

NUFFIELD DEPARTMENT OF
PRIMARY CARE
HEALTH SCIENCES

Childhood, Teenage and Young Adult Cancers: Evaluation of diagnostic pathways and their long-term effects

Defne Saatci

Green Templeton College

Nuffield Department of Primary Care Health Sciences

University of Oxford

A thesis submitted to the Division of Medical Sciences at the University of Oxford for the degree of Doctor of Philosophy

October 2024

Acknowledgements

I would like to start by expressing my eternal gratitude to my husband, Lahiru, for his unwavering love, support and steadfast belief in me. And the biggest thank you to my daughter, Ayla, for being the ray of sunshine that lit up my every day during this journey.

I am truly indebted to my parents for giving me the opportunity to get a head start in life and to my brother, for always reminding me to take things a little less seriously. I also want to express my deepest gratitude to my parents-in-law and sister-in-law, for their endless support and infectious positivity. An enormous thank you to my dearest friends Sila and Rachael for being there for me throughout my DPhil journey.

I am tremendously thankful to my supervisors, Professor Julia Hippisley-Cox and Professor Anthony Harnden, for their expertise, advice and exceptional mentorship during my DPhil. Their wisdom has inspired me and their encouragement has given me the confidence to continue to pursue my ambitions in clinical academia.

I would also like to thank my amazing colleagues at the Nuffield Department of Primary Care for their strong support. A special thank you to Dr. Jason Oke, for incredible statistical guidance, and to Dr. Ashley Clift, for his expert insight into risk prediction and overall amazing kindness.

I am grateful to my funders, Children with Cancer UK, Cancer Research UK and the University of Oxford Graduate School (Clinician Investigator Scholarship).

I would like to thank Aggie and Kaleb, for their invaluable personal insight that has helped shape this DPhil and to the Teenage Cancer Trust for getting me in touch with them.

Finally, I am extremely thankful to all the patients who provide their anonymised health records to QResearch Database, without whom this DPhil would not be possible.

Contribution of Work

I confirm that I led the work carried out in this thesis, which includes study design, generation of code, data analysis and write-up in each Chapter. My supervisors, Professors Anthony Harnden and Julia Hippisley-Cox, provided guidance and direction in study design and interpretation of data through regular meetings. Professor Julia Hippisley-Cox reviewed my analysis code to ensure its accuracy and completeness. My collaborator Dr. Jason Oke, reviewed my data analysis plan for Chapters 3 & 4, and offered statistical advice through meetings upon request.

Abstract

Background: Despite vast improvements in treatment and survival, cancer remains a leading cause of mortality and morbidity among children, teenagers and young adults in the United Kingdom. One contributing factor is thought to be delayed diagnosis. Given the rarity and diversity of cancers in this age group, delays are common. This challenge has been highlighted as a serious concern for young people with cancer and is an important research priority for the National Health Service.

Aims and Objectives: The overarching aims are to evaluate factors that influence diagnostic pathways of childhood, teenage and young adult cancers and to establish their association with long-term mortality and morbidity outcomes.

Specific objectives are to:

1. Identify early and late clinical features or feature combinations encountered in primary care, which are associated with a subsequent cancer diagnosis.
2. Develop clinical risk prediction models that can estimate the risk of an undiagnosed cancer in primary care.
3. Investigate person- and health-system factors that influence diagnostic pathways.
4. Explore the association between these pathways and mortality as well as mental health outcomes at 1- and 5-years following cancer diagnosis.

Methods: In order to address these aims, a nationally-representative cohort of children, teenagers and young adults between the ages of 0 and 24 years was generated using routinely-collected primary care-based electronic health records, available through the QResearch Database. Within this cohort, the commonest cancer subtypes in this age group

(leukaemias, lymphomas, central nervous system tumours, sarcomas, abdominal and gonadal germ cell tumours) were identified.

First, this cohort was used to carry out a year-of-birth, sex and general practice matched nested case-control study, using incidence density sampling. Associations between pre-specified clinical features encountered in primary care and a subsequent diagnosis of cancer were explored using conditional logistic regression. Median diagnostic intervals were used to identify relevant early symptoms. Symptom combinations were evaluated using both a clinically-driven approach as well as principal component analysis.

Second, the cohort was utilised to develop and validate risk prediction models for solid and haematological cancers, as well as models specific to leukaemias, lymphomas and central nervous system tumours. Logistic regression was used to predict cancer risk. Model discrimination and calibration were assessed following internal-external validation. The clinical utility of the models was evaluated and compared against existent guidance.

Third, the cohort of cancer cases was used to evaluate if a priori determined person-specific, cancer-specific, health system-specific, and societal factors influence cancer diagnostic pathways. Two outcomes were used to quantify diagnostic pathways: systemic diagnostic intervals and routes-to-diagnosis. The association between factors and the systemic diagnostic interval was explored using quantile regression, whilst the association between factors and routes-to-diagnosis was explored using logistic regression.

Finally, the cancer cohort was employed to investigate if variations in diagnostic pathways affect long-term outcomes, using Cox proportional hazards models. Two outcomes, mortality and mental health disorders in 1- and 5-year survivors, were explored, as they have been highlighted as areas of high-priority by young people with cancer.

Results: The cohort of young people generated was found to be nationally-representative. Cancer cases were most comprehensively captured using the primary care, secondary care datasets and the National Cancer Registry. The incidence of cancer cases in the cohort were comparable to national estimates.

In addition to supporting previously reported clinical features in primary care such as pain, head/neck lumps and seizures, the nested case-control study provided insight into novel associations in primary care, including torticollis, organomegaly, hemiparesis and limping. Furthermore, 17 symptom/signs in this study were associated with cancer at an earlier stage.

Risk prediction models identified several relevant key sociodemographic factors, clinical features and abnormal blood test findings associated with cancers in young people. Upon internal-external validation, models showed good discriminative ability, calibration and potential for clinical utility.

Age, cancer type, histological aggressiveness and symptomology were found to influence both systemic diagnostic intervals and emergency cancer presentations. The presence of blood test requests were found to prolong diagnostic intervals and regional differences were observed in emergency cancer presentations.

Emergency cancer presentations were associated with higher risk of mortality at 1- and 5-years. Exploratory work suggested that longer diagnostic intervals in restricted analyses may worsen mortality and mental health outcomes.

Conclusions: This thesis addresses key gaps in the understanding of cancer diagnostic pathways in young people. It advances knowledge of presenting clinical features of cancer; identifies person-, cancer- and healthcare-dependent factors that influence diagnostic pathways, and better characterises risk of cancer in primary care through the development of

statistical risk prediction models. It also provides insight into how delays may be linked to poorer outcomes, in order to emphasise the importance of achieving early diagnosis. Overall, these findings provide a platform to improve clinical and public awareness of cancer presentations in young people, aid clinical decision-making and inform future diagnostic pathway planning.

Table of Contents

Motivation to study this topic....12

Chapter 1 Introduction....13

1.1. What are Childhood, Teenage and Young Adult Cancers?....13

1.2. What is the burden of CTYA cancers?....15

1.2.1 The link between diagnostic delay and cancer burden....15

1.2.2 Strategies to mitigate cancer burden....16

1.3. What is known of early diagnosis in symptomatic CTYA cancers?....16

1.3.1 The extent of delays in CTYA cancers....16

1.3.2 What strategies have been tried out so far to reduce delays and aid early diagnosis of symptomatic cancer?....16

1.3.3 Modelling the diagnostic pathway....17

1.3.4 Factors that contribute to diagnostic delays in CTYA cancers....18

1.3.5 What is known of diagnostic delays of symptomatic CTYA cancers in the UK?....22

1.4. What data sources are available to study diagnostic pathways and delays?....22

1.5. Aims and Objectives....23

1.6. Thesis Outline....24

Chapter 2 Evaluating the representativeness and validation of electronic health records to study CTYA cancers, using the QResearch Database....26

2.1. Background to chapter....26

2.2. Aims and Objectives....28

2.3. Methods....29

2.3.1 Data Sources....29

2.3.2 Generation of Cohorts....30

2.3.3 Cancer Code Evaluation....33

2.3.4 Calculation of Incidence Rates....34

2.4. Results....35

2.4.1 Cancer diagnostic codes across datasets....35

2.4.2 Measures of Diagnostic Accuracy of GP and HES Datasets....37

2.4.3 Misclassification of diagnoses on GP and HES datasets....37

2.4.4 Evaluation of False Positive Diagnoses....38

2.4.5 Evaluation of false negative diagnoses....39

2.4.6 Evaluation of Dates of Cancer Diagnosis....40

2.4.7 Cohort Characteristics for Children, Teenagers and Young Adults....41

2.4.8 Cancer Cohort Generation....43

2.4.9 Cancer Cohort Characteristics....43
2.4.10 Cancer Incidence Rates....46
2.5. Cohorts used in the thesis....46
2.6. Discussion....50
2.6.1 Summary of Principal Findings....50
2.6.2 Findings and Implications....50
2.6.3 Strengths and Limitations....52
2.6.4 Future Directions....53
Chapter 3 Clinical features associated with childhood, teenage and young adult cancers: a population-based nested case-control study....55
3.1. Background to chapter....55
3.2. Methods....56
3.2.1 Data Source....56
3.2.2 Identification of Cases and Controls....57
3.2.3 Clinical Feature and risk factor selection....57
3.2.4 Coding of clinical features....61
3.2.5 Subgroup Analysis....61
3.2.6 Statistical Analysis....62
3.2.7 Sample Size Calculations....67
3.3. Results....67
3.3.1 Study Population Characteristics....67
3.3.2 All CTYA Cancers: Associated clinical features and their timeframes....70
3.3.3 Childhood Cancer: Associated clinical features and their timeframes....74
3.3.4 Teenage and Young Adult Cancers: Associated clinical features and their time-frames....78
3.3.5 Early and late clinical features associated with different CTYA cancer subtypes....82
3.3.6 Determining combinations of clinical features associated with a cancer diagnosis....94
3.4. Discussion....100
3.4.1 Summary of Principal Findings....100
3.4.2 Findings in Context....101
3.4.3 Strengths and Limitations....105
3.4.4 Future Directions....106
Chapter 4 Derivation and internal-external validation of clinical prediction models for childhood, teenage and young adult blood and solid cancers....108
4.1. Background to chapter....108
4.2. Aims & Objectives....109
4.3. Methods....110
4.3.1 Study Population....110

4.3.2 Outcome of Interest 110
4.3.3 Predictor variable selection 111
4.3.4 Handling Missing Data 114
4.3.5 Model Development 114
4.3.6 Model Evaluation 115
4.3.7 Sensitivity and Subgroup Analyses 117
4.3.8 Sample size calculations 117
4.4. Results 118
4.4.1 Population Characteristics 118
4.4.2 Models for Combined Solid Cancers 125
4.4.3 Models for Combined Haematological Cancers 136
4.4.4 Models for Leukaemia 147
4.4.5 Models for Lymphoma 157
4.4.6 Model for Central Nervous System Tumours 168
4.4.7 Clinical Scenarios 174
4.5. Discussion 176
4.5.1 Summary of Principal Findings 176
4.5.2 Findings in Context 176
4.5.3 Strengths and Limitations 179
4.5.4 Future Directions 181
Chapter 5 Identifying factors that affect diagnostic pathways in childhood, teenage and young adult cancer 182
5.1. Background to chapter 182
5.2. Aims and Objectives 183
5.3. Methods 183
5.3.1 Data Source & Study Population 183
5.3.2 Outcomes of Interest 183
5.3.3 Exposures of Interest 185
5.3.4 Statistical Analysis 189
5.3.5 Sample Size and Power Calculations 190
5.4. Results 192
5.4.1 Outcome 1: Diagnostic Intervals 192
5.4.2 The association between exposures and the cancer diagnostic interval 194
5.4.3 Outcome 3: Route-to-Diagnosis 202
5.4.4 The association between exposures and route-to-diagnosis 204
5.5. Discussion 208
5.5.1 Summary of Principal Findings 208
5.5.2 Findings in Context 208

5.5.3 Strengths and Limitations 210
5.5.4 Implications and Future Directions 211
Chapter 6 The association between diagnostic pathways and 1- and 5-year mortality in childhood, teenage and young adult cancers 213
6.1. Background to chapter 213
6.2. Aims and Objectives 214
6.3. Methods 214
6.3.1 Study Population 214
6.3.2 Exposures of Interest 215
6.3.3 Potential Confounders 216
6.3.4 Outcome of interest 217
6.3.5 Statistical Analysis 217
6.3.6 Handling Missing Data 218
6.3.7 Sample Size Calculations 219
6.4. Results 219
6.4.1 Exposure 1: Diagnostic Intervals 222
6.4.2 Exposure 2: Route-to-Diagnosis 229
6.5. Discussion 231
6.5.1 Summary of Principal Findings 231
6.5.2 Findings in Context 231
6.5.3 Strengths and Limitations 232
6.5.4 Implications and Future Directions 233
Chapter 7 The association between diagnostic pathways and mental health outcomes in 1- and 5-year survivors of childhood, teenage and young adult cancers 234
7.1. Background to chapter 234
7.2. Aims & Objectives 235
7.3. Methods 236
7.3.1 Data Source & Source Population 236
7.3.2 Study Population 236
7.3.3 Outcomes of interest 237
7.3.4 Exposures of Interest 238
7.3.5 Statistical Analysis 240
7.3.6 Handling Missing Data 241
7.3.7 Power Calculations 241
7.4. Results 243
7.4.1 Objective 1: Risk of mental health disorders in CTYA cancer survivors and controls 243
7.4.2 Objective 2: Diagnostic pathways and their association with the risk of mental health disorders 247

7.4.3 Diagnostic Intervals and anxiety-depressive disorders in 1-year and 5-year survivors.....247
7.4.4 Route-to-diagnosis and anxiety-depressive disorders in 1 year and 5-year survivors 255
7.4.5 Route-to-diagnosis and psychotic disorders in 1- and 5-year survivors.....257
7.5. Discussion.....259
7.5.1 Summary of Principal Findings.....259
7.5.2 Findings in Context.....259
7.5.3 Strengths and Limitations.....260
7.5.4 Implications and Future Directions.....261
Chapter 8 Discussion.....262
8.1. Summary of Findings and Implications.....262
8.1.1 Developing a platform to study CTYA cancer diagnostic pathways using EHR.....262
8.1.2 Advancing knowledge of presenting clinical features to better recognise CTYA cancer.....263
8.1.3 Developing a statistical model to ascertain the risk of CTYA cancer in primary care..264
8.1.4 Factors that influence diagnostic pathways.....264
8.1.5 Diagnostic pathways and subsequent cancer burden.....266
8.2. Strengths.....267
8.3. Limitations.....268
8.3.1 The limitations of clinical coding and EHR.....268
8.3.2 Limitations of study size.....270
8.4. Future Directions.....270
8.5. Concluding remarks.....271
Key Professional Development and Skills Acquisition.....272
Chapter 9 Supplementary Appendix.....274
Chapter 10 Tables and Figures.....456
Chapter 11 Glossary.....473
Chapter 12 References.....475

Motivation to study this topic

The one diagnosis paediatricians never want to miss is cancer, and, unfortunately, throughout my clinical training I have seen it missed time and time again. A 3 year old girl with abdominal pain mistaken for constipation who later was found to have a neuroblastoma; a 6 year old boy with a squint which later was considered relevant when he developed problems with his balance and was subsequently diagnosed with a medulloblastoma; and a 15 year old boy with fever re-presenting after being diagnosed with “flu” a few days prior, who was found to have acute lymphoblastic leukaemia. Each time a young person receiving a devastating diagnosis, having experienced a frustratingly protracted journey.

Missing a cancer diagnosis has a personal impact on young people and their families. I have spoken to several families who have been distressed by the process of diagnosis, blame themselves for not speaking up, and angry that their concerns were not acknowledged. This threatens the trust between families and healthcare professionals, which in turn is critical for their journey through cancer therapy. This trust is further jeopardised when families ask if this prolonged journey to diagnosis could impact long-term outcomes and receive “I do not know” as a response.Unfortunately, there is just not enough evidence to give an honest answer.

These challenging clinical experiences generated conflicting feelings for me during my training. Although I accepted that recognising cancer in young people is very hard as a doctor, I was also driven to find ways to improve the cancer diagnosis journey for young people. It was at this point in my clinical training that I decided to turn this drive into research questions and to develop research skills to be able to better answer 1) how we achieve earlier recognition of cancer in young people and 2) what are the long-term consequences of us clinicians missing cancer?

Chapter 1 Introduction

1.1. What are Childhood, Teenage and Young Adult Cancers?

Childhood, teenage and young adult (CTYA) cancers are a diverse group of cancer subtypes which affect young people from birth up to the age of 24 years [1]. They consist of 12 broad diagnostic groups and 88 distinct subgroups, as defined by the International Classification of Childhood Cancers-3 (ICCC-3) framework [2]. The underlying causes of CTYA cancers are poorly understood, with only a few risk factors identified [3]. The global incidence of CTYA cancer is low overall and varies considerably across countries, with the commonest diagnostic subtypes including leukaemias, central nervous system (CNS) tumours and lymphomas [4]. Notably, there have been vast advancements made in the treatment of cancers in young people over the last century, with 5-year survival rates increasing from below 20% to up to 81% in high-income countries [5].

In the United Kingdom (UK), approximately 4500 young people get diagnosed with cancer every year [6]. Although the overall incidence is low, it varies with age, being low risk in infants and moderately high risk in young adulthood (Figure 1-1)[7].

