke codes for each of the diagnostic groups was generated. This was an important step due to the often unique and varied 'alert' symptoms and signs identified within each specific diagnostic group, the ICD-10 codes were matched to the 'alert' symptoms and signs on a group by group basis. The broad cancer 'alert' symptoms and signs could be matched across the full dataset. All the extracted codes were labeled in-line with the following list of categories:

A. ICD-10 code relevant to a specific 'alert' symptom or sign of cancer in an ICCC diagnostic group.

B. ICD-10 code relevant to a broad ‘alert’ symptom or sign of cancer in the CYA population.
C. ICD-10 Cancer code.

The categorised lists of ICD-10 codes were then independently reviewed by two reviewers; SK an expert in CYA haematology and SP an expert in CYA oncology, in order to validate the selection process. Any disputed codes were discussed and agreement on inclusion was achieved between reviewers. Once a complete list of broad and specific ‘alert’ codes was identified the relevant codes were highlighted within the full HES data set using Stata. Episodes containing ‘alert’ codes were identified and the potential relevance of the episode to the subsequent diagnosis of primary cancer could be assessed.

4.8.4.2.1 Broad cancer diagnosis alert codes

Once the ICD-10 codes were extracted from the pre-diagnosis episodes for the study cohort, a list of alert codes relating to the broad diagnosis of cancer was generated from the awareness campaigns outlined above. The list of broad alert symptoms is included within appendix 2 with the corresponding ICD-10 codes.

4.8.4.2.2 Diagnostic group specific alert codes

The specific alert ICD-10 codes identified within the pre-diagnosis episodes from the renal tumour population, ICCC group VI, are outlined below and provides an example for the method applied across all diagnostic groups, Figure 4.3.

Figure 4.3 Children and young adults Renal tumour early diagnosis sign and symptoms (white cells) with the matched ICD-10 codes and descriptions identified (grey cells)

Early diagnosis & awareness resources				HES inpatient data	
Symptoms		**Source**	**Year**	**ICD-10 code**	**Description of code**
Painless mass	Persistent or progressive distension	NICE	2005	**R222**	**Localized swelling, mass and lump, trunk**
Swelling in the abdomen	Usually painless	NICE	2005	**R190**	**Intra-abdominal & pelvic swelling, mass & lump**
Haematuria		NICE	2005	**R31X**	**Unspecified haematuria**
Hypertension		Macmillan	2011	**I10X**	**Essential (primary) hypertension**
				I120	**Hypertensive renal disease with renal failure**
fever		Macmillan	2011	**R509**	**Fever, PUO**
Abdo pain		Macmillan	2011	**M545**	**Low back pain**
				R103	**Pain localized to other parts of lower abdomen**
				R104	**Abdominal & pelvic pain - Other & UN**
wt loss		Macmillan	2011	**R634**	**Abnormal weight loss**
Lack of appetite		Macmillan	2011		
Family History	1 in 100 cases	Macmillan	2011	**Z805**	**Family history of malignant neoplasm of urinary tract**

A full list of specific alert codes is provided in appendix 3.

4.8.4.3 Admission method

The method of admission to a secondary care hospital is coded via the 'admimeth' variable within the HES data. A two digit 'admimeth' code is assigned to every inpatient HES episode (FCE), these fall into four broad categories:

- Elective admissions include waiting list (11), booked (12), planned (13).
- Emergency admissions include A&E (21), GP (22), consultant clinic (24) and seven other codes.
- Maternity admissions include ante-partum (31) and post-partum (32).
- Other admissions include birth of a baby in this healthcare provider (82) or outside the healthcare provider (83.)

4.9 Data analysis

The linked data set was analysed using the statistical programme Stata (128). A copy of the Stata "do file" for each of the main commands relating to data analysis is included in appendix 4.

4.9.1 Summary statistics

The descriptive analysis of the study population within section 5.2 of the results chapter include summary data on: the demographics of the population including age, sex and geography; the distribution of cancer diagnoses across the study population and period; and the patterns of inpatient events across the study population and study period.

All data presented within this thesis or any publication relating to this thesis will be non-PID as discussed in section 4.7. Thus, all data are presented in a grouped format and contain a minimum of five cases.

4.9.1.1 Analysis of episodes for alert code

The time interval between the date of admission and the date of primary cancer diagnosis for each episode contained within the linked dataset was calculated using Stata.

Each episode identified as containing an ICD-10 ‘alert’ code was analysed and the TTD in days was generated as the number of days between episode start date and the date of definitive diagnosis. The frequency and distribution of alert code containing episodes was calculated. The population was therefore divided into those patients with alert code containing episodes and those with no alert code containing episodes. The cases with alert code events were further divided into those that appear to spend an appropriate time in secondary care and those that potentially have a more prolonged TTD, as defined by the 31 day threshold discussed in section 4.8.3.2. These sub-populations were then described in more detail and the variation in outcome between the populations analysed.

The route of the presentation for each inpatient alert code event was taken as the initial admission route for the CIPS in which the alert code episode occurred. The admission route for the alert code containing episode was not used as multiple episodes can occur within CIPS and once a case has entered a CIPS, the admission codes for each included episodes will be transfer codes. The transfer codes give little information into how the case accessed the healthcare system. From these results, the route and duration

of the primary alert code containing episode was generated for the group of patients with alert codes episodes.

4.9.2 Data cleaning

The full dataset was cleaned using a bespoke set of Stata command codes designed to remove duplicated entries and identify missing or incorrectly coded dates within the HES date prior to analysis. Specific cleaning tasks focused on:

- Duplicated episodes – Identified as episodes where the same admission date and discharge date contained the same ICD-10 codes
- Missing admission dates – Identified as episodes with either no date recorded within the admission date variable or an inappropriate date such as before the patient's date of birth or the date 15th October 1562, a date that incorrectly identifies a missing value in Stata.
- Discharge dates – Some episodes did not have a discharge date due to the patient being transferred within the hospital. In this senario, dates were ordered for each patient and the discharge date for the sequence of events was used where the dates were missing.
- Patients treated outside the FYRHA - The development of a visual representation of all episodes at a patient level was done to check for patients mistakenly recorded as receiving treatment within the FYRHA.

4.10 Survival analysis

Survival estimates are often summarised as the proportion of a population alive at a specified period following diagnosis (120). The specified time periods vary depending on the focus of the research or the relative time-scales afforded within the study period, in the majority of cases one, three or five-year survival rates are quoted. The longer measure of five-year survival is the most commonly quoted as it incorporates the effects of the disease, comorbidity, treatments and often overall patient management, therefore reflecting the long-term outcome or ‘cure’.

One-year survival has been used as a proxy for early/late diagnosis in adult cancers and as such is a proxy for stage at diagnosis (129). This has been widely adopted as an outcome measure in adult cancers, however the appropriateness of utilising short-term survival periods as proxy measures of early/late disease presentation in childhood and young adult cancers is highly debatable. As previously stated adult tumours are predominantly carcinomatous in origin, whilst most childhood tumours originate from embryonic tissue. The different origins result in different behaviour profiles and as such the rules and assumptions applied to adult populations are often not transposable to CYA cancers.

Conducting survival analysis for a population requires careful consideration, for example not all participants will enter the study at the same time, participants may leave follow-up at variable time intervals and participants will die at variable time points. Survival analysis is widely used in medical research to study time lapse from a defined entry time point (e.g. diagnosis date) to an outcome or end-point. Survival times are referred to as censored if the lifespan of a case is not fully observed by the study end point; for example, if a participant lives beyond the end of the study, their observed lifespan ends at the study end point, but the true lifespan is unknown (130)..

For the purpose of survival analysis within this project several assumptions are required:

- All participants are subject to the same circumstances and the survival probability is the same if recruited early or late.

- The event has occurred at the recorded time point, in other words, all deaths within the population studied are accurately recorded with the YSRCCYP.

Kaplan-Meier survival estimates (130), presented as a survival curve, is the most commonly used method to summarise time to event data within medical research. This is not a true curve but rather a step function where the steps represent the changes in survival probability over short time intervals within the study period. The probability of survival for each subsequent period is conditional on the probability of surviving all preceding time periods. The proportion of survivors therefore remains unchanged between events.

The data requirements for a survival analysis within the population of patients within this study are:

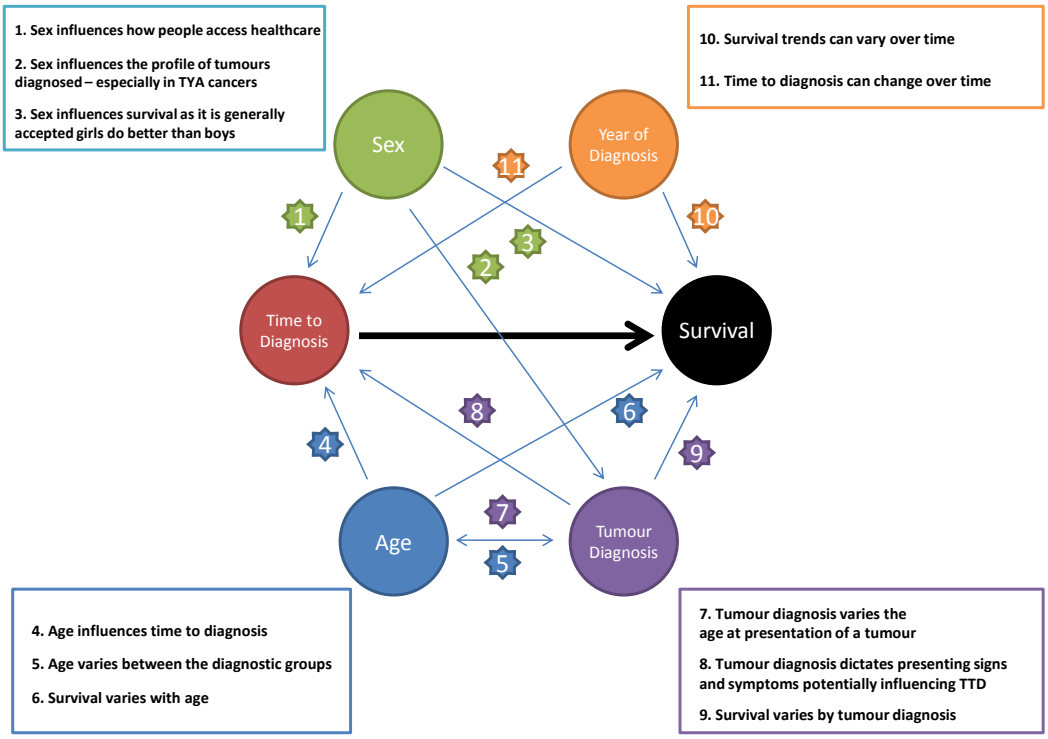
- Defined start date – date of definitive diagnosis of a primary cancer as recorded on the YSRCCYP.
- Censoring date – the date at which the study period terminates.
- The period of survival time to be studied – for this project this will be one and three-year survival.
- Participants who have died during the study period.
- Participants who have been lost to follow-up during the study period – due to the nature of registry data this is a very small proportion of cases estimated to be around 0.1% of cases.

The survival analysis within this project was conducted using Stata statistical software (128). 1098 cases from the YSRCCYP were included if diagnosed with a primary tumour between the start of 2004 and the end of 2009. The censoring date for this analysis was the 31st of December 2012 allowing a full three year period to lapse from the most contemporary date of diagnosis for the cohort. The date of death was available for patients who died prior to 31st December 2012, which was obtained from the YSRCCYP and an ONS flagging system. Anyone without a date of death was assumed to be alive on the censoring date. One and three-year survival probabilities were estimated using Kaplan-Meier estimates for each diagnostic group, and

Kaplan-Meier curves were assessed for the following variables which were identified from the literature review:

- Age at diagnosis
- Sex
- Route of engagement with secondary care relevant to subsequent cancer related event (defined by the admission method associated for the first alert code containing inpatient spell)
- Time period of study: 2004-2006 vs. 2007-2009
- Whether alert codes were present in the HES episodes prior to diagnosis (yes/no); did these alert codes occur within an appropriate time period preceding diagnosis (yes/no).

Figure 4.4 Time to diagnosis and confounding variables for analysing survival in children and young adults with cancer



The independent effect of TTD upon survival will be estimated using Cox proportional hazards regression, whilst adjusting for confounding factors. Figure 4.4 shows the potential confounding factors for survival in CYA

cancers, as shown by previous literature (see Chapter 3). These variables are therefore incorporated within the survival analysis model.

The Cox model makes no assumptions around the distributions of survival time, however it assumes the effects of different variables on survival are constant. Including age and year of diagnosis as continuous variables in the model will assume that these effects are linear in relation to survival. Therefore a likelihood ratio test will be used to determine if these assumptions are valid or not. If there is significant evidence against linearity then a categorical variable will be used instead.

4.11 Validation

4.11.1 Introduction

All clinical information analysed within this study was provided by HES data with the exception of the age, sex, date of diagnosis, the ICCC diagnosis, which were provided by the YSRCCYP. The clinical information within the HES data is collected by clinical coders from the medical records as discussed in section 0. Corroboration of the pre-diagnosis clinical information including the number of inpatient events, the clinical content of an event and the admission methods for an event was not possible for each case within the study cohort. Therefore a small sample of case-notes were reviewed across the range of diagnosis dates, diagnostic groups and ages contained within the study population.

The aim of the validation process was to assess the correlation between the clinical notes and HES records in terms of frequency of events and important clinical information such as date of diagnosis. The validation process was not designed to validate the alert codes identified as part of this study or provide additional information on diagnostic or patient intervals.

4.11.2 Methods

A limited sample of 30 cases were identified from the 1098 cases within the study. Ideally given more time a larger more representative sample of 5-10% of the study cohort would have been reviewed in order to test for clear correlation between the case notes and HES data. The selection process was not random in order to generate a representative sample of case notes across the CYA cancer population studied between 2004 and 2009. A pro forma was developed that specified key comparator information collected from the clinical notes, including:

- Date of Diagnosis
- Diagnosis
- Number of inpatient events

- Route of admission for referral for suspected cancer

The YSRCCYP was used to identify patient demographics and these were used to identify the location of case notes using the Leeds General Infirmary PAS system. LTHT is the principle treatment centre for paediatric and TYA haemato-oncology in the FYRHA and has clinical notes for all patient treated at the centre, including copies of the primary case notes in most instances.

The cases included leukaemia, lymphoma, CNS tumours, renal tumours, bone sarcomas, soft-tissue sarcoma, germ-cell tumours and carcinomas. The diagnosis dates reviewed included cases from each year between 2005-2009, however no cases were reviewed from 2004. The ages of cases at diagnosis ranged from less than 6 months to 19 years of age.

4.11.3 Findings

The case notes for 15 of the 30 cases identified were located and reviewed, of the 15 not reviewed 8 had been transferred to a digital storage system and 7 were either in other case note libraries or untraceable within the time limits of the study. The 15 case notes reviewed were located either within the paediatric oncology departments dedicated case note filing areas or in one of the Leeds General Infirmary filing areas. In 11 out of the 15 sets of case notes reviewed there was a clear correlation between the number of HES events and the inpatient events identified within the notes. In all the cases in which there was discord between the case note events and HES events: more HES events were recorded than were identified within the case notes. Upon review of the HES events it would appear this is due to short attendances either for minor procedures or to an assessment unit. The clinical notes for such events may not migrate to the full clinical notes, alternativley the clinical coders have the superior skill and experience to identify such events within the notes. There is also a potential for incomplete information of pre-diagnosis events within the oncology case notes. Most contained information pertaining to the clinical events leading up to the referral but not necessarily prior to this. One case had a clear difference in the number of HES events and the events in the clinical notes which was

due to a chronic illness and the full history of inpatient involvement was not maintained within the oncology notes.

The date of diagnosis, recorded as the date a specimen was taken that confirmed the diagnosis of cancer, was correctly recorded in all but two of the 15 cases. In one case there appeared to be a discrepancy of one day, while in the other case this was unclear as the medical records did not appear to be complete and I was unable to establish whether a biopsy had been taken prior to surgery.

4.11.4 Conclusions

The process of reviewing case notes was not straightforward and there were a number of issues that came to light including availability and completeness of the case notes and clinical information recorded within. From the limited review conducted it is difficult to draw clear conclusions on the correlation between the information in the case notes and HES data. Within the cases reviewed there appears to be a correlation between HES events and case note events and in some cases HES data appeared to record more events than the case notes, which is most likely due to the multiple sources of medical information that make up the case notes and not all sources merging successfully. There was a high correlation between the date of diagnosis recorded within the YSRCCYP and the specimen dates within the case notes. In order for clear conclusions to be drawn from the validation process a larger sample size of 5 to 10% of cases from the study population would be required, which was not possible within the time constraints of this study.

Chapter 5 Results

5.1 Introduction

The cohort of 1098 patients was identified from the Yorkshire Specialist Registry for Cancer in Children and Young People and these cases were linked to Hospital episodes statistics with a high degree of linkage (99.7%). Only 3 patients did not link and were not included within the study population. At this point it is important that the levels of data representation within this chapter are clearly defined. There are three main levels at which the data can be presented, outlined below:

1. Case-level – Data presented at the case level will reflect results at an individual cancer patient level. The descriptive statistics for cases are outlined in section 5.2 and the YSRCCYP is the main data source for these results.
2. Inpatient event level – As identified in Figure 4.1 the HES data can be divided into a number of different units depending on the healthcare structure being analysed. The dataset was analysed at the episode level data for the purposes of alert code identification given that each episode is summarised by a unique set of diagnosis codes. It therefore follows that each of the 1098 cases can have multiple episodes; the quantity and nature of which will be described within section 5.3. The dataset was analysed at a CIPS level for the purpose of admission route identification, a CIPS can be made up of multiple episodes and once the initial episode in a CIPS has passed, each subsequent episode will be punctuated by a transfer admission code. Therefore the admission code for the first episode within an alert code containing CIPS will be used to identify the admission method for that inpatient event.

3. Code-level – Each hospital episode contains a series of ICD-10 diagnostic codes which are used to describe the hospital event, and a sequence of up to 20 ICD-10 codes can be used to describe the episode. The individual codes have been analysed in conjunction with national CYA cancer awareness resources in section 5.4. The aim of this analysis was to identify either diagnoses or symptoms suggestive of cancer within hospital episodes occurring prior to the date of definitive diagnosis.

Combined analysis of codes, episodes and cases within this CYA cancer cohort will aim to identify those cases where there may have been hospital involvement suggestive of cancer prior to the date of definitive diagnosis. The time taken to receive a diagnosis for such highlighted cases will be described in section 5.5 and the effect on survival of a prolonged TTD within hospital care will be analysed in section 5.7. Finally the key findings are summarised in section 5.8.

5.2 Descriptive statistics

5.2.1 Demographic profile

Dividing the cohort into the 12 broad ICCC diagnostic groups leukaemia is the most common diagnosis constituting around a quarter of the study population followed by lymphoma (17.9%), CNS tumours (16.6%) and Germ-cell tumours (13.7%). The least common diagnosis being Other cancers (ICCC group XII), have been omitted from Table 5.1 due to the low level of cases, hepatic tumours and retinoblastoma were also very rare. The patterns of tumour incidence are generally in keeping with published data for the CYA population (24, 32).

Table 5.1 summarises the number of cases and the incidence rates per million population within the study cohort. Included is the breakdown of cases by sex, five-year age ranges and broad ICCC diagnostic groups with the overall case number, percentage of the cohort and incidence rate per million. Overall 1098 cases of CYA cancer were included within the study

population there is a higher proportion of males than females (1.35:1). The crude incidence rates of cancer were highest at the extremes of age in this cohort: 206 per million in the very young (0 to 4-years inclusive) and 169 per million 20 to 24-year olds, lowest crude incidence rates seen in the 5 to 9 year olds 102 per million. These crude incidence rates by sex and age are in keeping with those presented for children and young people in the current literature (3, 28, 36).

Table 5.1 Case distribution by frequency, percentage cohort and crude incidence per million of the general population by sex, five-year bands and diagnostic group

	Cases	Percentage of cohort	Crude Incidence Per million
Male	631	57.5	165.76
Female	467	42.5	129.02
Age 0 to 4	278	25.3	206.29
Age 5 to 9	134	12.2	101.71
Age 10 to 14	158	14.4	109.62
Age 15 to 19	235	21.4	147.84
Age 20 to 24	293	26.7	169.33
Leukaemia(I)	262	23.9	35.28
Lymphoma(II)	196	17.9	26.39
CNS(III)	182	16.6	24.51
Neuroblastoma(IV)	41	3.7	5.52
Retinoblastoma(V)	19	1.7	2.56
Renal(VI)	38	3.5	5.12
Hepatic(VII)	15	1.4	2.02
Bone(VIII)	62	5.6	8.35
STS(IX)	67	6.1	9.02
Germ-cell(X)	150	13.7	20.20
Carcinoma(XI)	63	5.7	8.48
Other(XII)	<5	<1	
All Ages	1098	100.0	147.85

The number of cancer cases diagnosed each year fluctuated across the period of study, the peak being in 2005 when 125 cases were diagnosed

and a trough in 2008 when 88 cases of CYA cancer where diagnosed. A clear pattern across the study population could not be established, though the relatively narrow timescale restricts any firm conclusions to be made relating to the temporal changes in incidence. Figure 5.1 presents the year on year percentage changes in the number of cases diagnosed by each ICCC grouping using 2004 as a baseline. Whilst the absolute figures are not important it is worth noting the general uniform rates of cases diagnosed for each ICCC group across the study period, with the exception of the group IV/neuroblastoma population. The apparent sharp rise in the number of cases of neuroblastoma diagnosed from 2005 onwards is an outlier within the data and likely reflects how small variations within a small population of cases can result in high percentage changes, no specific interpretation of this pattern is apparent. The data in Figure 5.1 was presented due to the low number of cases in a number of subgroups when separating the study population by year of diagnosis and diagnostic group, thus avoiding potentially patient identifiable data being presented.

Figure 5.1 Trends in incidence by diagnosis 2004-2009

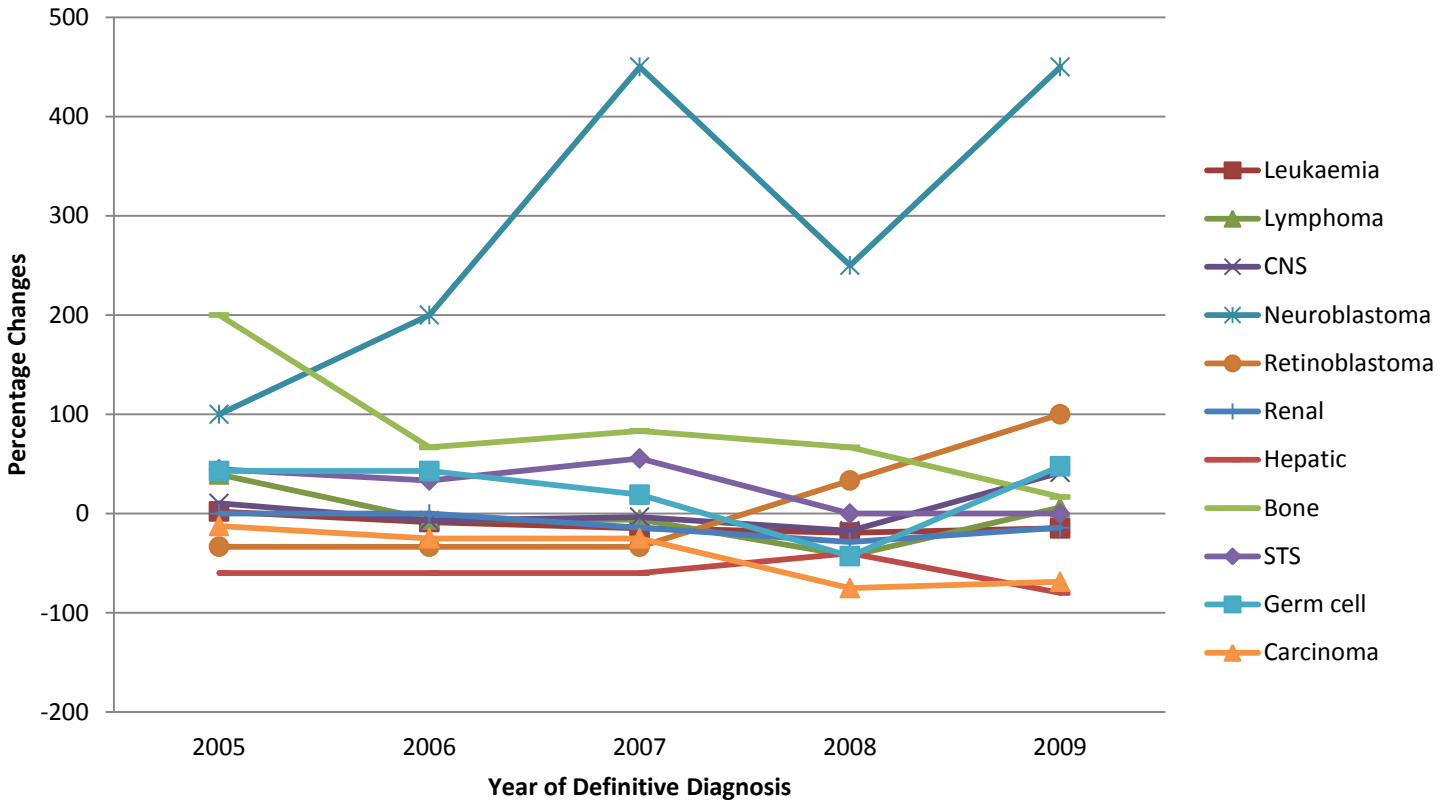


Figure 5.2 Cancer in 0 to 24 year-olds in Former Yorkshire Regional Health Authority by five year age bands (%)

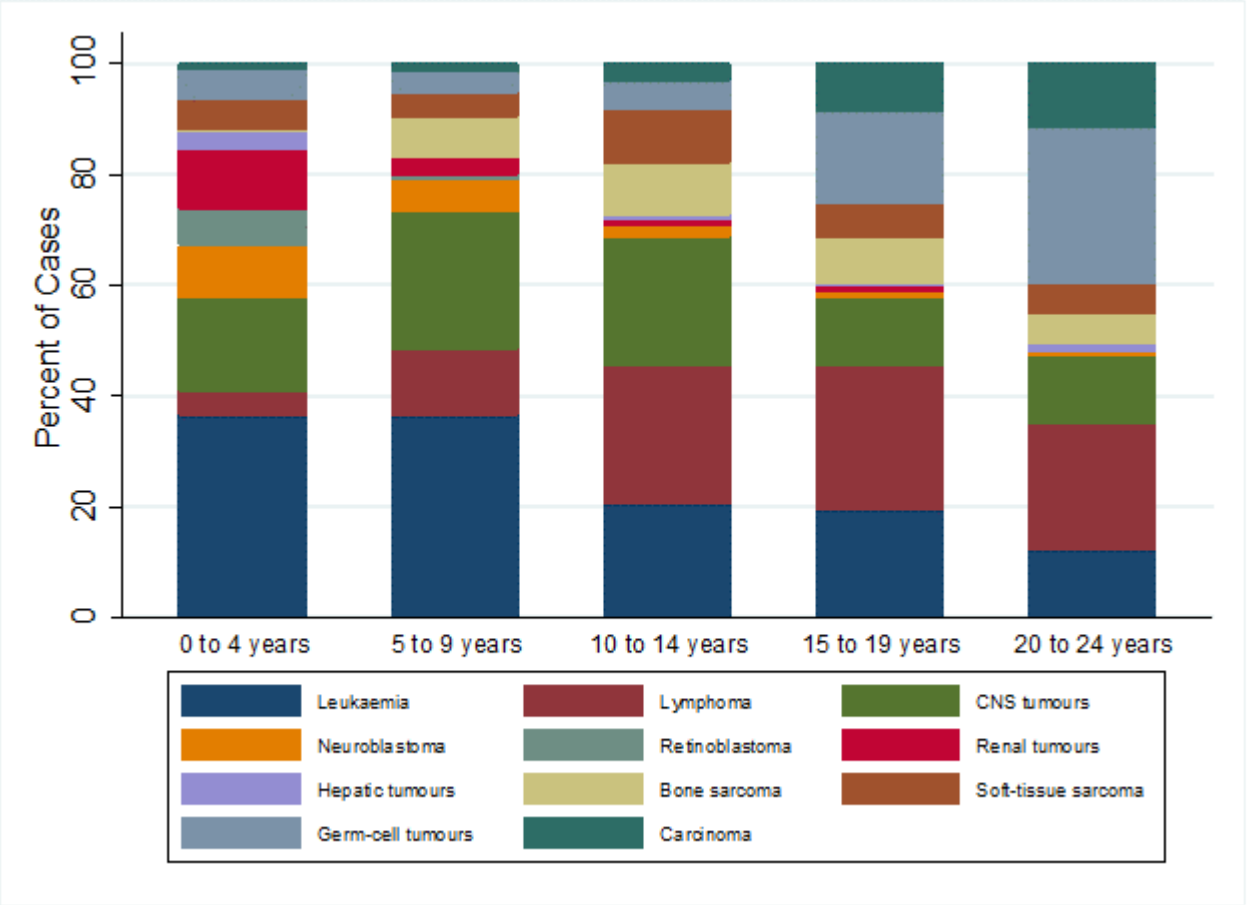


Figure 5.2 presents the breakdown of cases by diagnostic group within the five-year age bands across the study population, this graph elaborates the variation in case profile across age in the study cohort. In the younger population 0 to 4-years at diagnosis there is a clear predominance of leukaemia which contributes over a third of cancer diagnoses with other solid tumours such as CNS tumours, neuroblastoma, retinoblastoma, renal tumours and soft-tissue sarcoma also featuring. The predominance of leukaemia continues in the 5 to 9-years group, where there appears to be a higher proportion of CNS tumours and lymphoma with a reduction in some of the aforementioned paediatric solid tumours. The proportion of CNS tumours subsequently reduces with increasing age similar to leukaemia, however both these diagnoses are still major contributors to the cancer burden across all age ranges within the study population. The percentage of lymphoma diagnoses markedly increases within the 10 to 14 year age group and this increase is sustained in the older age-ranges. There is a steady increase in the contribution of germ-cell tumours and carcinoma within the 15 to 24 year population. Figure 5.2 highlights the challenge of analysing the CYA population and how clearly the profiles of cancer diagnosis and subsequently the biology and behaviour of tumours changes with age.

5.3 Hospital events

5.3.1 Inpatient events

The 1098 cases that met the inclusion criteria for this study were linked to nearly 30,000 inpatient HES episodes. These case were further restricted to only those episodes for which the episode start date preceded or coincide with the date of definitive diagnosis of a primary cancer, hereafter such episodes will be referred to as “pre-diagnosis episodes” (n=3,558). From the 1098 eligible cases, these included 47 cases which were identified as having no pre-diagnosis inpatient episodes. The median age at diagnosis for these 47 cases was 22 years, whilst a diagnosis of either a germ-cell tumour or carcinoma, more commonly seen in TYAs, was present in 23 cases. These cases were incorporated within the overall descriptive analysis and survival analysis but were excluded from episode and code level analysis.

The cases with no pre-diagnosis inpatient involvement included: 16 germ-cell tumours, 8 lymphomas, 7 carcinomas, 6 leukaemias and 5 or fewer cases with soft-tissue sarcoma, bone sarcoma and CNS tumours. No cases were represented in the neuroblastoma, retinoblastoma, renal and other tumour groups. The age of these 47 patients ranged from 2.6 to 24.9 years and the median age at diagnosis was 21.5 years with a mean of 12.7 years. This suggests a skewed distribution towards an older age at diagnosis for these cases.

5.3.1.1 Pre-diagnosis events

Figure 5.3 presents the frequency of pre-diagnosis inpatient episodes by the year according to the episodes starting date. The highest frequency of episodes occur during the years corresponding to the inclusion criteria for date of definitive diagnosis; 2004 to 2009 inclusive. No episodes were included beyond the end of 2009 as they will all occur after the date of diagnosis of the final case. The steady increase in episodes toward 2004 most likely reflects the year on year addition of cases as more children are born. It is worth noting the 13 year time span of pre-diagnosis inpatient events highlighted within the HES episodes for this study cohort.

Figure 5.3 Frequency of pre-diagnosis inpatient episodes by year (taken from the episode start date)

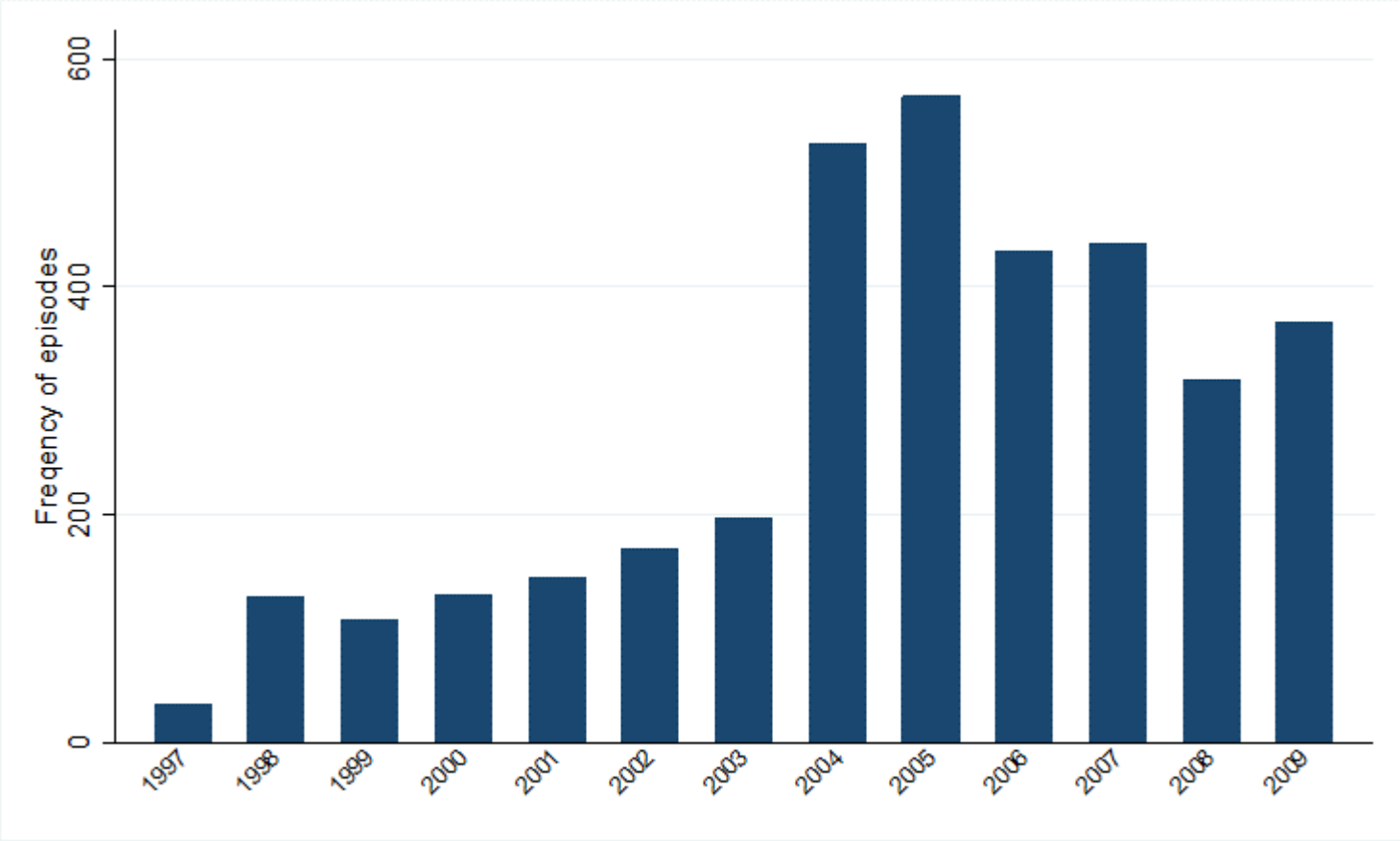


Figure 5.4: Year of pre-diagnosis inpatient episode occurrence (taken from episode start date) by year of definitive diagnosis of cancer

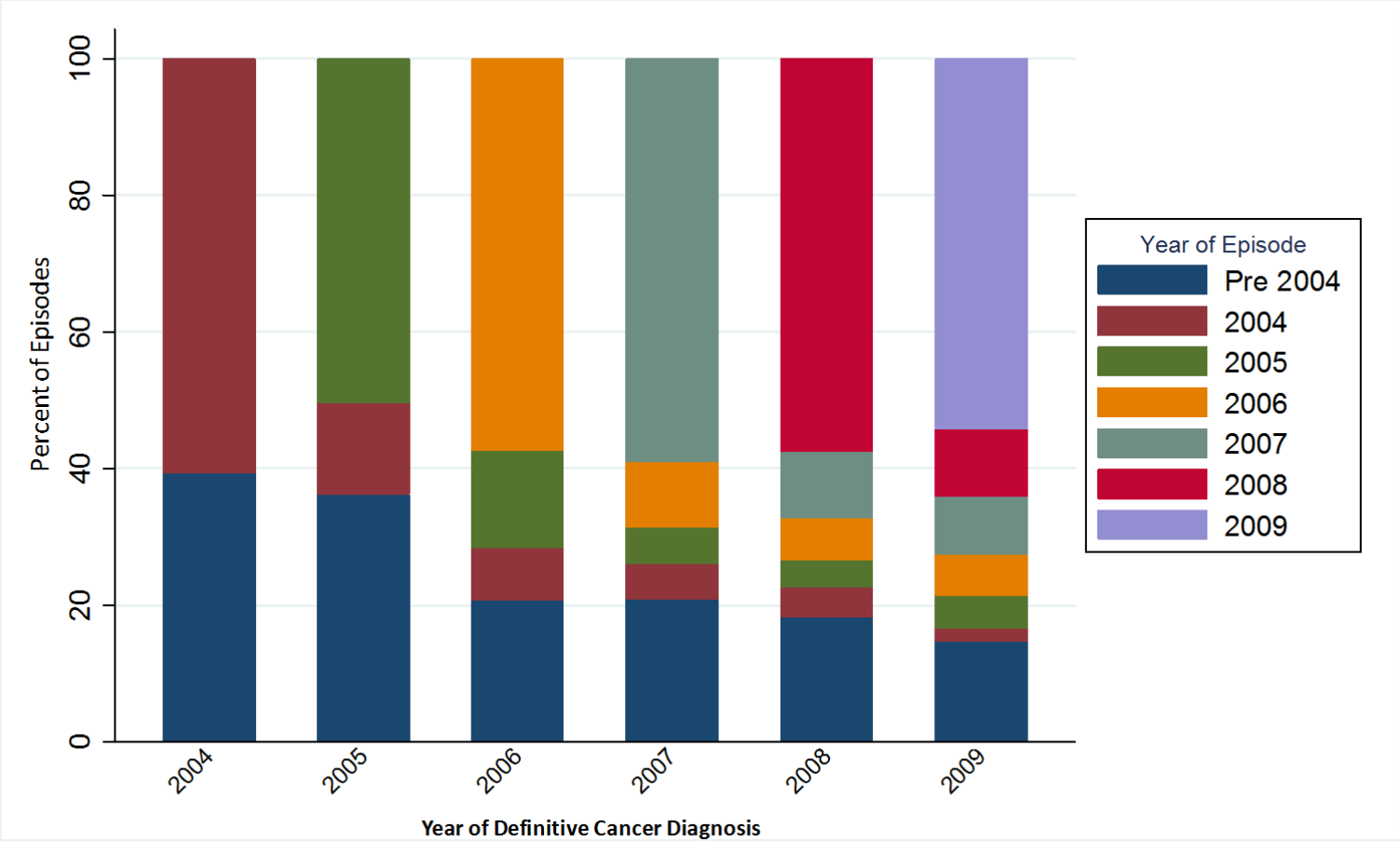


Figure 5.4 presents the stacked percentages for the year a pre-diagnosis episode occurred separated into the year of definitive cancer diagnosis. This highlights that the majority of pre-diagnosis inpatient episodes occurred in the same year as the case was diagnosed.