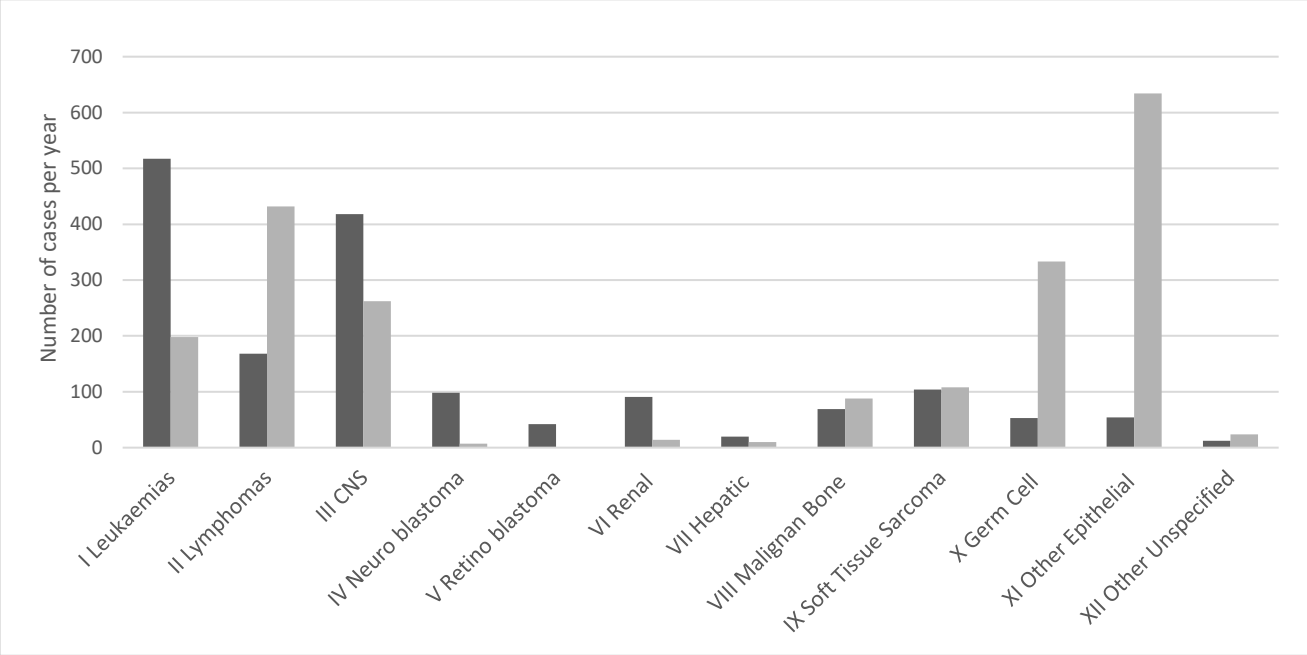
Figure 1-1 *Risk of cancer by age, adapted Walker et al., 2021[7].*



Cancers are commonly categorized into “childhood”, representing cancers in 0-14 year olds and “teenage and young adult (TYA)”, representing cancers in young people aged 15-24 [1]. The most frequent diagnostic cancer subtypes reported in the UK include leukaemias,

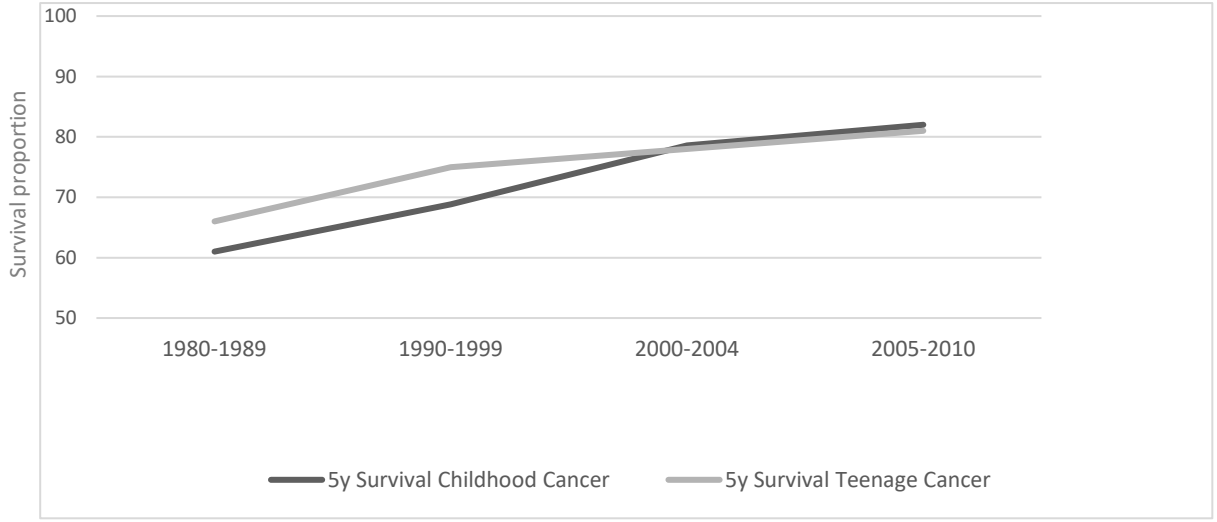
lymphomas and CNS tumours in childhood, and carcinomas, lymphomas, germ cell tumours as well as CNS tumours in TYA (Figure 1-2) [6].

Figure 1-2 *Annual average number of cases within the twelve broad diagnostic groups of CTYA cancers in the United Kingdom, adapted Cancer Research UK (CRUK) Cancer Statistics [6]. Black bar = childhood cancer, grey bar= teenager and young adult cancer.*



Survival from CTYA cancers has vastly improved in the UK and overall survival rate is reported to be 84% at 5-years and 80% at 10 years following diagnosis (Figure 1-3) [1].

Figure 1-3 *Survival following CTYA cancer from 1980 to 2010, adapted CRUK Cancer Statistics [4].*



1.2. What is the burden of CTYA cancers?

As detailed above, the global incidence of CTYA cancer is low, yet, it ranks as the 6th leading cause of cancer burden worldwide and is one of the commonest causes of mortality and morbidity in this age group [2, 8].

This high cancer burden is also strongly reflected in the UK. Despite its relative rarity, cancer remains the commonest cause of mortality in young people [9]. Furthermore, morbidity following survival is substantial in the UK, with reports of significant long-term physical and psychosocial complications compared to other high-income countries [10]. Several underlying factors are thought to contribute to the high cancer burden observed in CTYA, of which delayed diagnosis is a common and leading concern highlighted by young people with cancer in the UK [11].

1.2.1 The link between diagnostic delay and cancer burden

There is consistent evidence from adult cancers that suggest that the longer it takes to diagnose, the higher the risk of advanced cancer and subsequent mortality [12, 13]. This is further supported by basic science studies that follow the natural history of cancer and show that over time, if untreated, tumours grow and undergo local and metastatic spread [14]. Given cancer follows a similar natural history across all ages, it is postulated that diagnostic delays in young people are likely to result in advanced cancer stages as well as higher mortality, and this has indeed been shown in some studies but not all [15].

Delay in cancer diagnosis also has been shown to increase morbidity in young people, although studies exploring this association in CTYA is also limited. Only one population-based study in the UK has found that quality of life was lower in those who experienced diagnostic delay, but this was limited to 830 survivors of teenage and young adult cancers [16].

There is therefore an important need to carry out further epidemiological studies to establish this link between diagnostic delay and higher cancer burden in CTYA as this is likely

to highlight the impact of reducing delays and also help support clinicians and policymakers with future management and service delivery decisions.

1.2.2 Strategies to mitigate cancer burden

Accordingly, reducing delays and improving early diagnosis have been important strategies to mitigate disease burden following cancer [17]. It has had two successful approaches, specifically in adult cancers [17]. One approach aims to identify those who are asymptomatic and employs strategies such as population-level screening and biomarker discovery. It has had little success to date in CTYA cancers [3]. The second approach aims to identify ways to recognise symptomatic individuals early and has the biggest potential in CTYA cancers to improve diagnostic delays [17]. In the UK, achieving early diagnosis of symptomatic cancer is a key research and public health priority, as detailed in the National Health Service (NHS) Long Term Plan (2019) [18], UK Cancer Reform Strategy (2015) [19], and Childhood Cancer and Leukaemia Group (CCLG) Strategic Plan (2020) [20].

1.3. What is known of early diagnosis in symptomatic CTYA cancers?

1.3.1 The extent of delays in CTYA cancers

Delays in CTYA cancers have been found to be long, with reports from population-based studies identifying waiting times between 2-4 months on average for symptomatic CTYA cancers [21-23]. In addition, repeated contact with healthcare professionals prior to cancer diagnosis is particularly apparent in CTYA cancers [24-26]. In the UK, diagnostic delay of CTYA cancer is known to be a particularly significant challenge, with young people experiencing longer delays compared to other high-income countries [27].

1.3.2 What strategies have been tried out so far to reduce delays and aid early diagnosis of symptomatic cancer?

Opportunities to raise clinical and public awareness of symptoms/signs related to cancer is a central strategy to reduce delays and aid early diagnosis. One of the most successful initiatives

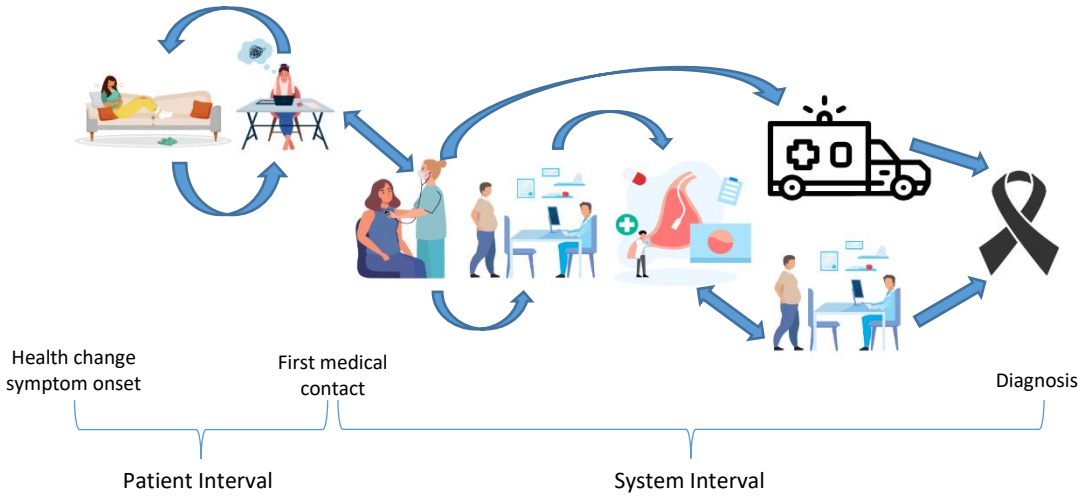
internationally to achieve this is the HEADSMART awareness initiative in the UK [28]. Providing age-stratified symptom lists for central nervous system tumour diagnoses to clinicians and the public, the initiative has been shown to result in a reduction of the diagnostic interval for brain tumours from 14.4 in 2006 to 6.7 weeks in 2013 in children [28, 29]. However, this initiative alone has not been able to fully reduce delays or resolve diagnostic uncertainty, particularly among general practitioners (GPs) in the UK, of which only 12% report confidence in being able to recognize a brain tumour in children after taking part in the initiative [30]. Thus, further strategies are still required to improve early diagnosis of symptomatic cancer.

In order to develop such strategies, it is first key to identify factors that may contribute to diagnostic delays. This initially requires knowledge of the diagnostic pathway, which is summarised in the following section. Next, current knowledge of the factors that contribute to diagnostic delays in CTYA cancers is summarised in order to highlight areas of ongoing research gaps.

1.3.3 Modelling the diagnostic pathway

Diagnostic delays, and factors associated with delays, can be identified through examination of the diagnostic pathway. This diagnostic pathway has already been modelled in order to provide a framework to study diagnostic delays and is summarised in the Aarhus Statement [31, 32]. This diagnostic pathway is complex and non-linear, beginning with a change in health status, such as the development of a new symptom, and ending when the diagnosis is made (Figure 1-4). It is dependent on patient knowledge and decision-making, as well as healthcare access, quality and provision in a given health system [32, 33].

Figure 1-4 *An illustration of a cancer diagnostic pathway, starting from symptom onset to confirmed diagnosis.*



Several components of the diagnostic pathway have been used in order to effectively study diagnostic delays [31]. In this thesis, the focus will be on three components: 1) the presenting clinical features (i.e., cancer onset), 2) the route-to-cancer-diagnosis; this is most frequently studied as emergency vs. non-emergency routes, where emergency route is most commonly defined as unplanned emergency presentation to hospital up to 30 days before confirmed cancer diagnosis [34], and 3) the length of the diagnostic process (i.e, from symptom onset to confirmed diagnosis), which is also referred to as the total diagnostic interval [32]. Diagnostic time intervals can be further subdivided into patient and system-intervals, where the patient-interval represents the time it takes for patients to seek medical care from the onset of their symptoms; while system-intervals represent the time it takes from first medical contact to diagnosis [32].

1.3.4 Factors that contribute to diagnostic delays in CTYA cancers

For CTYA cancers, three important factors play an important role in diagnostic delays.

First, as detailed above, CTYA cancers are uncommon, which has made undertaking research with sufficient sample size to characterise their clinical presentations difficult. Lack

of knowledge of clinical presentations, in turn, negatively affects recognition of cancer by young people as well as healthcare professionals and results in subsequent delays [17]. Most evidence of clinical presentations of CTYA cancer has been through hospital-based studies to date, which are summarized in systematic reviews [35-38]. These studies provide insight into the diverse symptomology of CTYA cancers but do not capture information relevant to the earlier stages of the disease onset. Only three population-based observational studies, two from the UK and one from Denmark, investigate the earlier presenting features of CTYA cancers in primary care [25, 26, 39] and all studies highlight pain and lumps as important features. However, in addition to being limited by smaller sample sizes and questionnaire-based designs which are subject to selection bias, these studies do not provide further granular information on the time-varying characteristics of symptoms or symptom-combinations that may be useful in identifying opportunities for early diagnosis. Furthermore, no population-based studies to date have explored other features which may be relevant to cancer presentations in CTYA, such as changes in blood markers, which have been previously reported in adult cancers [40, 41]. Thus, there is an ongoing need to better characterise clinical presentations to reduce delays in CTYA cancer diagnosis.

Second, cancer presentations in CTYA are diverse due to the wide variety of cancer subtypes that make up CTYA cancers. Presentations of these cancers are also largely non-specific with broad symptom signatures that overall have low predictive values, akin to multiple myeloma and pancreatic cancer presentations in adults [42]. Collectively, this makes quantifying the risk of cancer and differentiating it from other self-limiting disorders seen in young people difficult and, in turn, results in diagnostic uncertainty, misdiagnosis and consequent delays. One strategy to overcome this challenge in adult cancers is through risk prediction tools. These tools are increasingly being used in primary care settings in order to support medical decision-making and reduce diagnostic inaccuracies where diagnostic

uncertainty is present [43-45]. The potential for these tools in improving early cancer diagnosis have been highlighted as an “area of extraordinary opportunity”, with promising developments seen in several adult cancers [46]. Despite these advancements in adult cancers, however, clinical decision tools have not been explored in CTYA cancers. Thus, there is still a pressing need for novel approaches to better discriminate and ascertain risk of CTYA cancer.

Third, independent of cancer characteristics, other person-specific and country-specific healthcare factors have been found to play a role in diagnostic delays in cancer. The factors found to be relevant in CTYA cancers are summarised and presented in Table 1-1 [21, 23, 39, 47]. These include the age of the young person at diagnosis, their sex, the level of parental education and socioeconomic status, and the medical speciality consulted.

Table 1-1 *Person and healthcare factors with prior evidence of contributing to delays in CTYA cancers [21, 23, 39, 47].*

Factor	Description
Person-factors	
Age-at-diagnosis	Older age associated with longer diagnostic intervals.
Sex	Females likely to experience longer diagnostic delays compared to males.
Level of parental education	Lower level of parental education associated with longer diagnostic intervals
Socioeconomic status	Lower socioeconomic status was associated with higher number of medical contact prior to diagnosis
Healthcare factors	
Medical speciality consulted	Consulting a general practitioner compared to paediatrician was associated with longer diagnostic intervals
Number of medical contacts	Higher number of contacts with general practitioners were associated with longer diagnostic intervals

Although current evidence provides important insight into factors that may influence diagnostic pathways in CTYA cancer, the majority of studies are either hospital based or focus either on childhood (0-14 years) or TYA (15-24 years) populations separately. Additionally, other important person-specific factors, such as ethnicity and presence of comorbidities, known to influence adult cancer diagnostic delays, have not been explored in CTYA [48, 49]. The recent Coronavirus Disease 2019 (COVID-19) pandemic has also highlighted how access to healthcare services can impact cancer diagnostic intervals and contribute to long delays in adults [50, 51], but healthcare access and its impact on diagnostic delay has not been explored

fully in CTYA cancers. Thus, more comprehensive, population-based studies exploring person-specific and healthcare-specific factors are still needed to better understand contributors to diagnostic delay in CTYA cancers and provide opportunities to mitigate these delays.

1.3.5 What is known of diagnostic delays of symptomatic CTYA cancers in the UK?

Notable studies have explored factors that may contribute to long diagnostic delays in the UK. Cohorts recruited as part of the HEADSMART awareness initiative, for example, showed that better clinical and public knowledge of clinical features with high discriminatory potential result in reduced diagnostic uncertainty and consequent delays [52]. Furthermore, the BRIGHLIGHT cohort, a nationally representative cohort of 830 TYA from across the UK, diagnosed with cancer between 2012 and 2015, has provided important insight into how person-specific factors, such as sex, and cancer specific factors, such as tumour subtype, influence diagnostic delays in TYA [21, 53]. However, as both cohorts are limited either to a specific cancer subtype (i.e., brain tumours) or a specific age group (i.e., TYA), research gaps highlighted above still exist and need to be addressed in order to aid earlier cancer diagnosis across all CTYA in the UK.

1.4. What data sources are available to study diagnostic pathways and delays?

Various data sources have been used to model cancer diagnostic pathways in CTYA, including qualitative studies undertaking interviews with patients [54, 55], as well as quantitative studies gathering information either through questionnaire-based designs [21, 39] or through information collected from electronic healthcare records (EHR) [25, 26, 56]. All sources provide vital insight, from different and complementary angles, into the diagnostic journey.

Importantly, however, the growing availability of linked, nationally representative EHR, using information from primary care, secondary care and cancer registry level, have not been fully leveraged to study CTYA cancers. The important advantages of using such linked EHR in

early cancer diagnosis research has been shown through studies using adult cancers. First, linked datasets provide an effective opportunity to provide nationally/internationally representative and generalizable conclusions on cancer diagnostic pathways and associated delays. This is particularly the case in the UK, where there is a wealth of primary care based EHR linked to hospital and registry data [57]. Second, EHR-based cancer cohorts minimise the risk of selection, survivor and recall biases that may arise from questionnaire-based cohort designs [57]. Third, they use clinical coding practices, which can generate interoperable, internationally-recognised categorisations for cancer diagnoses for future studies. Finally, as EHR get routinely updated, they provide a cost-effective way to keep the cohort up-to-date [57].

1.5. Aims and Objectives

Accordingly, the overarching aims of this thesis are to first, determine factors that influence diagnosis of CTYA cancer in the UK and second, to establish their association with long-term mortality and morbidity outcomes.

Specific objectives are to:

1. Establish and validate an EHR-based, nationally-representative cohort of CTYA and accurately identify cancer cases within this population.

2. Identify early and late clinical features (or feature combinations) encountered in primary care associated with a subsequent diagnosis of CTYA cancer.

3. Develop and evaluate a clinical prediction tool to estimate the risk of an undiagnosed CTYA cancer in primary care.

4. Investigate which person- and health-system factors influence diagnostic pathways.

5. Explore whether diagnostic pathways are associated with mortality and long-term outcomes at 1- and 5-years following cancer diagnosis.

Linked, primary care-based EHR are leveraged to study early diagnosis of symptomatic CTYA cancers in this thesis. Specifically, the QResearch Database is utilised, which is a nationally representative primary care database consisting of 35 million anonymised primary care medical records linked to hospital, civil registration and cancer registry datasets, across all ages from approximately 1500 GPs across England, representing around 25% of the population [58].

1.6. Thesis Outline

Chapters of this thesis are outlined below, which address the aforementioned five objectives.

Chapter 2- "**Evaluating the representativeness and validation of electronic health records to study CTYA cancers, using the QResearch Database"**

This chapter describes QResearch Database as a data source, the validity of cancer diagnostic codes available on the database and its suitability to study diagnostic pathways of CTYA cancer.

Chapter 3 – "Clinical features associated with childhood, teenage and young adult cancers: a population-based nested case-control study"

This chapter reports on the clinical features and features combinations associated with a subsequent cancer diagnosis and establishes their timeframe through a nested case-control study.

Chapter 4 – "Derivation and internal-external validation of clinical prediction models for childhood, teenage and young adult cancers"

This chapter investigates opportunities for cancer risk stratification through the development of risk prediction models that identify those at higher risk of CTYA cancer in primary care.

Chapter 5 – "Identifying factors that affect diagnostic pathways in childhood, teenage and young adult cancer: a population-based cohort study"

This chapter investigates person-specific and health system-specific factors that influence CTYA cancer diagnostic pathways, quantified through the diagnostic interval and route-to-diagnosis, in a retrospective cohort study.

Chapter 6 – "The association between diagnostic pathways and 1- and 5-year mortality in childhood, teenage and young adult cancers: a population-based cohort study":

This chapter determines whether CTYA cancer diagnostic pathways, quantified through the diagnostic interval and route-to-diagnosis, affect long-term mortality outcomes in a retrospective cohort study.

Chapter 7 – "The association between diagnostic pathways and mental health outcomes in 1- and 5- year survivors of childhood, teenage and young adult cancers: a population-based cohort study":

This chapter evaluates whether CTYA cancer diagnostic pathways, quantified through the diagnostic interval and route-to-diagnosis, affect long-term mental health outcomes (depression, anxiety and acute psychosis) in a retrospective cohort study.

Chapter 2 Evaluating the representativeness and validation of electronic health records to study CTYA cancers, using the QResearch Database

2.1. Background to chapter

Cancers in CTYA are rare and diverse, with presentations that frequently mimic self-limiting conditions [23]. This poses several important research challenges. First, large representative samples are required in order to effectively study cancers in this age group. Second, comprehensive information across the full patient journey (through symptom-onset, primary care, secondary care to diagnosis) is needed to draw accurate conclusions. The current mainstay approach to study CTYA cancers, which is prospective cohort studies, have not been able to overcome these two challenges to date because of high-costs, need for long follow up periods and biases that arise from selection to the cohort and recall from participants. However, now, the availability of routinely collected EHR provides a crucial opportunity to tackle these challenges.

Globally, access to primary care-based EHR is limited and the UK is in a world-leading position to provide such rich community-based information through existence of nationally representative primary care databases [57]. Furthermore, these databases are linked to civil registry, secondary care (hospital and intensive care), cancer registry as well as radiotherapy/systemic anticancer therapy databases, and thus provide further advantages[57]. First, this makes it possible to combine valuable primary care data whilst also accurately identifying young people with cancer and their most up-to-date, detailed treatment data. Second, EHR-based cohorts minimise the risk of selection, survivor and recall biases that may arise from questionnaire-based cohort designs. Third, coding practices across these linked datasets provide the foundations to generate interoperable, internationally-recognised

categorisations for cancer diagnoses and treatment. Finally, this approach allows for on-going cohort enrolment as EHR get routinely updated, providing a cost-effective way to keep the cohort up-to-date and followed-up over time [57].

To date, knowledge of whether primary care-based EHR provide an accurate representation of young people, and specifically of young people with cancer, in the UK is limited. Accordingly, using one of the largest population-level primary care databases in the UK, the QResearch Database, the overarching aim of this chapter is to investigate if nationally representative cohorts of young people and young people with cancer can be generated using EHR.

The QResearch Database was selected for several reasons. First, it is one of the largest in the UK with 35 million anonymised health records from approximately 1500 GPs in England (~25% population in England)[58] and can thus provide a large representative sample of young people. Second, it has high-quality data linkages to 1) hospital data via Hospital Episode Statistics (HES), which contains admission, procedural and emergency attendance information from all National Health Service (NHS) hospitals in England [59]; 2) civil registration data (e.g., mortality data) through the Office for National Statistics (ONS)[60]; 3) National Cancer Registry, radiotherapy and Systemic Anti-Cancer Treatment (SACT) data through NHS England [61, 62].

2.2. Aims and Objectives

The overarching aim is to investigate if nationally representative cohorts of young people and young people with cancer can be generated using primary care-based EHR available through the QResearch Database. Specific objectives are:

1. Generate a cohort of young people aged 0-24 years.

- Compare population estimates and population structure to national reports provided through the ONS.

2. Generate a cohort of young people with a diagnosis of a priori selected, commonest cancers.

- Compare incidence rates to national reports provided through the National Cancer Registration and Analysis Service.

3. Evaluate the completeness and accuracy of cancer diagnoses recorded on primary care and secondary care databases by comparing to the National Cancer Registry as gold standard.

2.3. Methods

2.3.1 Data Sources

2.3.1.1 QResearch Database

QResearch Database is a nationally representative primary care database with over 35 million anonymised health records from approximately 1500 GPs in England (~ 25% population)[58]. Records contain patient-level demographic information (i.e., year-of-birth, sex, self-assigned ethnicity), as well as clinical information, including cancer diagnoses and clinical presentations. Primary care records are linked to hospital admission, civil registration and the National Cancer Registry data, where linkage is based on an individual patient's anonymized NHS number. This number is valid and complete in 99.8% of primary care/civil registry data and 98% of hospital admissions data[58].

2.3.1.2 Hospital Episode Statistics

HES is a national secondary care database that contains details for hospital admissions (including intensive care admissions) and emergency attendances to all NHS hospitals in England, provided by NHS Digital [59]. It is linked to the QResearch Database through an individual patient's anonymised NHS number.

2.3.1.3 National Cancer Registry

The National Cancer Registry collects data on malignant and pre-malignant neoplasms for all people living in England using an events-based registration model through NHS England [62]. Information is collected from multiple sources including multidisciplinary team meetings, pathology reports, molecular testing results, treatment records, national cancer audits, death certificates and HES records. The registry provides high-quality data (i.e., through a high proportion of microscopically verified cancer diagnoses), with 98% completeness and very low

rates of serious errors (0.1%) [62, 63]. It is linked to the QResearch Database through an individual patient's anonymised NHS number.

2.3.2 Generation of Cohorts

2.3.2.1 Cohort of Children, Teenagers and Young Adults

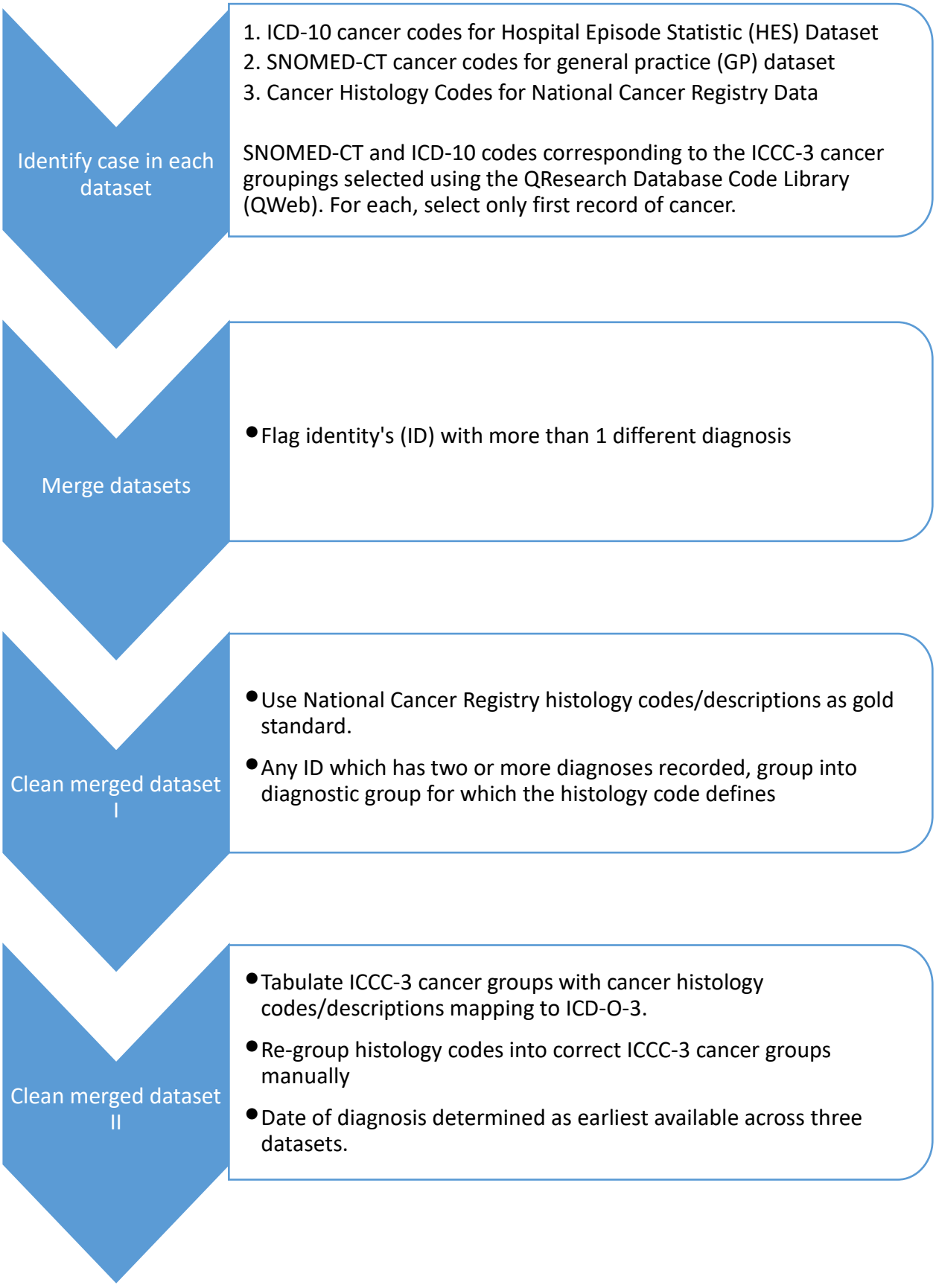
An open cohort of CTYA (0-24 years) registered with a practice in the QResearch Database between 1st January 1998 and 31st December 2019 was generated. Entry date was selected due to availability of linked data from 1998 on the QResearch Database and the exit date was selected due to availability of National Cancer Registry data up to 2019. Entry to the cohort was the latest of date of registration with the practice plus 1 year, date on which the practice computer system was installed plus 1 year, the study start date (1 January 1998). Exit from the cohort was earliest of date of turning 25, death, leaving the practice or study end date (31st December 2019).

Demographic factors including age at cohort entry, sex, ethnicity and deprivation level recorded for the cohort of CTYA were described and compared to the nationally reported mid-2019 ONS population estimates [60]. Ethnicity was explored using self-assigned ethnicity as defined by the 2021 Census [64]. Deprivation level was assessed using the Townsend Deprivation Score which is an area-level continuous score based on an individual's postcode; factors that included unemployment, non–car ownership, non–home ownership, and household overcrowding, are measured for a given area of approximately 120 households, via the 2011 Census of England and Wales and combined to give a Townsend score for that area, with the first quintile representing the lowest deprivation level and the fifth quintile representing the highest deprivation level [65].

2.3.2.2 Cohort of Children, Teenagers and Young Adults with Cancer

The open cohort of CTYA generated above was used to identify individuals with the commonest non-skin cancer diagnoses [6]. As cancer diagnostic codes are coded using either Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) (primary care records) [66], International Classification of Diseases-10 (ICD-10, secondary care records [67]) or International Classification of Diseases for Oncology (ICD-O-3, National Cancer Registry), codes were first harmonised within each coding system and subsequently grouped into subtypes defined by the National Cancer Registration and Analysis Service and in accordance with the International Classification for Childhood Cancers (3rd edition, ICCC-3) [2]. These subtypes were: 1) Leukaemias, myeloproliferative and myelodysplastic diseases, 2) lymphomas and reticuloendothelial neoplasms, 3) central nervous system and intraspinal tumours, 4) soft tissue and bone sarcomas, 5) abdominal tumours (renal tumours, neuroblastomas, hepatic tumours) and 6) gonadal germ cell tumours. The flowchart below outlines this process (Figure 2-1). ICD-10, SNOMED-CT and cancer histology codes used to define these cancer groups can be found in Supplementary Tables 9-1, 9-2 and 9-3 in the Appendix (Chapter 9).

Figure 2-1 *Flowchart for cancer diagnosis identification and classification. Only identification and merging steps are used for validation study described below.*



Specifically, two subtypes of cancer were excluded: retinoblastomas and "other epithelial carcinomas". Cases with a diagnosis of retinoblastoma were excluded because 1) a substantial proportion of retinoblastomas may be captured through screening (i.e., different presentation pathways) and 2) of those that present following the neonatal period, there is unlikely to be sufficient sample size to detect outcomes in subsequent chapters [68]. Cases with a diagnosis of "other epithelial carcinomas" were also excluded. This was because, although altogether these carcinomas represent a considerable proportion of TYA cancers, it was deemed unlikely to have sufficient sample size to detect associated outcomes with individual subgroups within this category. For example, for thyroid cancers, which is one of the commonest subtypes in the "other epithelial carcinoma" category in TYA, the average standardised incidence reported across all CTYA is approximately 8 in 1 million [6]. Furthermore, cases with a diagnosis prior to study start date were excluded.

2.3.3 Cancer Code Evaluation

SNOMED-CT, ICD-10 and ICD-O-3 histological codes were reviewed for each coding system and grouped into the cancer subtypes defined by ICCC-3, as explained in the flowchart above. Cancer subtypes based either on SNOMED-CT or ICD-10 were subsequently compared against the cancer subtypes generated from the ICD-O-3 histological codes on the National Cancer Registry records. Histological codes in the National Cancer Registry were taken as gold-standard given the substantial proportion of microscopically verified cancer diagnoses, demonstrating high-quality and very low rates of serious errors (0.1%) within the Registry [62]. Sensitivity, specificity and positive predictive values were calculated for cancers identified through SNOMED-CT and ICD-10, respectively (Figure 2-2). Subsequently, the proportion of misclassified diagnoses, false positives and false negatives were explored in further detail across cancer subtypes. Dates of cancer diagnoses were compared between the National Cancer Registry and GP as well as HES datasets.