The frequency of episodes only portrays parts of the picture and as stated in the introduction a single case can have multiple pre-diagnosis episodes. From the 1098 cases identified 47 were found to have no inpatient episodes in the linked HES dataset, out of the remaining 1051 cases with pre-diagnosis inpatient involvement the maximum number of episodes per patient was found to be 62. Figure 5.5 is a box (median and inter-quartile ranges Q1-Q3) and whisker (Q_3+1.5*(Q_3-Q_1) and Q_1-1.5*(Q_3-Q_1) plot of total inpatient episodes per case by sex for each of the five year age-groups, the data are summarised in such a manner due to the highly skewed distribution of the pre-diagnosis inpatient episodes per case. There were statistically significant differences in the median number of pre-diagnosis inpatient episodes between the five age-range groups based on the Kruskall-Wallis test (p=0.001): a median of three episodes per case in children under 10 years old at diagnosis compared to a median of two pre-diagnosis episodes per case in the older cases (10-14, 15-19, 20-24). The median number of pre-diagnosis episodes is the same between the sexes for each age group with the exception of 20 to 24 years olds for whom females have a higher median of three episodes compared to two for males. Females in the 20 to 24 year age group also have a markedly wider inter-quartile range (0 to 12 pre-diagnosis episodes) compared to all other age and sex categories.

The findings of variation in the number of pre-diagnosis episodes by age-bands could reflect the interaction between age and the underlying diagnosis, Figure 5.6 indicates variations in the median number of pre-diagnosis episodes per case by sex for each ICCC diagnostic group. The largest median number of pre-diagnosis episodes per case occurs in paediatric solid tumours such as neuroblastoma, hepatic tumours and renal tumours. The lowest median pre-diagnosis episodes per case are seen in tumours which predominate in the young adult populations such as carcinoma, germ-cell tumours, bone sarcoma and lymphoma. The reduced secondary care involvement in the aforementioned tumour groups may

reflect different patterns of healthcare engagement within TYAs. Another potential explanation for the differences in pre-diagnosis inpatient episodes by age-range and diagnosis is incomplete recording of pre-diagnosis inpatient events within increasing age. The earliest year of inpatient involvement was 1996, see Figure 5.3, suggesting there is incomplete data availability due to the poor quality of information available prior to this date. This variability in inpatient episodes with time is likely to reflect the evolution and implementation of HES and coding within healthcare structures. There is a roughly similar distribution presented between the sexes for each diagnostic group with the exceptions of neuroblastoma and soft-tissue sarcoma where there are wider inter-quartile ranges in the males compared to females, whilst the reverse is seen in the germ-cell tumour population.

The admission route, coded in HES by the ‘admimeth’ variable, may give an indication of the extent of missing data as age increases, particularly focusing on admissions assigned to birth and maternity involvement. The number of episodes by admission method is summarised in the Figure 5.7. Admission method codes relating to maternity services admissions (admimeth 31,32) and admission for birth (admimeth 82,83) are presented in Figure 5.7 as “Maternity” (admimeth 31, 32) and “Other” (admimeth 82,83). When considering these two admission methods for pre-diagnosis episodes between the age-ranges there are clear opposing patterns seen with maternity codes peaking in the older age group 20-24 and other codes in the 0-4 age**,** see table 5.2. The predominance of codes relating to a case being born and classified as “other” in the 0 to 4 years age range is the most likely explanation for the statistically significant difference in the distributions of total pre-diagnosis inpatient episodes by age range. This finding reflects the incomplete availability of data relating to birth for the older cases within the study cohort.

Table 5.2 Admission method by age at definitive diagnosis in five-year age bands, all pre-diagnosis episode level data

	5-year age bands				
	0 to 4	5 to 9	10 to 14	15 to 19	20 to 24
Emergency A&E	212	121	114	181	215
Emergency GP	149	73	65	63	91
Emergency Outpatient	19	9	18	26	31
Emergency Transfer	124	36	53	40	36
Elective	243	98	216	302	316
Transfer	134	71	48	36	26
Maternity	≤5	0	0	19	103
Other	250	53	≤5	0	0
Not Known	≤5	0	≤5	0	0

Figure 5.5 Box and whisker plot of number of pre-diagnosis episodes per case of children and young adults cancer by 5 year age bands

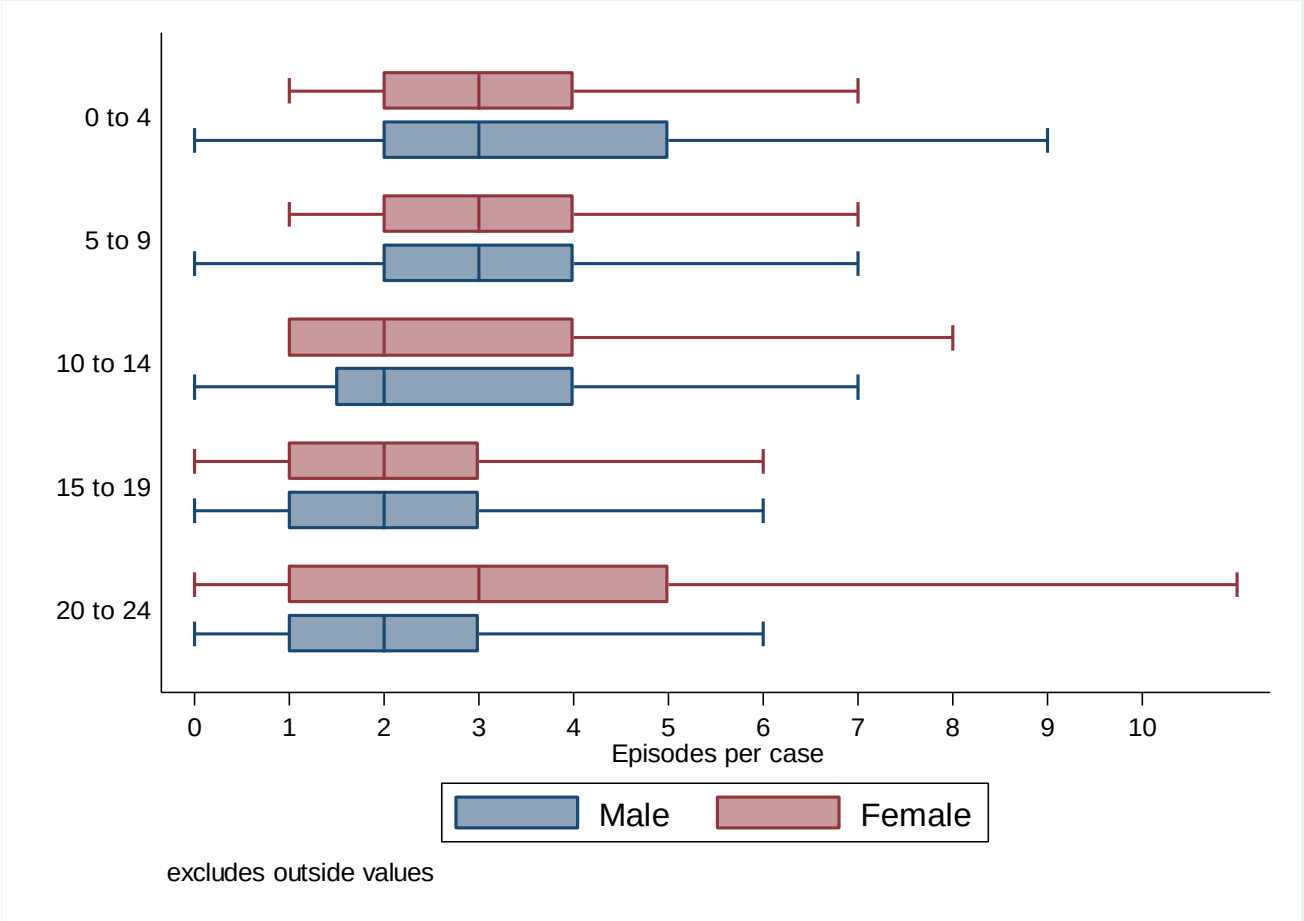


Figure 5.6 Box and Whisker plot of number of pre-diagnosis episodes per case of children and young adults cancer by sex and diagnostic group

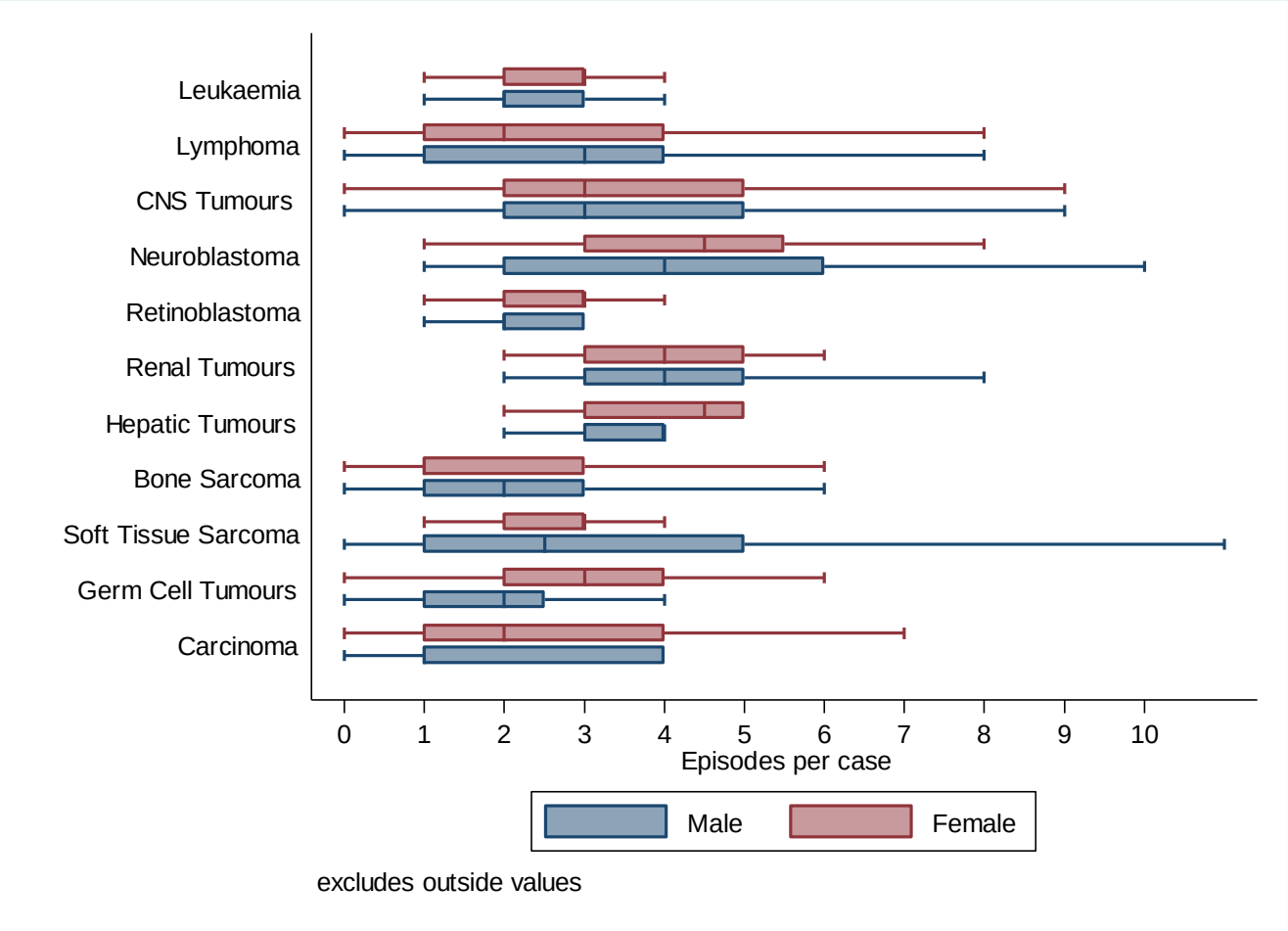
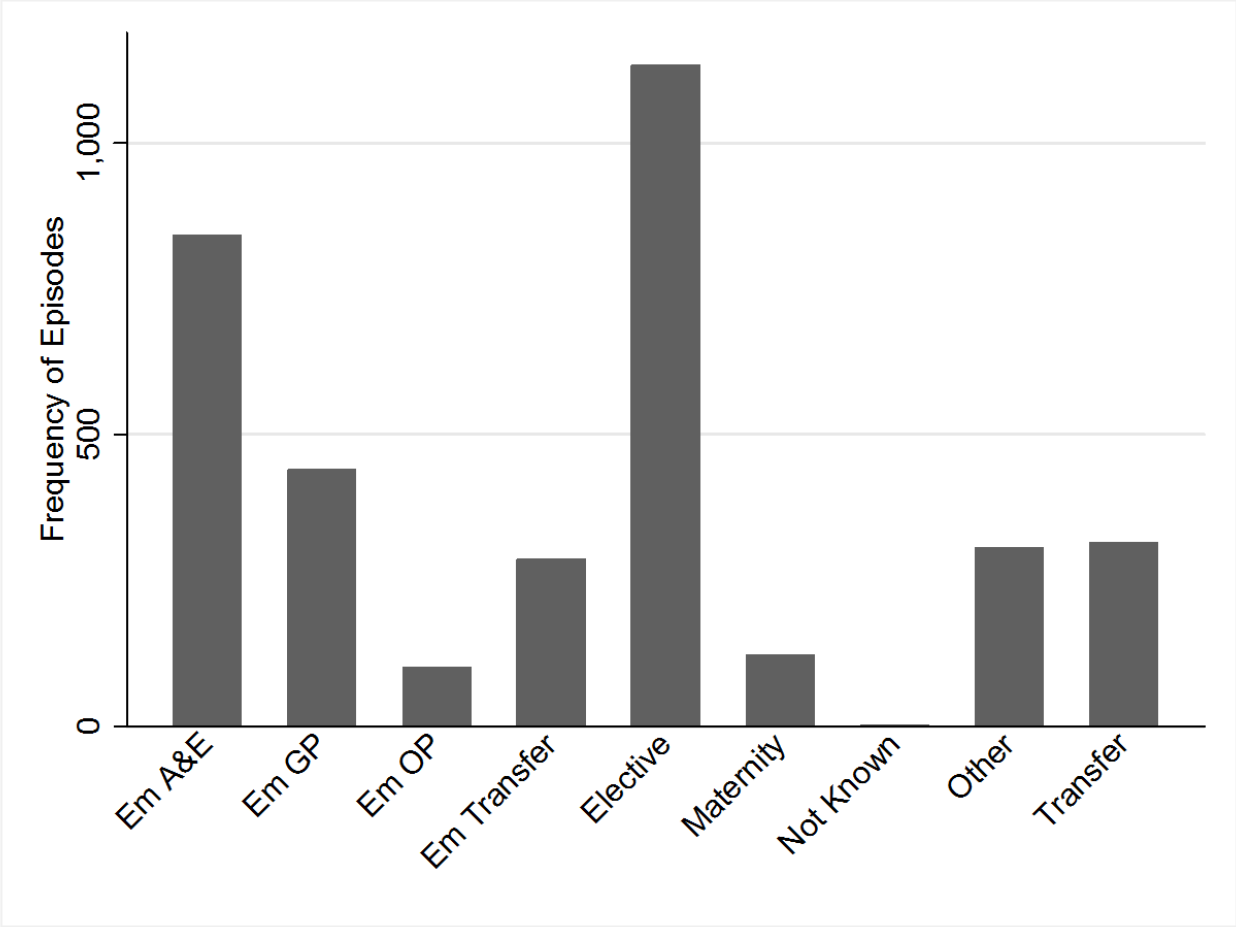


Figure 5.7 Frequency of pre-diagnosis inpatient events by admission method



5.3.2 Outpatient events

Outpatient HES episodes were linked to the YSRCCYP data, and are contained in a separate dataset to the inpatient data.

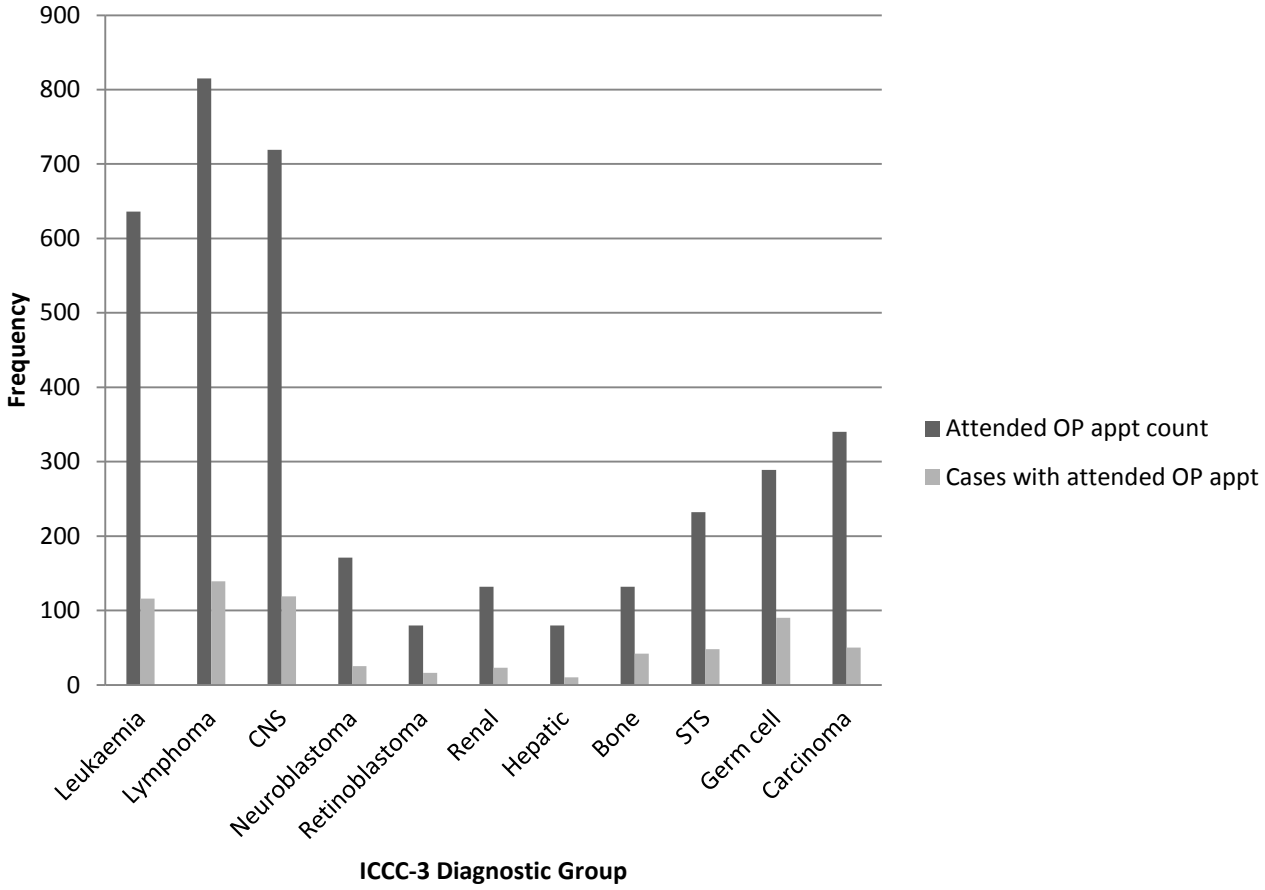
5.3.2.1 Pre-diagnosis events

The appointment date variable was used in a similar fashion to the episode start date and pre-diagnosis episodes were separated if the appointment date preceded or coincided with the date of definitive cancer diagnosis. There were 4,353 pre-diagnosis outpatient episodes identified within the HES data for the study population. Only 790 cases had pre-diagnosis outpatient HES events from the 1098 cases in the study cohort. The attendance code variable was used to highlight appointments that were attended, the exclusion of non-attended appointments reduced the overall number of pre-diagnosis outpatient appointments to 3,648. Figure 5.8 presents the number of pre-diagnosis outpatient appointments attended by ICCC diagnostic group. There is a predominance of the first three tumour groups as with the inpatient data, however there are fewer appointments seen in the paediatric solid tumour population compared to the germ-cell and carcinoma groups, which is in contrast to the inpatient data.

The priority of the requested outpatient appointment may help identify those appointments for cases where a cancer is suspected. The two week referral pathway for suspected cancer applies in the CYA population as with adults, however within this study population only 10 pre-diagnosis outpatient appointment were coded as two week cancer referrals. This identifies either an extremely poor uptake of the pathway by referring doctors in CYAs suspected of cancer or a deficiency in coding of this pathway.

In reviewing the diagnostic coding data within the outpatient events it is clear that there is no effort made by the coder to reflect the clinical events of the appointment as all codes are either R69X6 or R69X8 both falling into the category of “unknown or unspecified causes of morbidity”. The analysis of outpatient data for the presence of pre-diagnosis alert codes events was therefore not possible within this study.

Figure 5.8 Frequency of pre-diagnosis out-patient appointments and number of cases with pre-diagnosis out-patient appointments by diagnostic group (excluding group XII)



5.4 Diagnosis and symptom codes suggestive of cancer

This section will provide results of the analysis of ICD-10 codes contained within the pre-diagnosis inpatient episodes. This analysis will focus on ICD-10 codes identified as potentially relating to the subsequent diagnosis of cancer, as referenced against national CYA cancer awareness campaigns identified in section 4.8.4. The linked outpatient dataset will not feature within this section due to the lack of adequate presentation of ICD-10 codes within this data as discussed in section 4.8.3. The data within this section will be presented at a code level, codes per episodes or codes per case level. The section will comprise four main sub-sections: the first section will briefly look at ICD-10 cancer codes occurring prior to the definitive cancer diagnosis as recorded by the YSRCCYP, the second section is describes the identified pre-diagnosis codes relating to a broad diagnosis of cancer and as such will be referred to as “broad diagnosis alert codes”, the third section related to those codes relevant to each specific ICCC diagnostic group and as such will be referred to as “diagnosis specific alert codes”. Finally, the fourth section will bring together the broad diagnosis alert codes and the diagnosis specific alert codes to discuss the TTD for CYA cancer within the study population.

At the outset of this stage of analysis it is important to refer back to the background section. The ICD-10 classification system provides the codes for the HES episodes and up to 20 ICD-10 codes can be as used to provide a clinical description of an inpatient episode. The ICD-10 coding system is predominant based around diagnosis codes however there are symptom codes within the R section of the system. As discussed in section 4.8.4.2 the symptom codes are only present within episodes where a symptom or sign appears that cannot be directly attributed to the underlying diagnosis for that episode. For example, if a case has a headache recorded within an episode but the episode related to a lower respiratory tract infection the headache will be recorded by the coder.

Table 5.3 summarises the number of ICD-10 codes, the pre-diagnosis inpatient episodes and the number of cases by the ICCC diagnostic group.

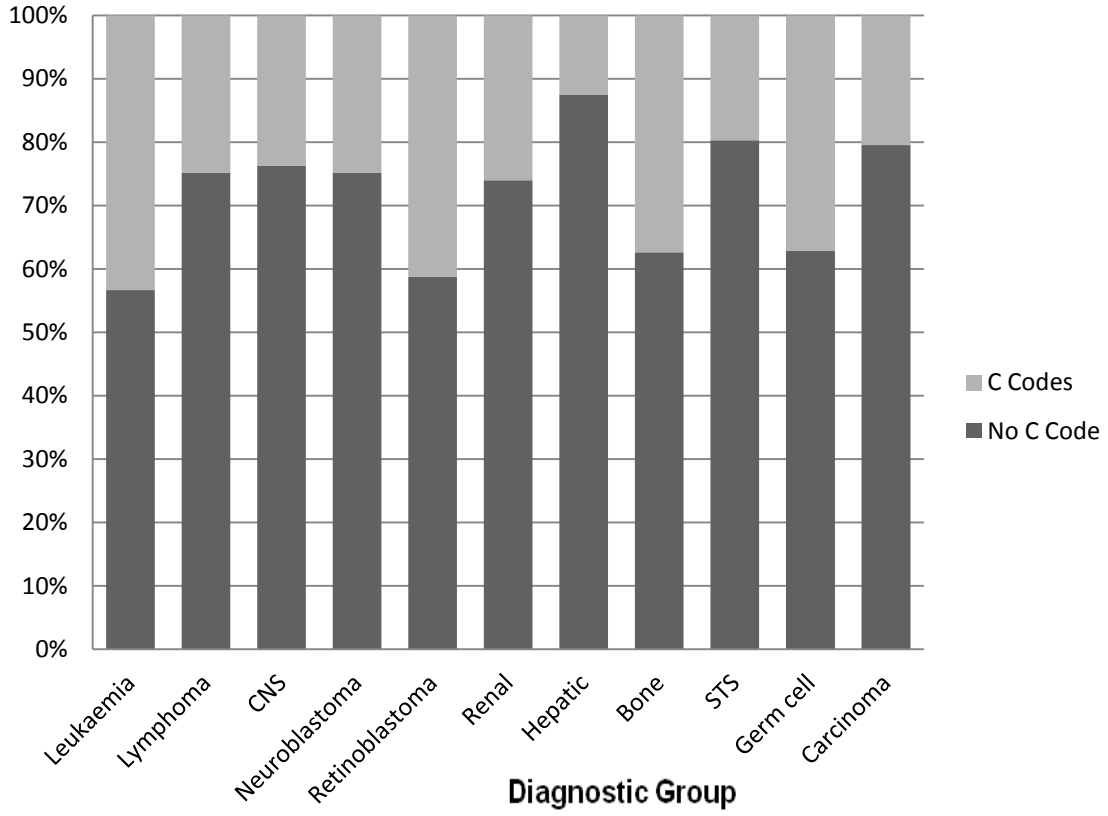
Table 5.3 Summary table of frequency of pre-diagnosis ICD-10 codes, inpatient episodes and cases by diagnostic groups

Diagnostic group	Number of Codes	Number of Episode	Number of Cases
Leukaemia	1808	736	262
Lymphoma	1419	707	196
CNS tumours	1296	686	182
Neuroblastoma	458	201	41
Retinoblastoma	85	63	19
Renal Tumours	490	196	38
Hepatic Tumours	291	96	15
Bone Tumours	230	139	62
Soft-Tissue Sarcoma	429	228	67
Germ-Cell Tumours	629	347	150
Carcinoma	405	200	63

5.4.1 ICD-10 codes relevant to a diagnosis of cancer

The difference in the date of diagnosis and a C-code date is given in section 4.8.3.2. A high proportion (29%) of the total number of episodes contained a C code: 1,018 C-code containing episodes from the 3,558 pre-diagnosis episodes. Figure 5.9 presents stacked percentages of episodes by C code status for each diagnostic group excluding group 12. Leukaemia, retinoblastoma, bone tumours and carcinoma all have C codes present in greater than 30% of the pre-diagnosis episodes. 95% of the C code containing episodes occur within the month preceding diagnosis though there are cases which have C code containing episodes as long as 2 years prior to the date of definitive cancer diagnosis. The lymphoma, CNS tumours, retinoblastoma and germ-cell cases all contain C codes at a duration of more than a month preceding the date of definitive cancer diagnosis.

Figure 5.9 The cancer codes (ICD-10 C codes) status (present or not) of pre-diagnosis inpatient episodes for children and young adult cancers in FYRHA by diagnostic group (stacked percentages of total pre-diagnosis episodes per diagnostic group)



5.4.2 Diagnostic and symptom codes suggestive of a general cancer diagnosis

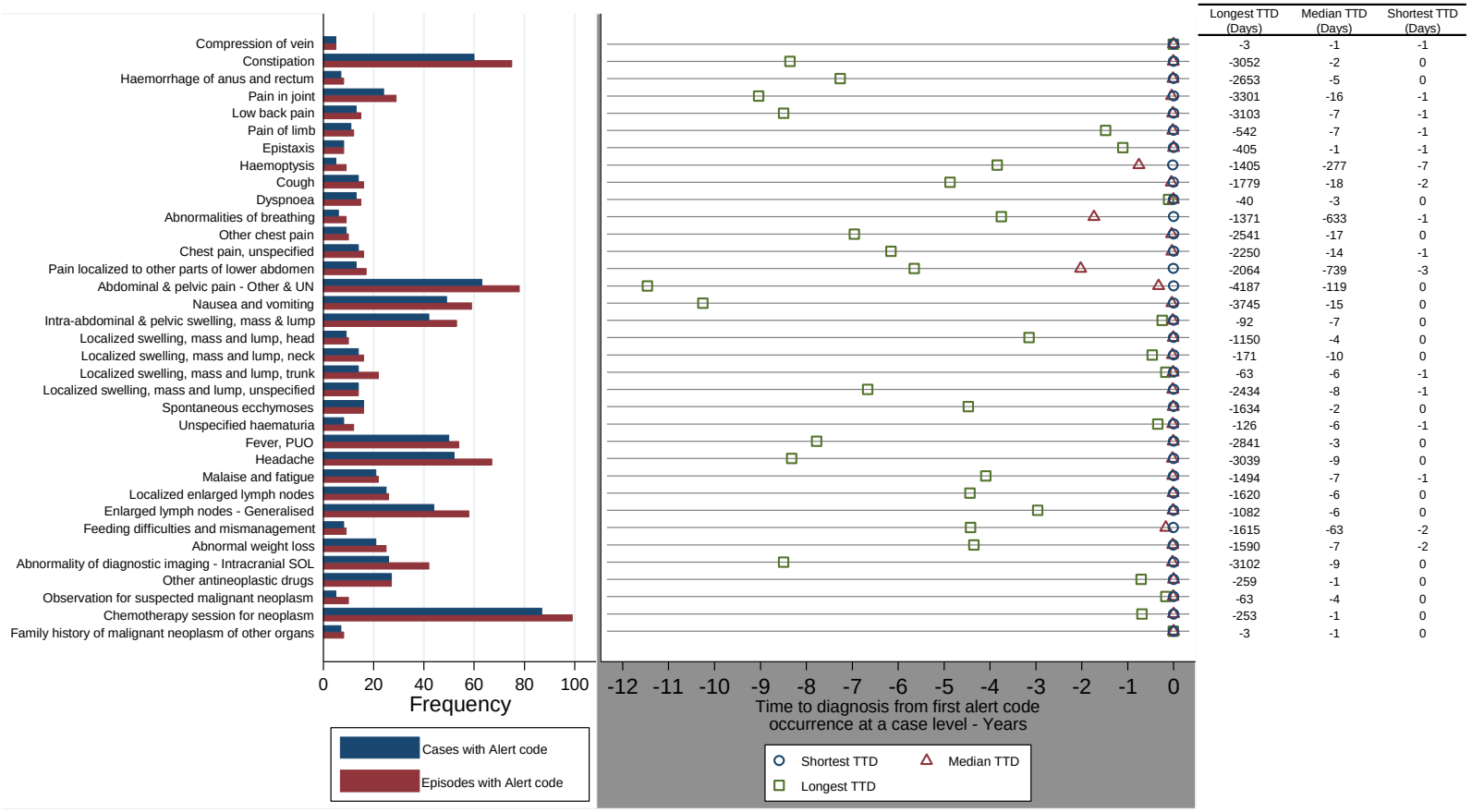
The NICE 2005 referral guidance and the TCT awareness campaign provide the majority of awareness resource relevant to a broad diagnosis of cancer, see appendix 2. 64 broad cancer alert codes were identified by cross referencing the awareness resource against the 1402 unique ICD-10 codes extracted from the pre-diagnosis episodes for the study population of 1051 cases with pre-diagnosis inpatient episodes. Of the 64 identified codes over half occurred in more than five cases within the population (n=34). Figure 5.10 summarises those broad cancer alert codes which occur in five or more cases within the study population. The ICD-10 symptom codes or “R” codes provided 24 of the 34 codes presented in Figure 5.10. There are four ICD-10 codes presented in Figure 5.10 that related directly to the investigation or treatment of cancer or a personal or family history of cancer. These codes were included within the analysis to provide quality markers for the HES data, identifying pre-diagnosis episodes where cancer was suspected or treatment for cancer was already underway. Such codes were included to identify cases where the diagnosis of cancer was known well before the definitive diagnosis was made, however the occurrence of such codes appear to coincide closely with the date of definitive diagnosis.

The three most common ICD-10 codes contained within the codes relevant to a broad cancer diagnosis are related to “Constipation”, “Abdominal & pelvic pain” and “Headache. It is also apparent that the number of cases with alert codes is similar to the number of episodes with alert codes as one can see by the close correlation of the two bars in general in Figure 5.10. This reflects the fact that each broad alert code only occurs once within the pre-diagnosis codes for each cases.

Figure 5.10 presents the longest, the shortest and the median duration of time from the point of initial occurrence of a broad alert code to the date of definitive diagnosis, this is presented as case level data. For a number of broad cancer alert codes there are outlier cases for whom that code is present long before date of definitive diagnosis, however in the majority of cases when a broad alert codes occurs it is within close proximity to the date

of definitive diagnosis, as reflected by the proximity of the median TTD to the date of diagnosis. Codes that appear to have a median TTD well before the date of definitive diagnosis include; “Heamoptysis”, “Abnormalities of breathing”, “ Pain localised to other parts of the lower abdomen” and “Abdominal and pelvis pain”.