Figure 2-2 *Validation study for cancer diagnoses with definitions for sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). SNOMED-CT/ICD-10 codes compared against National Cancer Registry, as gold standard.*

	National Cancer Registry **Cancer Present**	**National Cancer Registry** **Cancer Absent**	
SNOMED-CT/ICD-10 cancer present	True positive (a)	False Positive (b)	$PPV = \frac{a}{a+b}$
SNOMED-CT/ICD-10 cancer absent	False negative (c)	True negative (d)	$NPV = \frac{d}{c+d}$
	$Sensitivity = \frac{a}{a+c}$	$Specificity = \frac{d}{b+d}$	

2.3.4 Calculation of Incidence Rates

The crude incidence rate of cancer subtypes within the cohort were calculated with associated confidence intervals using *stptime* command on Stata [69]. This calculates incidence rate as follows:

$$Incidence\ Rate = \frac{\#\ new\ cases\ in\ follow\ up\ period}{total\ person-time\ at\ risk\ during\ follow\ up\ period}$$

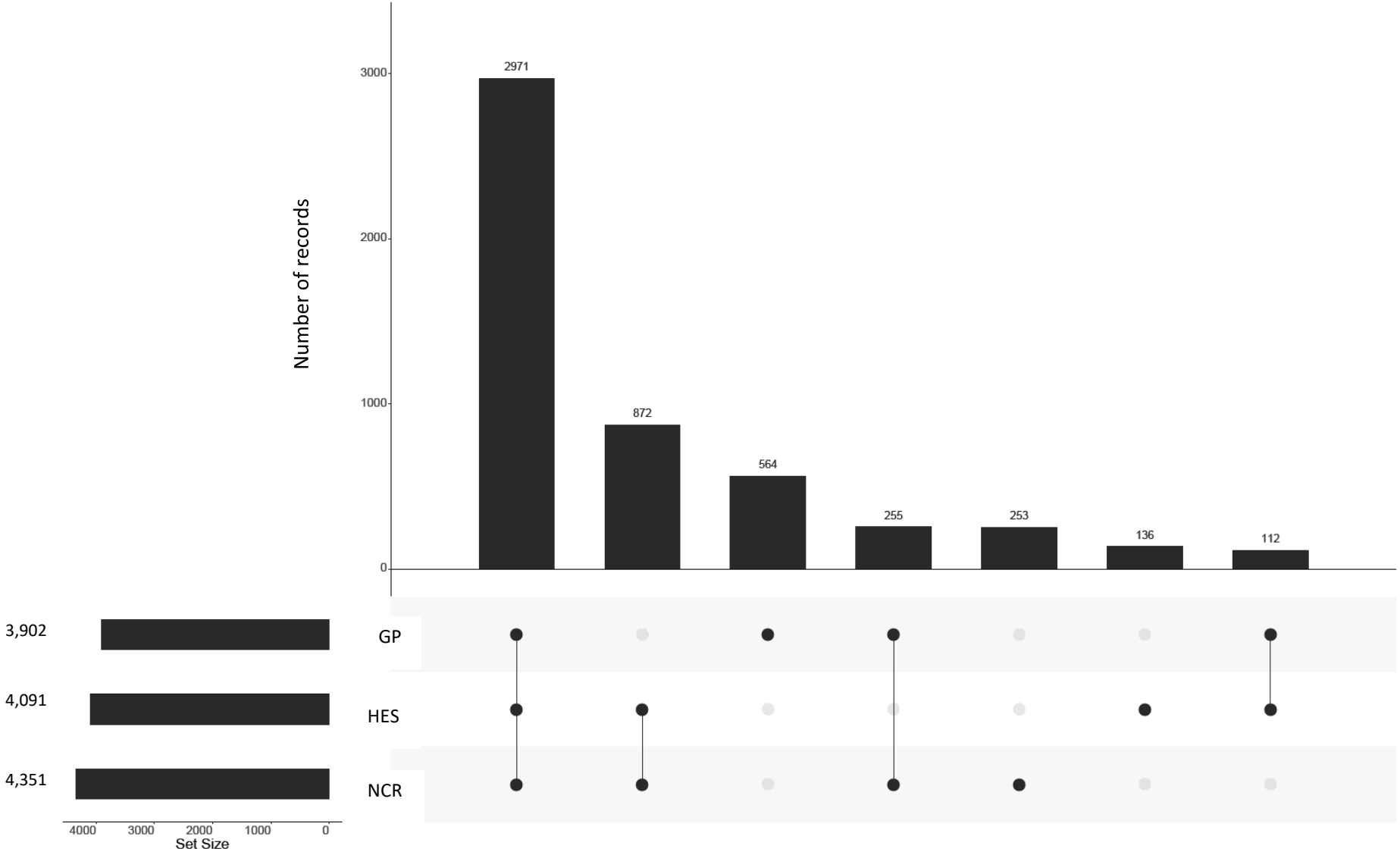
In order to make comparisons with nationally reported data provided by the National Cancer Registration and Analysis Service [6], incidence rates were calculated separately by sex across CTYA (0-24 years).

2.4. Results

2.4.1 Cancer diagnostic codes across datasets

Overall, within the study period and full CTYA population (n=5,239,307), there were 4,351 relevant cancer diagnoses for 0–24-year-olds identified on the linked National Cancer Registry, 3,902 relevant cancer diagnoses using SNOMED-CT codes (GP dataset) and 4,091 relevant cancer diagnosis using ICD-10 codes (HES dataset). 2,971 diagnoses were reported on all three datasets, whilst 253 were only present in the National Cancer Registry, 136 were only present on the HES dataset and 564 were only present on the GP dataset (Figure 2-3). Less than 5 cancer diagnoses were through death records only and thus not graphically illustrated. Source of diagnostic codes per cancer subtype are detailed in the Supplementary Appendix (Chapter 9, Supplementary Figure 9-1).

Figure 2-3 *Plot depicting number of cancer diagnoses identified in intersecting datasets. Dots represent diagnoses present in individual dataset, lines between dots represent diagnoses present in multiple datasets (74). Total number of cancer diagnoses in each dataset summarised on bar chart on bottom left. GP= General Practice Dataset, HES= HES Hospital Dataset, NCR= National Cancer Registry Dataset*



2.4.2 Measures of Diagnostic Accuracy of GP and HES Datasets

The measures of diagnostic accuracy of the GP and HES datasets compared to the gold-standard National Cancer Registry are detailed in Table 2-1. The HES dataset had higher sensitivity (87%) and positive predictive value (96%) compared to the GP dataset sensitivity (72%) and positive predictive value (85%). When cancers from both GP and HES datasets were considered together, the sensitivity increased to 94%.

Table 2-1 *Number of cancer diagnoses records in each dataset and diagnostic accuracy measures of SNOMED-CT and ICD-10 cancer codes using the National Cancer Registry as gold standard.*

Code Source	Total number present in source only	Total number present in source & Registry	Total number absent in source	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
SNOMED-CT	564	3,226	1,261	72	99	85	99
ICD-10	136	3,843	1,072	87	99	96	99
SNOMED-CT & ICD-10	112	4,098	253	94	99	97	99

When diagnostic accuracy was explored across different cancer subtypes, variability was seen in both HES and GP datasets, with higher sensitivity and positive predictive values for leukaemia and hepatic tumour diagnoses but lower sensitivity and positive predictive values for other cancer subtypes (Appendix Chapter 9, Supplementary Table 9-4).

2.4.3 Misclassification of diagnoses on GP and HES datasets

Within individuals identified to have a cancer on the datasets, there was a small proportion of misclassification of cancer subtypes within the GP dataset (n=15/3,226, 0.5%) and a larger proportion of misclassification in the HES datasets (n=100/3,843, 3%) when the National Cancer Registry was used as gold-standard (Table 2-2). The majority of HES misclassifications

were other cancer subtypes being incorrectly labelled as “leukaemias” (69/100, 69%); specifically lymphoma diagnoses being labelled leukaemias (50/69, 72%).

Table 2-2 *Alignment of cancer diagnoses between SNOMED-CT/ICD-10 and cancer diagnoses in the National Cancer Registry. All percentages are row %.*

Cancer Subtype	Alignment of SNOMED-CT generated cancer subtypes with National Cancer Registry (n, row %)	Alignment of ICD-10 generated cancer subtypes with National Cancer Registry (n, row %)
Leukaemias	881/881 (100)	1029/1029 (100)
Lymphomas	730/730 (100)	745/795 (93.8)
CNS Tumours	606/608 (99.7)	708/720 (98.3)
Sarcomas	391/394 (99.2)	415/426 (97.7)
Neuroblastomas	84/88 (95.4)	87/110 (78.8)
Renal Tumours	157/157 (100)	186/188 (98.9)
Hepatic Tumours	40/40 (100)	30/30 (100)
Gonadal Germ Cell Tumours	120/126 (95.2)	328/330 (99.4)

2.4.4 Evaluation of False Positive Diagnoses

Where a cancer diagnosis was not registered on the National Cancer Registry, there were records of 564 (11.3% of all identified cancer diagnoses) and 136 (3% of all identified cancer diagnoses) cancers within the GP and HES datasets, respectively (Table 2-3).

The majority of records only identified on the GP dataset were CNS tumour (n=150/564, 26.6%) and sarcoma (n=245/564, 43.4%) diagnoses. Specifically for sarcoma diagnoses, 137 (137/245, 56%) records had a SNOMED-CT codes of “fibrohistiocytic neoplasm (morphologic abnormality)” (Appendix Chapter 9, Supplementary Table 5). The majority of records only identified on the HES dataset were CNS tumour diagnoses (n=30/136, 22%).

Table 2-3 *Evaluation of false positives by cancer subtype on GP and HES datasets compared to National Cancer Registry used as gold standard. Any cell that has less than 5 records is depicted as <5.*

Cancer Subtype	**Cancer diagnoses only recorded on GP dataset (n=564) (column %)**	**Cancer diagnoses only recorded on HES dataset (n=136) (column %)**
Leukaemias	17 (3.0)	19 (14.0)
Lymphomas	30 (5.3)	19 (14.0)
CNS Tumours	150 (26.6)	30 (22.0)
Sarcomas	245 (43.4)	23 (16.9)
Neuroblastomas	39 (6.9)	17 (12.5)
Renal Tumours	51 (9.0)	<5
Hepatic Tumours	<5	<5
Gonadal Germ Cell Tumours	32 (5.7)	23 (16.9)

2.4.5 Evaluation of false negative diagnoses

Where a cancer diagnosis was not registered on the GP dataset, there were records of 1,126 out of 4,351 (25.9% of total) cancer diagnoses on the National Cancer Registry, with the majority belonging to leukaemia, sarcoma and CNS tumour diagnoses (Table 2-4). Where a cancer diagnosis was not registered on the HES dataset, there were records of 508 out of 4,351 (11.6% of total) cancer diagnoses on the National Cancer Registry, with the majority belonging to CNS tumour and sarcoma diagnoses (Table 2-4).

Table 2-4 *Evaluation of false negatives by cancer subtype on GP and HES datasets compared to National Cancer Registry used as gold standard*

Cancer Subtype	**Cancer diagnoses not recorded on GP dataset (n=1,126) (column %)**	**Cancer Diagnoses not recorded on HES dataset (n=508) (Column %)**
Leukaemias	219 (19.5)	72 (14.2)
Lymphomas	154 (13.6)	88 (17.3)
CNS Tumours	237 (21.1)	138 (27.2)
Sarcomas	170 (15.1)	133 (26.2)
Neuroblastomas	52 (4.6)	32 (6.3)
Renal Tumours	41 (3.6)	14 (2.8)
Hepatic Tumours	12 (1.1)	7 (1.4)
Gonadal Germ Cell Tumours	241 (21.4)	24 (4.7)

2.4.6 Evaluation of Dates of Cancer Diagnosis

There were 3,888 cancer diagnoses that had a recorded date on the National Cancer Registry and either the GP or HES dataset. The median difference of date-of-diagnosis between National Cancer Registry and the GP dataset was -7 days [IQR -20 days, 0 days] and between the National Cancer Registry and HES dataset was 0 days [IQR -6 days, +2 days]. There was variability in dates of cancer diagnoses across different subtypes (Table 2-5).

Table 2-5 *Dates of cancer diagnosis by cancer subtype on GP and HES datasets compared to the gold-standard National Cancer Registry. IQR= interquartile range. (+) represents subsequent to National Cancer Registry date of diagnosis and (–) represents earlier date to National Cancer Registry date of diagnosis.*

	Difference in cancer diagnosis dates on GP dataset (days, IQR)	**Difference in cancer diagnosis dates on HES dataset (days, IQR)**
Leukaemias	+2 (-1, +12)	-1 (-2, 0)
Lymphomas	+12 (+4, +23)	0 (-1, +17)
CNS Tumours	+7 (-1, +21)	-1 (-5, 0)
Sarcomas	+10 (0, +25)	+4 (-1, +27)
Neuroblastomas	+1 (-4, +16)	-2 (-5, +24)
Renal Tumours	+5 (-2, +17)	-2 (-4, 0)
Hepatic Tumours	+6 (-3, +22)	-3 (-6, 0)
Gonadal Germ Cell Tumours	+13 (0, +45)	0 (-1, +2)

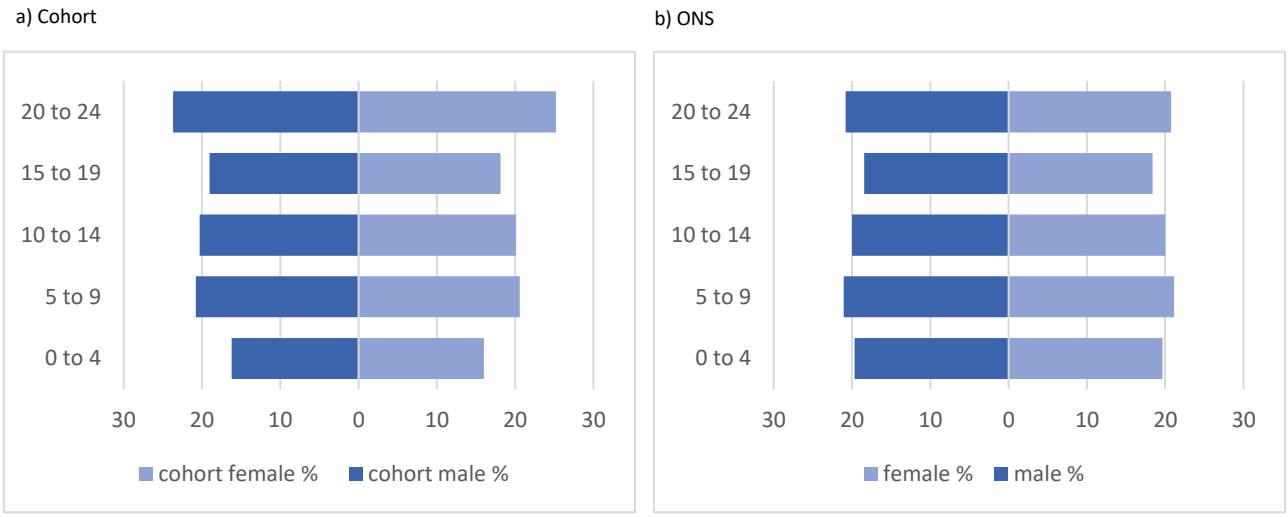
2.4.7 Cohort Characteristics for Children, Teenagers and Young Adults

Overall, there were 5,239,307 children, teenagers and young adults aged 0-24 years registered with a general practitioner between 1st January 1998 and 31st December 2019. The median follow-up time was 3.9 years [interquartile range (IQR) 2.0 to 8.0 years]. This is 31% of the mid-2019 population estimate for 0-24 year-olds reported by ONS (n=16,770,184) [60].

The baseline demographic characteristics of the cohort are summarised in Table 2-6 in the section titled *Cancer Cohort Characteristics*, below. The cohort population structure by age and sex is detailed in Figure 2-4 and show broadly similar population structure. Specifically, the percentage of males (n=2,678,405, 51.1%) and females (n=2,560,902, 48.9%) was representative of the mid-2019 population structure by sex reported by the ONS, in which the

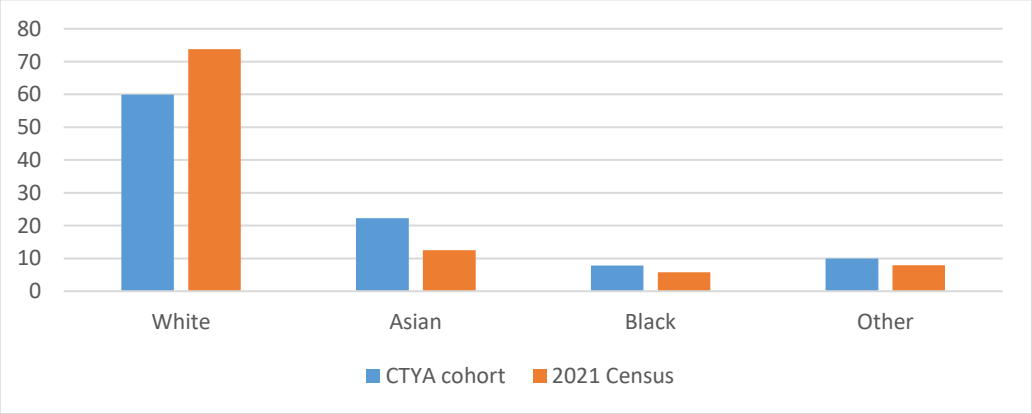
percentage of males (n=8,605,577) and females (n=8,164,607) aged 0-24 years was 51.3% and 48.7%, respectively [60].

Figure 2-4 *Proportion (%) by 4-year age band and sex for a) QResearch CTYA cohort, b) ONS population estimates.*



The proportion of each ethnic group within the CTYA cohort, alongside national estimates for this age group provided by the 2021 Census are summarised in Figure 2-5. Compared to national estimates [64], there was a lower proportion of the White ethnic group and a higher proportion of ethnic minority groups in the CTYA cohort. This persisted across the different geographical regions (Appendix Chapter 9, Supplementary Figure 9-2).

Figure 2-5 *Proportion (%) of ethnic categories for a) QResearch CTYA cohort, b) 2021 census population estimates [64].*



2.4.8 Cancer Cohort Generation

Given the significant proportion of false positives on the GP dataset (n=564, 11.3%) and broad SNOMED-CT codes that do not distinguish benign/malignant status, a decision was made to exclude any cancer diagnosis that was only recorded in the GP dataset. Accordingly, 4,389 had a diagnosis of one of the selected cancers within the full cohort out of 5,238,743 young people.

2.4.9 Cancer Cohort Characteristics

The baseline demographic characteristics of the 4,389 cancer cases in the cohort are summarised in Table 2-6. Compared to the full cohort of 0-24 year olds, cases were younger ($p<0.001$) and were more commonly observed in males than females ($p<0.001$), lower deprivation levels (e.g., lower Townsend score) compared to higher deprivation levels (e.g., higher Townsend score) ($p<0.001$) and the White ethnic group ($p<0.001$) compared with other ethnicities. This is in line with published national CTYA cancer statistics [6]. Of the cancer cases, 4,061 (92.5%) were 1-year survivors and 3,606 (82.1%) were 5-year survivors. These are also in line with national 5-year survival rates between 1997 and 2016, which was reported to be 82.6% [1].

Table 2-6 *Baseline characteristics of full cohort of 0-24 year olds and cancer cases. All percentages are column percentages. Age in years is rounded to the nearest integer above.*

Characteristics	**All population (n=5,238,743)**	**All cancers (n=4,389)**	**P-value**
Age at cohort entry			
Median ± IQR (years)	10 [2,20]	5 [1,13]	p<0.001
Sex			
Male	2,678,130 (51.1)	2,607 (59.4)	p<0.001
Female	2,560,613 (48.9)	1,782 (40.6)	
Ethnicity			
White	1,312,975 (25.1)	1,145 (26.1)	P<0.001
Asian	487,879 (9.3)	376 (8.6)	
Black	171,485 (3.3)	106 (2.4)	
Other	219,273 (4.2)	146 (3.3)	
Missing	3,047,131 (58.1)	2,616 (59.6)	
Townsend Deprivation Score			
1 (most affluent)	1,016,700 (19.4)	1,081 (24.6)	P<0.001
2	1,021,486 (19.5)	950 (21.7)	
3	1,042,070 (19.9)	883 (20.1)	
4	1,069,154 (20.4)	782 (17.8)	
5 (most deprived)	1,061,662 (20.3)	652 (14.9)	
Missing	27,671 (0.5)	41 (0.9)	

The proportion of cancer subtypes, alongside their baseline demographics, are summarised in Table 2-7. In line with national reports, leukaemia, lymphoma and CNS tumours are the most common three cancers in CTYA [6]. As per previous reports, all subtypes except renal tumours show male preponderance and leukaemias, lymphomas, CNS tumours and gonadal germ cell tumour subtypes have greater number of cases in the lower deprivation levels (Townsend Deprivation Score 1 & 2) compared to higher deprivation levels (Townsend Deprivation Score 4 & 5) [70, 71].

Table 2-7 *Baseline characteristics of cases by cancer subtype. All percentages except proportion of total cancer are column percentages. Age in years is rounded to the nearest integer above.*

	Leukaemia (n=1,127)	**Lymphoma (n=919)**	**CNS Tumours (n=933)**	**Sarcomas (n=637)**	**Neuro blastomas (n=164)**	**Renal Tumours (n=167)**	**Hepatic Tumours (n=38)**	**Gonadal Germ Cell Tumours (n=404)**
Proportion cancer (row %)	25.7	20.9	21.3	14.5	3.7	3.8	0.9	9.2
Age at cohort entry								
Median ± IQR (years)	2 [1,9]	10 [3,17]	4 [1,10]	8 [2,13]	2 [1,5]	2 [1,5]	3 [1,12]	15 [9,20]
Sex								
Male	641 (56.9)	544 (59.2)	538 (57.7)	351 (55.1)	97 (59.2)	64 (38.3)	21 (55.3)	351 (86.9)
Female	486 (43.1)	375 (40.8)	395 (42.3)	286 (44.9)	67 (40.8)	103 (61.7)	17 (44.7)	53 (13.1)
Ethnicity								
White	277 (24.6)	249(27.1)	253 (27.2)	173 (27.2)	45 (27.4)	43 (25.7)	12 (31.6)	93 (23.0)
Asian	116 (10.3)	93 (10.1)	59 (6.3)	51 (8.0)	12 (7.3)	5 (3.0)	6 (15.8)	34 (8.4)
Black	28 (2.5)	27 (2.9)	26 (2.8)	11 (1.7)	6 (3.6)	7 (4.2)	<5	<5
Other	42 (3.7)	31 (3.4)	26 (2.8)	16 (2.5)	6 (3.6)	14 (8.3)	<5	10 (2.5)
Missing	664 (58.9)	519 (56.5)	569 (61.0)	386 (60.5)	95 (57.9)	98 (58.6)	19 (50.0)	266 (65.8)
Townsend Deprivation Score								
1 (most affluent)	269 (23.9)	232 (25.2)	241 (25.8)	153 (24.0)	40 (24.4)	44 (26.3)	10 (26.3)	92 (22.8)
2	270 (24.0)	205 (22.3)	200 (21.4)	122 (19.2)	33 (20.1)	30 (17.9)	<5	87 (21.5)
3	237 (21.0)	174 (18.9)	191 (20.5)	120 (18.8)	38 (23.2)	31 (19.0)	8 (21.0)	84 (20.8)
4	169 (15.0)	160 (17.4)	177 (19.0)	119 (18.7)	33 (20.1)	30 (17.9)	11 (29.0)	83 (20.5)
5 (most deprived)	171 (15.1)	142 (15.5)	113 (12.1)	116 (18.2)	18 (11.0)	32 (19.2)	5 (13.1)	55 (13.6)
Missing	11 (1.0)	6 (0.7)	11 (1.2)	7 (1.1)	<5	<5	<5	<5

2.4.10 Cancer Incidence Rates

The crude incidence rates per million population by sex for each cancer subtype in the cohort are comparable to national reports of CTYA cancer incidence rates in England (Table 2-8) [1].

Table 2-8 *Crude incidence rates per million population by sex in generated cohort and nationally reported world age-standardised incidence rates per million population by sex in England.*

Cancer type	**Crude cohort incidence rate per million population in cohort (0-24 years)**	**World age-standardised incidence rate per million population (0-24 years)[6]**
Leukaemias, myeloproliferative disease and myelodysplastic diseases	Female: 35.0 (95%CI 32.0-38.2)	Female: 36.1
	Male: 40.6 (95% 37.6-44.0)	Male 44.1
Lymphomas & reticuloendothelial neoplasms	Female: 27.0 (95%CI 24.4-29.8)	Female: 24.2
	Male: 34.5 (95% 31.7-37.5)	Male 33.5
CNS Tumours & miscellaneous intracranial and intraspinal tumours	Female: 28.6 (95%CI 25.8-31.4)	Female: 33.6
	Male: 34.1 (95% 31.3-37.1)	Male 36.9
Sarcomas (bone, soft tissue and other extraosseous tumours)	Female: 20.6 (95%CI 18.3-23.0)	Female: 16.5
	Male:22.2 (95% 20.0-24.7)	Male 20.9
Neuroblastoma & other peripheral nervous system tumours	Female: 4.9 (95%3.8-6.2)	Female: 6.5
	Male: 6.2 (95%4.5-6.9)	Male 6.9
Renal tumours	Female: 7.4 (95%6.1-8.9)	Female: 7.0
	Male: 4.1 (95%3.2-5.2)	Male 6.0
Hepatic tumours	Female: 1.2 (95% 0.8-2.0)	Female: 1.6
	Male: 1.4 (95% 0.9-2.2)	Male 1.9
Gonadal Germ Cell Tumours	Female:3.8 (95%CI 2.9-5.0)	Female: 3.9
	Male: 22.3 (95%20.0-24.7)	Male: 22.8

2.5. Cohorts used in the thesis

Each study carried out in this thesis uses variations of the CTYA and cancer cohort. These variations per Chapter are detailed next and summarised in Figure 2-6 below to provide a framework for the rest of the thesis.

In Chapter 3, in order to explore clinical features associated with a subsequent cancer diagnosis, a nested case-control study was used. From the eligible full cohort of 4,389 cancer cases, 113 cases [TYA abdominal tumours (n=74) and childhood gonadal germ cell tumours (n=38)] were excluded given their rarity and thus lack of power to identify associations between clinical features and a diagnosis. Subsequently, from a total of 4,276 cancer cases, each cancer case was year-of-birth, sex and general-practice matched to up to 20 controls using incidence density sampling, where controls were selected from 5,238,743 CTYA (i.e., full cohort). This nested, matched case-control study design with incidence density sampling was chosen in place of using a full cohort/matched case-cohort for two reasons. First, by setting the index date as the date of cancer diagnosis, it better allowed to define the 6-month study period preceding the diagnosis to explore relevant clinical features. Second, it was computationally more effective whilst also still providing risk estimates that the full cohort analysis would provide.

In Chapter 4, in order to develop risk prediction models for CTYA cancers, the full CTYA cohort was utilised, where each CTYA entered the study either at the cohort entry date or the date in which they had a first-recorded cancer-relevant symptom. They were subsequently followed-up for 6 months. This was chosen over using a nested case-control study, as follow-up from symptom-onset or cohort entry could easily be defined and the much larger sample size of the full cohort provided sufficient power to develop risk prediction models with several predictor variables. As logistic regression analysis was used, CTYA who did not have 6 month follow up from their respective study entry date, were excluded (n=377,320). Accordingly, out of the full cohort of 5,238,743 CTYA, 4,861,423 CTYA were eligible and 617 developed a cancer within the 6 month follow-up.

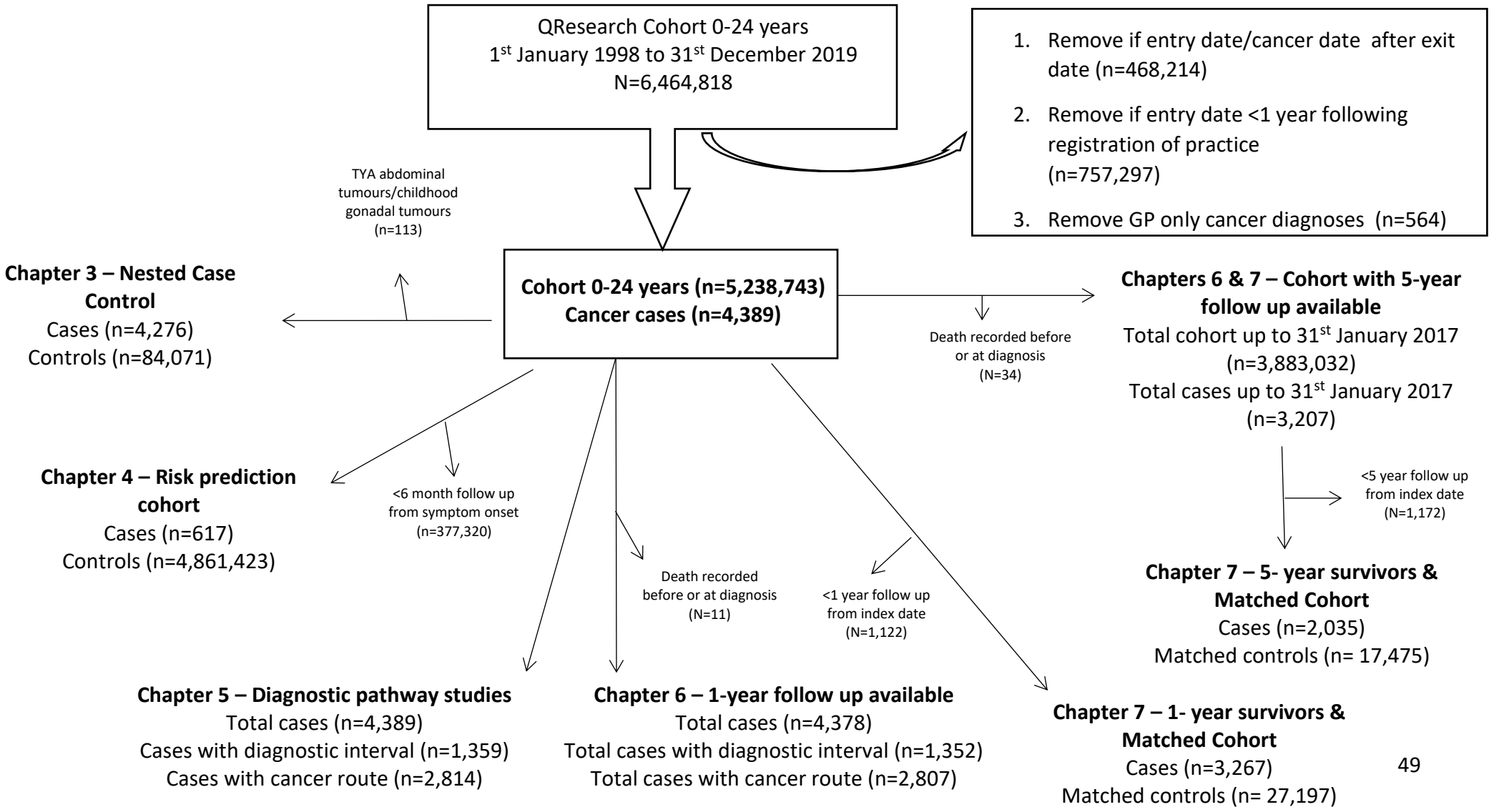
In Chapter 5, in order to explore factors associated with diagnostic pathways, the full cohort of 4,389 cancer cases were used. Of those, 1,359 had a recorded diagnostic interval (i.e.,

time interval from first recorded cancer-relevant symptom to diagnosis) and 2,814 had a recorded route-to-cancer-diagnosis.

In Chapter 6, in order to explore the association between diagnostic pathways and long-term mortality outcomes, cancer cases were followed up for either 1 year or 5 years following the date of cancer diagnosis. At 1 year, out of 4,389 cases in the full cohort, 11 had a date of death before or at diagnosis, and were thus excluded from analysis. Accordingly, there were 4,378 cases in analyses. At 5 years, out of 4,389 cases in the full cohort, 1,148 cases did not have 5-year follow-up and 34 had a date of death before or at diagnosis, and thus were excluded. Accordingly there were 3,207 cases in analyses.

In Chapter 7, in order to identify the risk of developing mental health disorders in cancer survivors compared to the non-cancer cohort, a year-of-birth and sex matched case-cohort study was carried out, where each case was matched to up to 10 controls. Out of the 4,389 cases, 3,267 were 1-year cancer survivors and had sufficient follow-up, and 2,035 were 5-year survivors and had sufficient follow-up. The controls were selected from either the full cohort with 1 year follow up (n=5,238,743) or the full cohort with 5-year follow up (n=3,883,032).

Figure 2-6 Summary flow diagram depicting the variations of the CTYA and cancer cohorts used in each chapter of the Thesis.



2.6. Discussion

2.6.1 Summary of Principal Findings

There are two important principal findings identified in this Chapter. First, the CTYA population generated through the QResearch Database appears to be representative of the CTYA population in England, and thus has the potential to be utilised in epidemiological studies to provide broadly generalizable conclusions for young people in England. Second, compared to the National Cancer Registry, GP and HES datasets individually appear to under-capture young people with a cancer diagnosis and there is a proportion of misclassification in those who are captured, particularly within the HES dataset. Thus, CTYA with cancer are most likely to be comprehensively studied by utilising both GP and HES datasets or all three linked datasets on the QResearch Database. This is supported by the observation in this Chapter that cancer incidence rates in the CTYA population generated through using all three linked datasets are closely reflective of national estimates.

2.6.2 Findings and Implications

Representativeness of primary care databases in the UK, including Clinical Practice Research Datalink (CPRD), QResearch Database and OpenSAFELY, have previously been reported [72-74]. The representative cohort of young people generated in this Chapter provides further support to these reports. Collectively, this suggests that routinely collected EHR can be utilised to provide population-level, broadly generalizable conclusions in UK-based epidemiological studies.

The cohort of young people generated by QResearch Database was 31% of the mid-2019 population estimates in England. This is slightly higher than previous QResearch Database reports, where approximately 25% of the population England has been found to be represented [58]. This may be due to the largest number of practices contributing to QResearch Database being in London and the North-West of England, where higher

proportions of young people reside [60]. The age and sex demographics were broadly similar to those reported by the ONS mid-2019 population estimates, providing reassurance of the representativeness of the cohort [60]. In the CTYA cohort, there was a lower proportion of the White ethnic group whilst slightly higher representation of the other ethnic minority groups compared to ONS Census 2021 estimates [64], which persisted across different geographical regions. This is most likely to be due to the variable recording of ethnicity in routinely collected EHR and similar results have been found in studies involving CPRD and HES datasets [75].

The completeness and accuracy of cancer diagnostic codes in GP and HES datasets, in relation to the National Cancer Registry, have only been explored in adult cancers and not been investigated for CTYA cancers in the UK, to date. Records of common adult cancers, such as lung and colorectal cancers, in the GP and HES datasets have been shown to have high sensitivity (>90%) and positive predictive value (>90%), however, less common adult cancers, such as CNS tumours and leukaemias, were found to have lower sensitivity (70-80%) and predictive values (80-90%) in these data sources(84-86). Similar to these less common adult cancers, the sensitivities of GP and HES datasets for CTYA cancers in this Chapter were found to be 72% and 87%, respectively. This suggests there is a proportion of cancers that are not being captured by each dataset (i.e., high false negatives). Possible explanations for lower sensitivity include administrative error (e.g., not recorded) and diagnoses made through private healthcare. Specifically for the GP dataset, there is also the possibility of misalignment of SNOMED-CT codes with ICD-O-3 codes where some cancers are not being accurately captured through the SNOMED-CT code alone. There is also the possibility that a patient's record is not being transferred correctly when changing practitioners within the GP dataset. Importantly, however, sensitivity improves substantially to 94% when GP and HES datasets are used together to capture CTYA cancers.

The validation study in this Chapter identified that the GP dataset had a significant proportion of false positives (11% total) with the majority found within CNS tumour and sarcoma subtypes. One explanation for this may be due to misalignment of SNOMED-CT codes with ICD-O-3 codes. For example, the majority of false positive sarcoma diagnostic codes were for "fibrohistiocytic neoplasm"; this does not allow to distinguish benign/malignant behaviour and in turn may result in the inclusion of benign neoplasms in the GP dataset that are not included within the ICCC-3 framework used in the National Cancer Registry. Furthermore, the validation study identified that the HES dataset had a notable proportion of misclassified diagnostic codes (3%), with the majority of lymphomas being incorrectly categorised as leukaemias. This is most likely to be due to administrative recording errors of individual haematopoietic neoplasms, which sometimes have overlapping definitions and treatment regimens, such as lymphoblastic lymphoma and acute lymphoblastic leukaemia [76].

Overall, given the higher proportion of false positives and difficulty in identifying benign/malignant status of cancers using the SNOMED-CT codes, a decision was made to remove CTYA cancers that were only captured through the GP dataset. Subsequently, the decision to include cancers captured through all three linked datasets (i.e., cancers recorded on 1. registry alone, 2. registry & HES, 3. registry & GP, 4. HES & GP, 5. HES alone) was made to identify and categorise CTYA cancers within the cohort. Comparable incidence rates in the cohort compared to national estimates provide strong support for these decisions.

2.6.3 Strengths and Limitations

The main strength of this Chapter is the use of high-quality linked electronic healthcare data provided through the QResearch Database, which has made comparisons across GP, hospital and registry datasets as well as with national estimates, possible. These comparisons have not previously been carried out specifically for CTYA and this work provides novel and comprehensive information on the representativeness of these linked datasets for young

people. Furthermore, this Chapter presents the first study to assess the completeness and accuracy of GP and HES datasets for CTYA cancers in relation to the National Cancer Registry, which will have important implications for epidemiological research in this field in the future. There are, however, several notable limitations. First, as with the majority of routinely collected healthcare data, there is missing data on the QResearch Database, specifically across ethnicity, which may introduce bias and limit comparisons with national estimates. Findings in this Chapter suggest that the CTYA recording of White ethnicity is less in young people compared to other ethnic minority groups in the QResearch Database, which is similar to findings in the CPRD [75]. A second limitation related to this is the possibility of incorrect recording, i.e., misclassification, which may further bias results. The comparable age and sex representation of the CTYA cohort to the national estimates provides evidence that misclassification for these demographic factors is likely to be minimal. The validation study within this Chapter also provides evidence that misclassification of cancer diagnoses is also likely to be minimal. It is important to note, however, that although the National Cancer Registry was used as gold standard in this Chapter, it also has a very small proportion of missing data [62]. Future validation studies, focusing on comparing datasets to individual GP records and hospital records, would overcome these limitations and provide further insight.