Figure 5.10 Frequency of pre-diagnosis broad cancer diagnosis alert code containing episodes by episodes and case presented along-side the longest, median and shortest duration of time between the first occurrence of such episodes at case level



5.4.3 Diagnostic and symptom codes suggestive of a specific cancer diagnosis

When analysing the 1402 unique ICD-10 codes for their potential relevance to a specific cancer diagnoses, the data had to be segregated by ICCC diagnosis, resulting in the generation of twelve separate datasets. This was necessary due to the constraints of Stata12 programming. From the twelve diagnostic groups 216 unique ICD-10 codes were identified across the study cohort, Table 5.4 presents the number of alert codes identified for each ICCC diagnostic group, group XII is not presented as there was no available awareness literature and less than five cases overall. The diagnosis specific alert codes occurring in five or more cases are presented in Table 5.4.

Table 5.4 The number of diagnosis specific alert codes in all cases and the number of diagnosis specific alert codes occurring in five or more cases by diagnostic groups

ICCC-3 group	Unique Diagnosis Specific Alert Codes	Codes occurring 5 or more cases
Leukaemia	28	12
Lymphoma	49	11
CNS tumours	58	8
Neuroblastoma	17	1
Retinoblastoma	4	0
Renal Tumours	11	3
Hepatic Tumours	4	0
Bone Tumours	19	1
Soft-Tissue Sarcoma	15	2
Germ-Cell Tumours	25	4
Carcinoma	6	0

Diagnosis specific alert codes for the 262 cases within the leukaemia population are presented in Figure 5.11, these codes predominantly relate to either pain or some degree of haematological deficiency. Eight of the twelve

codes presented in Figure 5.11 occur in close proximity to the date of definitive diagnosis, of the remaining four only “Down’s syndrome” appears to have a prolonged median time between first occurrence and diagnosis. The presence of Down’s syndrome within the specific alert codes for leukaemia is based around the NICE referral guidance that advises a doctor should have increased awareness of this cancer type in children with Down’s syndrome (125). This code applies to the diagnosis of Down’s syndrome which is made early in a child’s life, it therefore appears during all the inpatient events and will be coded in HES episodes from the point a diagnosis is made. Where present, this alert codes will result in an apparent long TTD for a case.

“Agranulocytosis” is by far the commonest diagnosis specific alert code within the leukaemia population and is applied to a finding of a severe deficiency of white blood cells called granulocytes. Within the ICD-10 system this code most commonly relates to the deficiency of a specific granulocyte called the neutrophil, severe deficiency of which can result in life threatening immune-suppression. Agranulocytosis can occur in certain autoimmune conditions common or viral and bacterial infections or as a side-effect of drug therapy, however it can also occur due to bone marrow infiltration often seen at diagnosis in the leukaemia’s and as a result of bone marrow suppression secondary to the effects of chemotherapy (131). This code occurs in 99 cases prior to the date of diagnosis, the median duration between the first occurrence of this code and the definitive diagnosis is just one day. The first occurrence within this population is over two years prior to diagnosis in a case where there is an underlying diagnosis of aplastic anaemia a condition which can predispose to the development of leukaemia.

Within the lymphoma cases (n=196) a total of 49 diagnosis specific alert codes were identified, see Figure 5.12. Eleven of the forty-nine codes occurred in five or more cases of which 10 were R codes relating to symptoms and signs not elsewhere classified, the other code being “agranulocytosis” a code discussed in the previous section on the leukaemia cases. The commonest pre-diagnosis specific alert codes in the lymphoma population was “enlarged lymph nodes-generalised” followed by “abdominal pain”, “agranulocytosis” and “localised swelling, mass and lump, neck”.

“Fever, PUO” and “abdominal & pelvic pain” were the only two codes for which the median duration from first occurrence to date of definitive diagnosis was greater than 14 days. The longest duration of time between first occurrence to definitive diagnosis for these two codes was greater than six-months prior to diagnosis, as was case for “enlarged lymph nodes-generalised” and “abnormal diagnostic imaging of the lung”. This all adds to the picture of common lymphoma specific alert codes occurring in close proximity to the date of definitive diagnosis. Though the predominance of R codes highlights the presence of unexplained symptoms or signs within the inpatient episodes for these case preceding diagnosis.

The CNS tumour population had the most diagnosis specific alert codes; 58 codes. This was possibly due to increased pre-diagnosis signs and symptoms suggestive of cancer within this population or alternatively the extent of the awareness literature in this tumour type relevant to CYA’s, see Figure 5.13. Indeed the HeadSmart campaign provides CNS tumour specific awareness advice for 0 to 18 year olds. Within the CNS specific alert codes codes a number relate to the diagnosis of epilepsy, headache, cranial nerve palsies, hydrocephalus as well as a number of R codes. There are also codes relevant to the underlying diagnosis of Tuberous Sclerosis and Neurofibromatosis. From the 58 diagnosis specific alert codes identified only eight appear in five or more cases. However, when grouping codes of a similar nature; codes relevant to epilepsy appeared in 29 cases, specific diagnoses of “headache or migraine” appeared in five cases, “focal neurological signs” were present in 24 case and symptoms or signs not elsewhere classified were present in 101 cases. The “epilepsy” and “convulsion” codes have a high ratio of episodes to cases, indicated by the height of the episodes bar comparative to the cases bar in Figure 5.13, suggesting these codes appear multiple times for each case prior to the date of definitive diagnosis. For “epilepsy” this finding is understandable as it is a known diagnosis, however “convulsion” is an R code which suggests its occurrence within an episode is unexplained. Couple this with the finding that the median duration between first occurrence of a “convulsion” codes and definitive diagnosis is more than one year identifies this code as a potentially concerning alert code within the CNS population. The codes

relevant to a diagnosis of epilepsy also appear to have a prolonged duration of time between first occurrence and definitive diagnosis, however the median duration for the other commonly occurring codes appears to be short.

From the 115 cases of solid tumours that predominate in childhood; neuroblastoma, retinoblastoma, renal tumours and hepatic tumours, only 35 disease specific alert codes were identified. Only four codes were contained within five or more cases, with no codes present from the retinoblastoma or hepatic tumour groups, see Figure 5.14. The 41 neuroblastoma cases only yielded a total of 17 diagnosis specific alert codes and from these only “intra-abdominal & pelvic swelling, mass & lump” was a common code in five or more cases. This code only occurred in close proximity to the date of definitive diagnosis. The renal tumour group comprised 38 cases and yielded 10 diagnosis specific alert codes with three occurring in five or more cases. “Intra-abdominal & pelvic swelling, mass & lump” was the most frequent diagnosis specific alert code featuring in 15 cases preceding diagnosis and at a frequency of once per case. All the commonly occurring renal tumour specific alert codes occurred in close proximity to the date of diagnosis.

The Bone Cancer Research Trust provides a specific awareness resource for bone tumours in CYA’s and in combination with the NICE 2005 referral guidance identified nineteen diagnosis specific alert codes in the 62 bone tumours cases. Codes relating to pain were present prior to diagnosis in eight cases, six cases had codes referencing a bony injury, four codes in four cases related to the mechanism of injury such as a fall and four codes present in five cases related to abnormalities of musculoskeletal imaging. Only “Pain in joint” appeared in five or more cases with a median duration from first occurrence to definitive diagnosis of 11 days, however the longest duration between the first occurrence of this code and definitive diagnosis being 1.6 years, however there was no further inpatient involvement for this particular case until the time immediately preceding diagnosis. This case highlights the fact that this study only provides a part of the healthcare pathway to diagnosis due to the focus on secondary care.

Soft-tissue sarcoma awareness resources were limited to the NICE 2005 guidance, which identified fifteen diagnosis specific alert codes present in the pre-diagnosis episodes of the 67 included cases. In a similar manner to the bone sarcoma population only one code appeared in five or more cases and was the “Intra-abdominal & pelvic swelling, mass & lump” code seen in several other tumour groups. Due to the varied primary presenting site of STS in CYA’s there was a wide range of anatomical sites included within the diagnosis specific alert codes included were codes relevant to pathology in the ear, nose and throat, cranial nerves, genitourinary system and musculoskeletal system.

Germ-cell tumours are the fourth most common diagnostic group contributing 150 cases to this study. This is a diagnosis that can affect any age, although the peak incidence is in the infant and then teenagers and young adults. Germ-cell tumours can present at a variety of sites around the body though they predominantly appear in the sex organs, along the midline of the body and in the central nervous system. The awareness information for germ-cell tumours was taken from NICE 2005 referral guidance, the TCT awareness campaign, Macmillan website and also incorporated awareness resources relating to CNS tumours. These resources identified 25 diagnosis specific alert codes within the codes extracted from the pre-diagnosis episodes: 19 cases contained codes relating to the central nervous system, 23 cases contained codes relating to the abdomen and 24 cases contained codes specifically relevant to either the male or female genitourinary systems. Only four codes were present in five or more cases, the most common being “other specified disorders of male genital organs” present in 14 cases, “other and unspecified ovarian cysts”, “abdominal & pelvic pain” and “intra-abdominal & pelvic swelling, mass & lump” each occurred in five cases, see Figure 5.14. The duration of time between first occurrence and diagnosis was prolonged for “abdominal & pelvic pain” with a median duration of more than four years preceding diagnosis. The longest duration between first occurrence of a code and diagnosis being over 10 years seen in “other specified disorders of the male genital organs”, however the median duration in cases containing this code was one day.

There was very limited awareness literature available for the carcinoma group, which is the predominant cancer in adults. The awareness literature for the adult cancers however is based on the primary site of a cancer and was not always relevant to this study population, as such the awareness resources used in the analysis were limited to the diagnosis of thyroid carcinoma. Only six diagnosis specific alert codes were identified using the applied methods and none of these occurred in five or more of the 63 carcinoma cases included within the study cohort.

Figure 5.11 Frequency of pre-diagnosis leukaemia specific alert code containing episodes occurring in five or more cases by episodes frequency and number of cases presented along-side the longest, median and shortest duration of time between the first occurrence of such episodes at case level

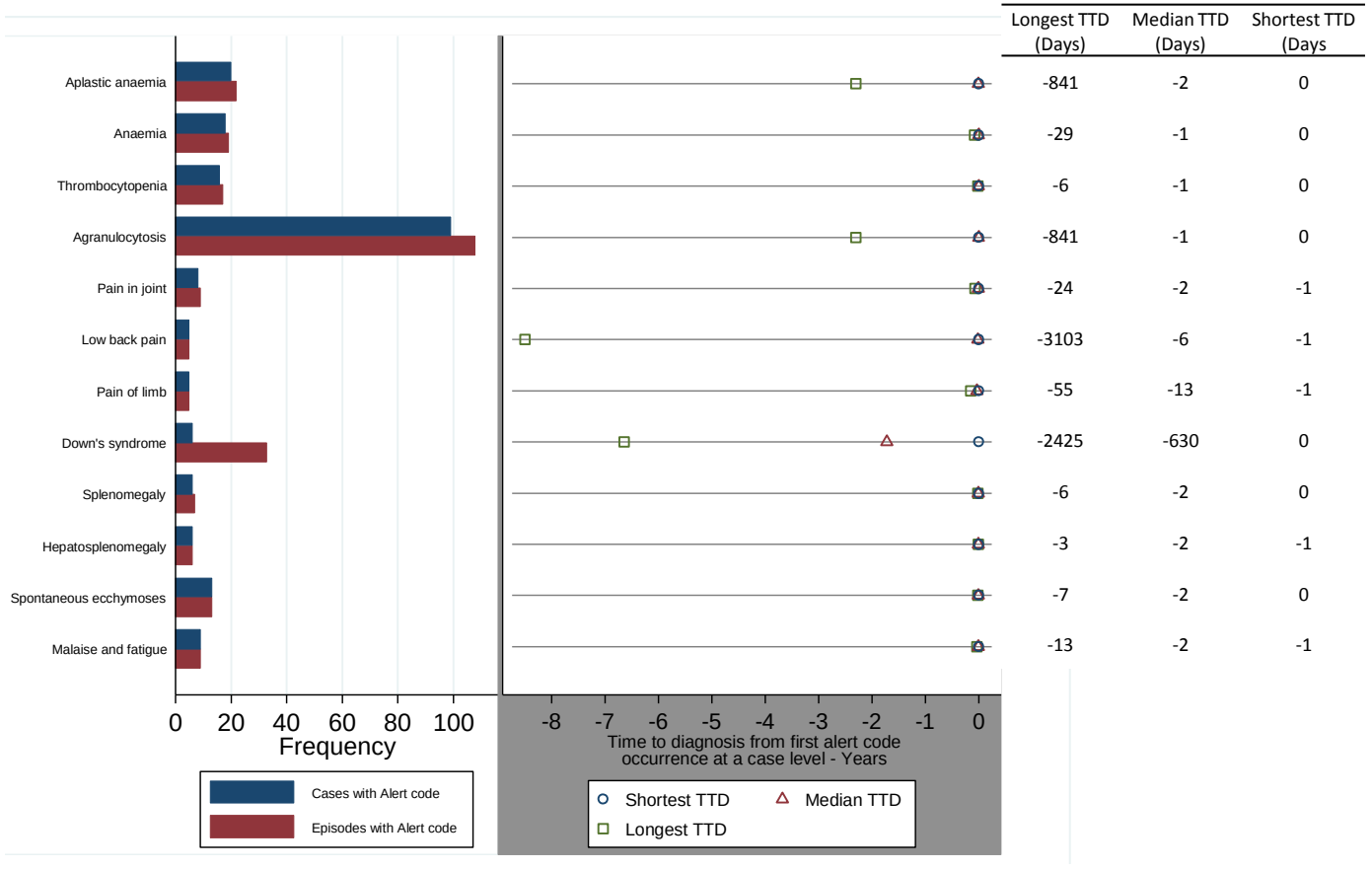
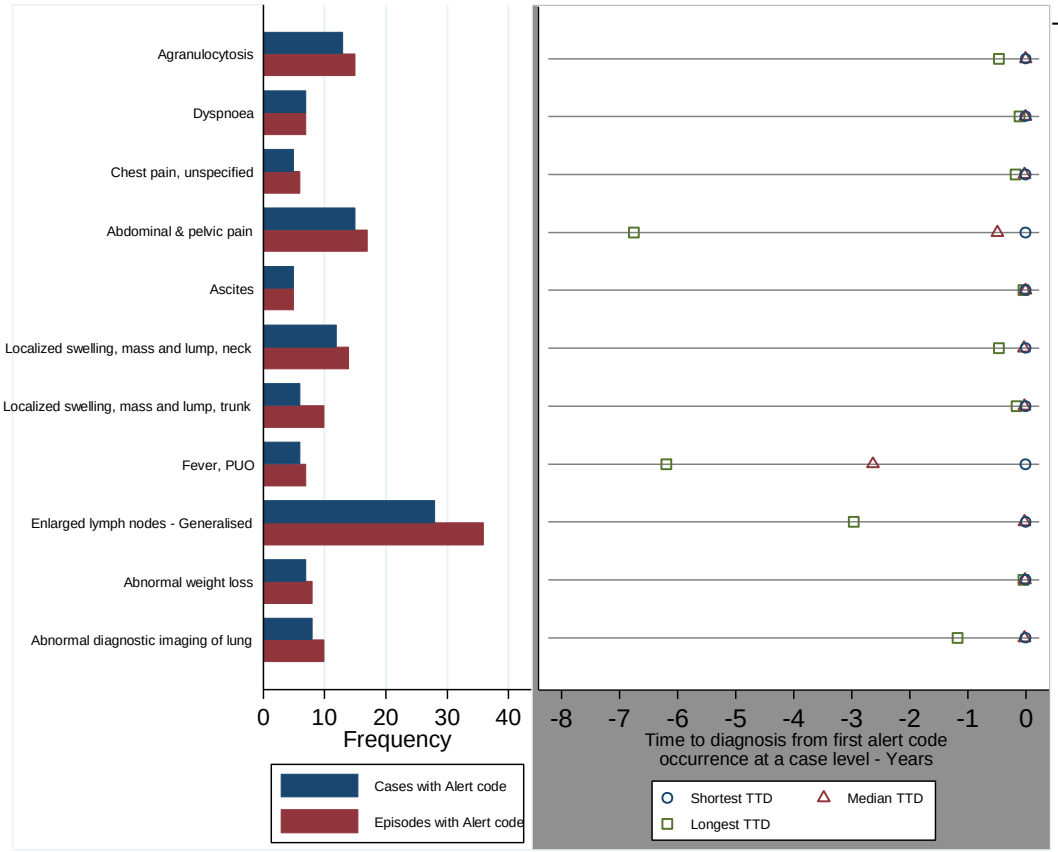


Figure 5.12 Frequency of pre-diagnosis lymphoma specific alert code containing episodes occurring in five or more cases by episodes frequency and number of cases presented along-side the longest, median and shortest duration of time between the first occurrence of such episodes at case level



	Longest TTD (Days)	Median TTD (Days)	Shortest TTD (Days)
Agranulocytosis	-171	-1	0
Dyspnoea	-40	-3	-1
Chest pain, unspecified	-63	-8	-1
Abdominal & pelvic pain	-2467	-179	-1
Ascites	-14	-3	-1
Localized swelling, mass and lump, neck	-171	-10	0
Localized swelling, mass and lump, trunk	-58	-8	-2
Fever, PUO	-2262	-963	-1
Enlarged lymph nodes - Generalised	-1082	-8	0
Abnormal weight loss	-14	-6	-2
Abnormal diagnostic imaging of lung	-431	-8	-2

Figure 5.13 Frequency of pre-diagnosis Central Nervous System tumour specific alert code containing episodes occurring in five or more cases by episodes frequency and number of cases presented along-side the longest, median and shortest duration of time between the first occurrence of such episodes at case level

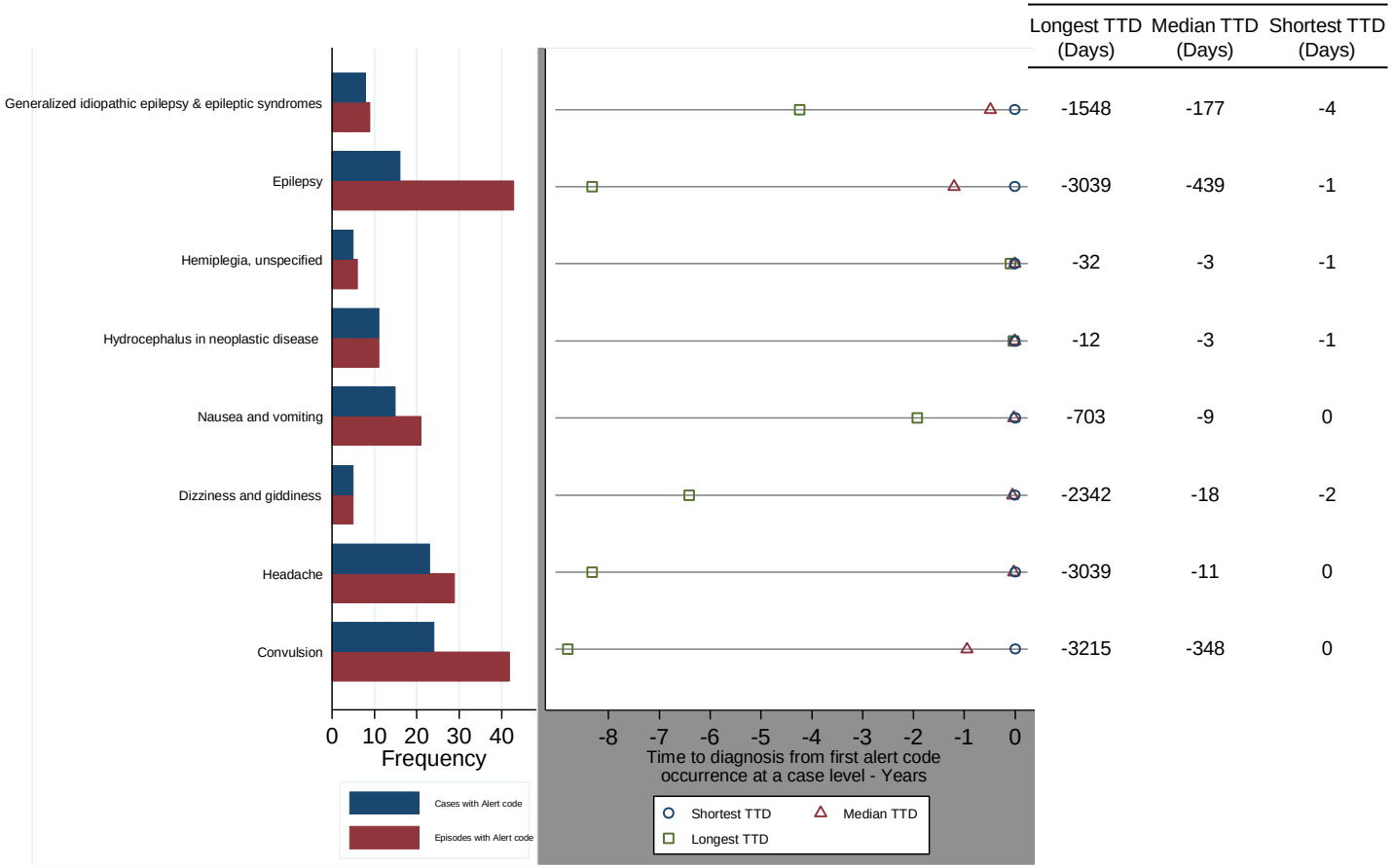
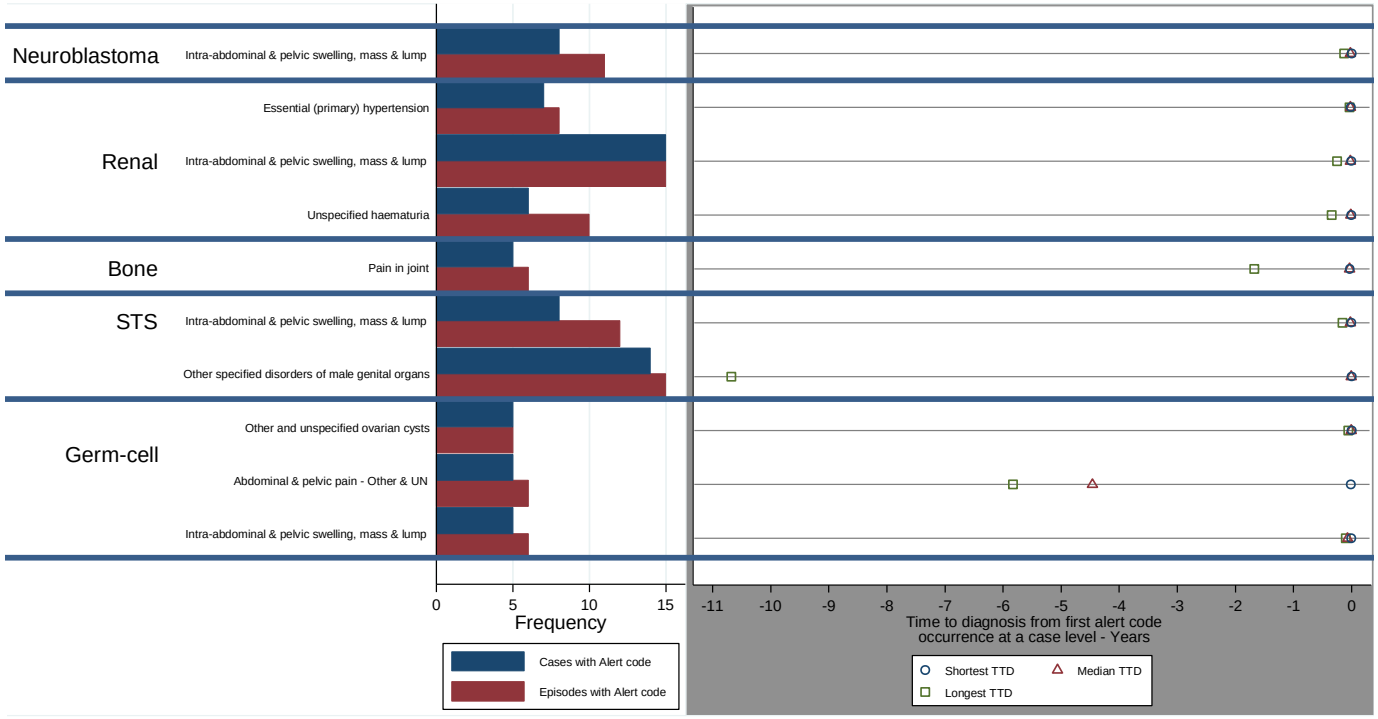


Figure 5.14 Frequency of pre-diagnosis specific alert code containing episodes occurring in five or more cases by episodes frequency and number of cases all other diagnostic groups presented along-side the longest, median and shortest duration of time between the first occurrence of such episodes at case level



Longest TTD (Days)	Median TTD (Days)	Shortest TTD (Days)
-45	-8	0
-12	-6	-4
-92	-8	-2
-126	-6	-1
-612	-12	-9
-56	-6	-1
-3900	-1	0
-20	-1	0
-2128	-1629	-4
-37	-24	-1

5.5 Time to diagnosis and codes suggestive of cancer

This section will bring together the three levels of analysis discussed in the results so far, to analyse TTD for those CYA cancer cases with alert code containing episodes preceding the date of definitive cancer diagnosis as recorded in the YSRCCYP. The section will also describe the broad cancer diagnosis alert code containing episodes and the diagnosis specific alert code containing episodes and refer to the combined as alert code episodes. It was not possible to preserve in a single combined data all the diagnosis specific alert code variables generated in the diagnosis specific datasets analysed in section 5.4. Therefore the episodes containing the diagnosis specific alert codes were tagged within each of the diagnosis specific datasets and this variable was added to the full cohort dataset, thus limiting the analysis to specific alert codes across the full cohort.

Figure 5.15 presents the episodes by alert code status for the 11 main ICCC diagnostic groups. This graph reflects the findings of the group specific analysis in section 5.4 and the fact that in all groups there was a predominance of non-alert code containing episodes.

Figure 5.15 Frequency of pre-diagnosis alert codes episodes and non-alert codes episodes by diagnostic group

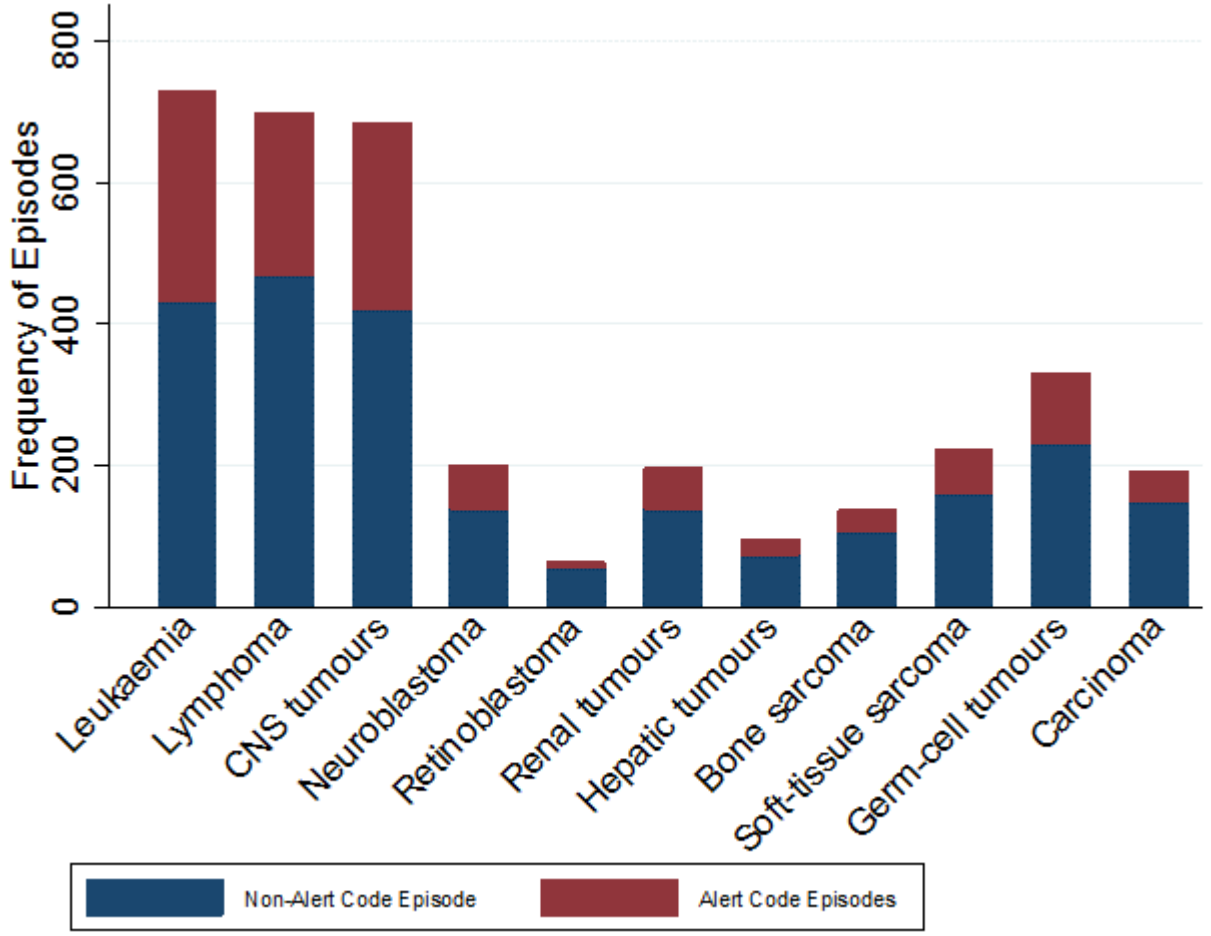


Figure 5.16 outlines the distribution of cases with alert codes and the proximity of alert code episodes to the date of definitive diagnosis within these cases. While 641 cases contain an alert code within an episode preceding the date of a definitive diagnosis of cancer, only 204 cases had alert codes more than a month prior to diagnosis and of these, in 70 cases the alert codes episodes only occurred at greater than 6 month preceding diagnosis.

Figure 5.16 Flowchart of case breakdown through the pre-diagnosis, and alert code analysis

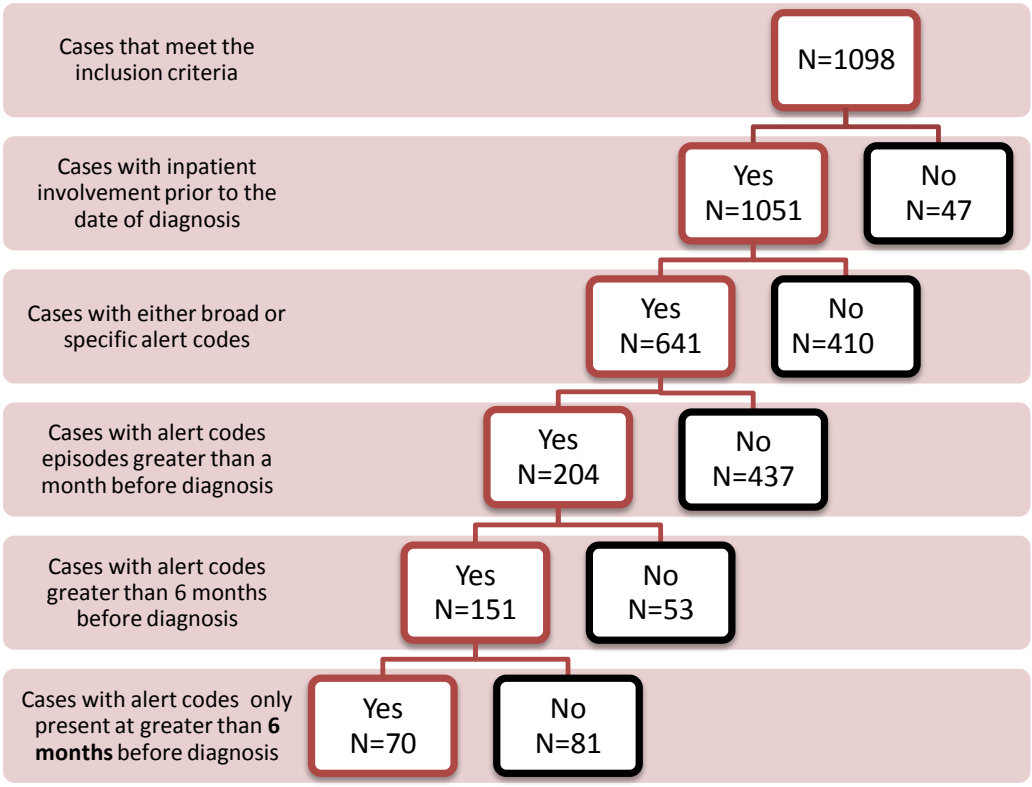


Figure 5.17 presents the time from the first alert code episode to the date of definitive diagnosis as a percentage of all cases by one-year interval preceding the date of definitive diagnosis. This graph identifies 41% of the study population as not having a pre-diagnosis alert code episode and over 50% of the study population with the first alert code episode confined to the year preceding diagnosis. Figure 5.16 identified 641 cases with pre-diagnosis alert codes. 63% of these cases have alert code episodes

confined to the month preceding diagnosis, thereafter 15.7% of alert code containing episodes occur in the remaining 11 months of the year and 21.3% of alert code episodes occur at more than one year prior to diagnosis, see Figure 5.18.