2.6.4 Future Directions

Findings of this Chapter suggest that the CTYA cohort generated by the QResearch Database is representative of the general population in England and the three linked datasets, National Cancer Registry, primary and secondary care, provide a comprehensive capture of cancers within this age group. Overall, this provides strong support for their use to address the aims and objectives of this thesis. Accordingly, the next four chapters will use this CTYA population and cancer cohort in order to address the following research questions:

1) Which early and late clinical features (or feature combinations) encountered in primary care are associated with a subsequent diagnosis of CTYA cancer (Chapter 3)?

2) Can the risk of an undiagnosed CTYA cancer in primary care be estimated comprehensively through risk prediction models (Chapter 4)?

3) Which person- and health-system factors affect diagnostic pathways (Chapter 5)?

4) Do changes in diagnostic pathways influence mortality and long-term outcomes at 1- and 5-years following a cancer diagnosis (Chapters 6 & 7)?

Publications from Chapter:

Saatci D., Harnden A., Hippisley-Cox J. *Validity of childhood, teenage and young adult cancer diagnoses in primary and secondary care records: a population-based study using linked electronic healthcare records (in submission)*

Chapter 3 Clinical features associated with childhood, teenage and young adult cancers: a population-based nested case-control study

3.1. Background to chapter

Evidence for the presenting clinical features of CTYA cancer is largely derived from hospital-based observational studies [35], and as CTYAs with underlying cancers more commonly first present to their general practitioner (GP) in the UK [21], it is unclear if these hospital-based clinical presenting features fully represent the symptoms encountered in primary care. This is also reflected by reports that a high level of diagnostic uncertainty in CTYA cancer among GPs exist [30]. Thus, improving the evidence base for symptoms and signs associated with CTYA cancers is a priority to improve early recognition.

To date, there have been a limited number of primary care-based observational studies investigating the presenting features of CTYA cancers [26, 39, 56]. Two studies based on GP data in the UK provide insight into symptoms and signs associated with CTYA cancers and highlight re-attendance as an important factor associated with a subsequent cancer diagnosis [26, 39, 56]. These findings provide a valuable starting point, however, more research is needed to establish which symptoms present at earlier stages and investigate the association between symptom combinations and a subsequent cancer diagnosis. Furthermore, both studies have important limitations. First, identification of cancer diagnoses is limited to GP medical records, for which medical codes have not been validated for CTYA cancers and therefore their accuracy is not established. Furthermore, cancer diagnoses which are only recorded in hospital records could not be captured. Second, the database used in these studies, CPRD, was only representative of 8% of the population at the time of publication and

may, therefore, be unable to capture all relevant and rarer symptoms associated with CTYA cancers.

In line with these limitations, concerns were raised by paediatric oncologists about the reliance of national guidance on these limited number of observational studies in the UK, as it was felt that they were unable to capture all relevant clinical presentations [77]. Indeed, the National Institute for Health and Care Excellence (NICE) suspected cancer guidelines introduced in 2015 in the UK specifically highlight the need for further observational studies to determine the symptomology of cancer in CTYA who attend primary care as an urgent research priority [78]. Accordingly, using a large representative population-based linked EHR from the UK, the overarching aim of this study is to identify primary care-based clinical features and feature-combinations associated with a diagnosis of CTYA cancers and establish a timeframe for these features, using a nested case-control study design.

The specific objectives were:

1. Quantify the association between clinical features that present to primary care and a subsequent diagnosis of CTYA cancer.
2. Investigate if combinations of clinical features provide stronger associations with a subsequent diagnosis of CTYA cancer.

3.2. Methods

3.2.1 Data Source

This Chapter utilised the open cohort of CTYA and selected cancer cases generated in Chapter 2. In brief, this is 4,389 young people with a cancer diagnosis from a cohort of 5,238,743 young people between the ages of 0-24.

3.2.2 Identification of Cases and Controls

Cancers were categorised into the commonest subtypes in this age group, which are defined by the National Cancer Registration and Analysis Service and in accordance with the ICCC-3[2]: 1) Leukaemias, myeloproliferative and myelodysplastic diseases (n=1,127), 2) lymphomas and reticuloendothelial neoplasms (n=919), 3) central nervous system and intraspinal tumours (n=933), 4) soft tissue and bone sarcomas (n=637), 5) abdominal tumours (renal tumours, neuroblastomas, hepatic tumours) (n=369) and 6) gonadal germ cell tumours (n=404).

Each case was matched to up to 20 controls without cancer from the cohort of 5,238,743 young people, by year-of-birth, sex, GP and calendar time using incidence density sampling [79]. Although 10 controls provided sufficient power for this study, an a priori decision was made to have 20 controls because of the rarity of both the outcome and relevant symptoms, such as organomegaly. Age and sex were used as matching variables to allow for adequate numbers in each age and sex strata and GP was selected as a matching variable in order to accommodate clustering of coding approaches. Incidence density sampling was used as this allows to obtain a direct estimate of the rate ratio (i.e., giving estimate of risk) and the estimates are not biased by differential loss to follow up among the exposed and unexposed [79]. The index date was the date of cancer diagnosis.

3.2.3 Clinical Feature and risk factor selection

Symptoms and signs that occurred up to 6 months prior to the index date were included. As previous evidence suggests cancer cases in this age group present more frequently up to 12 months before diagnosis and most commonly 3 months before diagnosis, an a priori mid-point was selected at 6 months [21, 25].

Clinical feature selection for each cancer type was carried out using a combination of literature review [35-37], patient representative input and national guidelines (HEADSMART[30], Grace

Kelly Childhood Cancer Trust [80], NICE Suspected Cancer)[78]. Patient representatives on the panel of this study were two CTYA cancer survivors, who had ongoing broader links with the Youth Advisory Group at the Teenage Cancer Trust. Representatives were provided the clinical features identified through the literature review/guidance and their broader experiences were discussed and additional clinical features were included. 51 clinical features were selected overall, 28 for leukaemias, 28 for lymphomas, 27 for CNS tumours, 20 for sarcomas, 15 for abdominal tumours and 14 for gonadal germ cell tumours (Table 3-1). One clinical feature, eczema, not considered to be relevant to cancers, was included to investigate recording bias.

Table 3-1 *Clinical features identified through literature search, patient input and national guidance, associated with CTYA cancer per cancer subtype*

Clinical Features	Leukaemias	Lymphomas	CNS Tumours	Sarcomas	Abdominal Tumours	Gonadal Germ Cell Tumours
Systemic / Endocrine	Fever Weight loss Anorexia Night sweats Lethargy/tiredness	Fever Weight loss Anorexia Night sweats Lethargy/tiredness Pruritus	Fever Weight loss Anorexia Lethargy/tiredness Irritability Developmental delay Precocious puberty	Fever Weight loss Anorexia Lethargy/tiredness	Fever Weight loss Anorexia Lethargy/tiredness	Fever Weight loss Anorexia Lethargy/tiredness Menstrual abnormalities Precocious puberty
Masses	Lymphadenopathy Swelling (head or body) Lump (head or body) Organomegaly Hepatomegaly Splenomegaly Hepatosplenomegaly Abdominal mass	Lymphadenopathy Swelling (head or body) Lump (head or body) Organomegaly Hepatomegaly Splenomegaly Hepatosplenomegaly Abdominal mass	-	Swelling (head or body) Lump (head or body) Abdominal mass	Abdominal mass	Abdominal mass Testicular mass/swelling
Cardiorespiratory	Cough Chest pain Breathlessness URTI/sore throat	Cough Chest pain Breathlessness Stridor URTI/sore throat	-	Cough Chest pain Breathlessness	-	Chest pain
Gastrointestinal	Abdominal pain Vomiting	Abdominal pain Vomiting	Vomiting	Abdominal pain Vomiting Constipation Haematuria	Abdominal pain Vomiting Constipation Haematuria	Abdominal pain Abdominal distension Vomiting Constipation Haematuria
Neurological	Headache Loss of consciousness	Headache Loss of consciousness	Headache Dizziness Loss of consciousness	Headache	Headache Cranial nerve palsies Unsteady gait	

			Unsteady gait Squint Visual acuity problems Abnormal eye movements Diplopia Papilloedema Optic atrophy Exophthalmos Hemiparesis Hemiplegia Macrocephaly/large head Seizures Torticollis Voice changes Cranial nerve palsies			
Haematological	Bruising/bleeding Anaemia/pallor	Anaemia/pallor	-	-	-	
Musculoskeletal	Limb pain Joint pain Limp	Limb pain Joint pain	Limp	Limb pain Joint pain Limp	Limb pain Joint pain Limp	
Dermatological	Rash Skin nodules/infiltrations	Rash Skin nodules/infiltrations	-	Rash Skin nodules/infiltrations	-	

Literature review also identified two risk factors, deprivation and ethnicity [81, 82]. The deprivation score used in the study is Townsend Deprivation Score/Quintile which is available on QResearch Database. As described in Chapter 2, Townsend Deprivation Score is an area-level continuous score based on an individual's postcode; factors that included unemployment, non–car ownership, non–home ownership, and household overcrowding, are measured for a given area of approximately 120 households, via the 2011 Census of England and Wales and combined to give a Townsend score for that area, with the first quintile representing the lowest deprivation level and the fifth quintile representing the highest deprivation level [65]. Self-assigned ethnicity was re-grouped into two categories White and Other due to the small number of events per ethnic group defined by the 2021 Census [64].

3.2.4 Coding of clinical features

All relevant SNOMED-CT codes were grouped together to define each clinical feature and are available in the Supplementary Appendix (Appendix Chapter 9, Supplementary Table 9-6).

3.2.5 Subgroup Analysis

Subgroup analyses were carried out for 1) pre-selected age categories (childhood 0-14 years, TYA 15-24 years) 2) each included cancer subtype and, 3) for each cancer subtype, if sufficient numbers, by the age categories (e.g., childhood leukaemia and TYA leukaemia). These categories were selected for two reasons: 1) to explore differing patterns of presentation and 2) to ensure comparisons with previous studies, which use these subgroups, is possible (25, 26). As most abdominal tumours occur in children and gonadal germ cells in TYA, these cancer types were explored only in their respective age group [6].

3.2.6 Statistical Analysis

For descriptive analyses, continuous variables were presented as means (medians if non-normal) and categorical variables were presented in frequencies and comparisons between cancers and the general population were carried out using t-tests (Mann-Whitney U if non-normal) and $\chi 2$ tests, respectively.

3.2.6.1 Objective 1: Quantification of associations between clinical features and subsequent cancer

The association between individual clinical features and a subsequent cancer diagnosis was initially explored by calculating odds ratios (OR) using univariable conditional logistic regression. Any feature with a p-value above 0.05 in univariable analysis was selected to be included in multivariable analyses. Subsequently a full model was fitted with all features identified through univariable analyses and were retained in the multivariable analysis only if a p-value threshold of 0.01 was met and the ORs were ≤0.9 or ≥1.1. Although there are several methods for variable selection, this approach was chosen in this Chapter as it provided a balance between clinical relevance and statistical magnitude of clinical features and accounted for multiple testing. First, as described in section 3.2.3, there was a high number of potential candidate clinical features identified through literature review. Thus an initial univariable step was taken to assess the individual associations between these candidate clinical features and a cancer diagnosis. Subsequently these individual associations were contextualised with clinical knowledge. Further, to consider weak individual associations between clinical features and a cancer diagnosis, the p-value threshold in univariable analyses was kept higher than multivariable analyses. Second, having regression coefficient thresholds as well as statistical p-value thresholds was decided to better integrate clinical relevance alongside statistical significance compared to automated selection methods, such as stepwise

selection [83]. Further, it was assumed a priori that some clinical features will be correlated to each other based on clinical knowledge (e.g., fever and lymphadenopathy), which may increase the risk of bias in automated selection methods[83]. To assess this, clinically plausible interactions based on paediatric and general practice expertise, were considered and included 1) anaemia/pale & bruising, 2)limb & joint pain, 3) fever & lymphadenopathy, 4) fever & cough, 5) lump on body & lymphadenopathy, 6) head/neck lump & lymphadenopathy, 7) chest pain & cough, 8) headache and visual acuity changes, 9) limb/joint pain & limp and 10) abdominal pain and constipation. To account for clustering of young people within GPs, clustered standard errors were used in regression analyses. Clinical features that were previously reported in population-based primary care studies [25,26] were categorised as "Already Known" and any unreported clinical features in primary care were categorised as "New Associations".

Positive predictive values (PPVs) were calculated using the inverse sampling fraction of the nested cohort, as follows [84]:

$$PPV = \frac{True\ Positive}{True\ Positive + \left(\left({}^{1}/_{sampling\ fraction}\right) \times False\ Positive\right)}$$

	Disease present	**Disease absent**
Symptom present	True positive (a)	False Positive (b)
Symptom absent	False negative (c)	True negative (d)

Associated confidence intervals for the predictive values were calculated using a bootstrap procedure described by van Zaane et al., 2012 [84]. This approach comprised of estimating the PPV in 500 bootstrap samples and the PPVs corresponding to the 2.5^{th} and 97.5^{th} percentile of the 500 bootstrap estimates were taken as the upper and lower bound of the 95th confidence interval. This approach has been shown to perform better compared to other confidence interval estimation for PPVs in nested case-control studies [85].

3.2.6.2 Objective 2: Diagnostic intervals and early clinical features associated with subsequent cancer diagnosis

The time interval between the date of statistically significant symptom/signs identified in Objective 1 and the index date (i.e. the diagnostic interval) was explored for each clinical feature and an overall and a cancer-specific median diagnostic interval was calculated and used to split symptoms into two timeframes: “late” and “early”. Thus, the early timeframe was defined as the time between the median and 6 months prior the index date, whilst the late timeframe was the time between the median and index date. Multivariable conditional logistic regression excluding the late timeframe was carried out to identify relevant early symptoms. Features in multivariable analyses were retained only if a p-value threshold of 0.01 was met and the ORs were ≤0.9 or ≥1.1, as per Objective 1.

3.2.6.3 Objective 3: Symptom combinations and a subsequent cancer diagnosis

Two methodological approaches were taken to explore if and how symptom combinations are associated with cancer diagnoses.

The first approach was clinically-driven to ascertain if any identified “early” symptom could have a stronger association if it was combined with other statistically significant symptoms. To achieve this, each “early” symptom identified in Objective 2 was combined to another statistically significant symptom identified in Objective 1 to generate paired combinations.

There were no restrictions to the timing of the symptoms, where the symptoms could co-occur on the same date, or be on different dates within the 6-month period. Odds ratios for paired combinations were calculated using conditional logistic regression.

The second approach was exploratory and utilised a two-step method: first using an unsupervised statistical method known as principal component analysis (PCA) to generate principal component scores and second using logistic regression analysis to explore the association between principal component scores and a cancer diagnosis. Briefly, PCA aims to reduce the dimension of data whilst still retaining all important information and can therefore allow for the identification of patterns within data [86]. This is achieved by generating new uncorrelated variables, known as principal components, from the linear combination of the original variables in the dataset whilst capturing the maximum amount of variance from the original data[86]. This is accomplished by computing eigenvectors and eigenvalues from the data's correlation matrix. Each variable is given a 'loading score' within the principal component. The 'contribution' of the variables from the original data to the principal component is considered through this loading score, where the higher the score, the larger the contribution. Exploring variables with the highest loading scores allows for the observation of patterns within the data.

Accordingly, all clinical features which were statistically significant in univariable regression models in Objective 1 were included as the original variables and PCA was used to reduce them into a smaller number of principal components. A proportion of variance plot, which plots the proportion of the total variance explained by each principal component, was used to identify the number of principal components required to explain 50% of the total variance within the original data and thus be evaluated further. Clinical features within each principal component with a loading score above 0.30 was deemed to provide significant contribution

to the principal component and were used to describe patterns in clinical features which may be relevant to cancers in this age group. Sampling adequacy was assessed using Kaiser-Meyer-Olkin (KMO) measure. This measure assesses whether there is sufficient commonality between the variables to warrant a PCA, with a value of 0.6 or above indicating sufficient commonality [87].

Next, loading scores of all clinical features were used to calculate the overall standardised principal component scores. The association between each principal component score and a subsequent cancer diagnosis was then explored using univariable conditional logistic regression, adjusted for ethnicity, and deprivation level.

In order to gain further insight into clinical utility, the symptoms contributing most to the selected principal components were subsequently combined (e.g., for principal component 1, limb and joint pain were categorised as a symptom combination). Proportions of cases and controls with these symptom combinations were summarised and then used to calculate positive predictive values, using the same methodology described in Objective 1.

3.2.6.4 Handling Missing Data

Previously published studies using QResearch Database consistently report missing data for ethnicity and the Townsend Deprivation Score [88, 89]. In this dataset, ethnicity was missing for 58% and Townsend Quintile was missing for 0.4% of the total population. Thus, multiple imputation with chained equations was carried out to impute missing values for ethnicity and Townsend Quintile under the missing at random assumption [90]. 50 imputed datasets were generated to account for % missingness [90]. Imputation models included all candidate variables, and the primary outcome variables; statistical models were developed on each of the 50 imputed data sets and the estimates pooled using the Rubin rules[91].

All analyses were carried out using Stata versions 17 and 18 [92]. Analyses adhere to the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) [93] and Reporting of studies Conducted using Observational routinely collected health Data (RECORD) guidelines [94].

3.2.7 Sample Size Calculations

Prior to carrying out analyses, sample size calculations to identify the minimum required sample size to detect odds ratios of 2 at power of 0.8 and significance level of 0.05 for 1) a single clinical feature and, 2) two interacting clinical features were carried out for the full study sample and separately for childhood (0-14 years), TYA (15-24 years), and each cancer subtype (Appendix Chapter 9, Supplementary Tables 9-7 & 9-8). An odds ratio of 2 was selected as this was the smallest reported odds ratio in the most recent primary care-based studies exploring symptomology of CTYA cancers [26, 56]. Given the total sample size of 88,436 in this study, there is sufficient sample size to detect the selected odds ratio.

3.3. Results

3.3.1 Study Population Characteristics

Out of 4,389 cancer cases in the full cohort generated in Chapter 2, 4,276 cancer cases were included within this nested case-control study, following exclusion of TYA abdominal tumours (n=74) and childhood gonadal germ cell tumours (n=38). The baseline characteristics of CTYA with cancer are presented in Table 3-2. Characteristics between cases and controls were comparable. This was expected for age- and sex- given the matching process. Ethnic groups and Townsend Deprivation Scores were also comparable (p=0.2 and p=0.08, respectively).