Figure 5.17 Time from the first alert code episode to the date of definitive diagnosis as a percentage of all cases by one-year interval preceding the date of definitive diagnosis

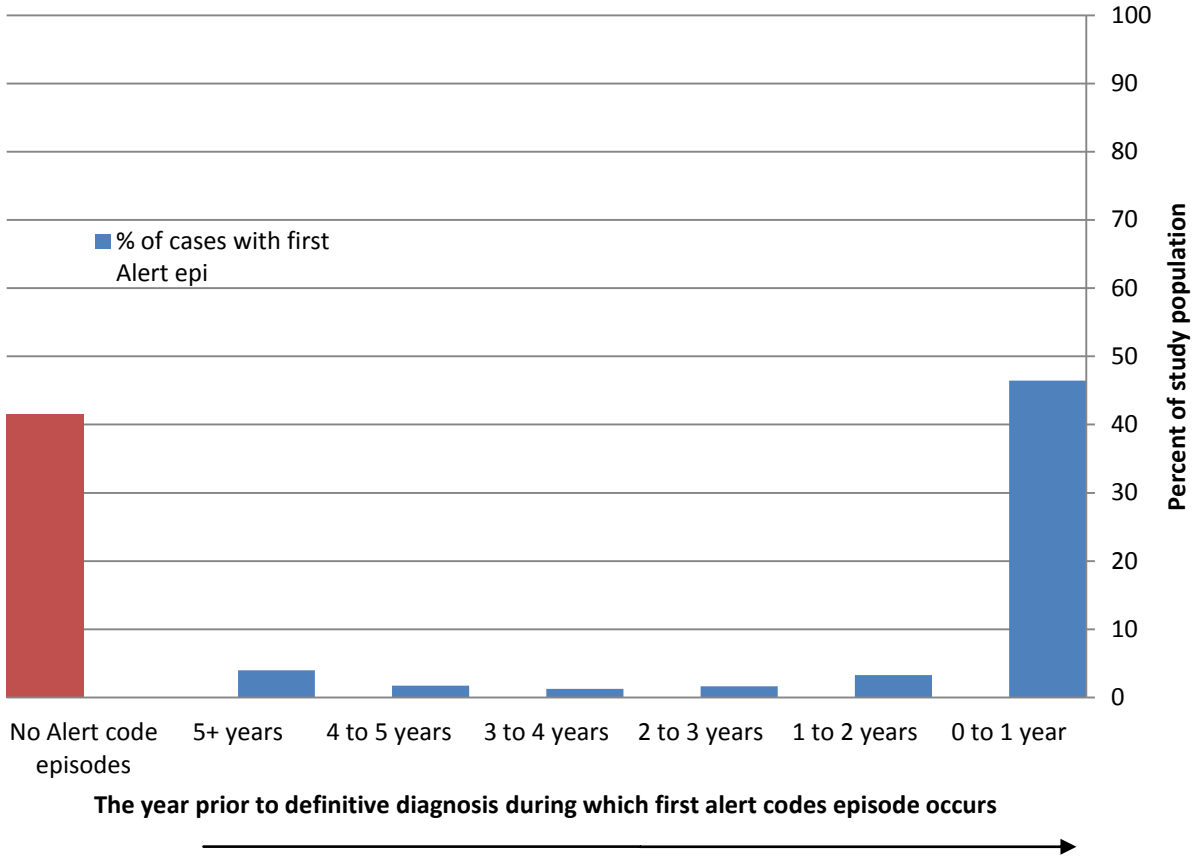
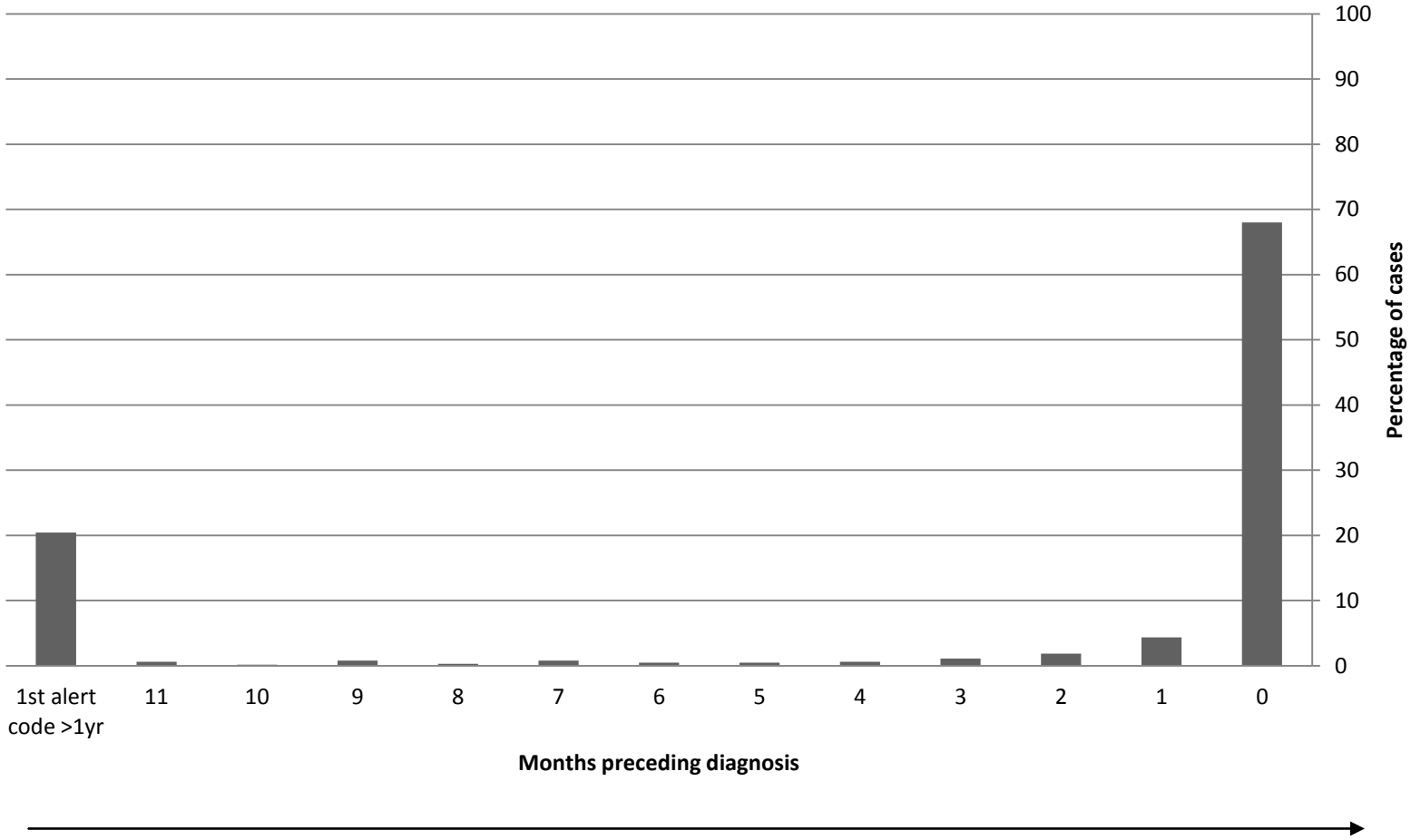


Figure 5.18 Time from the first alert code episode to the date of definitive diagnosis as a percentage of all alert code cases by month intervals in the year preceding the date of definitive diagnosis



There are three distinct population sub-groups identified from this preliminary analysis:

1. Cases with no alert code involvement preceding the date of definitive cancer diagnosis (N=457), comprising those cases either with no inpatient involvement (n=47) or those cases with no alert code containing episodes (n=410) preceding the date of a definitive diagnosis of cancer.
2. Cases with alert code containing episodes only within the month before diagnosis (N=437).
3. Cases with alert code containing episodes more than a month prior to diagnosis (N=204).

The three populations outlined above were defined within the TTD categorical variable within the dataset. For the rest of the results and discussion Group 1 is referred to as the **“no alert code”** group, group 2 as the **“alert codes immediately prior to diagnosis”** group and group 3 as the **“potentially prolonged time to diagnosis”** group. The distinction between groups 2 and 3 is important. Cases in group 2, identified as having alert codes only within the month preceding diagnosis, fall within the adapted standard guidance for referral and diagnosis of cancer in use at the time of study. Group 3 cases, for whom alert codes appear prior to the month preceding diagnosis, are highlighted as the cases with a potentially more prolonged TTD in secondary care. Analysis of the cases within group 3 may identify characteristic features in disease type, age, sex or admission routes for those cases at risk of a more prolonged TTD in secondary care. Furthermore, survival differences between the three groups and within the three groups may identify cases for whom a prolonged TTD in secondary care could have an impact on survival outcome.

There was evidence of significant differences in the TTD status between the age groups and diagnostic groups, using the chi-squared test both tests had a p-value=0.000. However there was no difference in TTD status by sex, p-value=0.488, TTD is therefore summarised by age groups and diagnosis but not sex.

Figure 5.19 summarises TTD by age group, with the absolute numbers of cases within each group are presented at the top of each of the percentage stacks. All three TTD groups are presented within each age group, though there is variation in the proportion of cases within each TTD group between each age group. The 15-19 and 20-24 age groups appear to have the highest percentage of cases without alert code episode prior to diagnosis, around 45% and 60% respectively. The proportion of cases with alert codes present during the month preceding diagnosis reduces as age increases across the five age groups, nearly 60% in 0-4 year olds compared to 20% in 20-24 year olds. The older age groups 10-14, 15-19 and 20-24 have a higher proportions of cases with more prolonged TTD, over 20% of population, compared to the 0-4 and 5-9 age groups.

Figure 5.20 summarises the TTD variable by diagnostic group, excluding retinoblastoma, hepatic tumours and other tumours due to few cases within each population. The three alert code categories are represented within each of the common diagnostic groups, however there are variations in the number of cases between the diagnostic groups. More the 60% of cases with bone sarcoma, germ-cell tumour and carcinoma groups had no alert codes preceding the date of diagnosis. Over 60% of cases with renal tumours and leukaemia have alert codes within the month preceding diagnosis. CNS tumours, neuroblastoma and carcinoma have the highest percentage of cases within the potentially prolonged TTD group with more than 25% of cases within this TTD group.

Figure 5.19 Time to diagnosis groups by five year age groups

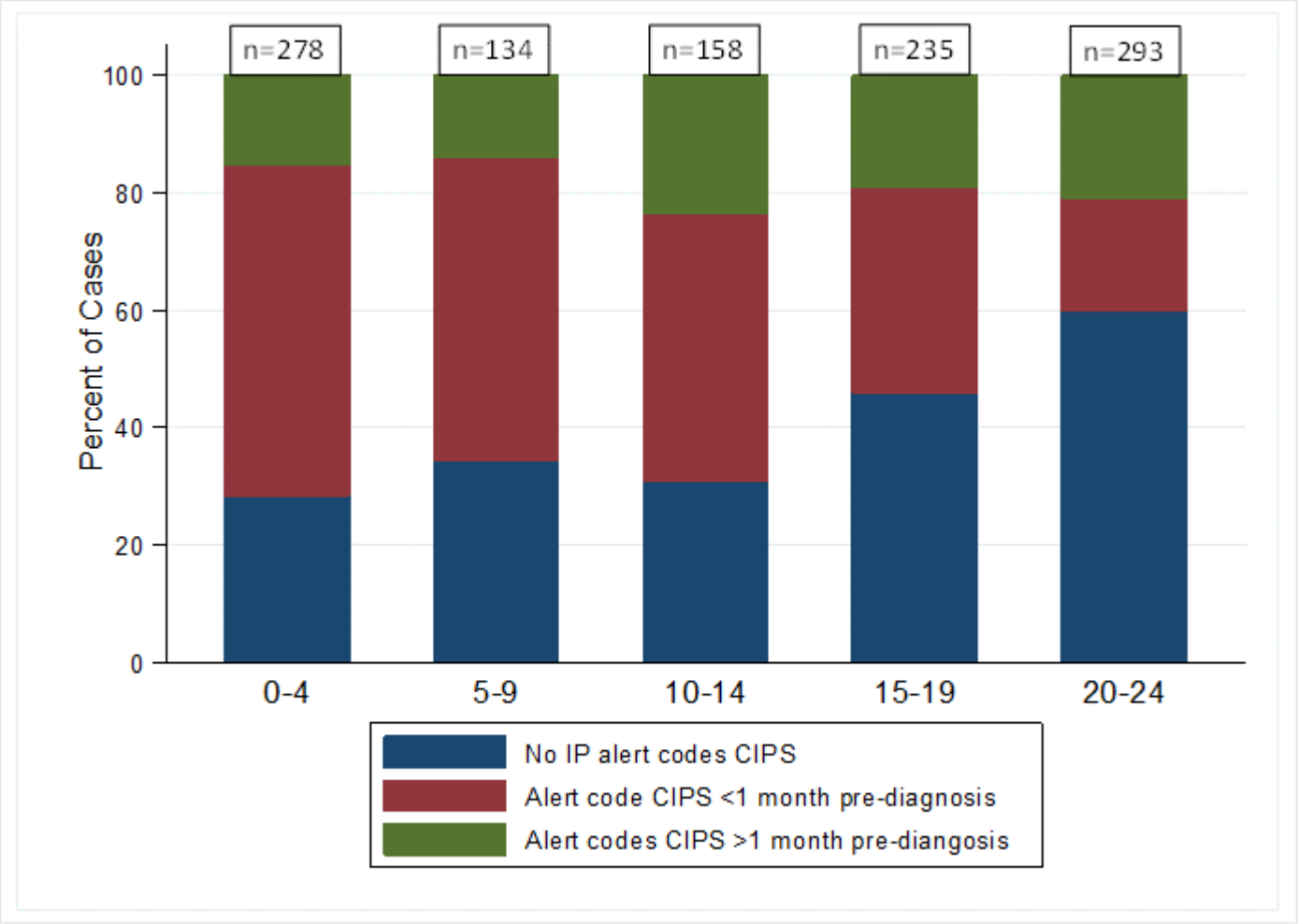
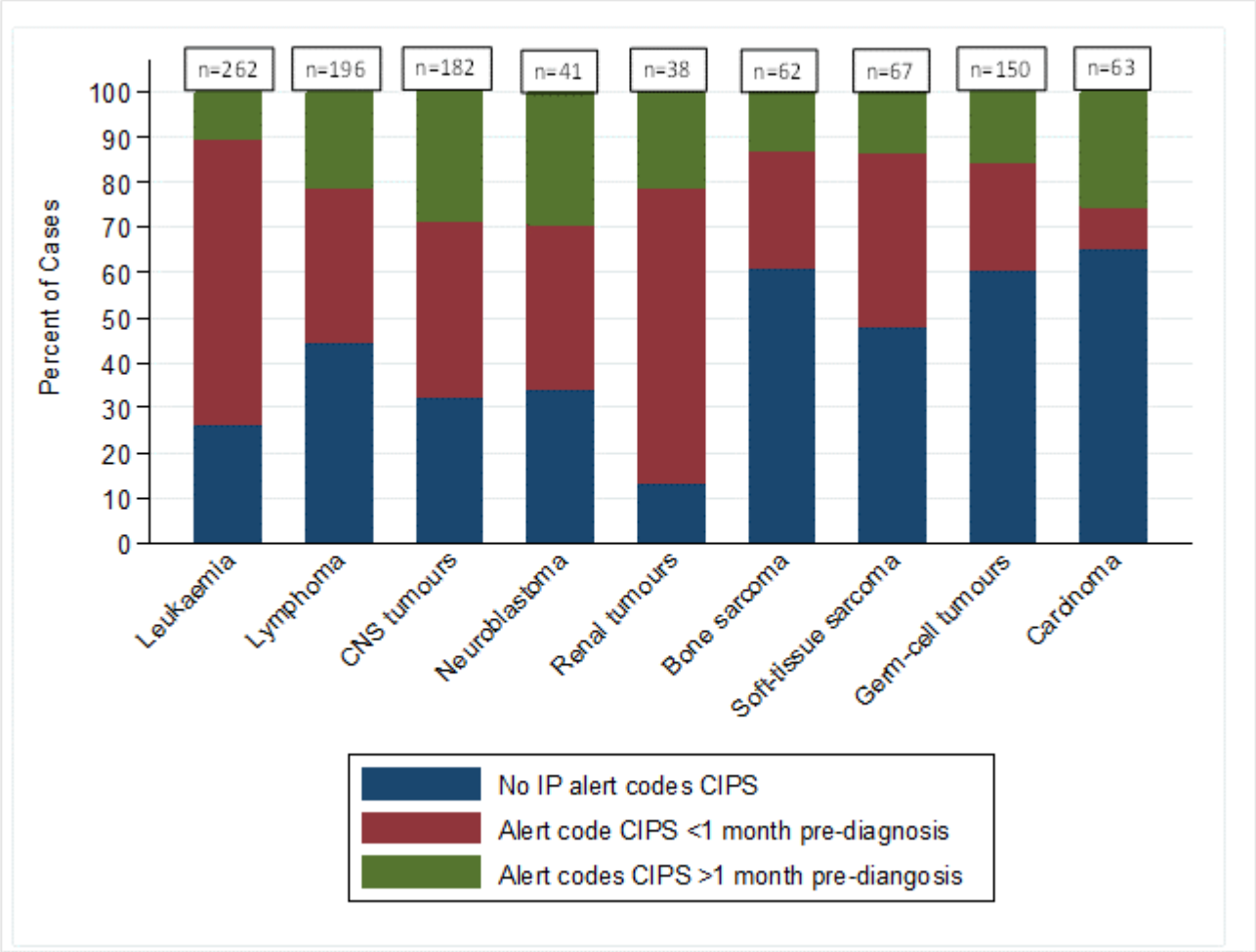


Figure 5.20 Time to diagnosis groups by diagnostic groups (retinoblastoma, hepatic tumours and other tumours excluded)

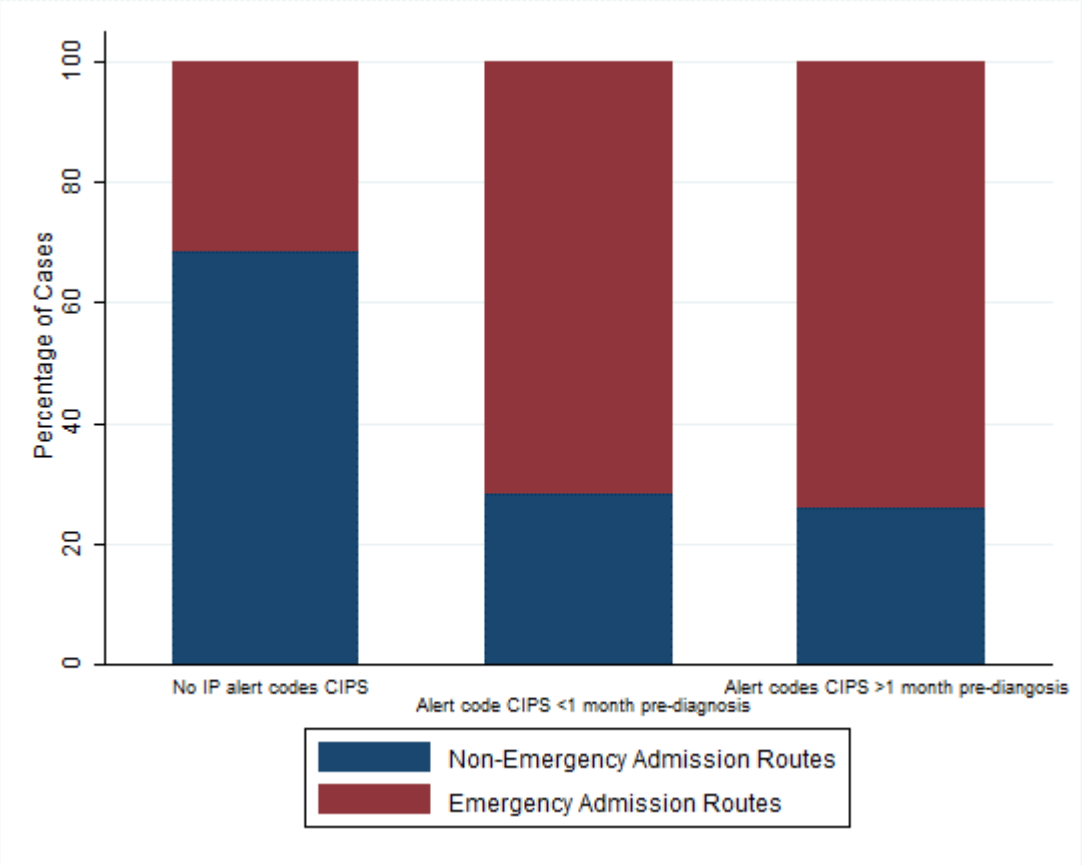


5.6 Routes to Diagnosis

Table 5.2 summarises the admission codes for each pre-diagnosis inpatient episode identified within the linked dataset. Pre-diagnosis inpatient events that contain cancer alert codes have been identified through the study of episode level HES, however consecutive episodes can contribute to a continuous inpatient event, known as CIPS, see Figure 4.1. The initial admission code in a sequence of episodes that make up a CIPS will record the point of entry into secondary care, subsequently each episode within that CIPS are punctuated by a transfer admission code. Therefore, as discussed in the Methods section 4.9, when considering the point of entry for a pre-diagnosis secondary care inpatient event the initial admission code for a CIPS is extremely useful at demonstrating the route to diagnosis.

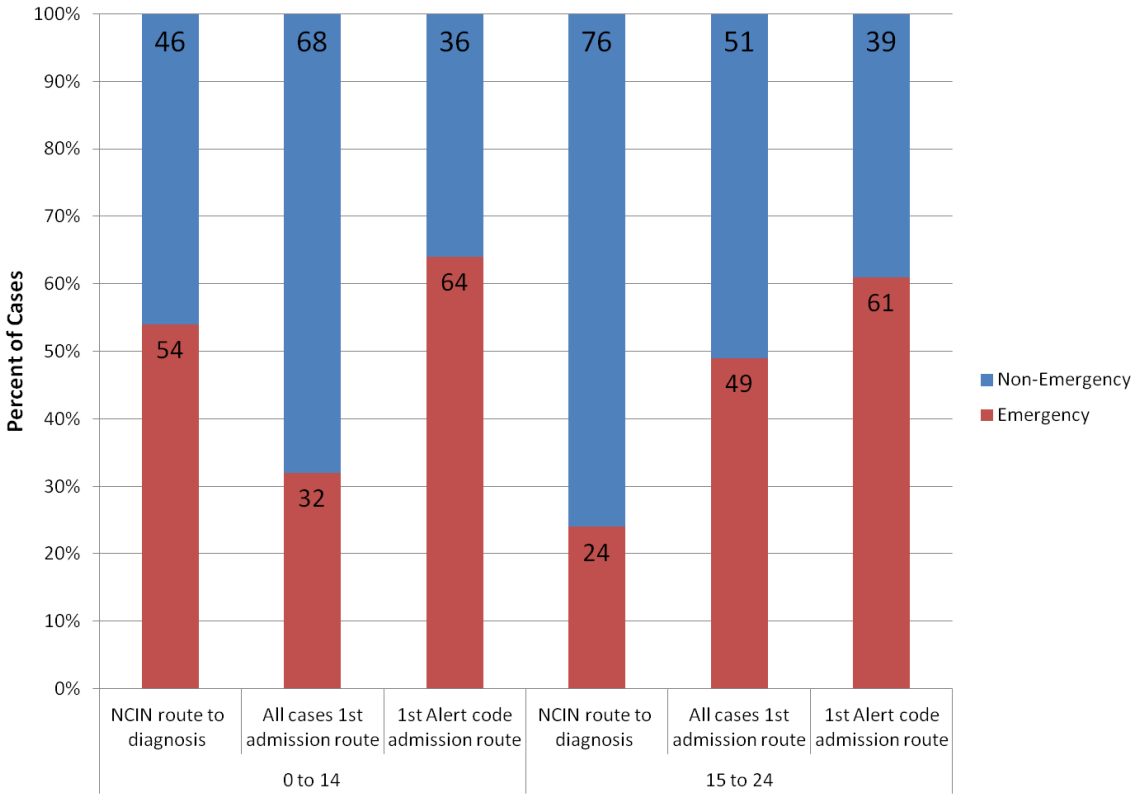
There are a number of admission codes that make up the admimeth variable in HES data, which can be divided into nine groups as identified in table 5.2. These nine groups can be further combined into two main categories; ‘emergency’ and ‘non-emergency’ routes to admission. Figure 5.21 summarises the initial admission route for CIPS as a percentage stack for the three TTD categories identified in the previous section. It is clear that the emergency admission route predominates in alert code containing inpatient events. There is no appreciable difference in emergency verses non-emergency admission route for cases with alert code immediately preceding diagnosis and those cases with potentially a more prolonged TTD.

Figure 5.21 The first admission route, emergency versus non-emergency, by the time to diagnosis groups



Routes to diagnosis for CYA cancers previously studied within the NCIN routes to diagnosis work, separated the 0 to 24 year age group into childhood cancers defined as 0 to 14 years and teenage and young adult cancers defined as 15 to 24 years. Figure 5.22 compares the percentage of emergency and non-emergency admissions for cancer cases in the NCIN study with results from this study. Within the NCIN results there are clear differences by age group: the childhood cancer cases had a higher percentage of emergency admissions (54%) compared to TYA cancer cases (24%). The reverse of this was found for the first admission route for all cases regardless of alert code status (n=1051) within this study: 32% of admissions were emergencies compared to 49% for TYA cancers. For the cases with alert codes (n=641) emergency admission routes were the predominant route into inpatient secondary care for both children and TYAs , with around two-thirds of cases with alert codes being admitted to inpatient through this route.

Figure 5.22 Emergency versus non-emergency admission routes for the alert code cases within the Yorkshire study population compared to the NCIN routes to diagnosis study (71)



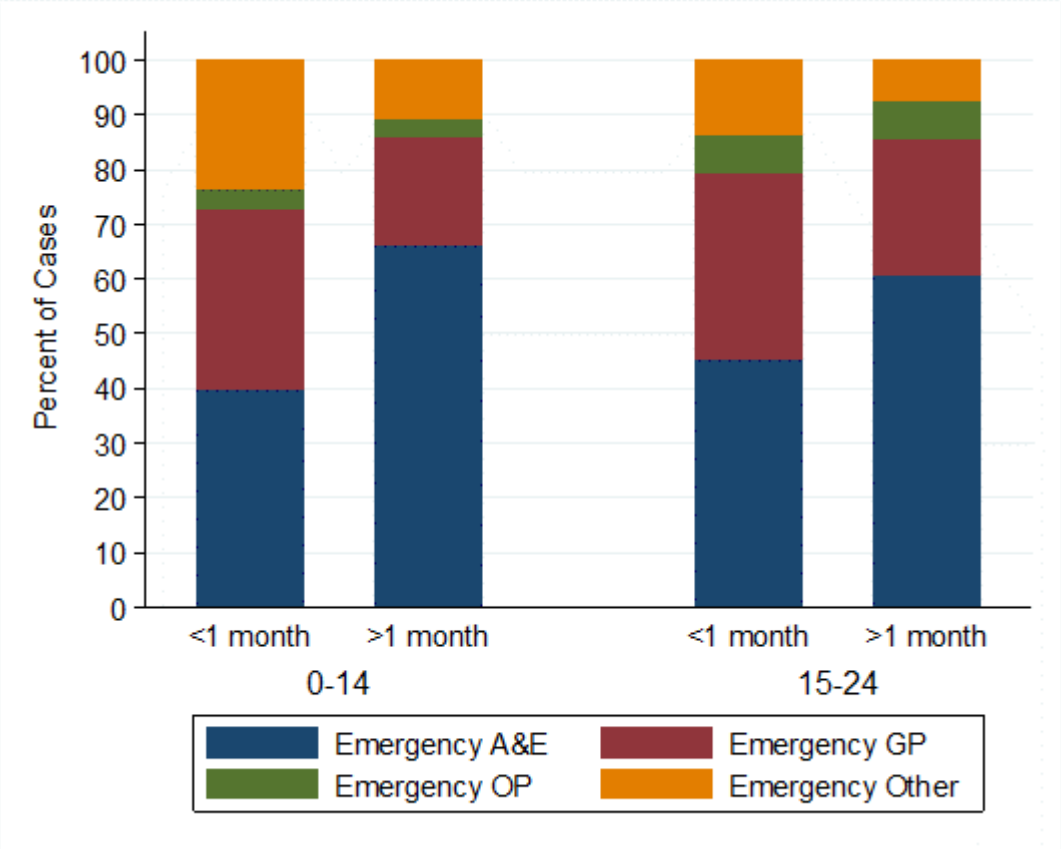
In Figure 5.22 the results for 0 to 24 year olds in the NCIN routes to diagnosis study were taken from all newly diagnosed malignant cancers, excluding non-melanoma malignant skin cancer, diagnosed between 2006-2008 resident in England. The method for defining the route to diagnosis did not take into account clinical information contained within the HES data, however it did sequence the inpatient and outpatient events preceding the diagnosis of cancer. Therefore the study period, population sampled and method of analysis applied within this study differs from the NCIN routes to diagnosis work, as such any comparison of results in the two studies must be interpreted with this in mind.

There were four types of emergency admission routes: emergency A&E, emergency GP, emergency outpatients (OP) and emergency other. The pattern of emergency admission routes for the two groups with alert codes were similar between the age groups (Figure 5.23). Emergency A&E was the predominant emergency route for admission to inpatient care for cases with

alert codes, and was the most common route for cases with potentially prolonged TTD in both age groups. Emergency GP made up a larger proportion for the 15 to 24 year olds with alert codes immediately prior to diagnosis compared to the 0 to 14 year olds. Emergency outpatient admissions appeared to be less prominent in the 0 to 14 year olds compared to 15-24 year olds.

Variations in emergency admission routes for each diagnostic group were explored but are not presented due to the small numbers. The data are therefore difficult to interpret at a regional level and may require a national cohort.

Figure 5.23 The specific emergency admission routes for alert code cases



5.7 Survival

The results within this section are presented either as univariable Kaplan-Meier survival curves or multivariable Cox proportional hazards regression estimates. One-year survival is presented as a proxy for late or early diagnosis and three-year survival is presented as a measure of long-term survival.

5.7.1 Overall survival

Figure 5.24 presents the overall survival for this study population.

There were 204 deaths overall from the 1098 cases studied, representing 18.6% of the study population. The survival at one-year following diagnosis was 91.5% compared to 83.6% at three-years.

5.7.2 Survival by sex

Figure 5.25 presents the Kaplan-Meier survival patterns by sex. The figures for one-year survival were very similar for males (91.2%) and females (91.0%) as were the three-year survival figures (males 83.2%, females 84.1%).

5.7.3 Survival by year of diagnosis

Figure 5.26 presents the survival by year of definitive diagnosis grouped according to the period of diagnosis. The survival percentages at both one and three years for cases diagnosed between 2007 to 2009 (92.6%, 85.1%) was higher than that for cases diagnosed between 2004 to 2006 (90.6%, 82.3%). However the difference in survival for the two groups was not statistically significant (p=0.151).

5.7.4 Survival by age

Figure 5.27 presents the survival curve by five-year age groups.

There was variation across the five age groups with one year survival of 89.2% in cases diagnosed at 0 to 4 years compared to 93.7% for cases aged 10 to 14 years. Three-year survival was highest in the 10 to 14 year olds (87.3%) followed by the 0 to 4 year olds (84.9%) and lowest in 15 to 19 year olds (80.4%). The differences however are not statistically significant using univariable Cox analysis. Figure 5.28 presents the survival curves for 0-14 year olds and 15-24 year olds. This shows a more favourable survival immediately following diagnosis for TYA cases compared to childhood cases. The survival trends then converge at one year and by three years TYA survival percentages are poorer than childhood cases.

5.7.5 Survival by diagnosis

Figure 5.29 presents the survival curves by the ICCC diagnostic group, with the exception of retinoblastoma and other tumours, both excluded due to the small number of cases. Figure 5.30 presents the one and three-year survival percentage with confidence intervals for the 10 ICCC groups included within the survival curves.

From the 10 ICCC groups presented in Figure 5.29, six have a one-year survival above 90%; lymphoma highest at 95.9%, renal tumours 94.7%, germ-cell 93.3%, leukaemia 92.8%, carcinoma 92.1% and neuroblastoma 90.2%. The lowest one-year survival is seen for hepatic tumours 73.3%, followed by soft-tissue sarcoma 86.6%, CNS tumours 86.8% and bone tumours 88.7%. The number of deaths in some of the ICCC groups are very few, reducing the reliability of the results and is reflected in the wide confidence intervals particularly within the hepatic tumour population.

Three-years the survival differences between the 10 groups were more marked. Survival for lymphoma (91.3%), renal tumours (92.1%) and germ-cell tumours (90%) remained at 90% or above. Three-year survival for leukaemia and carcinoma fell to 85.1% and 82.5% respectively. whilst neuroblastoma displayed the largest difference between one and three years

dropping nearly 20% to 70.7% at three years. CNS tumour survival at three-years only fell 7% to 79.1%, whilst in contrast bone tumours and soft-tissue sarcoma survival fell markedly to 71% and 70.2% respectively. Hepatic tumours had the worst three-year survival falling to 60%.

There was no statistically significant difference in survival between diagnostic group based on the univariable Cox regression analysis.