Table 3-2 *Demographic characteristics of all cancers, all controls and individual cancer subtypes. Numbers are count (col %) unless otherwise specified. IQR= interquartile range. CI= confidence interval.*

	All Cancers (n=4,276)	**Controls (n=84,071)**	**Leukaemia (n=1,127)**	**Lymphoma (n=919)**	**CNS (n=933)**	**Sarcoma (n=637)**	**Abdominal (n=295)**	**Gonadal germ cell (n=366)**
Age at cohort entry								
Median ± IQR (years)	5 [1,13]	5 [1,13]	2 [1,9]	10 [3,17]	4 [1,10]	8 [2, 13]	1 [1,2]	16 [11,20]
Sex								
Male	2,546 (59.5)	50,246 (59.7)	641 (56.9)	544 (59.2)	538 (57.7)	351 (55.1)	140 (47.5)	333 (90.8)
Female	1,730 (40.5)	33,825 (40.2)	486 (43.2)	375 (40.8)	395 (42.3)	286 (44.9)	155 (52.5)	33 (9.0)
Self-assigned Ethnicity								
White	1,112 (26.2)	21,099 (25.1)	277 (24.6)	248 (27.0)	253 (27.1)	173 (27.2)	84 (28.5)	87 (23.8)
Other	607 (14.2)	12,195 (14.5)	186 (16.5)	152 (16.5)	111 (11.9)	78 (12.2)	40 (13.6)	41 (11.2)
Missing	2,547 (59.6)	50,777 (60.4)	664 (58.9)	519 (56.5)	569 (61.0)	386 (60.6)	171 (58.0)	238 (65.0)
Townsend Deprivation Score								
1 (most affluent)	1,058 (24.7)	19,501 (23.2)	269 (23.9)	232 (25.3)	241 (25.8)	153 (24.0)	81 (27.5)	82 (22.4)
2	925 (21.6)	18,066 (21.5)	270 (24.0)	205 (22.3)	200 (21.4)	122 (19.2)	52 (17.6)	76 (20.8)

3	859 (20.1)	17,021 (20.3)	237 (21.0)	174 (19.0)	191 (20.5)	120 (18.8)	62 (21.0)	75 (20.5)
4	756 (17.7)	15,707 (18.7)	169 (15.0)	160 (17.4)	177 (19.0)	119 (18.7)	53 (18.0)	78 (21.3)
5 (most deprived)	639 (14.9)	13,306 (15.8)	171 (15.2)	141 (15.4)	113 (12.1)	116 (18.2)	45 (15.3)	53 (14.5)
Missing	39 (1.0)	470 (0.6)	11 (1.0)	7 (0.7)	11 (1.2)	7 (1.1)	<5	<5

3.3.2 All CTYA Cancers: Associated clinical features and their timeframes

2,234 cases (52.2%) and 14,926 (17.8%) of controls had at least one of the selected clinical features recorded within 6 months prior to diagnosis (p<0.001). The frequencies of each clinical feature in cancer cases and controls are summarised in Table 3-3. Cancer cases were significantly more likely to present two or more times, with any relevant symptom, prior to diagnosis compared to controls (OR 7.0 [95%CI 6.4-7.6], p<0.001) (Table 3-3).

In univariable analyses, 34 of the 40 symptom/signs were found to be statistically significantly associated with a CTYA cancer diagnosis (Table 3-3). Of note, for exophthalmos and papilledema odds ratios could not be calculated as there were no events in the control group. Overall, there was no significant association between the negative control symptom (eczema) and a diagnosis of cancer in univariable analysis (OR 0.49 [95%CI 0.1-3.4], p=0.5).

Table 3-3 *Frequency of selected symptom/signs in cancer cases and controls. Association between individual clinical feature and a subsequent cancer diagnosis in univariable conditional logical regression analyses. Any counts <5 were suppressed for confidentiality. All percentages are column %.*

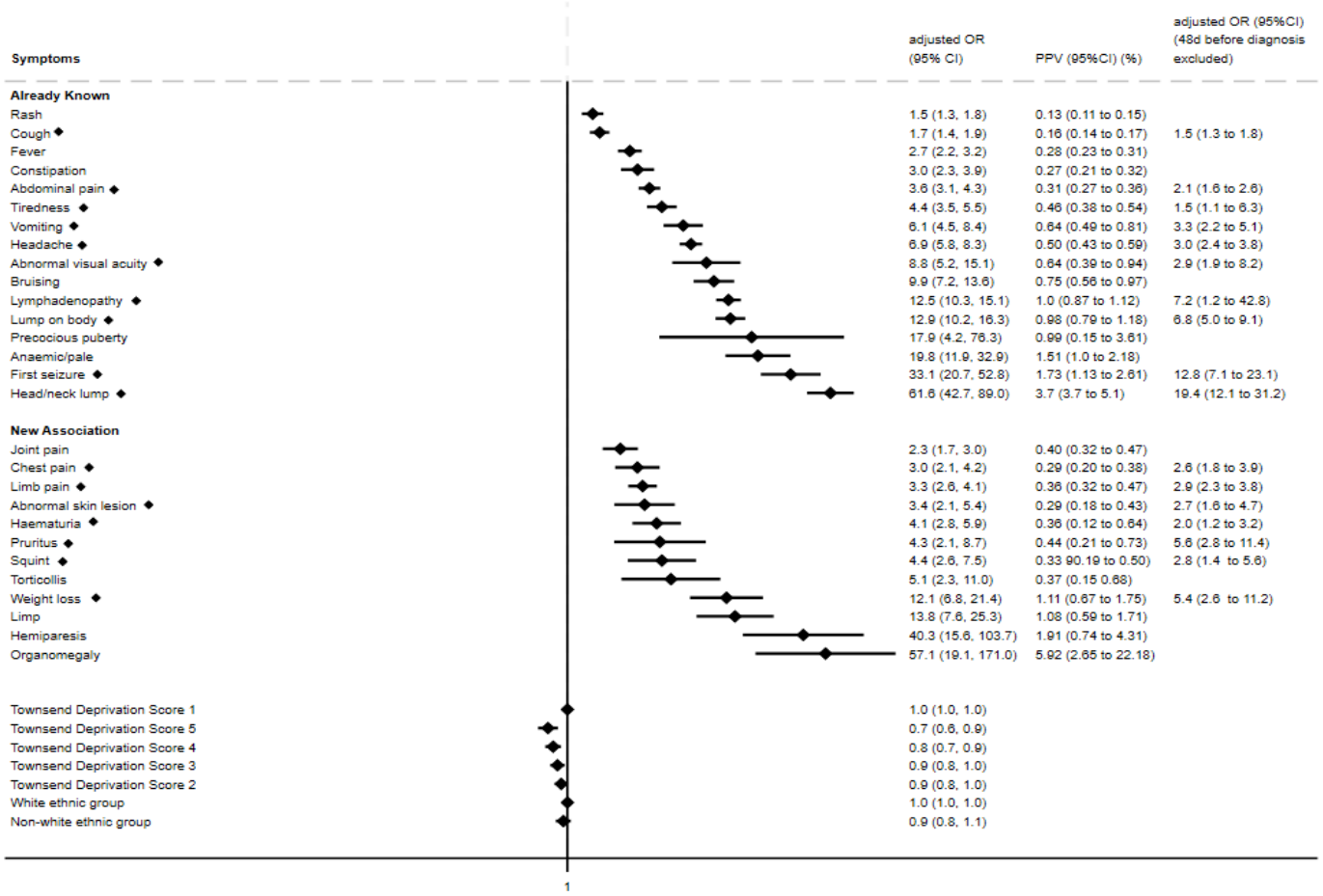
Clinical Feature	Cases (n, %) N=4,276	Controls (n, %) N=84,071	OR (95%CI) in univariable analysis	p-value
Organomegaly	21 (0.5)	5	82.97 (31.3-220.1)	<0.001
Lymphadenopathy	277 (6.5)	399 (0.5)	16.1 (13.6-19.1)	<0.001
Fever	223 (5.2)	1193 (1.4)	4.4 (3.8-5.2)	<0.001
Upper respiratory tract infection (URTI)	631 (14.7)	5953 (7.0)	1.9 (1.7-2.1)	<0.001
Limb pain	212 (4.9)	875 (1.0)	5.1 (4.4-5.5)	<0.001
Joint pain	137 (3.2)	208 (0.3)	5.5 (4.6-6.7)	<0.001
Limp	24 (0.6)	33 (0.04)	14.2 (8.3-24.2)	<0.001
Anaemic/pallor	51 (1.2)	50 (0.06)	27.4 (17.8-42.3)	<0.001
Tiredness	150 (3.5)	488 (0.6)	6.4 (5.4-7.8)	<0.001

Abdominal pain	231 (5.4)	1,108 (1.3)	4.4 (3.8-5.1)	<0.001
Anorexia	13 (0.3)	40 (0.05)	6.6 (3.5-12.4)	<0.001
Rash	203 (4.7)	2288 (2.7)	1.8 (1.6-2.1)	<0.001
Abnormal skin lesion	26 (0.6)	135 (0.2)	3.8 (2.5-5.8)	<0.001
Chest pain	55 (1.3)	287 (0.3)	3.9 (2.9-5.3)	<0.001
Cough	275 (6.4)	2646 (3.1)	2.2 (1.9-2.5)	<0.001
Breathlessness	42 (1.0)	336 (0.4)	2.0 (1.4-2.7)	<0.001
Headache	225 (5.2)	665 (0.8)	7.2 (6.2-8.4)	<0.001
Bruising	79 (1.8)	157 (0.2)	10.2 (7.8-13.4)	<0.001
Vomiting	83 (2.0)	192 (0.2)	8.8 (6.8-11.2)	<0.001
Abdominal mass	57 (1.3)	5 (0.005)	219 (87.8-547.1)	<0.001
Head/neck lump	132 (3.1)	51 (0.06)	55.6 (29.7-77.8)	<0.001
Other lump	160 (3.7)	243 (0.3)	13.9 (11.3-17.2)	<0.001
Weight loss	33 (0.8)	44 (0.05)	14.9 (9.4-23.7)	<0.001
Pruritus	16 (0.4)	54 (0.06)	6.2 (3.5-10.9)	<0.001
Papilledema	17 (0.4)	<5	-	-
Hemiparesis	13 (0.3)	10 (0.01)	30.2 (12.5-73.0)	<0.001
Torticollis	10 (0.2)	40 (0.05)	4.8 (2.4-9.8)	<0.001
Squint	21 (0.5)	94 (0.1)	4.4 (2.7-7.1)	<0.001
Abnormal eye movements	7 (0.2)	24 (0.03)	5.8 (2.4-13.8)	<0.001
Visual acuity abnormalities	30 (0.7)	70 (0.08)	10.1 (6.3-16.2)	<0.001
Diplopia	13 (0.3)	<5	122.7 (27.6-544.3)	-
Exophthalmos	<5	<5	-	-
Behavioural Changes	14 (0.3)	186 (0.2)	1.6 (0.9-3.5)	0.1
Irritability	<5	22 (0.03)	0.9 (0.1-6.4)	0.8

Developmental delay	9 (0.2)	77 (0.09)	2.6 (1.1-5.5)	0.02
Abnormal cranial nerve signs (any)	<5	<5	-	-
First seizure	47 (1.1)	40 (0.05)	23.4 (15.2-35.7)	<0.001
Precocious puberty	<5	6 (0.01)	13.3 (3.7-47.0)	<0.001
Haematuria	45 (1.1)	187 (0.2)	5.0 (3.6-7.0)	<0.001
Constipation	90 (2.1)	502 (0.6)	3.7 (2.8-5.0)	<0.001
Testicular mass/swelling	91 (2.1)	21 (0.03)	100.3 (59.7-168.5)	<0.001
2 or more presentations	903 (21.1)	3645 (4.3)	7.0 (6.4-7.6)	<0.001

Of the 34 significant clinical features in univariable analyses, 28 remained in the final multivariable model. Of these, the strength of association of 11 clinical features have not previously been quantified in a primary care setting (Figure 3-1). The positive predictive values for 8 clinical features (lymphadenopathy, anaemic/pale, first seizure, head/neck lump, hemiparesis, limp, weight loss, organomegaly) were above 1% (Figure 3-1). The median diagnostic interval for the 34 clinical features which remained in the final multivariable model was 48 days [IQR 12-109]. When the last 48 days prior to diagnosis were excluded from multivariable analyses, 17 symptoms/signs remained significant and accordingly, were classified as “early” (Figure 3-1).

Figure 3-1 *Clinical features encountered in primary care associated with any cancer in CTYA in multivariable conditional logistic regression models. Early symptoms denoted with ◆ . OR= Odds ratio, CI= Confidence interval, PPV=Positive Predictive Value.Already Known= previously reported in primary-care based studies. New Association= no previous report of association between symptom and cancer in primary care based studies.*



A priori determined clinically plausible interactions between features were considered in the multivariable conditional logistic regression model and no significant interactions were identified.

3.3.3 Childhood Cancer: Associated clinical features and their timeframes

The nested case-control study consisted of 2,299 childhood cancer cases and 45,047 controls. Univariable analyses identified 32 statistically significant clinical features (Table 3-4), of which 26 remained in the multivariable model. Cancer cases were significantly more likely to present two or more times prior to diagnosis compared to controls (OR 5.4 [95%CI 3.8-6.0], p<0.001). The association between 11 symptom/signs and a cancer diagnosis had not been quantified in previous primary care-based studies (Figure 3-2).

Table 3-4 *Frequency of selected clinical features in childhood cancer cases and controls. Association between individual clinical feature and a subsequent childhood cancer diagnosis in univariable conditional logical regression analyses. Any counts <5 were suppressed for confidentiality. OR= Odds Ratio. CI=Confidence interval. All percentages are column %.*

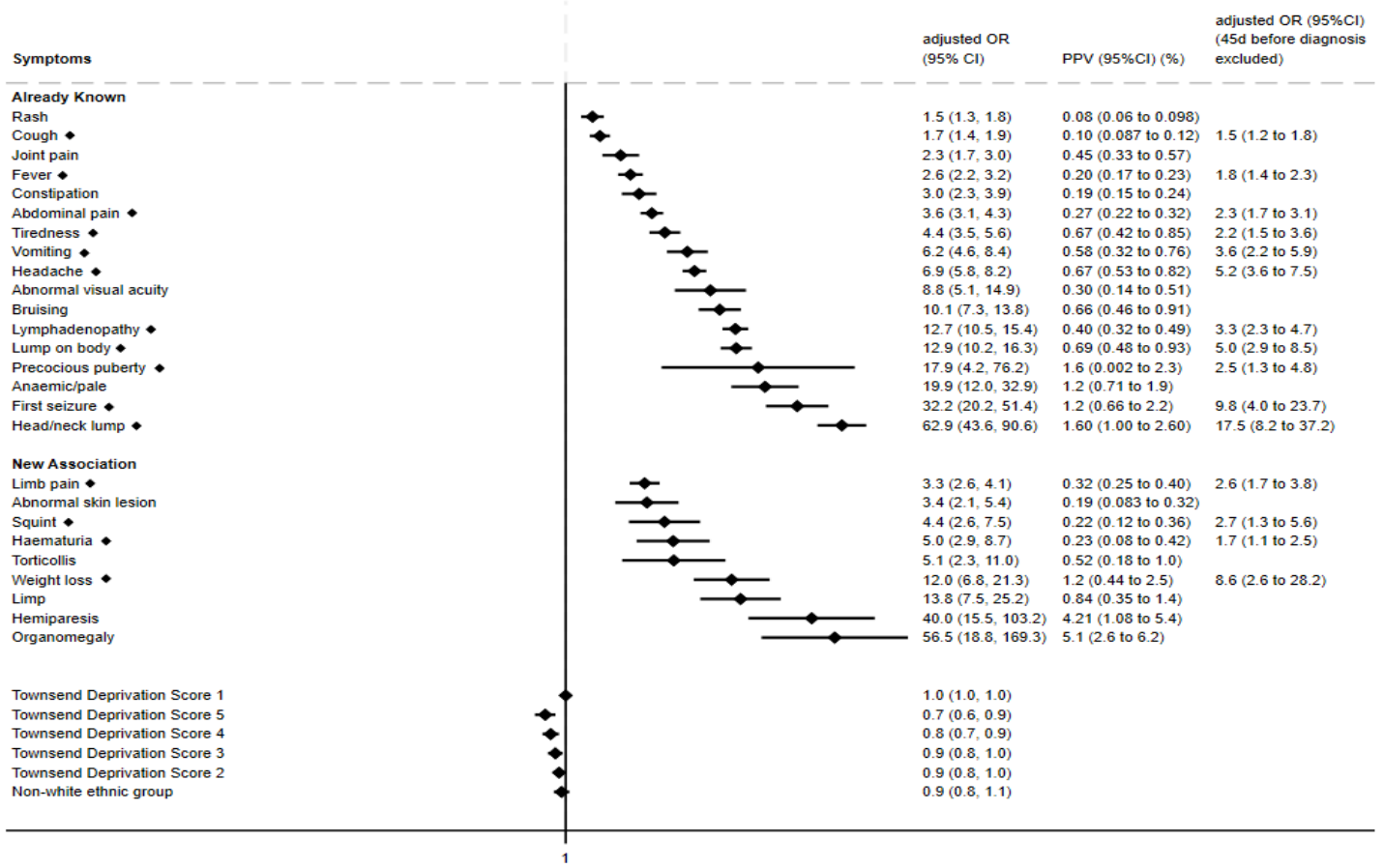
Clinical Feature	**Cases (n, %) N=2,299**	**Controls (n, %) N=45,047**	**OR (95%CI) in univariable analysis**	**p-value**
Organomegaly	16 (0.8)	<5	157.2 (36.1-684.1)	<0.001
Lymphadenopathy	104 (4.5)	283 (0.6)	8.5 (6.7 to 10.8)	<0.001
Fever	191 (8.3)	1057 (2.3)	4.3 (3.7-5.2)	<0.001
URTI	450 (19.5)	4584 (10.1)	1.8 (1.6-1.9)	<0.001
Limb pain	106 (4.6)	359 (0.8)	6.3 (5.0-7.9)	<0.001
Joint pain	74 (3.2)	180 (0.4)	8.7 (6.5-11.4)	<0.001
Limp	23 (1.0)	30 (0.07)	15.4 (8.9-26.6)	<0.001
Anaemic/pallor	35 (1.5)	31 (0.07)	33.2 (19.1-57.7)	<0.001
Tiredness	108 (4.7)	176 (0.4)	12.9 (10.1-16.6)	<0.001
Abdominal pain	134 (5.8)	561 (1.3)	5.2 (4.2-6.3)	<0.001