Figure 5.24 Survival all cases

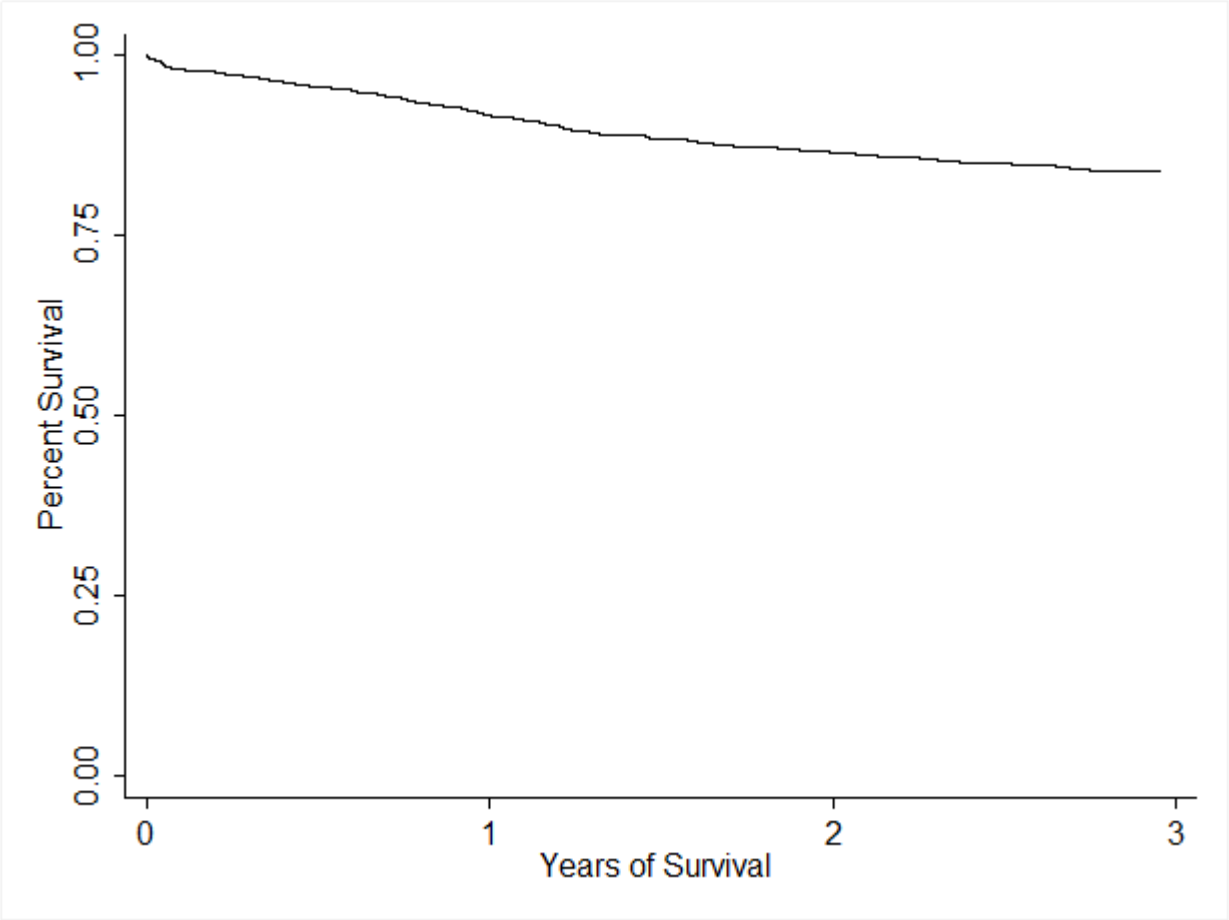


Figure 5.25 Survival by sex

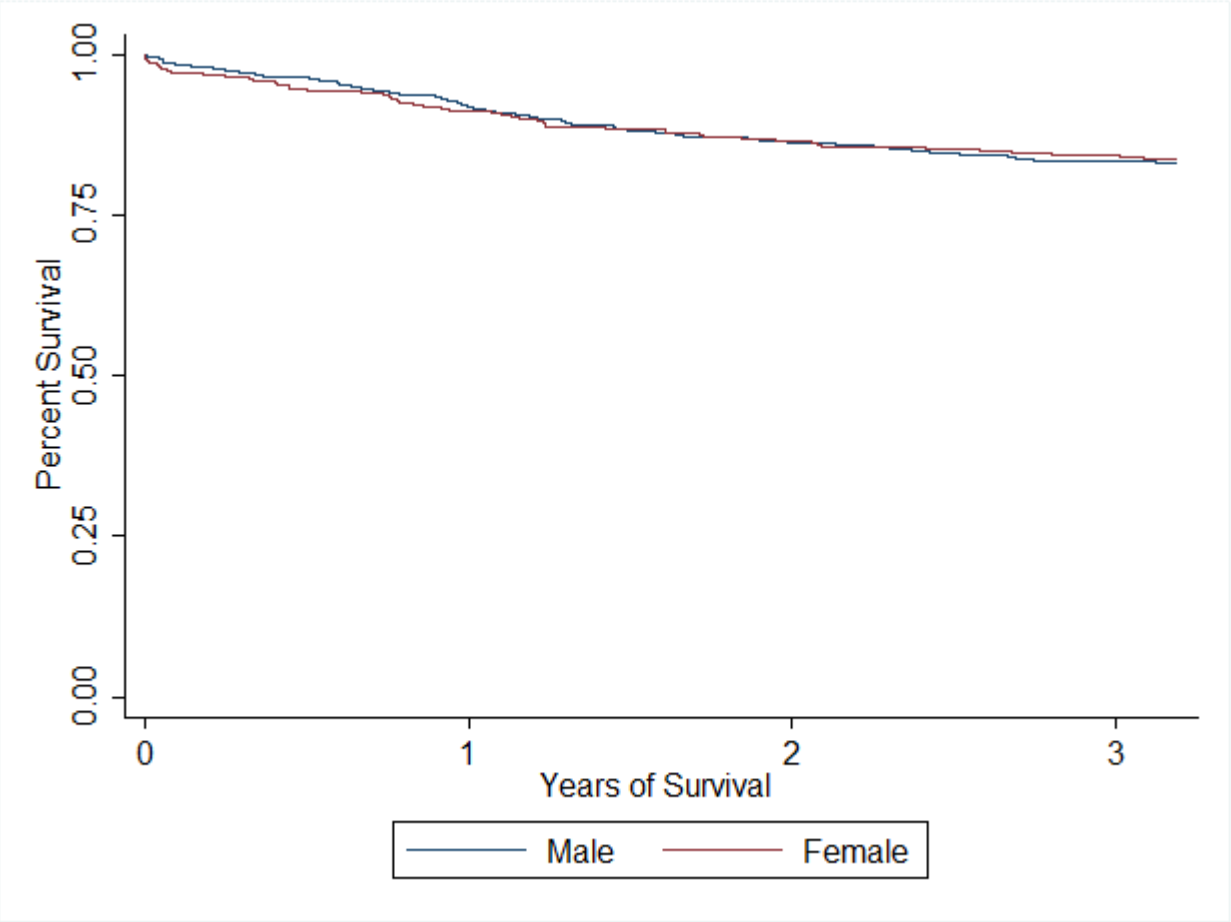


Figure 5.26 Survival by year of definitive diagnosis

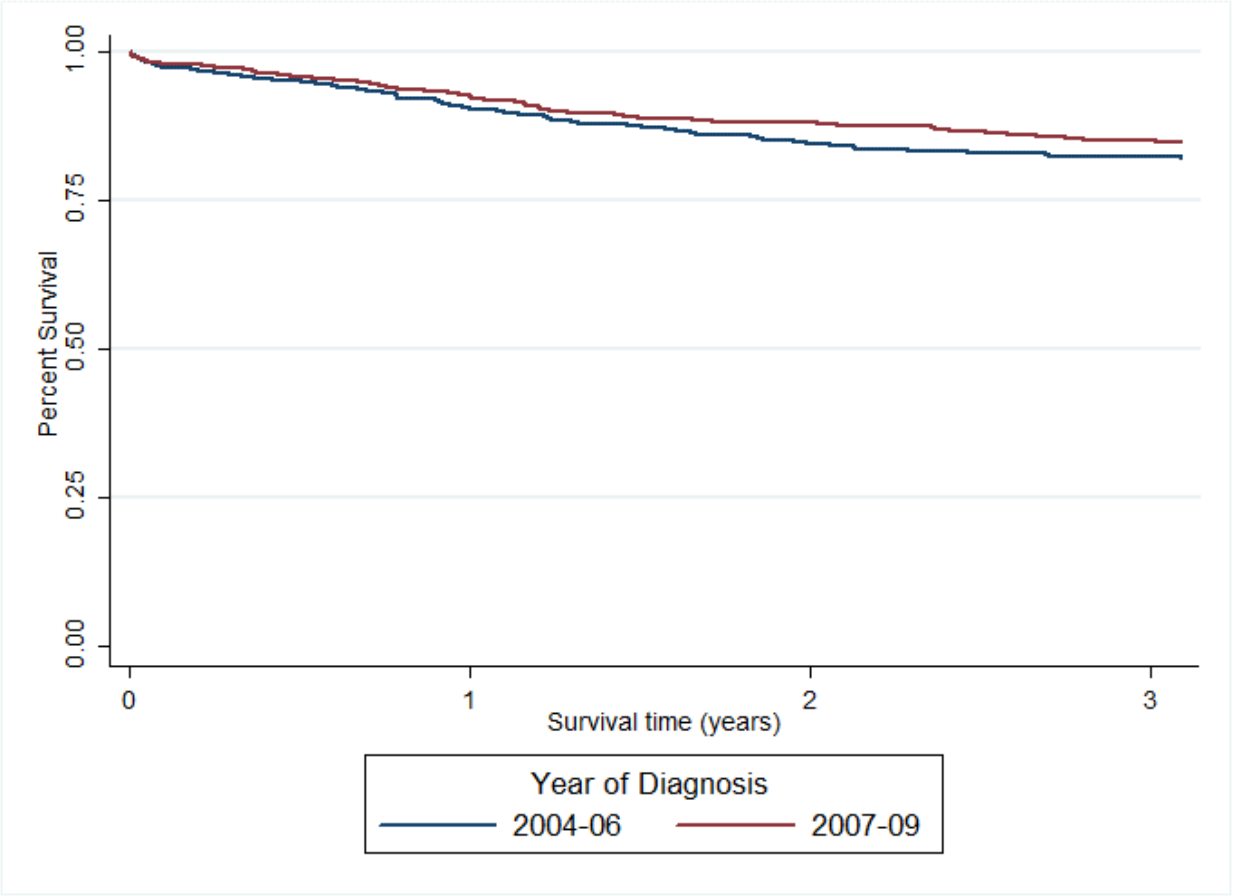


Figure 5.27 Survival by five year age groups

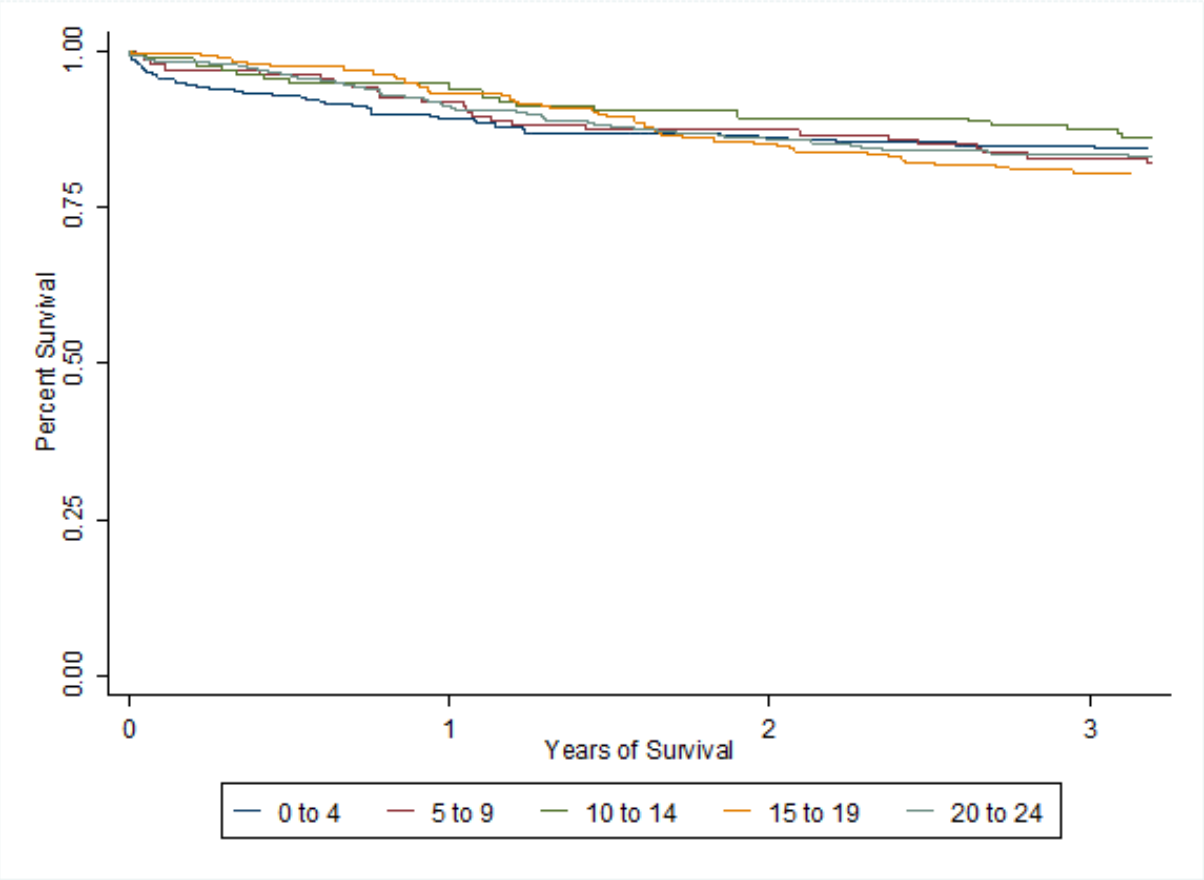


Figure 5.28 Survival by age groups 0-14 and 15 to 24

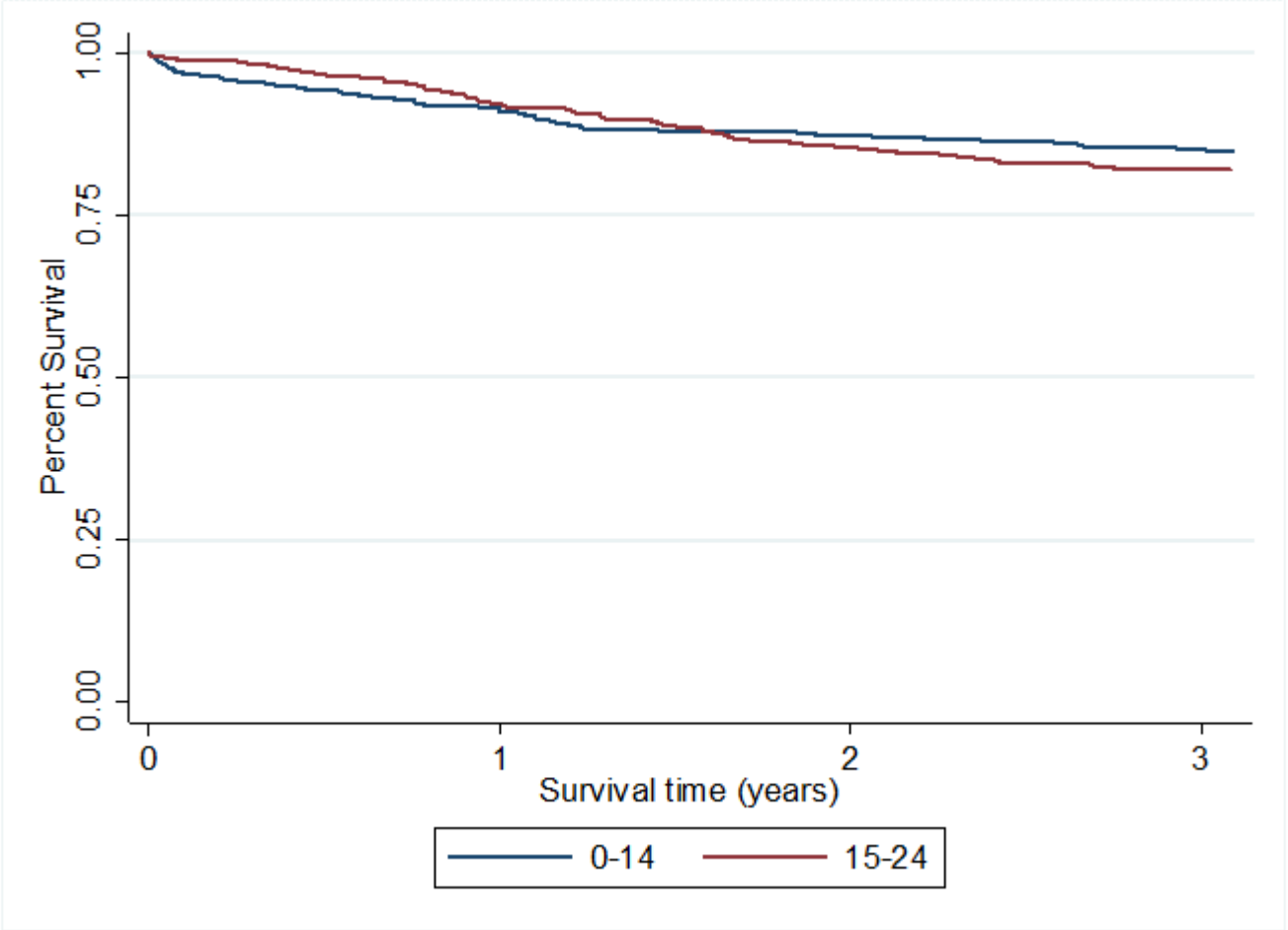


Figure 5.29 Survival by International Classification of Childhood Cancer diagnostic group (groups V & XII are not presented)

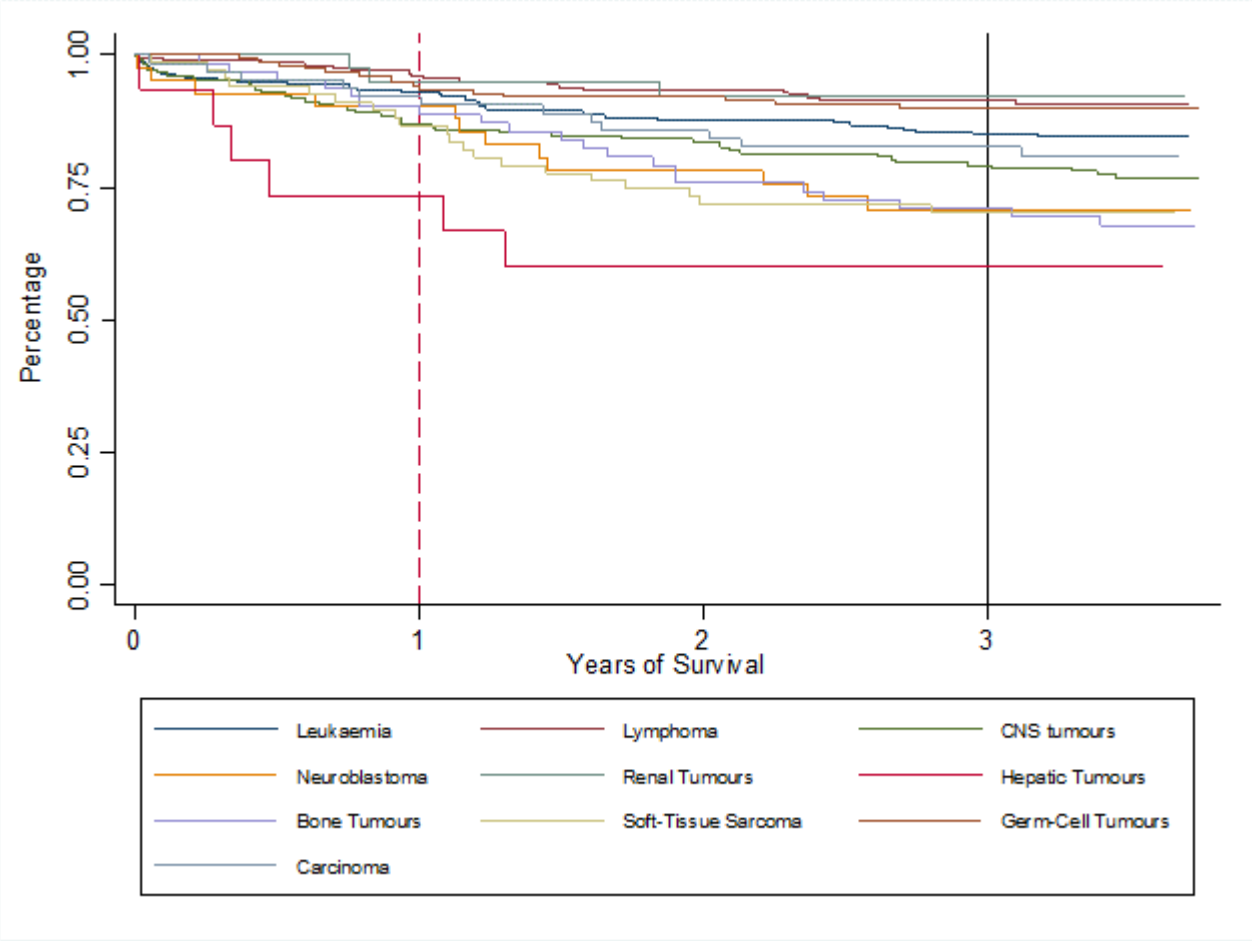
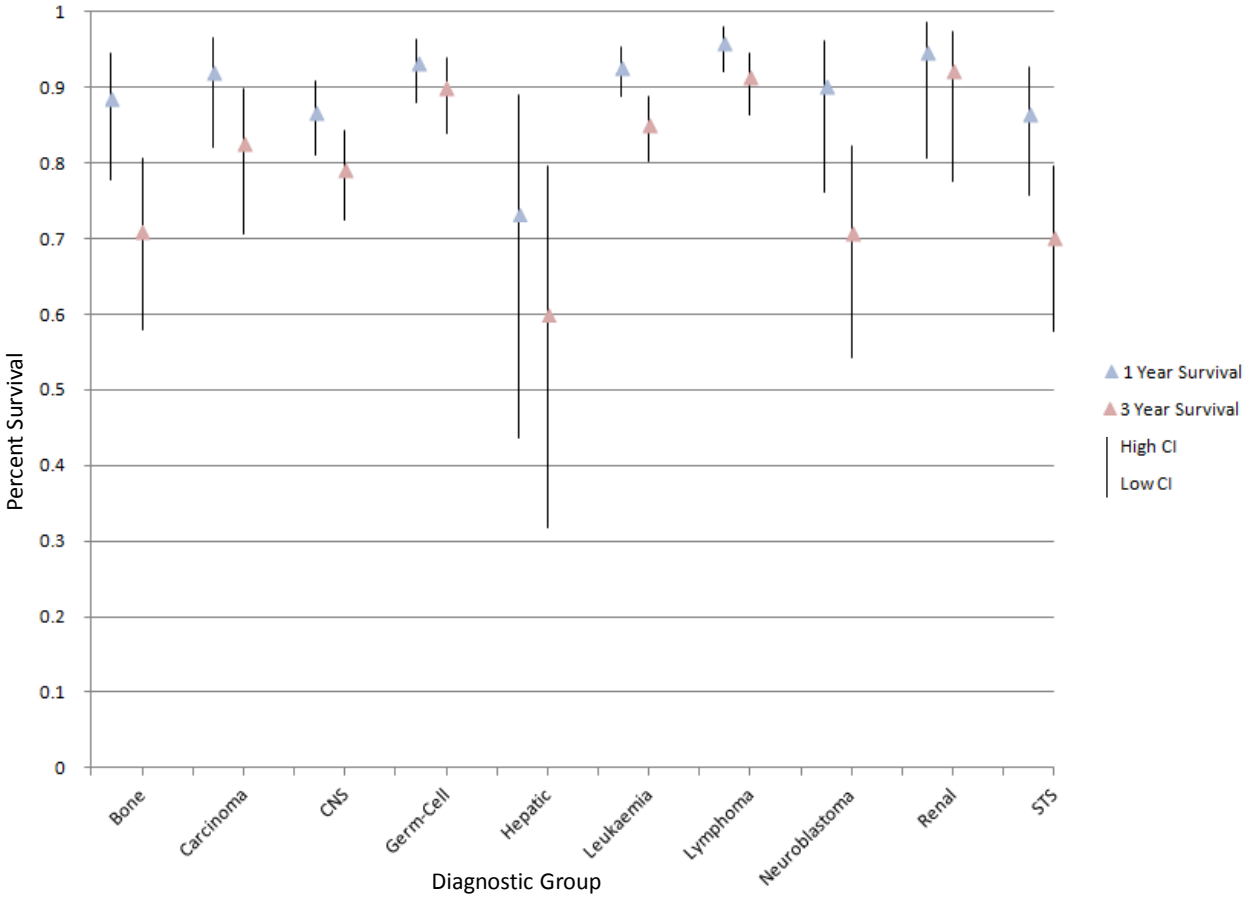


Figure 5.30 Percentage of cases surviving to one and three years by diagnostic group



5.7.5.1 Unadjusted survival by alert code status preceding a diagnosis of cancer in children and young adults

Figure 5.31 presents the survival curves to three years by alert code status.

The proportion of cases surviving at both one and three-years was highest for those cases without alert code episodes preceding the date of diagnosis, being 94% and 87% respectively. The cases with alert codes immediately prior to diagnosis had a lower proportion of cases surviving at one-year compared to cases with a potentially prolonged TTD (89% vs 91%), however at three years the proportional survival for these two group was similar at 81%.

The hazard ratio of cases with alert codes immediately prior to diagnosis was 1.41, indicating a 41% increased risk of death in this group and the p-value of 0.029 indicates this was a statistically significant effect. The hazard ratio for the group of cases with a potentially prolonged TTD also showed an increased risk of death compared to those cases without alert codes. However, this effect was not statistically significant.

There was no statistically significant difference in survival for each of the TTD groups by gender, using the log-rank test. Comparing TTD by age group, showed a significant difference in survival function for those cases with potentially prolonged TTD (0-14 HR 0.98, 15-24 HR 2.48) (Figure 5.32). TYA cases had a poorer survival estimates compared to the childhood cases with potentially prolonged TTD; one-year survival 89% for TYA's and 94% for childhood cases, three-year survival 75% for TYA's and 87% for childhood cases.

Comparison of the survival functions for the diagnostic groups by each TTD group did not yield robust results due to the low number of cases. Figure 5.33 presents the percentage of cases alive or dead within the TTD groups for each diagnostic group. The low number of cases within the multiple groups reduces the reliability of the analysis of survival function and consequently the results are not included.

Figure 5.31 Survival by alert code status

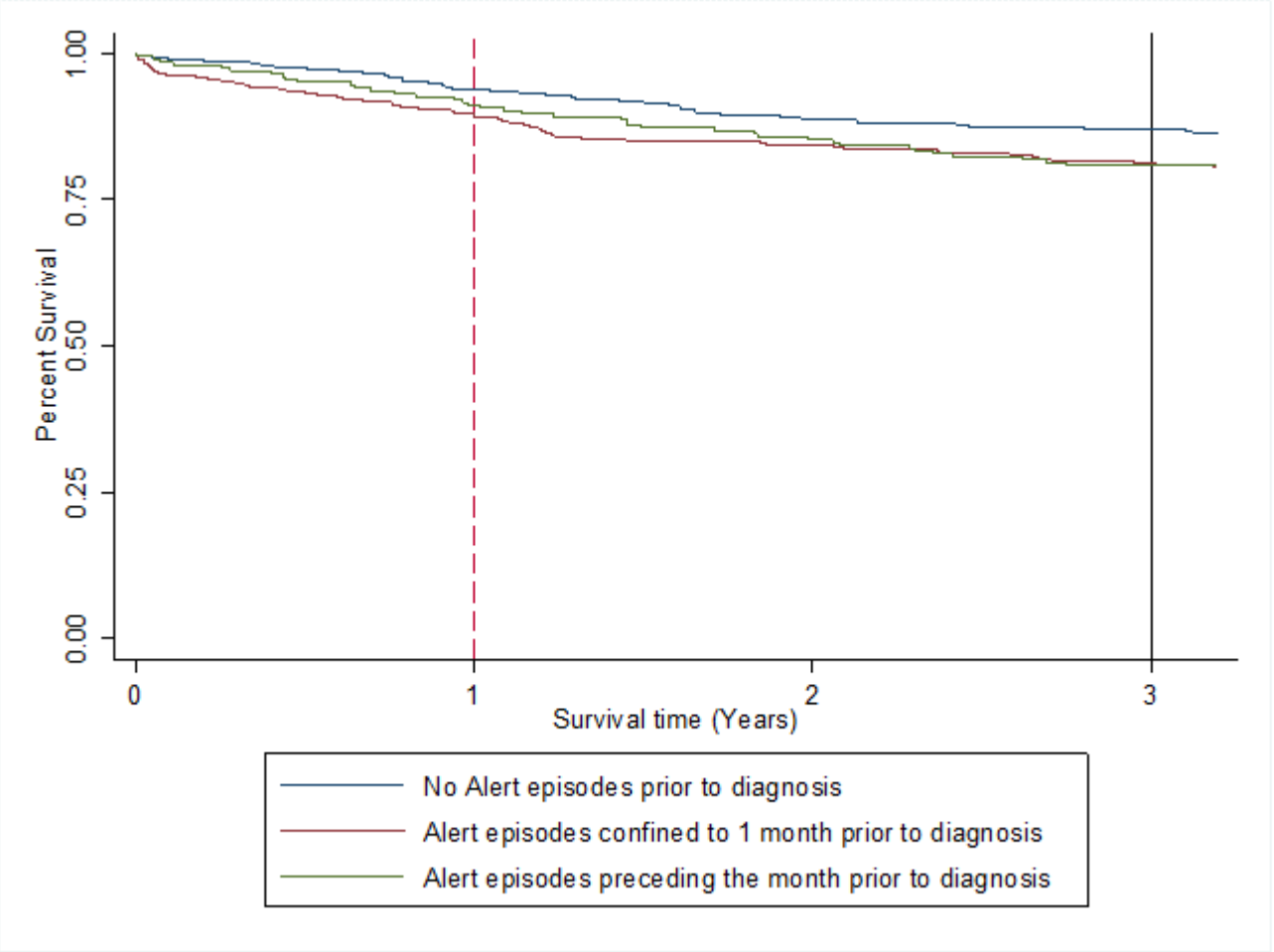


Figure 5.32 Survival for cases with potentially more prolonged time to diagnosis by age groups at diagnosis

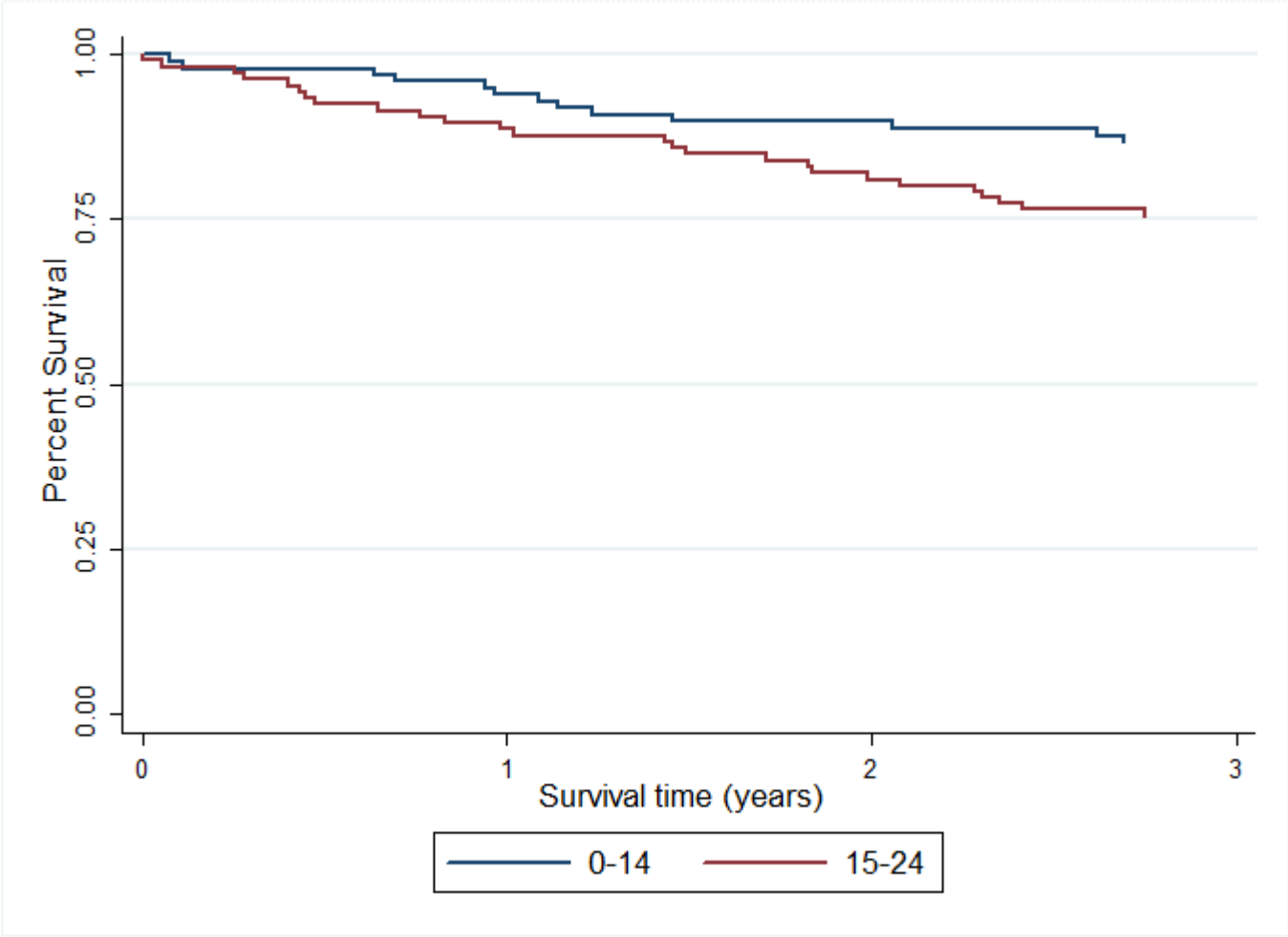
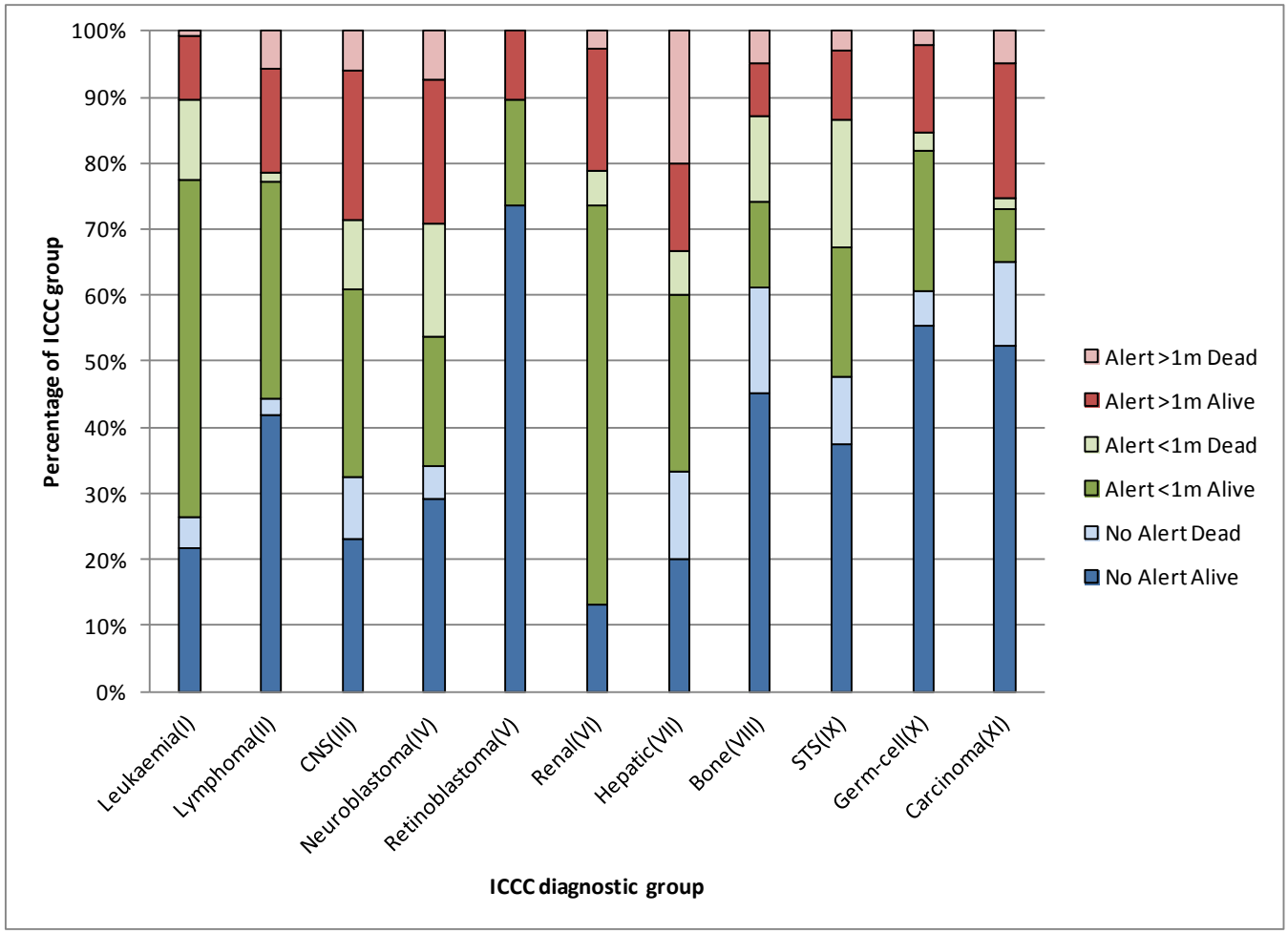


Figure 5.33 Survival (%) at three years from diagnosis for the time to diagnosis status by the diagnostic group



5.7.5.2 Multivariable analysis of survival by alert code status preceding a diagnosis of cancer in children and young adults

Univariable analysis of TTD does not take into account how other important factors such as age, sex or diagnosis influence survival. This section will develop a multivariable Cox proportional hazard model to analysis survival by alert code status. The initial model will include alert code status and additionally adjust for the following confounding factors (as determined in Figure 4.4):

- Age at diagnosis – 0 to 14 versus 15 to 24; evidence against linearity (Likelihood ratio test p=0.0194)
- Sex
- Diagnosis
- Year of diagnosis – continuous; no evidence against linearity (likelihood ratio test p=0.7447)

Results of model 1 as described above are presented in table 5.5. Subsequently, interactions between TTD and age at diagnosis were tested to assess whether any effect on survival from TTD differed by age at diagnosis, the hazard ratios are summarised in model 2 (table 5.6). An interaction between TTD and sex was not included as preliminary analysis showed no difference in the TTD by sex (see 5.7.5.1).

Table 5.5 Model 1: Cox regression model for survival in CYA cancer

Variable		Hazard Ratio	P>z	95% Confidence Intervals
TTD				
	No Alert	1.00		
	Alert <1 month	1.67	0.003	1.20-2.34
	Alert >1 month	1.34	0.142	0.91-1.98
Age group				
	0-14	1.00		
	15-24	1.64	0.002	1.21-2.23
Sex				
	Male	1.00		
	Female	0.86	0.316	0.65-1.15
ICCC group				
	Leukaemia(I)	1.00		
	Lymphoma(II)	0.50	0.014	0.29-0.87
	CNS(III)	1.70	0.014	1.12-2.58
	Neuroblastoma(IV)	2.46	0.007	1.28-4.74
	Retinoblastoma(V)	0.00	.	.
	Renal(VI)	0.48	0.214	0.15-1.54
	Hepatic(VII)	3.42	0.005	1.45-8.06
	Bone(VIII)	2.20	0.004	1.29-3.76
	STS(IX)	2.28	0.002	1.35-3.83
	Germ-cell(X)	0.51	0.032	0.28-0.94
	Carcinoma(XI)	1.07	0.844	0.55-2.09
	Other(XII)	3.00	0.282	0.41-22.16
Year of Diagnosis		0.93	0.08	0.85-1.01

Model 1 shows a significant increased risk of death of 67% for cases with alert codes less than one month prior to diagnosis compared to those without alert codes. Although the model shows an increased risk of death for cases with more prolonged alert code involvement, this result is not statistically significant within model 1 when adjusting for age at diagnosis, sex, diagnosis and year of diagnosis.

Table 5.6 Model 2: Cox regression model for survival in CYA cancer

Variable			Hazard Ratio	P>z	95% Confidence Interval
TTD x Age group					
	No Alert	0-14	1		
	Alert <1 month	0-14	1.71	0.030	1.05-2.79
	Alert >1 month	0-14	0.98	0.964	0.50-1.94
	No Alert	15-24	1.56	0.091	0.93-2.61
	Alert <1 month	15-24	2.41	0.001	1.40-4.13
	Alert >1 month	15-24	2.48	0.001	1.43-4.29
Sex					
	Male		1		
	Female		0.85	0.282	0.64-1.14
ICCC group					
	Leukaemia(I)		1		
	Lymphoma(II)		0.51	0.015	0.29-0.88
	CNS(III)		1.68	0.016	1.10-2.56
	Neuroblastoma(IV)		2.45	0.007	1.27-4.72
	Retinoblastoma(V)		0.00	1.000	.
	Renal(VI)		0.47	0.212	0.15-1.53
	Hepatic(VII)		3.46	0.005	1.47-8.16
	Bone(VIII)		2.20	0.004	1.28-3.76
	STS(IX)		2.23	0.003	1.33-3.76
	Germ-cell(X)		0.51	0.031	0.28-0.94
	Carcinoma(XI)		1.06	0.858	0.54-2.09
	Other(XII)		2.96	0.288	0.40-21.85
Year of Diagnosis					
			0.93	0.078	0.85-1.01

The interaction term in model 2 shows that the effect of TTD on survival differs between age groups. In model 2 there is a 71% increased risk of death in the 0 to 14 year age group, compared to those cases without alert codes which was statistically significant (p=0.030); the finding for cases with alert codes more than a month prior to diagnosis compared to those without alert codes was not statistically significant. These findings are consistent with results from model 1 which looked at all age groups combined. For 15 to 24 year olds, model 2 shows a statistically significant difference in survival for TYAs with alert codes within the month preceding diagnosis, similarly to 0-14 year olds, however, the effect is much larger (HR=2.41) in this age

group. Furthermore, for 15 to 24 year olds there is also an increased risk of death (HR=2.48) for those with alert codes occurring more than a month preceding diagnosis.

5.8 Summary Points

- The incidence of cancer in 0 to 14 year olds within the study population is in keeping with nationally accepted figures. However, comparisons with national incidence rates in the 15 to 24 year-olds are more difficult due to marked geographical variations across the UK and inconsistent age definitions within previous studies.
- The profile of cancer cases across this population of children and young adults varied with age and is in keeping with nationally accepted figures.
- The majority of CYAs with cancer have some secondary care inpatient involvement prior to their date of definitive diagnosis. Only 4% of the study population did not have any pre-diagnosis inpatient involvement, of those case with pre-diagnosis inpatient involvement 39% had inpatient involvement confined to the month preceding diagnosis.
- Early cancer diagnosis awareness resource can be used to identify alert codes within HES inpatient data, however this approach was not applicable to outpatient HES data due to a lack of clinical information coded within the outpatient HES data .
- The analysis of outpatient data is limited by the quality of the coding within episodes, however from the study population very few cancers were referred to tertiary oncology services through the two-week cancer referral route.
- The date of definitive diagnosis of cancer is often preceded by the presence of an episode containing an ICD-10 cancer code with more than 95% of such episodes occurring within a month of definitive diagnosis.
- From the 216 diagnosis specific alert codes identified from the pre-diagnosis inpatient events only 41 codes occurred in five or more cases within each ICCC diagnostic group.
- The majority of diagnosis specific alert codes occurred within the month immediately prior to diagnosis, however there were frequently outlying cases with codes at many years preceding diagnosis.

- Emergency admission were the predominant routes into secondary care for cases with alert code events preceding diagnosis.
- In two-thirds of the cases with pre-diagnosis alert codes the episodes occur only within the month preceding diagnosis.
- A TTD of more than one month in secondary inpatient care is not significantly associated with a poor survival for cancer in children.
- A TTD of more than one month in secondary inpatient care was significantly associated with a poor survival for cancer in TYAs

Chapter 6 Discussion & Conclusions

6.1 Introduction

This study has primarily examined the period of time preceding a diagnosis of cancer for a cohort of children and young adults in Yorkshire, focusing on their hospital involvement prior to diagnosis. This is the first study to focus on secondary care services in this population and through the use of linked health data aimed to identify potential early warning signs for CYA cancer and investigate the association between time spent in secondary care pre-diagnosis and survival up to three years after diagnosis. The study also aimed to examine the application of HES data in the investigation of TTD within secondary care for CYA cancer. To date in the UK, the CYA cancer awareness literature has been focused on patients and primary care services. There is often an assumption that secondary care services have a minimal influence on the TTD for cancer (12, 25, 125).

This chapter will cover the following areas: firstly, an evaluation of the results presented in Chapter 5, providing a discussion of how the study population relates to the overall UK population, pre-diagnosis secondary care involvement, pre-diagnosis CYA cancer early warning signs and symptoms, TTD and its association with survival as well as a discussion of the application of HES data within CYA early diagnosis research. Secondly, reflecting on the strengths and limitations of this study, along with the future application of current CYA cancer awareness literature in secondary care services. The third sub-section sets out considered recommendations for improving TTD within secondary care services for CYA cancers informed by the results of the analysis and literature review. This third section will also include health service research recommendations to aid researchers utilising linked data sets in future studies. The chapter will end with a conclusive summary of the thesis and outline of suggested future CYA early diagnosis work.

6.2 Evaluation of study findings

6.2.1 Key findings

This study identified increasing inpatient involvement towards the date of diagnosis for CYA cancer cases, primarily occurring within the month preceding diagnosis. Increasing healthcare involvement for CYA cancer cases toward the date of diagnosis has been highlighted in the primary care setting in a large population-based Danish study (132). This pattern of healthcare engagement was further emphasised by the CYA cases identified as having cancer alert codes within inpatient events preceding diagnosis within this Yorkshire cohort.

Less than 20% of cases (n=204) within this study were identified as potentially having cancer signs and symptoms further than a month from diagnosis. The period of time between first inpatient alert code event and diagnosis in these cases therefore potentially exceeded UK guidance for referral, diagnosis and treatment of suspected cancer (124), thus highlighting a group for whom early diagnosis interventions within secondary care could improve TTD. Further analysis of the study population identified differences in the TTD status of cases by age group and diagnosis, though no differences between gender.

For the cases identified as having alert code inpatient events prior to diagnosis, there were more cases entering hospital via an emergency route compared to any other referral method for cases diagnosed aged 0-14 and 15-24 years. This finding is similar to that shown within the NCIN routes to diagnosis work for 0-14 year olds. However the NCIN study showed fewer emergency routes into hospital for the older TYA cases (ages 15-24 years) in contrast to the observations within this study.

Within the overall study population of CYAs there was no clear association between a prolonged time spent in secondary care and a worse outcome. Sub-group analysis by age and TTD revealed a poorer outcome for cases aged 15 to 24 years with a more prolonged TTD compared to their counterparts with no alert codes and all 0 to 14 year-olds. There was a significantly poorer survival outcome in cases with alert codes events within

a month of diagnosis, potentially identifying cases presenting with clearly identifiable cancer signs and symptoms and therefore more advanced disease. Unfortunately there was a lack of reliable staging data within the health services data used within this study and the association between TTD, stage and outcome could not be explored within this work.

A novel approach was the analysis of linked electronic health data; cross-referencing of early awareness literature with diagnostic information contained within HES inpatient data to identify cancer alert codes preceding diagnosis. However, the sensitivity of referencing symptom and sign based awareness literature with the diagnosis based ICD-10 coding system applied within HES inpatient data is not clear and further work must be done to refine the method. One potential approach would be to focus on the occurrence of unexplained signs and symptoms, known as R codes, within HES episodes. The R codes made up a large proportion of the identified alert codes within this study. The method was not applicable to HES outpatient data due to the lack of diagnostic information recorded within these data. There was a paucity of awareness literature available for the common TYA cancers such as germ-cell tumours and carcinomas as well as the rarer childhood cancers such as hepatic tumours.

The study assessed the reliability and accuracy of HES data through the analysis of a sample of case records. There was high level consistency between HES data and case records in the recording of pre-diagnosis inpatient episodes, the diagnosis and the date of diagnosis. However limited availability of case records resulted in only half the case notes for the sample population actually being reviewed. The sample was chosen to reflect a wide range of the case population, however the case notes review was limited to a principal treatment centre. Therefore future work should involve wider case record review, including secondary care centres other than the principal treatment centre.

The aims and objectives set out in section 1.2 were met and TTD was described according to three distinct groups within the study population: cases with no signs suggestive of cancer preceding diagnosis, those with signs and symptoms suggestive of cancer within the month preceding

diagnosis who therefore fall with accepted guidance; and cases with signs and symptoms of cancer outside the accepted referral and diagnosis guidance. This population-based regional CYA cancer cohort study demonstrated variations in TTD by age and diagnosis and identified differences in survival by TTD most notably by age.

6.2.2 Cancer in children and young adults in Yorkshire

The study cohort was identified from cases within the population-based Yorkshire specialist registry of cancer in children and young people (YSRCCYP) including only cases recorded as living within the FYRHA. The register receives notifications from a variety of source including national and other regional registries as well as specialist NHS services (e.g. neuropathology) resulting in virtually complete case ascertainment. We can therefore be confident of the representative nature of the study cohort in relation to the general population of CYA cancer cases in Yorkshire, certainly in comparison to the predominantly institution based studies identified within the systematic review in Chapter 3.

The overall incidence of childhood cancer (ages 0-14) in this study population was 139 per million per year; the variation in incidence rates by age, sex and diagnosis are summarised in Table 5.1. The incidence rates for cancer in 0-14 year-olds is in keeping with previously published regional and national rates. The incidence rates in this study are similar to those published for childhood cancer within Yorkshire between 1990-2001 (115). The crude incidence of childhood cancer (excluding non-malignant skin cancers) in England between 2008 and 2010 was reported as 137 cases per million per year (3).The established distribution of childhood cancer by diagnostic group within this study matches that seen in the wider childhood cancer literature (3, 67). To some extent this study population is representative of the national childhood cancer population and the conclusions drawn here are potentially directly applicable to the wider childhood cancer population.

The crude incidence rate for 15 to 24 year olds within this study population was 159 cases per million population. Unlike the childhood cancer

population, the incidence rate of cancer in 15 to 24 years in the literature is harder to define, due to variations in the age boundaries used to define the TYA population. CRUK reports a crude incidence rate of 267 cases per million for all cancers in 15-24 year olds in England between 2008-2010 (3). There is considerable discrepancy which in part can be explained by regional variations in cancer incidence within England. In 2007 Alston *et al* published an overview of regional variations in incidence of cancer in TYAs, showing the lowest incidence rates in the north of England and rates increasing towards the south of the country (28). Alston *et al* published incidence rates for Yorkshire and Humber of 185 cases per million between 1979 and 2000, however the age boundaries ranged from 13 to 24 years (28). The variations in incidence for TYAs described may reflect different boundaries set for inclusion within a population such as the diagnoses, the age limits or the geographical region covered or different tumour classification systems used between studies, which are eluded to in Figure 5.1. Direct comparisons of incidence rates between this study population and national incidence rates for TYAs are therefore difficult, raising uncertainty around how representative this study population is of the TYA cancer population across England.

6.2.3 Pre-diagnosis secondary care involvement

This study provides a unique focus on secondary care services via the linkage of HES data sets for inpatient and outpatient events to the YSRCCYP.

6.2.3.1 Inpatient

Only 47 cases from the group of 1098 eligible study individuals were found to have no inpatient episodes prior to the date of definitive diagnosis. This finding could be either due to failure in linkage of events or alternative routes to diagnosis excluding inpatient activity. These cases arose throughout the study period and were distributed across the region according to residential postcode. This population may highlight cases in which the route of

diagnosis is atypical and require more in depth investigation that is beyond the scope of the information provided by the linked dataset. Such information may be contained within the case notes or maybe gleaned from patient interviews.

In the remaining 1051 cases with pre-diagnosis inpatient involvement, the number of pre-diagnosis inpatient episodes varied by age and diagnosis (Figure 5.5 and Figure 5.6). Age at diagnosis influenced the approach to healthcare with younger children reliant on their parents and older TYAs practising more independent healthcare engagement. This may be reflected in the reduced number of median inpatient episodes for the older age ranges (15 to 19 and 20 to 24) compared to 0 to 14 years. Nonetheless, this finding raises important questions of completeness of information within the HES records, which will be discussed further in section 6.3.

Within the study cohort, 48% of pre-diagnosis inpatient events occurred within the month preceding diagnosis involving 91% of the study population. The number of cases and episodes increased dramatically immediately prior to the date of diagnosis. This finding draws parallels with the pattern of engagement prior to diagnosis seen in primary care studies of childhood cancer (132). The findings of this study and those of primary care when taken together suggest increasing contact with the healthcare system leading up to the point of diagnosis.

The results show a higher proportion of admissions were categorised as emergency routes (38%) compared to elective routes (33%) within the pre-diagnosis episodes. Higher levels of emergency admissions are also observed by the routes to diagnosis work done in both primary and secondary care services in cancer patients across all ages in the UK healthcare service (16, 71). A comparison between the figures in the NCIN routes to diagnosis cohort and this study population for emergency and non-emergency admissions for 0 to 24 year olds with cancer is summarised in Figure 5.22. However this study did not include any control group data to act as a comparison and this limits the conclusions which can be drawn. Further analysis of the admission route will be included later in this section within the discussion of alert codes episodes.

6.2.3.2 Outpatient

As stated within the Results Chapter 5 the quality of the diagnostic coding in the outpatient data was very poor and limits the application of the data within this study. The inadequate presentation of diagnosis coding in relation to each outpatient appointment during the period of this study most likely reflects the fact that diagnostic coding is not a mandatory field in outpatient HES data (122).

72% of the study population had pre-diagnosis outpatient appointments and 30 cases from the 47 without pre-diagnosis inpatient episodes had pre-diagnosis outpatient appointments. The frequency of pre-diagnosis outpatient appointments displays a similar distribution by diagnostic group as seen in the pre-diagnosis inpatient episodes. Only 84% of outpatient appointments were recorded as having been attended, thus the overall study population who attended pre-diagnosis outpatient appointments was 62%. The median duration between an attended outpatient appointment and the date of diagnosis was 254 days which is a longer duration than the 41 days seen in the inpatient episodes. Lower case involvement in outpatient care pre-diagnosis and increased median TTD in outpatient care comparative to inpatient care highlights the latter as the most frequently accessed form of healthcare within hospital services preceding the diagnosis of CYA cancers.

Outpatient HES records the priority variable, which indicates the urgency required for consultant input in the outpatient setting. Within the attended appointments 20% were classed as urgent and only 12 pre-diagnosis outpatient appointments were given two-week wait priority; a referral route used commonly in adult care to fast-track suspected cancer cases through out-patient services. Limited use of this cancer referral pathway in children has been identified in the NCIN routes to diagnosis work and a study by Mant *et al* 2012 who found only 35 two-week urgent referrals were made in a three and a half year period from a study of referral patterns in a district general hospital with a shared care interest (69, 71). Only one out of 48 cases diagnosed with cancer over the period of the study were referred by the two week wait route (69). The extremely low representation of the two

week referral pathway within this study population suggests there may be incomplete or incorrect coding of this referral type. However the findings from the wider literature suggest either a lack of awareness in the application of the two week referral pathway in childhood cancer or that this pathway has limited application to the age group. The latter is the most likely explanation as most parents of a child with suspected cancer would be unlikely to wait for a two week appointment.

6.2.4 Pre-diagnosis cancer signs and symptoms

The vast majority of widely accessible awareness literature relevant to CYA cancers is aimed at the patient or the primary care professional, it is therefore applied within this study to set a standard for CYA cancers awareness for secondary care professionals. There are a number of resources that cover various tumour types such as the NICE suspected cancer referral guidance, TCT awareness 2012 and MacMillan cancer signs and symptoms (12, 17, 125, 126). Certain campaigns focus on specific cancers such as the HeadSmart campaign (CNS tumours) and literature published by the Bone Cancer Research Trust (bone tumours) (12). There is a bias towards the promotion of early diagnosis of CNS tumours in children and teenagers, which has a dedicated awareness literature and the common tumour type in resources that cover multiple tumour types (17, 125). There is a paucity of CYA specific guidance for certain tumours, most notably carcinoma and germ-cell tumours. There is however widely published site specific guidance aimed at older adults with carcinomas and the CNS tumour literature is relevant to intra-cranial germ-cell tumours.

From the 235 disease specific codes identified only 42 appeared in five or more cases and no disease specific alert codes occurred in five or more cases in hepatic tumour or carcinoma cases within the study population, reflecting the rare nature of hepatic tumours and the sparseness of awareness literature for carcinoma in CYA's. The most common disease specific alert codes appeared in CNS tumours with 58 codes identified, however only eight codes appeared in five or more cases, potentially reflecting the wealth of awareness literature available or the heterogeneity of

the clinical manifestations. “Intra-abdominal & pelvic swelling, mass & lump” is a common alert code that appeared in the disease specific codes for four diagnostic groups suggesting common clinical manifestations for certain CYA solid tumours. Many of the disease specific alert codes could be grouped into common themes such as epilepsy, focal neurological signs, headache and migraine and general signs and symptoms in CNS tumour cases. A summary of the disease specific codes is presented in table 5.4 and extensive lists of broad cancer codes and disease specific codes for the ICCC diagnostic groups is included in appendix 3.

There was also considerable overlap between the broad cancer alert codes highlighted in the full study cohort and the disease specific alert codes highlighted in the individual diagnosis cohorts see Figure 5.10 to Figure 5.15. This overlap is reflected within the code level analysis. However it doesn’t translate in the cases level analysis as alert codes were analysed simply for their presence within an episodes, whether they were broad or specific.

The awareness literature relates to signs and symptoms suggestive of a diagnosis of cancer, however the ICD-10 classification system used to assign codes to the clinical findings in an inpatient event is a diagnosis based system. The R code section of ICD-10 is relevant to “Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified”. This section of ICD-10 codes contributed 102 out of the 235 disease specific codes and 29 of the 64 broad cancer diagnosis codes. When reviewing these codes in more depth it is clear they predominantly occur within close proximity to the date of diagnosis. It is worth highlighting that the coders only include R codes if a sign, symptom or finding cannot be explained by the other diagnosis codes applied to the events within an episodes. Therefore if an R code appears in a HES episode it relates to a significant unexplained clinical finding that is not related to the overriding diagnosis for that episode. Thus the R codes are of particular interest for the analysis of misinterpreted or missed signs and symptoms suggestive of cancer preceding the eventual diagnosis. Specific analysis of this population of codes was not feasible within this study but could form part of future work.

The development of symptom libraries for CYA cancers is challenging and the heterogeneous patterns of clinical manifestation in this differing population of tumours hinders the development of clear red flag identifiers. Wilne *et al* 2007 published a systematic review and meta-analysis of the features of childhood CNS tumours presentation, identifying 56 signs and symptoms present at the time of diagnosis, taken from 74 papers (80). They highlighted the varied pattern of presenting features of CNS tumours by age, primary site and presences of raised intra-cranial pressure. No universal features were highlighted, although more common items such as headache, nausea and vomiting, motor and sensory deficits and signs of raised intracranial pressure were highlighted (80). Wilne *et al* (2007) concluded that it is important to pay close attention to children with multiple symptoms and signs at presentation and this systematic review contributed significantly to the development of the HeadSmart campaign (12, 80). The development of guidance relevant to clustering of signs and symptoms may improve the specificity of guidance aiming to identify CYA cancers, although the clustering of clinical features may result in failure to identify cases with rare presentation or those with isolated signs and symptoms.

The study method highlighted a number of alert codes occurring many years prior to the eventual cancer diagnosis (figures 5.10-5.14). A clear method for differentiating symptoms and signs of CYA cancer from those relating to more common illnesses within the pre-diagnosis episodes has still to be identified, given the low positive predictive value of ‘red-flag’ symptoms in CYA cancer (72, 117). The noise created by symptoms and signs related to other illnesses and not the eventual cancer diagnosis within pre-diagnosis episodes impedes the identification of clear time-lines to diagnosis within the HES data for this CYA population.

6.2.5 Time to diagnosis in secondary care

In the majority of cases the first pre-diagnosis alert code containing episode occurred within the year preceding diagnosis (Figure 5.17) and of the cases for whom the first alert code occurred within the year preceding diagnosis, just under 70% appeared within a month of definitive diagnosis (Figure

5.18). This finding suggests that the identifiable clinical manifestation of CYA cancer within secondary care services are predominantly seen in the time period immediately prior to diagnosis, indicating that for the majority of CYA cancers secondary care services have a minimal role to play prior to diagnosis. A high proportion of cases had their first alert code episodes and the first cancer code episodes within the month preceding the date of diagnosis. This suggests many of the identified cases may have a clinical suspicion of cancer in the month prior to the date a tumour specimen is taken that confirms the diagnosis of cancer, which is then recorded as the date of diagnosis.

The duration of time between the start of an alert code episode and the date of definitive diagnosis was calculated and the population divided into three cohorts as defined in section 5.5 and illustrated in Figure 5.19. Inpatient events with an alert code present preceding the date of definitive diagnosis were identified in 58% of the study population; in only a third of these cases was the duration between first alert code appearance and the date of definitive diagnosis longer than a month. As discussed previously there is a 31 day period accepted within NHS cancer referral pathways for the urgent referral and starting of treatment in childhood cancer. This pathway model has been adapted and applied within this study. Therefore the cases with first alert code confined to the month prior to diagnosis would fall within this present standard.

The profile of pre-diagnosis inpatient involvement varied by diagnostic group (Figure 5.17). It is generally accepted that the presentation of leukaemia whilst often vague has a short presentation pathway and exponential course which results in widespread disease at presentation, as discussed in Chapter 2. This is potentially reflected in the finding that nearly two-thirds of alert codes for leukaemia cases were identified within the month preceding diagnosis. A little over 10% of leukaemia cases had a duration of time between first alert code appearance and diagnosis of more than a month, a group likely to consist of a number of cases with a predisposing illness identified as carrying an increased risk of leukaemia, such as Down’s syndrome (Figure 5.11). Alert codes relating to a predisposing illness can therefore be present from the point of initial diagnosis, which in the case of

Down's syndrome could be at birth resulting in what appears to be a prolonged TTD.

The highest levels of prolonged secondary care alert code involvement were seen in CNS tumours, neuroblastoma and hepatic tumours. The hepatic tumour group comprised only 15 cases, 7 of whom had pre-diagnosis alert codes thus limiting the ability to draw firm conclusions concerning the significance of the aforementioned pattern of pre-diagnosis involvement. There were 41 neuroblastoma cases in the study population of whom 32 had pre-diagnosis alert codes and none of these related to an underlying diagnosis that predisposed to the development of neuroblastoma within the awareness literature. A prolonged TTD in secondary care was suggested in a third of these cases with associated pre-diagnosis alert codes, suggesting that neuroblastoma may be amenable to an early diagnosis intervention in secondary care services. Neuroblastoma cases were also shown to have the highest median number of pre-diagnosis inpatient events along with other predominantly paediatric solid tumours such as hepatic and renal tumours.

There is a considerable amount of awareness literature published for the CNS tumour population compared to most other CYA cancers. The high mortality and morbidity associated with the diagnosis of a CNS tumour means a high potential to improve outcomes by minimising TTD. Within this study a number of alert codes appeared well before the date of diagnosis with the most common codes relating to epilepsy and convulsions. Almost 30% of the 182 CNS cases within the study population appeared to have a prolonged time between first alert code and definitive diagnosis. CNS tumours therefore present a clear opportunity for targeting early diagnosis interventions. The HeadSmart campaign had not been published until after the most recent diagnosis included within the study population, therefore the TTD for CNS cases in more contemporaneous cases may have subsequently improved.

The lowest frequency of alert codes was seen in the carcinoma, germ-cell tumour, bone tumour and retinoblastoma populations. The retinoblastoma group consisted of very few cases with limited alert codes and as for hepatic tumours, the discussion of this group is limited. There was a paucity of early

diagnosis awareness literature identified that was relevant to germ-cell tumours and carcinoma in CYAs and this is likely to explain the low levels of alert codes in these tumour groups. In contrast, a dedicated resource was identified for bone tumours. The pattern of alert code involvement seen within the bone tumour group may reflect a paucity of secondary care involvement associated with this disease preceding diagnosis, only 26 out of 62 cases had alert code involvement preceding the date of diagnosis; indeed, out of the 19 alert codes identified only one occurred in five or more cases: “pain in joint”. The fact that there is a lot of early diagnosis information for bone tumours but a paucity of pre-diagnosis alert code identified within this study would suggest that prolonged secondary care involvement preceding the diagnosis is not an issue for bone tumours. Early diagnosis interventions should instead focus on pre-hospital services and patients in this specific tumour group.

Admission methods for cases with alert codes were presented at the continuous inpatient spells (CIPS) level (Figure 4.1). The findings of this study suggest that CYA cancer cases with signs and symptoms of cancer preceding their date of diagnosis present to secondary care via emergency routes. This is in agreement with published results of other NHS routes to cancer diagnosis work (16, 71). There is a higher proportion of admissions through the emergency GP route in those cases where involvement occurs within the month prior to diagnosis only compared to cases with more prolonged TTD. A higher proportion of emergency A&E admissions was seen in cases with more prolonged TTD. The high proportion of emergency routes for admission associated with alert code status indicates a high degree of urgency relating to these inpatient events, which may reflect admissions associated with a concerning clinical picture.

6.2.6 Survival for children and young adults with cancer in Yorkshire

The overall survival probability for this study population at one-year was 91.5% (95% CI 89.7-93.0%) compared to 83.6% (95% CI 81.3-85.7%) three years after diagnosis. One-year survival has been promoted within the DoH

as a marker for late/early diagnosis and reflects the mortality related to the severity of disease at presentation and immediate disease and treatment related complications (11). Three-year survival is used in this study as a longer-term measure of outcome, rather than the usual five-year follow-up period, due to censoring of individuals diagnosed within the study period (2004-2009). The 8% difference between one-year and three-year survival suggests a significant impact of treatment and patient management on the outcome for CYA cancer within the FYRHA. Stiller *et al* 2007 reported one-year and three-year survival estimates for childhood cancer in Britain between 1991-2000 of 88% and 78%, compared to 91% and 85% for 0 to 14 year olds within this study population (133). The more favourable survival seen at both one and three-years survival and the narrower gap between one and three-year survival within this study population compared to Stiller *et al* 2007 perhaps suggests improvements in earlier diagnosis and treatment effects between 1991 to 2000 and 2004 to 2009 (133). It must be noted that Stiller *et al* 2007 is based on a national population and additionally that there have been improvements in the case ascertainment and diagnostic sensitivity of cancer registration between the two study periods (133). There appears to be a slightly more favourable outcome at one-year for 15 to 24 year olds with 92% alive at one-year post diagnosis, with three-year survival falling to 82%. This 10% fall in survival may reflect the reduced involvement of 15 to 24 year olds within trials as these cases may fall between childhood and adult services (134).

No difference in survival was observed by sex at one or three years within the study population. However within the wider literature, females are more likely to survive than males, emphasised by the prolonged treatment course for males with ALL (135). Nonetheless sex was retained in the multivariable survival model as it potentially impacts on the route to diagnosis especially in the older cases; girls are more likely to have contact with their GP for routine care and contraception (66, 134).

The results of the EUROCARE-4 study indicated that overall teenagers and young adults with cancer had better five-year survival outcomes compared to children with cancer (5). However the survival for certain cancers, such as ALL, bone tumours and soft-tissue sarcomas is worse in TYAs compared to

children (4, 5). The 15 to 24 year olds within this study population had slightly better one-year survival rates compared to their younger counterparts, however further analysis is required as the difference was not statistically significant.

Survival across the diagnostic groups varied widely, results for retinoblastoma (V) and other tumours (XII) were not considered due to the very low numbers of cases involved. Lymphoma and germ-cell tumours showed at least a 50% reduced risk of death compared to the leukaemia when correcting for age and sex, which was statistically significant ($p<0.05$). Several tumour groups showed an increased risk of death including CNS tumours (70%), osteosarcoma and soft-tissue sarcoma (120%). neuroblastoma (150%) and hepatic tumours (240%); all achieved statistical significance ($p<0.05$) when correcting for age at diagnosis and sex (Table 5.5). The variable pattern of survival by diagnosis seen in this study reflects the generally accepted patterns within the wider literature, however due the limited number of cases further subgroup analysis within this study population was not conducted. The significant differences in the risk of death by diagnostic group identified within this study population supports the inclusion of diagnostic group within the multivariable Cox proportional hazards model used to analyse the association between survival and TTD.

6.2.7 Time to diagnosis and survival outcomes

The second aim of this study is to assess how variations in TTD within secondary care affect outcomes for CYA patients. Three groups of patients were highlighted within the previous Results Chapter:

- Cases with no alert code involvement preceding the date of definitive cancer diagnosis (N=457), which is made up of those cases with no inpatient involvement (n=47) and those cases with no alert code episodes (n=410) preceding the date of a definitive diagnosis of cancer.
- Cases with alert code episodes only within the month before diagnosis (N=437).

- Cases with alert code episodes more than a month prior to diagnosis (N=204).

Cases with alert codes present immediately prior to diagnosis (group 2) were 67% significantly more likely to die compared to all other cases across the study population, see table 5.5. This increased risk remained after correcting for age at diagnosis, sex, diagnostic group and year of diagnosis. A 30% increased risk of death was seen in cases 0 to 24 years with potentially more prolonged TTD (group 3). These findings indicate an increased risk of death for the cases where the clinical manifestation of their cancer are identified within secondary care preceding the definitive diagnosis. The cases with alert codes confined to the month preceding diagnosis have the worst outcome, potentially identifying those cases where there is an acute presentation or clinically apparent severe manifestation of disease. The limited inpatient involvement of cases with poorer outcomes raises the question of whether these are higher-grade fast growing tumours or is generally evidence of late presentation with advanced stage disease due to prolonged time spent within the PI or other healthcare settings. Dang-Tan *et al* 2009 provided a potential answer to the above question, they described a low risk of prolonged healthcare service delay associated with a prolonged patient delay in a population of in Canadian children and adolescence with leukaemia and lymphoma (114).

Interaction tests identified a significantly worse outcome in the older age group for cases with potentially more prolonged TTD (group 3) compared to children within this group and all other TTD groups by age. The increased risk of death in TYAs with a prolonged TTD, identified within this study, was also a finding within a number of early diagnosis studies within the systematic review (Chapter 3). Despite the limited awareness literature for the common TYA cancers identified within this study, those TYA cases identified with potentially more prolonged TTD had a poor outcome, indicating improving TTD in secondary care for TYA could influence outcome.

Gender did not appear to have any influence on survival within the three TTD groups. Once the population of cases were divided by the diagnostic

groups the number of deaths within each subgroup was very small (Figure 5.33). This therefore limited the assessment of how variations in the TTD affected survival for each diagnostic group.

The group with potentially more prolonged TTD were likely to include a higher proportion of cases for whom an underlying disease was identified within the awareness literature, such as Down's syndrome, tuberous sclerosis or neurofibromatosis. The alert codes identified within this group may also be relevant to clinical manifestations of more indolent or lower-grade tumours. The specificity of the alert codes may decrease as time intervals between alert code occurrence and definitive diagnosis increases, with alert codes relating to other and unrelated illness due to the often vague and varied signs and symptoms seen in CYA cancers. This point will be discussed further within the limitation section.

The high proportion of emergency admission routes in cases with alert codes preceding definitive diagnosis provides a valuable insight into healthcare engagement for CYAs presenting to secondary care. CRUK in 2013 described one in four new cancer cases presenting via Accident and Emergency as unacceptable and associated this fact with thousands of preventable deaths, a statement relating to all cancers across all ages (136). The high proportion of emergency routes to diagnosis in CYAs, especially A&E, identified within this study and the NCIN routes to diagnosis work pertain to different patterns of healthcare engagement within CYAs compare to older adults. This highlights the value of emergency care for the diagnosis of CYA cancer and the need to work with rather than to avoid the emergency healthcare structures.

There are other potential confounding factors which could have been included within the survival model, such as the initial secondary care centre accessed by the patient. The key findings outlined in this section and limitations discussed within the next section will be considered together and recommendations for early diagnosis in CYA cancer within secondary care developed in section 6.4, along with potential future work.

6.3 Strengths and Limitations

6.3.1 Strengths

This study introduces a novel population-based approach to analysis of objective health services data. The registry data used within this study has high case ascertainment and there was a high level of data linkage between the YSRCCYP and HES data sets.

There were clear definitions within the study design for the time-intervals within the TTD informed by a systematic review of early diagnosis literature (Chapter 3), based upon an established theoretical framework for early diagnosis research.

The study method introduces a reproducible coding scheme of relevant alert codes for CYA cancer applicable to both secondary and primary care. The scheme was developed from established and widely available early diagnosis resources for CYA cancers. The alert codes were scrutinised by experienced clinicians in the field of paediatric oncology and paediatric haemato-oncology prior to the inclusion within the analysis.

6.3.2 Limitations

In order for clear and appropriate conclusions to be drawn from this study the limitations of the study methods must be considered. This section will discuss the limitations of the data sources and applied methods, including the referencing of the early diagnosis alert codes and the intrinsic limitations of research within the CYA cancer population. It is important to consider the feasibility of achieving the aims and objectives of the study within the limits of the data and methods applied.

The date of definitive diagnosis is the first and perhaps most fundamental point of consideration within the limitations of this study. This date was extracted from the YSRCCYP and is defined as the date when a specimen was taken that confirmed the diagnosis of cancer. This date provided the censoring point for the duration of time spent within inpatient care prior to diagnosis within this study. The process of making a diagnosis of cancer

begins before the date of pathological confirmation and from the tertiary care perspective begins with a clinical diagnosis based on the initial history and examination. It is possible to crudely identify the date of clinical diagnosis within the HES records as the date of first cancer code occurrence reflected by the presence of an ICD-10 C code within an episode. In section 5.4.1 the frequency of C code containing episodes preceding the date of definitive diagnosis is discussed, identifying over 1000 such episodes within the dataset. However, these C code episodes are mostly confined to the month preceding the date of definitive diagnosis. Only 4% of C code episodes preceding the definitive diagnosis occur more than a month prior to the date of definitive diagnosis. Given this, correcting the date of diagnosis for the date of the first C code episode may not have a huge impact on the duration to diagnosis. However, correcting for first C code episode could significantly impact on the number of cases with alert codes preceding the date of definitive diagnosis since poorer survival rates were observed in cases with alert codes immediately prior to the diagnosis for the overall study cohort. This population may consist of cases of CYA cancers with clinically apparent cancer diagnoses presenting acutely to secondary care with highly suspicious and obvious signs and symptoms potentially associated with more advanced or aggressive disease.

A number of limitations were identified within the HES records used to analyse the inpatient and outpatient events preceding a diagnosis of cancer within this study. The pre-diagnosis inpatient involvement for the older cases was incomplete, a finding identified by the fact that the date of the first HES inpatient episode occurred in 1996. This means a maximum inpatient history of 8 years for cases diagnosed in 2004. This could impact on the number of episodes per case in the older cases and the TTD for those cases with an underlying condition that predisposes to the eventual diagnosis of cancer.

The information contained within a HES episode will be heavily influenced by the coders reviewing the medical notes. The methods of coding have become more standardised due to the influence of payment by results, however the implementation of PbR occurred during the course of the time period for this study, the impact of this is most likely to be seen after 2006 (122). Time spent within the coding department in Leeds Teaching Hospitals

Trust demonstrated the extent of the highly standardised approach and consistent audit of coding in current practise. However concerns remain about the quality and reliability of coding with regard to historical practises and inter-hospital variations.

This study aimed to highlight misinterpreted or overlooked signs and symptoms of cancer within CYA's and is reliant on completeness of records relating to such clinical events. The CYA cancer awareness literature used within this study highlights symptoms and signs indicative of cancer. Yet the ICD-10 classification system used to code the clinical features within an inpatient episode are based on morbid entities which are mostly defined as diagnoses. ICD-10 does contain the R section describing unexplained symptoms and signs, which provided over 100 of the 235 alert codes identified (32). The discordance in focus between the referenced resources and the ICD-10 classification system applied to code the HES data does raise concerns regarding the appropriate utilisation of HES records for this study and the completeness of information presented. The validation of HES records with the information within the medical records is therefore an important process within this study. Incomplete presentation of the medical manifestations of cancer within the HES records will result in underrepresentation of early diagnosis signs in the identified cases with alert codes and underestimation of the number of cases with pre-diagnosis inpatient events suggestive of cancer.

As previously stated the majority of the early diagnosis literature used within this study is aimed at patients and primary care professionals. The primary care professional is most often identified as making the predominant contribution to the doctor interval preceding a cancer diagnosis. The awareness literature used therefore acts as a base-line for secondary care doctors and as such could result in potential under-estimation of alert signs and symptoms identifiable within secondary care episodes. CYA cancer diagnosis guidance focused on secondary care professionals is not available at present and it is beyond the scope of this study to create this.

The methods applied within this study rely on the completeness and accuracy of recorded data within the HES records for the identification of

pre-diagnosis alert codes. As previously discussed, the omission of key clinical information relevant to a diagnosis of cancer due to the fact it is ascribed an alternative misdiagnosis or missing because of failure to recognise it as significant would result in omission from analysis for the alert codes. Such information may be recorded in the clinical notes and the process of clinical coding may have inadvertently filtered this out. Alternatively, the medical professional may omit information from the medical records due to it being incongruous with an initial misdiagnosis, thus leading to over-estimation of cases without alert codes preceding diagnosis.

As identified in the systematic review (Chapter 3), the relationship between TTD, stage and survival is the missing piece of the early diagnosis jigsaw. One of the main aims of this study was to investigate the association between TTD, stage and outcome for CYA cancers. There was a lack of available data on stage or grade limiting the ability to tackle this question in its entirety, and limited data on treatment meant this could not be used as a proxy for stage. At the time of this study the only means of achieving clear staging data was to conduct a complete case note review, again reliant on availability of complete records documenting staging. Alternatively one could conduct a prospective study on a new cohort with clearly define staging criteria from the outset. There are plans to improve the recording of stage in cancer registries across the UK and the Cancer Outcomes Service Database has made this a requirement for cancer registries. The challenge for specialist CYA cancers registries is maintaining a high level of data accuracy given the varied and ever evolving nature of cancer staging in CYA cancer. There must also be consideration of cancer grade alongside stage, the grade provides an indication of the aggressiveness of a tumour and reflects the speed of growth and tumour development. In the CNS tumour literature there is an established link between shorter times to diagnosis in higher grade tumours (84). It may be more practical in future to think of the TTD relative to the tumour grade when considering the impact of prolonged TTD in CYA tumours.

The method also failed to identify clustering of signs and symptoms within episodes and across episodes and time. Disease specific alert codes were identified using Stata12 commands for each of the diagnostic groups

separately. Due to overlap in the alert codes between diagnostic groups, the Stata12 programming could not be easily written across for the entire population at once, which therefore didn't allow me to identify clusters of codes in the entire dataset. This is a code writing limitation and not an intrinsic limitation of the data used. Therefore with more time and advanced programme writing capabilities could overcome this issue.

The rare nature of CYA cancers and the diverse patterns of presentation result in a low predictive value of alert symptoms for cancer within this population and marks a key obstacle to the development of 'red-flag' symptoms and signs. This issue was clearly highlighted by Dommett *et al* (2012) in a large case-control study of alert symptoms in primary care for CYA cancers, identifying that …"*of 10000 children with a recorded alert symptom, approximately 6 would be diagnosed with cancer within 3 months*" (72). Within the setting of primary care there have been a number of recent publications which draw clear and important conclusions from CYA early diagnosis research through the use of case-control study methods (72, 132). The lack of control group data within this study limits the conclusions that can be drawn regarding the occurrence of alert codes, the admission routes and the frequency of inpatient involvement preceding the date of diagnosis.

CYA cancers are a heterogeneous group of tumours with a diverse tumour biology and are rare. This study has struggled to conduct multi-level subgroup analysis due to the sample size and this was especially noted when assessing how variations in TTD affect survival for each diagnostic groups. A larger sample size may have allowed a more sophisticated modelling approach, such as the use of multi-level methods allowing us to account for the natural nesting of patients within diagnostic groups.

Only a limited number of variables were used in the survival analysis models within this project, the variables used were well recorded and had all been implicated as influential to survival in CYA cancer. Future work could incorporate variables such as treatment, stage, geographical location, GP, deprivation, ethnicity, willingness to seek medical help (compliance).

6.4 Recommendations and future work

6.4.1 Health care recommendation

Hospital care plays an important role in routes to diagnosis for CYAs with cancer, often as the final step in the diagnostic pathway. This is reflected within this study as 96% of cases had pre-diagnostic inpatient involvement and more than a third of cases had inpatient involvement isolated to within a month preceding diagnosis. This study also identified evidence that secondary care is not just the final step for healthcare involvement leading to a diagnosis of CYA cancer. Just under 20% of CYA cancers in a large regional population based registry had identifiable signs and symptoms suggestive of cancer at more than a month preceding diagnosis within hospital care and their survival varied by age.

No difference in survival was seen for children (0-14) with signs and symptoms suggestive of cancer more than a month preceding diagnosis compared to children without identifiable signs and symptoms suggestive of cancer. The survival was significantly worse for those children with symptoms and signs isolated within the month prior to diagnosis, suggesting interventions aimed at improving the TTD for childhood cancers in secondary care services would not impact on survival. This supports the hypothesis that interventions should be aimed at the pre-hospital admission and patient level in childhood cancer.

TYA's (15-24) with potentially more prolonged TTD within secondary care had a worse survival compared to their counterparts without alert codes and also childhood cases with potentially more prolonged TTD. This study identified a lack of TYA specific awareness literature especially in relation to germ-cell tumours and carcinomas, there is however a large library of literature available for adult cancers by site that could be adapted for use in these more common TYA cancers. Adaptation and not application of adult literature is important as TYA carcinoma's can often behave differently to adult cancers and TYA's themselves will respond differently to awareness interventions.

In section 5.4.2 and 5.4.3 alert codes occurring in five or more cases pre-diagnosis were relatively infrequent, especially the disease specific alert codes. It is also clear that in the majority of cases these codes occurred for the first time in close proximity to the date of diagnosis, although there were commonly extreme outliers in duration. The frequency of primary care attendances for CYA with cancer increases towards diagnosis, a pattern reflected within secondary care services in this study (132). It is therefore important to consider the clustering of alert codes within a certain time span for a case during the pre-diagnosis time period. In many accident and emergency departments in the UK there is a system of recording the number of previous attendances for a patient. This system could be adopted in the acute paediatric assessment room setting and possibly coupled with a means of producing previous attendance sheets. Highlighting previous or multiple attendances with related or concerning features potentially identifying cases for whom an undiagnosed or miss-diagnosed process may be ongoing. The emphasis should be on establishing the time line for each CYA seen in secondary care and the promotion of systems that record attendances will help in this effort. As electronic records continue to develop, we may also be able to develop linked attendance records across healthcare systems.

Identification of an abdominal mass was a common first alert code within a number of tumour types as well as within the broad cancer alert codes. The duration to diagnosis from the first occurrence of this sign was often short. There is a suggestion in the literature that the UK has a worse outcome for tumours that present in childhood with an abdominal mass such as neuroblastoma and Wilms' tumours (47). In the case of the former, the abdominal mass may be an incidental finding; in the case of the latter, the TTD within the literature was consistently amongst the shortest duration. Both points reflect the fact that abdominal tumours can have an insidious onset and the disease may be extensive by the time it is eventually detected. It is therefore important that at the point of contact, a reviewing doctor has the experience and clinical skills to examine a child in a comprehensive manner. The abdominal examination must be a mandatory part of the assessment in every child seen in secondary care services. Education and

development of the clinical skills for medical professionals working with children could form part of a healthcare intervention to improve TTD for this specific tumour presentation.

This study has utilised a number of CYA cancer awareness resources to highlight alert codes, the majority of which focused on the symptoms and signs that could indicate a potential cancer diagnosis. The HeadSmart campaign also provides the professional working with children with advice regarding the investigation of a child or teenager with a suspected CNS tumour (12). The campaign advocates the use of an MRI scan in any child suspected of having a brain tumour and only CT with contrast if MRI is not available. This advice is readily available online to professionals and there is an education module that accompanies the awareness campaign. This study has highlighted a number of alert codes, especially relating to convulsions and epilepsy, within the CNS tumour population which are present long before the date of definitive diagnosis. There is a suggestion from this study that cases with CNS tumours may have more prolonged secondary care involvement (Figure 5.13) especially for epilepsy and convulsions related admissions. The advice given within the HeadSmart campaign provides the secondary care services with a challenge as the provision of resources for MRI scans for all suspected CNS tumours would require major financial investment given the extremely low positive predictive value of headache and neurological signs published by Dommett *et al* 2012 (72). Improved access and resource for appropriate and indicated imaging techniques such as MRI for the investigation of suspected CNS tumours maybe key to improving the TTD in certain CYA cancer. However, this needs to be implemented in conjunction with improved access to healthcare services for CYAs and improved history and examination skills for point of contact junior staff engaging with CYA.

Outpatient data was of limited use to this project due to the lack of diagnostic codes applied to each appointment, although it was clear that the two-week referral pathway for cancer is poorly recorded. It has been suggested that there is poor uptake of this referral pathway for suspected CYA cancers, although there have been limited studies of the use of this referral pathway in CYAs (69). This finding may reflect the inappropriate nature of this

pathway in CYA suspected of cancer, for whom even a two week wait is deemed unacceptable to parents and referring professionals. There is scope for reassessment of this pathway in CYA cancers as well as any resource attached, followed by development of more specific and targeted referral structures for CYAs suspected of cancer. This may take the form of development of communication structures across the region, for example improving links between specialists and primary and secondary care professionals who have contact with CYAs.

6.4.2 Health services research recommendations & future work

The systematic review section (Chapter 3) identified several requirements for robust and reproducible early diagnosis research in all age groups, including a theoretical framework from the outset, clear definitions for the time intervals and hierarchical approach to defining the time points for the study.

This study has focused on TTD for CYA cancer in secondary care. Nonetheless, this is merely part of the complex picture of pre-diagnosis healthcare involvement for these young people. A key finding from this study was the association between prolonged secondary care involvement suggestive of cancer and worse survival rates in the 15 to 24 year olds. Unfortunately, due to the limited sample size a clear assessment of how variations in TTD affected survival within TYA specific diagnostic groups. The findings of this study support the need for future work in TYA cancers to identify how variation in TTD affect outcome for specific diagnoses, identifying at risk diagnostic groups and facilitating TYA specific early diagnosis strategies. A potential approach for future studies could involve inter-regional analysis of pathways to diagnosis for cancer; potentially identify variations in approach to TYA care and highlighting areas of best practise.

Inpatient involvement only tells part of the story of secondary care and the outpatient and emergency contact also needs to be considered. Unfortunately the quality of outpatient HES as a resource for this type of health services research is poor as the diagnostic coding is currently

inadequate. The quality of outpatient HES data is unlikely to improve given that diagnostic coding is not a mandatory field, unlike inpatient HES. A combined data set of inpatient HES, outpatient HES and A&E HES would facilitate an enhanced view point of secondary care, however the lack of diagnostic information in outpatient HES data restricts its use to the sequencing of events, illustrated by the NCIN routes to diagnosis work (71). In order to provide a more complete picture of pre-diagnosis involvement secondary care data could be linked to primary care data. However, at present the national primary care data resource of Clinical Practise Research Datalink only covers a limited proportion of the UK population. Linkage to electronic health records for primary care such as SystmOne may provide a mechanism to include primary care data in the future, however the coverage and feasibility of exploiting such resources across a regional or national study would need investigating.

This study aimed to assess how reliable and accurate HES data are in the analysis of TTD and survival. Unfortunately, this study was unable to provide an insight into the association between TTD and stage within secondary care services for CYA cancers. Due to the lack of staging data within the routine health datasets utilised this aim was not achieved. Findings from the systematic review show that the TTD is influenced by the biology of a tumour. It is therefore recommended that the grade and stage of the tumour be recorded for future TTD studies and the TTD be considered relative to the tumour biology. In high-grade fast growing tumours there is a more pressing need to achieve a prompt diagnosis and the key features that identify these aggressive tumours should be investigated and promoted within CYA early diagnosis interventions.

At present early diagnosis research often defines the time of symptom recognition as the point of initiation (Figure 2.1), however there is a period of pre-symptomatic disease development that precedes this. The Background section (Chapter 2) discussed some limited knowledge of how the rate of tumour progression affects the TTD and this was further explored as part of the systematic review (Chapter 3). Investigation of markers for tumour proliferation such a Ki-67 and mitotic index may help inform this gap in knowledge regarding variations in tumour development and could form part

of a combined study of tumour biology and TTD. Markers for tumour proliferation could be incorporated into theoretical models of tumour development, helping to identify the point of initiation for a tumour, thus giving a definitive TTD.

Survival is not the only outcome measure for CYA cancers. In fact, as survival increases the emphasis shifts more to the reduction of morbidity, for example in the UKALL 2011 trial currently underway in the UK (137). The challenge for the CYA early diagnosis researcher investigating morbidity as an outcome is the heterogeneity of endpoints being measured due to variation in disease biology and treatment modalities across CYA cancers, further compounded by the rarity of the disease. The association between TTD long-term outcome measures such as survival or morbidities may be difficult to clearly define due to the influence of confounding factors such as treatment. The stage at diagnosis could be the end-point for future TTD studies in CYA cancers, allowing investigation of more contemporaneous data for patients recently diagnosed who are more likely to have clear staging recorded.

Beyond an internal assessment within diagnostic groups of the cohort, the study lacked a comparator group for CYA secondary care involvement in the general population. This would have improved the strength of conclusions, for example adoption of a case-control design, similar to other CYA early diagnosis research done in the primary care setting (72, 132). Challenges of this approach would include defining clear endpoints for analysis for those case not diagnosed with cancer.

6.5 Conclusions

This is the first study to focus on the TTD for CYA cancers in secondary care services within the UK, using linked datasets to identify clinical manifestation of CYA cancers preceding the definitive diagnosis and investigate associations with survival. The use of the YSRCCYP allowed a population based approach for the study, resulting in a high level of case ascertainment. This study has a number of limitations that must be considered when drawing conclusions, relating mainly to the methods applied and the resources used. In particular, the limitations of analysis of clinical information contained within health data that has been collected and coded by non-standardised methods for a purposes other than health services research must be considered.

This study shows increasing numbers of CYA cancer cases with inpatient involvement towards the date of definitive diagnosis. Maximal inpatient involvement occurred in the month immediately preceding diagnosis for both general inpatient episodes and alert codes containing inpatient episodes. This pattern of increased healthcare engagement prior to diagnosis is has also been identified in primary care studies of CYA cancers (72, 132).

In this large regional population of CYA cancer cases a number of clinical features suggestive of the subsequent diagnosis of cancer have been identified from the codes assigned within individual HES inpatient episodes. When cases have clinical features suggesting a cancer diagnosis prior to the date of diagnosis they are most commonly admitted to inpatient services through emergency routes. A&E and GP emergency routes predominate, with the emergency GP route utilised mostly in admissions proximal to the date of definitive diagnosis. The utilisation of emergency pathways for CYA with cancer has been highlighted in previous primary care and secondary care routes to diagnosis work. The utilisation of emergency routes coupled with the lack of utilisation of the two-week cancer referral pathway shown in the outpatient data within this study highlights the potential need to reassess cancer referral structures in CYAs.

Cases with alert symptoms and signs isolated to the month preceding definitive diagnosis have a significant worse outcome compared to cases without alert codes. This suggest that these cases have marked and easily identifiable signs of disease and signs potentially indicating rapidly progressive disease or a prolonged period of development prior to engaging with inpatient care, which may reflect a prolonged primary care interval or patient interval. A significant or detrimental effect of a prolonged TTD in secondary for CYA cancers was identified within the 15-24 year olds within this study, but not the 0-14 year olds. This suggests secondary care specific early diagnosis interventions in the older population may be helpful and improve survival.

This study suggests that secondary care services in Yorkshire appear to be organised effectively to deal timely diagnosis of cancer in CYAs. However, further investigation of TTD for CYA cancer across all healthcare services is required and the inclusion of staging data is imperative. The TTD relative to the tumour biology is a key component in future studies in order to facilitate clear conclusions with regard to the effect of TTD on outcome.

Reference List

1. Cancer Research UK. CancerStats Incidence UK 2009. 2012.

2. Birch JM, Pang D, Alston RD, Rowan S, Geraci M, Moran A, et al. Survival from cancer in teenagers and young adults in England, 1979-2003. Br J Cancer. 2008;99(5):830-5.

3. Cancer Research UK. CancerStats - Childhood Cancer - Great Britain & UK. 2010.

4. Eden T, Barr R, Bleyer A, Whiteson M, editors. Cancer and the Adolescent. Second ed: Blackwell Publishing & BMJ Books; 2005.

5. Gatta G, Zigon G, Capocaccia R, Coebergh JW, Desandes E, Kaatsch P, et al. Survival of European children and young adults with cancer diagnosed 1995-2002. Eur J Pediatr. 2009;45(6):992-1005.

6. Department of Health. Cancer Reform Strategy. 2007.

7. Geraci M, Birch JM, Alston RD, Moran A, Eden TOB. Cancer mortality in 13 to 29-year-olds in England and Wales, 1981-2005. Br J Cancer. 2007;97(11):1588-94.

8. National Cancer Survivorship Initiative (NCSI). Children and young people living with and beyond cancer. Designing and implementing pathways to benefit patient aftercare: Continuing to build the evidence. 2011.

9. Department of Health. The NHS Cancer Plan: A plan for investment, A plan for reform. 2000.

10. National Institute for Health and Clinical Excellence (NICE). Guidance on Cancer Services: Improving Outcomes in Children and Young People with Cancer, The Manual. 2005.

11. Department of Health. Improving Outcomes: A Strategy for Cancer, First Annual Report. 2011.

12. RCPCH, SDBTT, CBTRC, The University of Nottingham and The Health Foundation. HeadSmart. 2011 [updated Date accessed: 28th June 2012]; Available from: http://www.headsmart.org.uk/.

13. Gill PJ, Goldacre MJ, Mant D, Heneghan C, Thomson A, Seagroatt V, et al. Increase in emergency admissions to hospital for children aged under 15 in England, 1999–2010: national database analysis. Archives of Disease in Childhood. 2013.

14. Elliss-Brookes L, McPhail S, Ives A, Greenslade M, Shelton J, Hiom S, et al. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. Br J Cancer. 2012;107(8):1220-6.

15. Purkayastha D, O' Hara, C., Moran, T. Routes to diagnosis: investigating the different pathways for cancer referrals in England for Teenagers and Young Adults. In: National Cancer Intelligence Network, editor.: NHS North West Cancer Intelligence Service; 2013.

16. Royal College of General Practitioners, National Health Service National Cancer Action Team, National Cancer Intelligence Network. National Audit of Cancer Diagnosis in Primary Care. In: Heath Do, editor. 2011.

17. Teenage Cancer Trust. Teenage Cancer Awareness Week. 2012; Available from: http://www.teenagecancertrust.org/workspace/documents/TeenageCancerTrustAwarenessWeek_Pack.pdf.

18. Olesen F, Hansen RP, Vedsted P. Delay in diagnosis: the experience in Denmark. Br J Cancer. 2009;101 Suppl 2:S5-8.

19. Weller D, Vedsted P, Rubin G, Walter FM, Emery J, Scott S, et al. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. Br J Cancer. 2012;106(7):1262-7.

20. Walter F, Webster A, Scott S, Emery J. The Andersen Model of Total Patient Delay: a systematic review of its application in cancer diagnosis. Journal of Health Services & Research Policy. 2012;17(2):110-8.

21. Safer MA, Tharps QJ, Jackson TC, Leventhal H. Determinants of three stages of delay in seeking care at a medical clinic. Medical Care. 1979;17(1):11-29.

22. Andersen BL, Cacioppo JT. Delay in seeking a cancer diagnosis: delay stages and psychophysiological comparison processes. British Journal of Social Psychology. 1995;34(Pt 1):33-52.

23. Hargrave D. Pontine glioma. To biopsy or not to biopsy: that is the question. British Journal of Neurosurgery. 2008;22(5):624.

24. European Network of Cancer Registries. Recommendations for coding Incidence Date.

25. Teenage Cancer Trust. Teenage Cancer Trust - What we do/about us. 2013; Available from: http://www.teenagecancertrust.org/who-we-are/about-us/.

26. Martin S, Ulrich C, Munsell M, Taylor S, Lange G, Bleyer A. Delays in cancer diagnosis in underinsured young adults and older adolescents. Oncologist. 2007;12(7):816-42.

27. Pollock BH, Krischer JP, Vietti TJ. Interval between symptom onset and diagnosis of pediatric solid tumors. J Pediatr. 1991;119(5):725-32.

28. Alston RD, Rowan S, Eden TOB, Moran A, Birch JM. Cancer incidence patterns by region and socioeconomic deprivation in teenagers and young adults in England. Br J Cancer. 2007;96(11):1760-6.

29. Birch JM, Pang D, Alston RD, Rowan S, Geraci M, Moran A, et al. Survival from cancer in teenagers and young adults in England, 1979-2003. Br J Cancer. 2008;99(5):830-5.

30. Croucher C, Whelan JS, Moller H, Davies EA. Trends in the incidence and survival of cancer in teenagers and young adults: regional analysis for South East England 1960-2002. Clinical Oncology (Royal College of Radiologists). 2009;21(5):417-24.

31. Tobias J, Hochhauser D, Souhami R. Cancer and its management. 6th edition ed: Blackwell publishing; 2009.

32. World Health Organisation. ICD-10, International Statistical Classification of Diseases and Related Health Problems. 2010.

33. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, third edition. Cancer. 2005;103(7):1457-67.

34. Birch JM, Alston RD, Kelsey AM, Quinn MJ, Babb P, McNally RJQ. Classification and incidence of cancers in adolescents and young adults in England 1979-1997. Br J Cancer. 2002;87(11):1267-74.

35. Kaatsch P, Steliarova-Foucher E, Crocetti E, Magnani C, Spix C, Zambon P. Time trends of cancer incidence in European children (1978-1997): report from the Automated Childhood Cancer Information System project. European Journal of Cancer. 2006;42(13):1961-71.

36. Alston RD, Geraci M, Eden TOB, Moran A, Rowan S, Birch JM. Changes in cancer incidence in teenagers and young adults (ages 13 to 24 years) in England 1979-2003. Cancer. 2008;113(10):2807-15.

37. Stiller C. Epidemiology of cancer in adolescents: Medical and Pediatric Oncology. 39 (3) (pp 149-155), 2002. Date of Publication: 2002.; 2002.

38. Stiller CA, Bielack SS, Jundt G, Steliarova-Foucher E. Bone tumours in European children and adolescents, 1978-1997. Report from the Automated Childhood Cancer Information System project. European Journal of Cancer. 2006;42(13):2124-35.

39. Peris-Bonet R, Martínez-García C, Lacour B, Petrovich S, Giner-Ripoll B, Navajas A, et al. Childhood central nervous system tumours – incidence and survival in Europe (1978–1997): Report from Automated Childhood Cancer Information System project. European Journal of Cancer. 2006;42(13):2064-80.

40. Spix C, Pastore G, Sankila R, Stiller CA, Steliarova-Foucher E. Neuroblastoma incidence and survival in European children (1978–1997): Report from the Automated Childhood Cancer Information System project. European Journal of Cancer. 2006;42(13):2081-91.

41. Pastore G, Znaor A, Spreafico F, Graf N, Pritchard-Jones K, Steliarova-Foucher E. Malignant renal tumours incidence and survival in European children (1978–1997): Report from the Automated Childhood Cancer Information System project. European Journal of Cancer. 2006;42(13):2103-14.

42. Arora RS, Alston RD, Eden TOB, Moran A, Geraci M, O'Hara C, et al. Cancer at ages 15-29 years: the contrasting incidence in India and England. Pediatric Blood & Cancer. 2012;58(1):55-60.

43. Lewis I, Lenehan C. Report of the Children and Young People's Health Outcomes Forum. 2012.

44. DEPARTMENT OF HEALTH. The Children and Young People’s Health Outcomes Forum In: Health Do, editor. 2011.

45. Estlin EG, R. Wynn, R., editor. Pediatric hematology and oncology : scientific principles and clinical practice: Oxford : Wiley-Blackwell, ; 2010.

46. Bleyer A, Whiteson, M., Eden, T., Barr, R., editor. Cancer and the Adolescent. 2nd ed: Malden, Mass. : Blackwell, 2005; 2005.

47. Powell JE, Esteve J, Mann JR, Parker L, Frappaz D, Michaelis J, et al. Neuroblastoma in Europe: differences in the pattern of disease in the UK. SENSE. Study group for the Evaluation of Neuroblastoma Screening in Europe. Lancet. 1998;352(9129):682-7.

48. Hachitanda Y, Ishimoto K, Hata J, Shimada H. One hundred neuroblastomas detected through a mass screening system in Japan. Cancer. 1994;74(12):3223-6.

49. Ishimoto K, Kiyokawa N, Fujita H, Yabuta K, Ohya T, Miyano T, et al. Problems of mass screening for neuroblastoma: analysis of false-negative cases. Journal of Pediatric Surgery. 1990;25(4):398-401.

50. Kerbl R, Urban CE, Ambros IM, Dornbusch HJ, Schwinger W, Lackner H, et al. Neuroblastoma mass screening in late infancy: insights into the biology of neuroblastic tumors. J Clin Oncol. 2003;21(22):4228-34.

51. Schilling FH, Berthold F, Erttmann R, Michaelis J, Spix C, Sander J, et al. Population-based and controlled study to evaluate neuroblastoma screening at one year of age in Germany: interim results. Medical & Pediatric Oncology. 2000;35(6):701-4.

52. Woods WG, Gao R-N, Shuster JJ, Robison LL, Bernstein M, Weitzman S, et al. Screening of infants and mortality due to neuroblastoma. New England Journal of Medicine. 2002;346(14):1041-6.

53. Pastore G, Peris-Bonet R, Carli M, Martínez-García C, de Toledo JS, Steliarova-Foucher E. Childhood soft tissue sarcomas incidence and survival in European children (1978–1997): Report from the Automated Childhood Cancer Information System project. European Journal of Cancer. 2006;42(13):2136-49.

54. Bretherick KL, Bu R, Gascoyne RD, Connors JM, Spinelli JJ, Brooks-Wilson AR. Elevated circulating t(14;18) translocation levels prior to diagnosis of follicular lymphoma. Blood. 2010;116(26):6146-7.

55. Ferry JA. Burkitt's lymphoma: clinicopathologic features and differential diagnosis. Oncologist. 2006;11(4):375-83.

56. Morley-Jacob C, Gallop-Evans E. An update on lymphoma in children and young adults: Paediatrics and Child Health. 22 (3) (pp 92-97), 2012. Date of Publication: March 2012.; 2012.

57. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, et al. The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathologica. 2007;114(2):97-109.

58. Arora RS, Alston RD, Eden TOB, Estlin EJ, Moran A, Birch JM. Age-incidence patterns of primary CNS tumors in children, adolescents, and adults in England. Neuro-oncol. 2009;11(4):403-13.

59. McKinney PA. Central nervous system tumours in children: epidemiology and risk factors. Bioelectromagnetics. 2005;Suppl 7:S60-8.

60. Packer RJ, Vezina G. Pediatric glial neoplasms including brain-stem gliomas. Seminars in Oncology. 1994;21(2):260-72.

61. Glauser TA, Packer RJ. Cognitive deficits in long-term survivors of childhood brain tumors. Childs Nervous System. 1991;7(1):2-12.

62. Robinson KE, Kuttesch JF, Champion JE, Andreotti CF, Hipp DW, Bettis A, et al. A quantitative meta-analysis of neurocognitive sequelae in survivors of pediatric brain tumors. Pediatric Blood & Cancer. 2010;55(3):525-31.

63. Bielack S, Carrle D, Casali PG, Group EGW. Osteosarcoma: ESMO clinical recommendations for diagnosis, treatment and follow-up. Ann Oncol. 2009;20 Suppl 4:137-9.

64. Ladenstein R, Potschger U, Le Deley MC, Whelan J, Paulussen M, Oberlin O, et al. Primary disseminated multifocal Ewing sarcoma: results of the Euro-EWING 99 trial. J Clin Oncol. 2010;28(20):3284-91.

65. Pritchard-Jones K, Kaatsch P, Steliarova-Foucher E, Stiller CA, Coebergh JWW. Cancer in children and adolescents in Europe: developments over 20 years and future challenges. European Journal of Cancer. 2006;42(13):2183-90.

66. Fern LA, Campbell C, Eden TO, Grant R, Lewis I, Macleod U, et al. How frequently do young people with potential cancer symptoms present in primary care?.[Erratum appears in Br J Gen Pract. 2011 Jun;61(587):382]. Br J Gen Pract. 2011;61(586):e223-30.

67. Stiller CA, Desandes E, Danon SE, Izarzugaza I, Ratiu A, Vassileva-Valerianova Z, et al. Cancer incidence and survival in European adolescents (1978–1997). Report from the Automated Childhood Cancer Information System project. European Journal of Cancer. 2006;42(13):2006-18.

68. Bragonier R, Kenyon C. Two-week urgent referrals for suspected childhood cancer: experience within a large tertiary centre. Archives of Disease in Childhood. 2012;97(7):674.

69. Mant J, Nanduri V. Role of the 2-week urgent referral pathway in childhood cancer. Archives of Disease in Childhood. 2012;97(3):233-5.

70. Cheung CRLH, Gray JAM. Unwarranted variation in health care for children and young people. Archives of Disease in Childhood. 2013;98(1):60-5.

71. NCIN. Routes to Diagnosis 2006-2008: Technical Document. London: National Cancer Intelligence Network; 2010.

72. Dommett RM, Redaniel MT, Stevens MCG, Hamilton W, Martin RM. Features of childhood cancer in primary care: A population-based nested case-control study. Br J Cancer. 2012;106(5):982-7.

73. Neal RD. Do diagnostic delays in cancer matter? Br J Cancer. 2009;101(Supl 2):S9-S12.

74. Macleod U, Mitchell ED, Burgess C, Macdonald S, Ramirez AJ. Risk factors for delayed presentation and referral of symptomatic cancer: evidence for common cancers. Br J Cancer. 2009;101 Suppl 2:S92-S101.

75. Mitchell E, Macdonald S, Campbell NC, Weller D, Macleod U. Influences on pre-hospital delay in the diagnosis of colorectal cancer: A systematic review. Br J Cancer. 2008;98(1):60-70.

76. Ramirez AJ, Westcombe AM, Burgess CC, Sutton S, Littlejohns P, Richards MA. Factors predicting delayed presentation of symptomatic breast cancer: a systematic review. Lancet. 1999;353(9159):1127-31.

77. Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ. Influence of delay on survival in patients with breast cancer: a systematic review. Lancet. 1999;353(9159):1119-26.

78. Dang-Tan T, Franco EL. Diagnosis delays in childhood cancer: A review. Cancer. 2007;110(4):703-13.

79. Brasme J-F, Morfouace M, Grill J, Martinot A, Amalberti R, Bons-Letouzey C, et al. Delays in diagnosis of paediatric cancers: a systematic review and comparison with expert testimony in lawsuits. Lancet Oncology. 2012;13(10):e445-59.

80. Wilne S, Collier J, Kennedy C, Koller K, Grundy R, Walker D. Presentation of childhood CNS tumours: a systematic review and meta-analysis. Lancet Oncology. 2007;8(8):685-95.

81. Macdonald S, Macleod U, Campbell NC, Weller D, Mitchell E. Systematic review of factors influencing patient and practitioner delay in diagnosis of upper gastrointestinal cancer. [Review] [36 refs]. Br J Cancer. 2006;94(9):1272-80.

82. Ferrari A, Miceli R, Casanova M, Meazza C, Favini F, Luksch R, et al. The symptom interval in children and adolescents with soft tissue sarcomas. Cancer. 2010;116(1):177-83.

83. Goyal S, Roscoe J, Ryder WDJ, Gattamaneni HR, Eden TOB. Symptom interval in young people with bone cancer. Eur J Cancer. 2004;40(15):2280-6.

84. Halperin EC, Watson DM, George SL. Duration of symptoms prior to diagnosis is related inversely to presenting disease stage in children with medulloblastoma. Cancer. 2001;91(8):1444-50.

85. Kukal K, Dobrovoljac M, Boltshauser E, Ammann RA, Grotzer MA. Does diagnostic delay result in decreased survival in paediatric brain tumours? Eur J Pediatr. 2009;168(3):303-10.

86. Saha V, Love S, Eden T, Micallef-Eynaud P, MacKinlay G. Determinants of symptom interval in childhood cancer. Arch Dis Child. 1993;68(6):771-4.

87. Dang-Tan T, Trottier H, Mery LS, Morrison HI, Barr RD, Greenberg ML, et al. Delays in diagnosis and treatment among children and adolescents with cancer in Canada. Pediatr Blood Cancer. 2008;51(4):468-74.

88. Shay V, Fattal-Valevski A, Beni-Adani L, Constantini S. Diagnostic delay of pediatric brain tumors in Israel: A retrospective risk factor analysis. Child's Nervous System. 2011;28(1):93-100.

89. Bai S, Ren R, Li B, Xu X, Zhao B, Gao F, et al. Delay in the diagnosis of retinoblastoma in China. Acta Ophthalmol. 2011;89(1):e72-e4.

90. Haimi M, Perez-Nahum M, Stein N, Ben Arush MW. The role of the doctor and the medical system in the diagnostic delay in pediatric malignancies. Cancer Epidemiology. 2011;35(1):83-9.

91. James BO, Ajayi SO, Ogun OA, Oladokun RE. Factors influencing time to diagnosis of childhood cancer in Ibadan, Nigeria. African Health Sciences. 2009;9(4):247-53.

92. Rodrigues KES, Latorre MDRDO, De Camargo B. Delayed diagnosis in retinoblastoma. Jornal de Pediatria. 2004;80(6):511-6.

93. Cecen E, Gunes D, Mutafoglu K, Sarialioglu F, Olgun N. The time to diagnosis in childhood lymphomas and other solid tumors. Pediatr Blood Cancer. 2011;57(3):392-7.

94. Edgeworth J, Bullock P, Bailey A, Gallagher A, Crouchman M. Why are brain tumours still being missed? Arch Dis Child. 1996;74(2):148-51.

95. Stefan DC, Siemonsma F. Delay and causes of delay in the diagnosis of childhood cancer in Africa. Pediatr Blood Cancer. 2011;56(1):80-5.

96. Loh AHP, Aung L, Ha C, Tan AM, Quah TC, Chui CH. Diagnostic delay in pediatric solid tumors: A population based study on determinants and impact on outcomes. Pediatr Blood Cancer. 2012;58(4):561-5.

97. Brasme JF, Chalumeau M, Doz F, Lacour B, Valteau-Couanet D, Gaillard S, et al. Interval between onset of symptoms and diagnosis of medulloblastoma in children: Distribution and determinants in a population-based study. Eur J Pediatr. 2012;171(1):25-32.

98. Hayashi N, Kidokoro H, Miyajima Y, Fukazawa T, Natsume J, Kubota T, et al. How do the clinical features of brain tumours in childhood progress before diagnosis? Brain Dev. 2010;32(8):636-41.

99. Klitbo DM, Nielsen R, Illum NO, Wehner PS, Carlsen N. Symptoms and time to diagnosis in children with brain tumours. Danish Medical Bulletin. 2011;58(7):1-5.

100. Goddard AG, Kingston JE, Hungerford JL. Delay in diagnosis of retinoblastoma: Risk factors and treatment outcome. Br J Ophthalmol. 1999;83(12):1320-3.

101. Thulesius H, Pola J, Hakansson A. Diagnostic delay in pediatric malignancies: A population-based study. Acta Oncol. 2000;39(7):873-6.

102. Chotel F, Unnithan A, Chandrasekar CR, Parot R, Jeys L, Grimer RJ. Variability in the presentation of synovial sarcoma in children: A plea for greater awareness. J Bone Joint Surg Br. 2008;90(8):1090-6.

103. Yang JYK, Cheng FWT, Wong KC, Lee V, Leung WK, Shing MMK, et al. Initial presentation and management of osteosarcoma, and its impact on disease outcome. Hong Kong Med. 2009;15(6):434-9.

104. Crawford JR, Santi MR, Vezina G, Myseros JS, Keating RF, LaFond DA, et al. CNS germ cell tumor (CNSGCT) of childhood: Presentation and delayed diagnosis. Neurology. 2007;68(20):1668-73.

105. Flores LE, Williams DL, Bell BA, O'Brien M, Ragab AH. Delay in the diagnosis of pediatric brain tumors. Am J Dis Child. 1986;140(7):684-6.

106. Mehta V, Chapman A, McNeely PD, Walling S, Howes WJ, Sutton LN, et al. Latency between symptom onset and diagnosis of pediatric brain

tumors: An Eastern Canadian geographic study. Neurosurgery. 2002;51(2):365-73.

107. Reulecke BC, Erker CG, Fiedler BJ, Niederstadt TU, Kurlemann G. Brain tumors in children: Initial symptoms and their influence on the time span between symptom onset and diagnosis. J Child Neurol. 2008;23(2):178-83.

108. Wilne S CJ, Kennedy C, Jenkins A, Grout J, Mackie S, Koller K, Grundy R, Walker D. Progression from first symptom to diagnosis in childhood brain tumours. Eur J Pediatr. 2011;171(1):87-93.

109. Wallach M, Balmer A, Munier F, Houghton S, Pampallona S, Von Der Weid N, et al. Shorter time to diagnosis and improved stage at presentation in Swiss patients with retinoblastoma treated from 1963 to 2004. Pediatrics. 2006;118(5):e1493-e8.

110. Wirix M, Parys-Vanginderdeuren R, Casteels I, Uyttebrouck A. Delayed diagnosis of retinoblastoma. Bull Soc Belge Ophtalmol. 2000(278):37-41.

111. LaQuaglia MP, Heller G, Filippa DA, Karasakalides A, Vlamis V, Wollner N, et al. Prognostic factors and outcome in patients 21 years and under with colorectal carcinoma. J Pediatr Surg. 1992;27(8):1085-90.

112. Schnurr C, Pippan M, Stuetzer H, Delank KS, Michael JWP, Eysel P. Treatment delay of bone tumours, compilation of a sociodemographic risk profile: A retrospective study. BMC Cancer. 2008;8(22):1-10.

113. OECD. OECD Better Policies for Better Lives. 2013 [1st July 2013]; Available from: http://www.oecd.org/about/membersandpartners/.

114. Dang-Tan T, Trottier H, Mery LS, Morrison HI, Barr RD, Greenberg ML, et al. Determinants of delays in treatment initiation in children and adolescents diagnosed with leukemia or lymphoma in Canada. International Journal of Cancer. 2010;126(8):1936-43.

115. Feltbower RG, Lewis IJ, Picton S, Richards M, Glaser AW, Kinsey SE, et al. Diagnosing childhood cancer in primary care--a realistic expectation? Br J Cancer. 2004;90(10):1882-4.

116. Wilne S, Liu J, Clough L, Dudley J, Lakhanpaul M, Kennedy C, et al. P01 A Service Evaluation of Implementation of RCPCH Brain Pathways Clinical Guideline Linked to HeadSmart – Be Brain Tumour Aware (HeadSmart). A Health Foundation, Closing the Gap Project. Archives of Disease in Childhood. 2013;98(Suppl 1):A1.

117. Dommett RM, Redaniel MT, Stevens MCG, Hamilton W, Martin RM. Features of cancer in teenagers and young adults in primary care: a population-based nested case-control study. Br J Cancer. 2013;108(11):2329-33.

118. Torring ML, Frydenberg M, Hansen RP, Olesen F, Hamilton W, Vedsted P. Time to diagnosis and mortality in colorectal cancer: a cohort study in primary care.[Erratum appears in Br J Cancer. 2011 Jun 7;104(12):1930]. Br J Cancer. 2011;104(6):934-40.

119. National Patient Safety Agency. Delayed diagnosis of cancer: Thematic review. National Health Service; 2010.

120. Webb P, Bain C. Essential Epidemiology: An Introduction for Students and Health Professionals 2nd ed: Cambridge University Press; 2011.

121. National Health Service, The Information Centre for Health and Social Care. HES online - About HES - Background. 2012; Available from: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=457.

122. DEPARTMENT OF HEALTH 2011. Payment by results - Background. 2010; Available from: http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Managingyourorganisation/NHSFinancialReforms/DH_077259.

123. United Kingdom Association of Cancer Registries (UKACR). Release of potentially identifiable patient information. 2013; Available from: http://www.ukcancassoc.ismysite.co.uk/content/release-potentially-identifiable-patient-information?phpMyAdmin=628c0920c0ef2bb3163c7b56b06f3b9c.

124. National Cancer Action Team. National Cancer Action Team part of the National Cancer Programme - Ensuring Better Treatment - Cancer Waiting Times. 2013 [1st July 2013]; Available from: http://ncat.nhs.uk/our-work/ensuring-better-treatment/cancer-waiting-times#.

125. National Institute for Health and Clinical Excellence (NICE). Referral guidelines for suspected cancer. 2005.

126. MacMillan Cancer Support. Types of children's cancers 2013; Available from: http://www.macmillan.org.uk/Cancerinformation/Cancertypes/Childrenscancers/Typesofchildrenscancers/Typesofchildrenscancers.aspx.

127. Bone Cancer Research Trust. BCRT Be Aware. 2010 [1st July 2013]; Available from: http://www.bcrt.org.uk/bci_be_symptom_aware.php.

128. StataCorp. Stata Statistical Software: Release 12. College Station, TX, USA: Stata Corp LP; 2011.

129. Thomson CS, Forman D. Cancer survival in England and the influence of early diagnosis: what can we learn from recent EUROCARE results? Br J Cancer. 2009;101 Suppl 2:S102-9.

130. Bland M. An Introduction to medical statistics. Third ed: Oxford University Press; 2000.

131. Andersohn F, Konzen C, Garbe E. Systematic review: agranulocytosis induced by nonchemotherapy drugs. Ann Intern Med. 2007;146(9):657-65.

132. Ahrensberg JM, Fenger-Gron M, Vedsted P. Use of Primary Care during the Year before Childhood Cancer Diagnosis: A Nationwide Population-Based Matched Comparative Study. PLoS ONE. 2013;8(3):e59098.

133. Stiller C, Kroll M, Eatock E. 5 Survival from Childhood Cancer. In: Stiller C, editor. Childhood Cancer in Britain: Incidence, Survival, Mortality, Volume 1: Oxford University Press; 2007.

134. Fern L, Davies S, Eden T, Feltbower R, Grant R, Hawkins M, et al. Rates of inclusion of teenagers and young adults in England into National Cancer Research Network clinical trials: report from the National Cancer Research Institute (NCRI) Teenage and Young Adult Clinical Studies Development Group. 2008.

135. Goulden N, editor. UKALL 2011 Trial - United Kingdom National Randomised Trial For Children and Young Adults with Acute Lymphoblastic Leukaemia and Lymphoma 2011: Cancer Research UK Clinical Trials Unit; 2011.

136. Cancer Research UK. Our campaign to encourage earlier diagnosis of cancer. 2013; Available from: http://scienceblog.cancerresearchuk.org/2013/07/02/our-campaign-to-encourage-earlier-diagnosis-of-cancer/.

137. UK Clinical Trials Research Network. UKALL 2013. 2013; Available from: http://public.ukcrn.org.uk/search/StudyDetail.aspx?StudyID=11319.

Appendix 1 - NICE Guideline for suspected cancer in children and young adults (125)

General		
1	CYA presenting with symptoms and signs of cancer should be referred to a paediatrician or a specialist children's cancer service	
2	CYA presenting several times with the same problems, but with no clear diagnosis, urgent referral should be made	
3	The parent is the best observer of the CYA's symptoms, as such parental insight should be taken into consideration when considering urgent referral	
4	Persistent parental anxiety should be a sufficient reason for referral	
5	Persistent back pain in CYA should prompt urgent examination, investigation with FBC, film and consider referral	
6	Associations such as Down's and leukaemia, NF and CNS tumours and other rare syndromes and some cancers should alert to a potential diagnosis in CYA patients with unexplained symptoms	
7	Primary HC prof should convey info to parents and CYA about reason for specialist referral	
8	Primary HC prof should develop good communication with parents and CYA to provide a supportive care relationship	
Leukaemia		
9	Has a relatively short history in weeks rather than months.	
	Presence of >=1 symp & sign should prompt FBC & film	Pallor
		Fatigue
		Unexplained irritability
		Unexplained fever
		Persistent or recurrent URTI
		Generalised lymphadenopathy
		Persitent or unexplained bone pain
		Unexplained Bruising
10	Immediate referral in the presence of either of the following	Unexplained petechiae
		Hepatomegaly
Lymphoma		
11	Typically present with non-tender cervical +/or suprclavicular lymphadenopathy.	
	The natural history is long except in NHL	
	NHL may present with lymphadenopathy, breathlessness, SVC obst, abdo distension	
	Lymphadenopathy in the presence of no evidence of local infection	LN are non-tender, firm ot hard
		LN >2cm
		LN progressively enlarging

			Other features of general ill health, fever or weight-loss
			Axillary nodes (in theabsence of local infection or dermatitis)
			Supraclavicular nodes
12	The presence of hepatosplenomegaly requires immediate referral		
13	SOB can indicate chest involvement and if assoc with above symp should prompt urgent referral		
14	CYA with mediastinal or hilar mass on CXR should be referred immediateley		
CNS >=2			
15	Persistent headache in CYA requires primary HC prof to conduct urgent exam and unable to undertake adequate exam they should refer		
16	Headache and vomiting that occur at or lead to early morning waking require immediate referral		
17	The presence of the following neuro symp and signs should prompt urgent referral		New-onset seizures
			Cranial nerve abnormalities
			Visual disturbance
			Gait abnormalities
			Motor or sensory
			Unexplained deterioration in school performance of developmental milestones
			Unexplained behaviour +/or mood changes
18	CYA with reduced level of consciousness requires immediate referral		
CNS <2			
19	Symp & signs prompting referral	Immediate	New-onset seizures
			Bulging fontanelles
			Extensor attacks
			Persistent vomiting
		Urgent	Abnormal increase in head-size
			Arrest or regression of motor development
			Altered behaviour
			Abnormal eye movements
			Lack of visual following
			Poor feeding/FTT
		Urgent contingent on other factors	Squint
Neuroblastoma			
20	Most CYA with neuroblastoma have symptoms of metastatic disease at presentation, symp & signs similar to 9 should prompt FBC & film		

21	Other symp which should raise concern		Proptosis
			Unexplained back pain
			Leg weakness
			Unexplained urinary retention
22	If symp suggest neuroblastoma an abdo exam, CXR and FBC should be considered		
23	Immediate referral in children <1 with localized abdo or thoracic masses, some babies have skin lesions		
Wilms' Tumours			
24	Commonly present with painless masses, urgent referral in the presence of persistent or progressive distension		
25	Haematuria in CYA, although a rarer presentation should prompt urgent referral		
Soft Tissue Sarcoma			
26	CYA with an unexplained mass at any site with one or more of the following		Deep to fascia
			Non-tender
			Progressively enlarging
			Associated with a regional LN that is enlarging
			>2cm in diameter
27	Unusual Location	Head & Neck	Proptosis
			Persistent unexplained unilateral nasal obstruction +/- discharge +/or bleeding
			Aural polyps/discharge
		Genitourinary	Urinary retention
			Scrotal swelling
			Bloodstained vaginal discharge
Bone Sarcoma			
28	Limbs are most common site, especially knees in the case of osteosarcoma, persistent localized pain +/or swelling requires and X-ray and if tumour suspected an urgent referral		
29	History of an injury should not exclude the possibility of a bone tumour		
30	Rest pain, back pain or unexplained limp may all point to a bone tumour		
Retinoblastoma			
31	Leukocoria spotted by parent, in photo or on exam should prompt urgent referral		
32	A new squint or change in visual acuity should be referred, if cancer suspected urgently		
33	A family history of retinoblastoma in a child presenting with eye signs should prompt the primary HC prof to ? retinoblastoma. Offspring of an affected parent or siblings to affected child should be screened from birth		

Table adapted from the National Institute for Health and Clinical Excellence (NICE). Referral guidelines for suspected cancer. 2005

Appendix 2 – Broad alert codes

D689	**Coagulation defect, unspecified**
G952	**Cord compression, unspecified**
I871	**Compression of vein**
K590	**Constipation**
K625	**Haemorrhage of anus and rectum**
K920	**Haematemesis**
K921	**Melaena**
M254	**Effusion of joint**
M255	**Pain in joint**
M436	**Torticollis**
M542	**Cervicalgia**
M545	**Low back pain**
M549	**Dorsalgia, unspecified**
M796	**Pain of limb**
M844	**Pathological fracture, Not elsewhere classified**
N62X	**Hypertrophy of breast**
Q850	**Neurofibromatosis (nonmalignant)**
R040	**Epistaxis**
R042	**Haemoptysis**
R05X	**Cough**
R060	**Dyspnoea**

R065	**Abnormalities of breathing - mouth breathing (snoring)**
R068	**Other and unspecified abnormalities of breathing**
R073	**Other chest pain**
R074	**Chest pain, unspecified**
R101	**Pain localized to upper abdomen**
R103	**Pain localized to other parts of lower abdomen**
R104	**Abdominal & pelvic pain - Other & UN**
R11X	**Nausea and vomiting**
R190	**Intra-abdominal & pelvic swelling, mass & lump**
R220	**Localized swelling, mass and lump, head**
R221	**Localized swelling, mass and lump, neck**
R222	**Localized swelling, mass and lump, trunk**
R227	**Localized swelling, mass and lump, multiple sites**
R229	**Localized swelling, mass and lump, unspecified**
R233	**Spontaneous ecchymoses**
R31X	**Unspecified haematuria**
R509	**Fever, PUO**
R51X	**Headache**
R521	**Chronic intractable pain**
R53X	**Malaise and fatigue**
R590	**Localized enlarged lymph nodes**
R591	**Enlarged lymph nodes - Generalised**
R633	**Feeding difficulties and mismanagement**
R634	**Abnormal weight loss**
R69X	**Unknown & unspecified causes of morbidity**

R900	**Abnormality of diagnostic imaging - Intracranial SOL**
Y431	**Antineoplastic antimetabolites**
Y433	**Other antineoplastic drugs**
Z031	**Observation for suspected malignant neoplasm**
Z112	**Special screening examination for neoplasm of respiratory organs**
Z115	**Special screening examination for neoplasm of prostate**
Z128	**Special screening examination for neoplasms of other sites**
Z129	**Special screening examination neoplasm, UN**
Z510	**Radiotherapy session**
Z511	**Chemotherapy session for neoplasm**
Z800	**Family history of malignant neoplasm of digestive organs**
Z801	**Family history of malignant neoplasm of trachea, bronchus and lung**
Z803	**Family history of malignant neoplasm of breast**
Z806	**Family history of leukaemia**
Z807	**Family history of other malignant neoplasms of lymphoid, haematopoietic and related tissues**
Z808	**Family history of malignant neoplasm of other organs or systems**
Z834	**Family history of other endocrine, nutritional and metabolic diseases**
Z856	**Personal history of leukaemia**
Z858	**Personal history of malignant neoplasms of other organs and systems**
Z877	**Personal history of congenital malformations, deformations and chromosomal abnormalities**

Appendix 3 – Specific alert codes by International Classification of Childhood Cancer group

Leukaemia

D471	**Refractory anaemia with excess of blasts**
D610	**Constitutional aplastic anaemia**
D619	**Aplastic anaemia, unspecified**
D630	**Anaemia in neoplastic disease**
D649	**Anaemia, unspecified**
D696	**Thrombocytopenia, unspecified**
D699	**Haemorrhagic condition, unspecified**
D70X	**Agranulocytosis**
D721	**Eosinophilia**
J351	**Hypertrophy of tonsils**
J353	**Hypertrophy of tonsils with adeniods**
K921	**Melaena**
K922	**Gastrointestinal haemorrhage, unspecified**
M255	**Pain in joint**
M545	**Low back pain**
M549	**Dorsalgia, unspecified**
M796	**Pain of limb**
Q909	**Down's syndrome, unspecified**
R160	**Hepatomegaly, not elsewhere classified**
R161	**Splenomegaly, not elsewhere classified**
R162	**Hepatomegaly with splenomegaly, not elsewhere classified**

R233	**Spontaneous ecchymoses**
R53X	**Malaise and fatigue**
R72X	**Abnormality of white blood cells, not elsewhere classified**
R91X	**Abnormal findings on diagnostic imaging of lung**
T810	**Haemorrhage and haematoma complicating a procedure, not elsewhere classified**
Y420	**Glucocorticoids and synthetic analogues**
Y430	**Antiallergic and antiemetic drugs**

Lymphoma

D70X	**Agranulocytosis**
D728	**Other specified disorders of white blood cells**
D803	**Selective deficiency of immunoglobulin G [IgG] subclasses**
D804	**Selective deficiency of immunoglobulin M [IgM]**
D823	**Immunodeficiency following hereditary defective response to Epstein-Barr virus**
D831	**Common variable immunodeficiency with predominant immunoregulatory T-cell disorders**
D839	**Common variable immunodeficiency, unspecified**
D849	**Immunodeficiency, unspecified**
I871	**Compression of vein**
J929	**Pleural plaque without asbestos**
J985	**Diseases of mediastinum, not elsewhere classified**
R060	**Dyspnoea**
R061	**Stridor**
R160	**Hepatomegaly, not elsewhere classified**
R161	**Splenomegaly, not elsewhere classified**
R162	**Hepatomegaly with splenomegaly, not elsewhere classified**
R190	**Intra-abdominal & pelvic swelling, mass & lump**
R221	**Localized swelling, mass and lump, neck**
R222	**Localized swelling, mass and lump, trunk**
R224	**Localized swelling, mass and lump, lower limb**
R229	**Localized swelling, mass and lump, unspecified**
R630	**Anorexia**

R634	**Abnormal weight loss**
R91X	**Abnormal findings on diagnostic imaging of lung**
R932	**Abnormal findings on diagnostic imaging of liver and biliary tract**
R933	**Abnormal findings on diagnostic imaging of other parts of digestive tract**
R938	**Abnormal findings on diagnostic imaging of other specified body structures - Skin, Subcut, mediastinal shift**

CNS tumours

E230	**Hypopituitarism**
E232	**Diabetes insipidus**
E309	**Disorder of puberty, unspecified**
G400	**Localization-related (focal)(partial) idiopathic epilepsy and epileptic syndromes with seizures of localized onset**
G401	**Localization-related (focal)(partial) symptomatic epilepsy and epileptic syndromes with simple partial seizures**
G403	**Generalized idiopathic epilepsy and epileptic syndromes**
G409	**Epilepsy UN**
G431	**Migraine with aura [classical migraine]**
G439	**Migraine, unspecified**
G442	**Tension-type headache**
G500	**Trigeminal neuralgia**
G510	**Bell's palsy**
G810	**Flaccid hemiplegia**
G819	**Hemiplegia, unspecified**
G823	**Flaccid tetraplegia**
G941	**Hydrocephalus in neoplastic disease**
H471	**Papilloedema, unspecified**
H490	**Third [oculomotor] nerve palsy**
H492	**Paralytic strabismus - 6th N palsy**
H501	**Divergent concomitant strabismus**
H509	**Strabismus, unspecified**
H532	**Diplopia**
H534	**Visual field defects**

H538	**Other visual disturbances**
H540	**Blindness, binocular**
H547	**Visual impairment including blindness (binocular or monocular)**
H55X	**Nystagmus and other irregular eye movements**
L813	**Café au lait spots**
M415	**Other secondary scoliosis**
M625	**Muscle wasting and atrophy, not elsewhere classified**
N319	**Neuromuscular dysfunction of bladder, unspecified**
Q753	**Macrocephaly**
Q850	**Neurofibromatosis (nonmalignant)**
Q851	**Tuberous sclerosis**
Q878	**Other specified congenital malformation syndromes, not elsewhere classified**
Q998	**Other specified chromosome abnormalities**
R202	**Paraesthesia of skin**
R208	**Other and unspecified disturbances of skin sensation**
R251	**Tremor, unspecified**
R253	**Fasciculation**
R258	**Other and unspecified abnormal involuntary movements**
R268	**Abnormalities of gait & mobility - Unsteadiness on feet**
R270	**Ataxia, unspecified**
R278	**Other and unspecified lack of coordination**
R401	**Stupor**
R410	**Disorientation, unspecified**
R42X	**Dizziness and giddiness**
R470	**Dysphasia and aphasia**
R51X	**Headache**
R55X	**Syncope and collapse**
R568	**Convulsion, UN**
R628	**Other lack of expected normal physiological development**
R629	**Lack of expected normal physiological development, unspecified**

R633	**Feeding difficulties and mismanagement**
R900	**Abnormality of diagnostic imaging - Intracranial SOL**
R930	**Abnormal findings on diagnostic imaging of skull and head, not elsewhere classified**
R933	**Abnormal findings on diagnostic imaging of skull and head, not elsewhere classified**

Neuroblastoma

G253	**Myoclonus**
H052	**Exophthalmic conditions**
H492	**Paralytic strabismus - 6th N palsy**
H500	**Convergent concomitant strabismus**
H509	**Strabismus, unspecified**
K590	**Constipation**
R061	**Stridor**
R13X	**Dysphagia**
R18X	**Ascites**
R190	**Intra-abdominal & pelvic swelling, mass & lump**
R229	**Localized swelling, mass and lump, unspecified**
R53X	**Malaise and fatigue**
R681	**Nonspecific symptoms peculiar to infancy**
R91X	**Abnormal findings on diagnostic imaging of lung**
R934	**Abnormal findings on diagnostic imaging of urinary organs**
R935	**Abnormal findings on diagnostic imaging of other abdominal regions, including retroperitoneum**
R937	**Abnormal findings on diagnostic imaging of other parts of MSK**

Retinoblastoma

H102	**Other acute conjunctivitis**
H118	**Other specified disorders of conjunctiva**
H409	**Glaucoma, unspecified**
H509	**Strabismus, unspecified**

Renal tumours

I10X	**Essential (primary) hypertension**
I120	**Hypertensive renal disease with renal failure**

M545	**Low back pain**
R103	**Pain localized to other parts of lower abdomen**
R104	**Abdominal & pelvic pain - Other & UN**
R190	**Intra-abdominal & pelvic swelling, mass & lump**
R222	**Localized swelling, mass and lump, trunk**
R31X	**Unspecified haematuria**
Z805	**Family history of malignant neoplasm of urinary tract**

Hepatic tumours

R160	**Hepatomegaly, not elsewhere classified**
R161	**Splenomegaly, not elsewhere classified**
R18X	**Ascites**

Malignant bone tumours

M255	**Pain in joint**
M545	**Low back pain**
M796	**Pain of limb**
M844	**Pathological fracture, Not elsewhere classified**
M895	**Osteolysis**
M899	**Disorder of bone UN**
M907	**Fracture of bone in neoplastic disease**
R268	**Abnormalities of gait & mobility - Unsteadiness on feet**
R509	**Fever, PUO**
R936	**Abnormal diagnostic imaging of limbs**
R937	**Abnormal findings on diagnostic imaging of other parts of MSK**
R938	**Abnormal findings on diagnostic imaging of other specified body structures - Skin, Subcut, mediastinal shift**
S568	**Injury to other & unspecified muscles & tendons of forearm level**
S724	**Fractures of lower end of femur**
W008	**Fall on same level invloving ice or snow**
W100	**Fall on & from stairs & steps**
W192	**Unspecified Fall**
W213	**Striking against or struck by sports equipment**

Z016	**Radiological examination, not elsewhere classified - Routine CXR, mammogram**

Soft-tissue sarcoma

G510	**Bell's palsy**
H024	**Ptosis of eyelid**
H653	**Chronic muciod otitis media**
H744	**Polyp of middle ear**
N328	**Other specified disorders of bladder**
N508	**Other specified disorders of male genital organs**
N509	**Disorder of male genital organs, unspecified**
N938	**Other specified abnormal uterine and vaginal bleeding**
N939	**Abnormal uterine and vaginal bleeding, unspecified**
R190	**Intra-abdominal & pelvic swelling, mass & lump**
R220	**Localized swelling, mass and lump, head**
R222	**Localized swelling, mass and lump, trunk**
R33X	**Retention of urine**
R599	**Enlarged lymph nodes, unspecified**
R934	**Abnormal findings on diagnostic imaging of urinary organs**

Germ-cell tumours

H474	**Disorders of optic chiasm**
H539	**Visual disturbance, unspecified**
H545	**Severe visual impairment, monocular**
N44X	**Torsion of testis**
N508	**Other specified disorders of male genital organs**
N832	**Other and unspecified ovarian cysts**
N939	**Abnormal uterine and vaginal bleeding, unspecified**
R190	**Intra-abdominal & pelvic swelling, mass & lump**
R31X	**Unspecified haematuria**

Carcinoma

H653	**Chronic muciod otitis media**
H919	**Hearing Loss UN**

H921	**Otorrhoea**
J352	**Hypertrophy of adenoids**
L040	**Acute lymphadenitis of face, head and neck**
R221	**Localized swelling, mass and lump, neck**

Appendix 4 – Stata do files

```
****** INCLUSION OF CASES BY YEAR OF DIAGNOSIS ******

gen Inclusion=1 if (( d_diag1>=td(01Jan2004)) & ( d_diag1<td(31Dec 2009)))
keep if Inclusion==1

****** INCLUSION BY AGE ******

gen Age= ((d_diag1 - dob_yctr)/365.25)
keep if Age <=25
gen Age_group=0 if Age<15
replace  Age_group=1 if  Age>=15
gen age_ranges=5 if Age>=20
replace age_range=4 if Age<20
replace age_range=3 if Age<15
replace age_range=2 if Age<10
replace age_range=1 if Age<5

****** GENERATING ICCC GROUPS ******

gen icccgroup=1 if iccc_1<20
replace icccgroup=2 if iccc_1<30 & iccc_1>19
replace icccgroup=3 if iccc_1<40 & iccc_1>29
replace icccgroup=4 if iccc_1<50 & iccc_1>39
replace icccgroup=5 if iccc_1<60 & iccc_1>49
replace icccgroup=6 if iccc_1<70 & iccc_1>59
replace icccgroup=7 if iccc_1<80 & iccc_1>69
replace icccgroup=8 if iccc_1<90 & iccc_1>79
replace icccgroup=9 if iccc_1<100 & iccc_1>89
replace icccgroup=10 if iccc_1<110 & iccc_1>99
replace icccgroup=11 if iccc_1<120 & iccc_1>109
replace icccgroup=12 if iccc_1<130 & iccc_1>119

****** IDENTIFYING THE FIRST EPISODE ******
```

```
gen duration=(epistart-d_diag1)

sort yctr_id epistart

by yctr_id: gen epiindex=_n if duration<=0

gen first_epi=1 if epiindex==1

gen first_epi_date=epistart if first_epi==1

format  first_epi_date %td

by yctr_id: egen first_epi_date_pt=total(first_epi_date)

format first_epi_date_pt %td

replace first_epi_date_pt=. if first_epi_date_pt==0

by yctr_id:  egen total_epis1=count( first_epi_date_pt )


****** LABELLING EPISODES WITH A CANCER DIAGNOSIS CODE ******


gen Diag_Cancer = strpos( diag_01, "C")  | strpos( diag_02, "C") | strpos( diag_03, "C") | strpos( diag_04, "C")

| strpos( diag_05, "C") | strpos(diag_06, "C") | strpos( diag_07, "C") | strpos( diag_08, "C") | strpos( diag_09, "C")

| strpos( diag_10, "C") | strpos( diag_11, "C") | strpos( diag_12, "C") | strpos( diag_11, "C") | strpos( diag_13, "C")

| strpos( diag_14, "C") | strpos( diag_15, "C") | strpos( diag_16, "C") | strpos( diag_17, "C") | strpos( diag_18, "C")

| strpos( diag_19, "C") | strpos( diag_20, "C")


****** IDENTIFYING THE FIRST CANCER CODE EPISODE ******


sort yctr_id  Diag_Cancer epistart

bysort yctr_id Diag_Cancer: gen index=_n if duration<=0

gen first_cancer_epi=1 if index==1 &  Diag_Cancer==1

gen first_cancer_epi_date=epistart if first_cancer_epi==1

format first_cancer_epi_date %td

by yctr_id: egen first_cancer_epi_date_pt=total(first_cancer_epi_date)

format first_cancer_epi_date_pt %td

replace first_cancer_epi_date_pt=. if first_cancer_epi_date_pt==0

by yctr_id:  egen total_cancer_epis=total(Diag_Cancer)


****** LABELLING ADMISSION CODES ******


gen AdmiCode="L" if admimeth==11 | admimeth==12 | admimeth==13

replace AdmiCode="EA" if admimeth==21

replace AdmiCode="EG" if admimeth==22
```

```
replace AdmiCode="EO" if admimeth==24

replace AdmiCode="Eot" if admimeth==28 | admimeth==23

replace AdmiCode="M" if admimeth==31 | admimeth==32

replace AdmiCode="T" if admimeth==81

replace AdmiCode="B" if admimeth==82 | admimeth==83

replace AdmiCode="O" if admimeth==84 | admimeth==89

replace AdmiCode="N" if admimeth==98 | admimeth==99

gen Tag=1 if AdmiCode!="X"


****** IDENTIFYING ALERT CODES: ICCC GROUP 4 (EXAMPLE DO FILES, 1 OF 11 SPECIFIC AND 1 BROAD ALERT CODE DO FILE) ******


forvalues i=1(1)9{

rename diag_0`i' diag_`i'

}

forvalues i=1(1)20{

gen diag_`i'_4=substr(diag_`i',1,4)

}

gen A_diag_binary=0

foreach x of varlist diag_1_4 diag_2_4 diag_3_4 diag_4_4 diag_5_4 diag_6_4 diag_7_4 diag_8_4

diag_9_4 diag_10_4 diag_11_4 diag_12_4 diag_13_4 diag_14_4 diag_15_4 diag_16_4 diag_17_4 diag_18_4 diag_19_4 diag_20_4 {

#delimit ;

replace A_diag_binary=1 if (`x'=="G253" | `x'=="H052" | `x'=="H492" | `x'=="H500" | `x'=="H509" | `x'=="K590"

| `x'=="R061" | `x'=="R13X" | `x'=="R18X" | `x'=="R190" | `x'=="R229" | `x'=="R53X" | `x'=="R681"

| `x'=="R91X" | `x'=="R934" | `x'=="R935" | `x'=="R937" );

#delimit cr

}


****** THE PROCESS OF COUNTING THE NUMBER OF ALERT CODES PER CASE ******


forvalues i = 1(1)20{

      generate x`i' = ""

            foreach x of varlist diag_`i'_4 {

#delimit ;

                                    replace x`i'=`x' if (A_diag_binary==1 |  A_diag_binary==2) & (`x'=="G253" | `x'=="H052" | `x'=="H492"

| `x'=="H500" | `x'=="H509" | `x'=="K590" | `x'=="R061" | `x'=="R13X" | `x'=="R18X" | `x'=="R190" | `x'=="R229" | `x'=="R53X"
```

```
| `x'=="R681" | `x'=="R91X" | `x'=="R934" | `x'=="R935" | `x'=="R937" );

#delimit cr

}

foreach x of newlist G253 H052 H492 H500 H509 K590 R061 R13X R18X R190 R229 R53X R681 R91X R934 R935 R937{

gen `x'=0

}

foreach x of varlist G253 H052 H492 H500 H509 K590 R061 R13X R18X R190 R229 R53X R681 R91X R934 R935 R937{

forvalues i = 1(1)20{

replace `x'=1 if x`i'=="`x'"

}

foreach x of varlist G253 H052 H492 H500 H509 K590 R061 R13X R18X R190 R229 R53X R681 R91X R934 R935 R937{

bysort yctr_id: egen `x'_tot= total(`x')

}

egen total_A_diags=rsum(G253_tot H052_tot H492_tot H500_tot H509_tot K590_tot R061_tot R13X_tot R18X_tot R190_tot R229_tot R53X_tot R681_tot R91X_tot R934_tot R935_tot R937_tot)

egen total_distinct_A_diags = anycount(G253_tot H052_tot H492_tot H500_tot H509_tot K590_tot R061_tot R13X_tot

R18X_tot R190_tot R229_tot R53X_tot R681_tot R91X_tot R934_tot R935_tot R937_tot), values(1(1)100)


****** GENERATING A TIME TO DIAGNOSIS FOR EACH CODE & A MINIMUM TIME FOR EACH CASE ******

foreach x of newlist x1_time x2_time x3_time x4_time x5_time x6_time x7_time x8_time x9_time x10_time x11_time x12_time x13_time x14_time x15_time x16_time x17_time x18_time x19_time x20_time{

gen `x'=.

}

forvalues i=1(1)20{

foreach x of varlist x`i'_time {

replace `x'=(epistart-d_diag1)/365.25 if x`i'!="" & icccgroup==4

}

foreach x of varlist G253 H052 H492 H500 H509 K590 R061 R13X R18X R190 R229 R53X R681 R91X R934 R935 R937{

by yctr_id: gen `x'_time=x1_time if `x'_tot!=0 & x1=="`x'" & icccgroup==4

forvalues i=2(1)20{

replace `x'_time=x`i'_time if `x'_tot!=0 & x`i'=="`x'" & `x'_time>x`i'_time & icccgroup==4

}

foreach x of varlist G253 H052 H492 H500 H509 K590 R061 R13X R18X R190 R229 R53X R681 R91X R934 R935 R937{

by yctr_id: egen `x'_min=min(`x'_time)

}
```

****** GENERATING CIPS FOR ADMISSION ROUTE ANALYSIS ******

```
gen transit = 0

replace transit = 1 if ((admisorc<51 | admisorc>53) & admimeth!=81) & (disdest>=51 & disdest<=53)

replace transit = 3 if ((admisorc>=51 & admisorc<=53) | admimeth==81) & (disdest<51 | disdest>53)

replace transit = 2 if ((admisorc>=51 & admisorc<=53) | admimeth==81) & (disdest>=51 & disdest<=53)

gen procode3 = substr(procode, 1,3)

sort hesid epistart epiorder epiend transit epikey

gen admidisdate = admidate-disdate[_n-1] if hesid==hesid[_n-1]

tab adm_cfl

replace adm_cfl=2 if admidate==.

replace adm_cfl=2 if admidate>disdate

replace adm_cfl=0 if adm_cfl==.

rename adm_cfl adm_cfl

gen dis_cfl=0

replace dis_cfl=2 if disdate<admidate

replace dis_cfl=2 if disdate<dob_yctr

gen long cips =_n

replace cips= cips[_n-1] if (epiorder!=1 & dismeth[_n-1]>5 & hesid==hesid[_n-1]) | ((admidate==admidate[_n-1] & adm_cfl==0 & adm_cfl[_n-1]==0) & hesid==hesid[_n-1] & admidisdate<0)

| (hesid==hesid[_n-1] & epiorder==1 & adm_cfl==0 & dis_cfl[_n-1]==0 & admidisdate<=2 & ((disdest[_n-1]>=51 & disdest[_n-1]<=53) | (admisorc>=51 & admisorc<=53) | admimeth==81))

gen long spell =_n

replace spell = spell[_n-1] if (((epiorder != 1 & dismeth[_n-1]>5 & hesid==hesid[_n-1]) | (admidate==admidate[_n-1] & adm_cfl==0 & adm_cfl[_n-1]==0 & hesid==hesid[_n-1] & admidisdate<0))

 & (procode3==procode3[_n-1]))

codebook cips spell

count if cips!=cips[_n-1] & cips!=cips[_n+1]
```

****** COUNTING EPISODES & CIPS ******

```
bysort yctr_id: egen number_episodes=count(epiorder)

sort cips

gen c=0

replace c=1 if cips!=cips[_n-1]

bysort yctr_id: egen number_cips= total(c)

sort spell

gen s=0
```

```
replace s=1 if spell!=spell[_n-1]
bysort yctr_id: egen number_spells = total(s)

****** GENERATING DATE FOR FIRST CIPS & FIRST ALERT CODE CONTAINING CIPS ******

gen cipstart=epistart if c==1
by yctr_id cips: egen cipstartdate=max(cipstart)
format %td cipstartdate
sort yctr_id cipstartdate
by yctr_id: gen index_cips=_n
gen first_cipS_date= cipstartdate if  index_cips==1
format %td first_cipS_date
gen first_cips_admi=admimeth if  first_cipS_date!=.
gen first_cipS_duration_pt= first_cips_pt- d_diag1
sort yctr_id cips epistart epiorder
by yctr_id cips: egen AB_diag_binary_cips=max(AB_diag_binary)
sort yctr_id cips  AB_diag_binary_cips
gen alert_cips_date= cipstartdate if  AB_diag_binary_cips==1
sort yctr_id  alert_cips_date
by yctr_id: gen index_cipsAB=_n if  AB_diag_binary_cips==1
gen first_AB_cips_date= alert_cips_date if  index_cipsAB==1
format %td first_AB_cips_date
by yctr_id: egen  first_AB_cips_date_pt=max( first_AB_cips_date)
format %td  first_AB_cips_date_pt

****** GENERATING TIME TO DIAGNOSIS VARIABLE ******

sort yctr_id  A_diag_binary epistart
by yctr_id  A_diag_binary: gen AlertAindex=_n
gen first_alertA_epi=1 if AlertAindex==1 &  A_diag_binary>0
gen first_alertA_date=epistart if first_alertA_epi==1
format  first_alertA_date %td
by yctr_id: egen first_alertA_date_pt=total(first_alertA_date)
format first_alertA_date_pt %td
replace first_alertA_date_pt=. if first_alertA_date_pt==0
sort yctr_id  B_diag_binary epistart
by yctr_id  B_diag_binary: gen AlertBindex=_n
gen first_alertB_epi=1 if AlertBindex==1 &  B_diag_binary>0
gen first_alertB_date=epistart if first_alertB_epi==1
```

```
format  first_alertB_date %td
by yctr_id: egen first_alertB_date_pt=total(first_alertB_date)
format first_alertB_date_pt %td
replace first_alertB_date_pt=. if first_alertB_date_pt==0
gen AB_diag_binary=0
replace  AB_diag_binary=1 if  A_diag_binary>0
replace  AB_diag_binary=1 if  B_diag_binary>0
sort yctr_id  AB_diag_binary epistart
by yctr_id  AB_diag_binary: gen AlertABindex=_n
gen first_alertAB_epi=1 if AlertABindex==1 &  AB_diag_binary>0
gen first_alertAB_date=epistart if first_alertAB_epi==1
format  first_alertAB_date %td
by yctr_id: egen first_alertAB_date_pt=total(first_alertAB_date)
format first_alertAB_date_pt %td
replace first_alertAB_date_pt=. if first_alertAB_date_pt==0
by yctr_id:  egen total_alertAB=count(first_alertAB_date)
gen First_Epi_Duration=  first_epi_date_pt-d_diag1
gen first_ab_Duration= first_alertAB_date_pt- d_diag1
gen TTD=0 if  first_ab_duration==.
replace TTD=1 if  first_ab_duration>-31 & TTD!=0
replace TTD=2 if  first_ab_duration<=-31

****** GENERATING SURVIVAL TIME DATA SET ******

gen dead=1 if  d_death!= .
replace  dead=0 if  d_death==.
gen survtime=( d_death- d_diag1)/365.25
replace  survtime=(td(31Dec2012)- d_diag1)/365.25 if dead==0
replace survtime=survtime+0.0027

****** Final Models ******

/*Model 1*/

stcox i.lab0 i.age_group2 i.sex_yctr i.icccgroup Year_Dx

/*Model 2*/
```

```
stcox i.lab0##i.age_group2 i.sex_yctr i.icccgroup Year_Dx
lincom 0.age_group2 + 0.lab0 + 0.age_group2#0.lab0, hr
lincom 0.age_group2 + 1.lab0 + 0.age_group2#1.lab0, hr
lincom 0.age_group2 + 2.lab0 + 0.age_group2#2.lab0, hr
lincom 1.age_group2 + 0.lab0 + 1.age_group2#0.lab0, hr
lincom 1.age_group2 + 1.lab0 + 1.age_group2#1.lab0, hr
lincom 1.age_group2 + 2.lab0 + 1.age_group2#2.lab0, hr

 /*Likelihood Ratio Test*/

stcox i.lab0 i.age_group2 i.sex_yctr i.icccgroup Year_Dx
estimates store model1
stcox i.lab0 i.age_group2 i.sex_yctr i.icccgroup Year_Dx i.Year_Dx
estimates store model2

lrtest model1 model2
```