Anorexia	10 (0.4)	30 (0.07)	6.8 (3.3-14.0)	<0.001
Rash	139 (6.0)	1802 (4.0)	1.6 (1.3-1.9)	<0.001
Abnormal skin lesion	12 (0.5)	71 (0.2)	3.4 (1.8-6.2)	0.001
Chest pain	8 (0.3)	76 (0.2)	2.1 (1.0-4.4)	0.04
Cough	195 (8.5)	2113 (4.7)	2.0 (1.7-2.3)	<0.001
Breathlessness	25 (1.0)	310 (0.7)	11.6 (1.0-2.5)	0.02
Headache	118 (5.1)	192 (0.4)	12.7 (10.1-16.1)	<0.001
Bruising	53 (2.3)	87 (0.2)	12.6 (8.9-17.8)	<0.001
Vomiting	69 (3.0)	130 (0.3)	11.1 (8.2-15.0)	<0.001
Abdominal mass	41 (1.8)	<5	387.5 (93.6-1603.0)	<0.001
Head/neck lump	39 (1.7)	26 (0.06)	36.1 (21.2-61.4)	<0.001
Other lump	55 (2.4)	87 (0.2)	12.6 (9.0-17.8)	<0.001
Weight loss	12 (0.5)	11 (0.02)	24.3 (10.2-58.0)	<0.001
Pruritus	<5	32 (0.07)	2.6 (0.9-7.4)	0.07
Papilledema	7 (0.3)	<5	-	-
Hemiparesis	8 (0.4)	<5	76.2 (16.2-359.4)	<0.001
Torticollis	9 (0.4)	19 (0.04)	9.4 (4.2-20.7)	<0.001
Squint	18 (0.9)	87 (0.2)	4.1 (2.4-6.8)	<0.001
Abnormal eye movements	5 (0.2)	17 (0.04)	5.8 (2.1-16.1)	0.001
Visual acuity abnormalities	15 (0.6)	54 (0.1)	7.1 (3.7-13.5)	<0.001
Diplopia	7 (0.3)	<5	137.0 (16.9-1116.0)	<0.001
Exophthalmos	<5	<5	-	-
Behavioural Changes	13 (0.5)	163 (0.3)	1.7 (0.8-3.3)	0.1
Irritability	<5	7 (0.01)	2.5 (0.3-20.0)	0.4

Developmental delay	9 (0.4)	75 (0.2)	2.7 (1.3-5.7)	0.01
Abnormal cranial nerve signs (any)	<5	<5	-	-
First seizure	20 (0.9)	23 (0.04)	23.8 (13.0-43.8)	<0.001
Precocious puberty	<5	<5	30.0 (5.01 -179.5)	<0.001
Haematuria	28 (1.2)	69 (0.2)	8.3 (5.3-13.0)	<0.001
Constipation	75 (3.2)	423 (0.9)	3.7 (2.7-5.1)	<0.001
2 or more presentations	489 (21.3)	2510 (5.6)	5.4 (4.8-6.0)	<0.001

The median diagnostic interval for the 26 symptoms/signs which remained in the final multivariable model was 45 days [IQR 6-83]. When the last 45 days prior to diagnosis were excluded from multivariable analyses, 15 symptom/signs remained significant and accordingly, were classified as “early” features (Figure 3-2).

Figure 3-2 *Clinical features encountered in primary care associated with any childhood cancer in multivariable conditional logistic regression models. Early symptoms denoted with ◆ . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value. All percentages are column %. Already Known= previously reported in primary-care based studies. New Association= no previous report of association between symptom and cancer in primary care based studies.*



3.3.4 Teenage and Young Adult Cancers: Associated clinical features and their time-frames

There were 1,977 selected TYA cancers and 39,024 controls within the nested case-control study. Univariable conditional logistic regression analyses identified 30 statistically significant clinical features (Table 3-5), of which 21 remained in the multivariable model. The association between 9 features and a cancer diagnosis were not previously quantified in primary care-based studies (Figure 3-3).

Table 3-5 *Frequency of selected clinical features in TYA cancer cases and controls. Association between individual clinical feature and a subsequent TYA cancer diagnosis in univariable conditional logical regression analyses. Any counts <5 were suppressed for confidentiality. OR= Odds Ratio. CI=Confidence interval. All percentages are column %.*

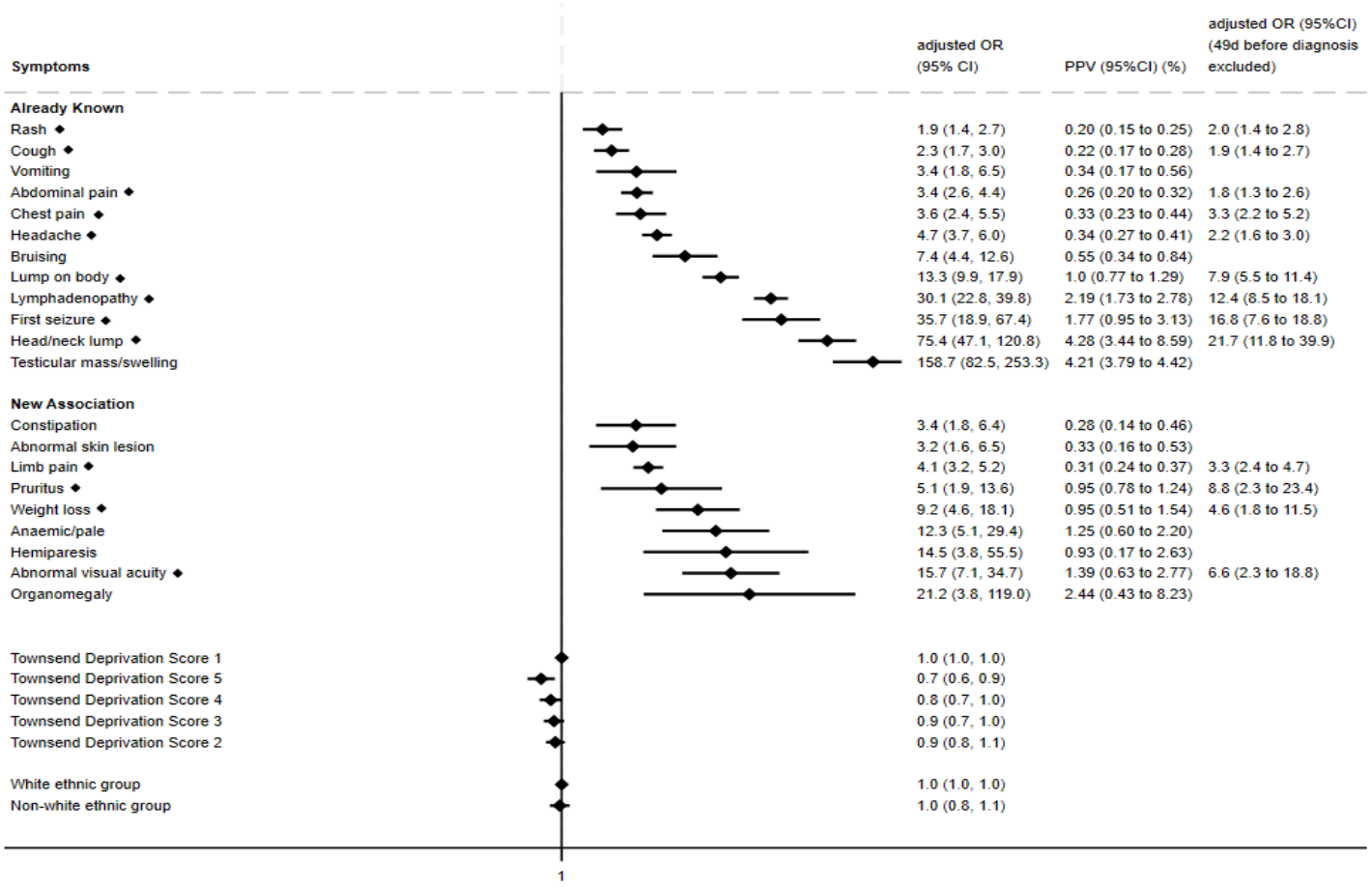
Clinical Feature	**Cases (n, %) N=1,977**	**Controls (n, %) N=39,024**	**OR (95%CI) in univariable analysis**	**p-value**
Organomegaly	5 (0.3)	<5	33.3 (8.0-139.5)	<0.001
Lymphadenopathy	173 (8.8)	116 (0.3)	32.4 (25.2-41.7)	<0.001
Fever	32 (1.6)	136 (0.3)	5.0 (3.3-7.5)	<0.001
URTI	181 (9.1)	1700 (4.3)	2.3 (1.9-2.7)	<0.001
Limb pain	106 (5.4)	516 (1.3)	4.3 (3.4-5.3)	<0.001
Joint pain	63 (3.2)	334 (0.9)	3.9 (2.9-5.1)	<0.001
Limp	<5	<5	-	-
Anaemic/pallor	16 (0.8)	19 (0.05)	19.8 (9.7-40.3)	<0.001
Tiredness	42 (2.1)	312 (0.8)	2.7 (2.0-3.9)	<0.001
Abdominal pain	97 (4.9)	561 (1.4)	3.7 (2.9-4.6)	<0.001
Anorexia	<5	10 (0.03)	6.0 (1.7-21.8)	0.02
Rash	64 (3.2)	486 (1.2)	2.7 (2.1-3.6)	<0.001

Abnormal skin lesion	14 (0.7)	64 (0.2)	4.3 (2.4-7.8)	<0.001
Chest pain	47 (2.3)	211 (0.5)	4.6 (3.3-6.4)	<0.001
Cough	80 (4.0)	533 (1.4)	3.1 (2.4-3.9)	<0.001
Breathlessness	17 (0.9)	121 (0.3)	2.9 (1.7-4.8)	<0.001
Headache	107 (5.4)	473 (1.2)	4.7 (3.8-6.0)	<0.001
Bruising	26 (1.3)	70 (0.2)	7.3 (4.7-11.6)	<0.001
Vomiting	14 (0.8)	62 (0.2)	4.6 (2.5-8.0)	<0.001
Abdominal mass	16 (0.8)	<5	106 (31.1-366.0)	<0.001
Head/neck lump	93 (4.7)	25 (0.06)	72.0 (46.3-112.0)	<0.001
Lump on body	105 (5.3)	156 (0.4)	14.7 (11.3-19.1)	<0.001
Weight loss	21 (1.0)	33 (0.08)	12.3 (7.0-21.4)	<0.001
Pruritus	12 (0.6)	22 (0.06)	11.5 (5.6-23.5)	<0.001
Papilledema	8 (0.4)	<5	-	-
Hemiparesis	5 (0.3)	8 (0.02)	15.0 (4.5-49.5)	<0.001
Torticollis	<5	21 (0.05)	0.9 (0.2-6.6)	0.8
Squint	<5	7 (0.02)	9.4 (2.3-37.7)	0.002
Abnormal eye movements	<5	7 (0.02)	5.9 (1.1-29.4)	0.03
Visual acuity abnormalities	15 (0.6)	16 (0.04)	16.4 (8.0-33.6)	<0.001
Diplopia	6 (0.3)	<5	108.3 (13.0-903.4)	<0.001
Exophthalmos	<5	<5	-	-
Behavioural Changes	<5	23 (0.06)	0.9 (0.1-6.7)	0.8
Irritability	<5	15 (0.05)	-	-
Developmental delay	<5	<5	-	-
Abnormal cranial nerve signs (any)	<5	<5	-	-
First seizure	20 (1.0)	24 (0.05)	22.9 (12.7-41.6)	<0.001

Precocious puberty	<5	<5	-	-
Haematuria	17 (0.9)	118 (0.3)	3.0 (1.8-5.0)	<0.001
Constipation	15 (0.8)	79 (0.2)	4.0 (2.3-7.0)	<0.001
Testicular swelling	91 (4.6)	21 (0.05)	104.9 (62.6-176.0)	<0.001
2 or more presentations	414 (20.9)	1,135 (2.9)	10.0 (8.7-11.4)	<0.001

The median diagnostic interval for the 21 symptom/signs which remained in the final multivariable model was 49 [IQR 1, 90] days. When the last 49 days prior to diagnosis were excluded from multivariable analyses, 13 symptom/signs remained significant and accordingly, were classified as “early” (Figure 3-3).

Figure 3-3 *Clinical features encountered in primary care associated with any TYA cancer in multivariable conditional logistic regression models. Early symptoms denoted with ◆ . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value. Already Known= previously reported in primary-care based studies. New Association= no previous report of association between symptom and cancer in primary care based studies.*



3.3.5 Early and late clinical features associated with different CTYA cancer subtypes

Clinical features were further investigated across each cancer subtype (leukaemias, lymphomas, CNS tumours, sarcomas and abdominal tumours) overall, as well as in children and TYA. The frequencies of selected clinical features in cases and controls per cancer subtype can be found in Appendix, Chapter 9 Supplementary Tables 9, 10 and 11. Table 3-6 summarises the median diagnostic intervals for each cancer subtype overall and across each age category. The median diagnostic intervals for lymphomas and sarcomas were longer than the median diagnostic intervals for leukaemias, CNS and abdominal tumours.

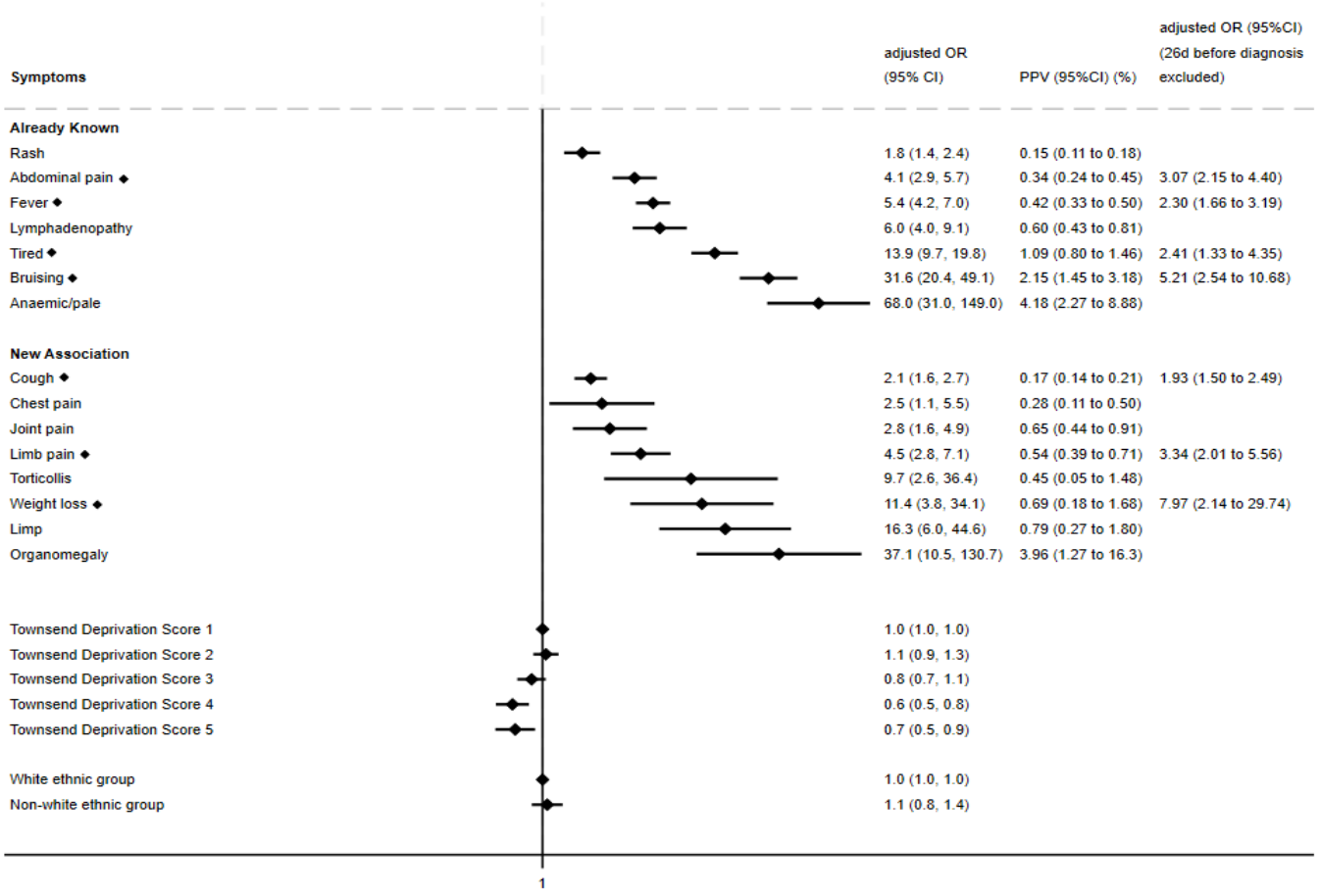
Table 3-6 *Median diagnostic intervals for each cancer subtype overall and across each age category. IQR= interquartile range*

Cancer Subtype	**Median Diagnostic interval ± IQR (days)**
Leukaemias	
Overall	26 (2,98)
Childhood	25 (2, 105)
TYA	25 (1, 90)
Lymphomas	
Overall	51 (18, 105)
Childhood	52 (15,88)
TYA	49 (20,106)
CNS Tumours	
Overall	21 (6, 68)
Childhood	17 (5, 63)
TYA	25 (8, 71)
Sarcomas	

Overall	57 (20, 113)
Childhood	41 (14, 100)
TYA	67 (23, 115)
Abdominal Tumours	
Overall (childhood only)	22 (3, 79)

For leukaemias, there were 22 associated symptom/signs in univariable models and 15 remaining in the final multivariable model (Figure 3-4). Anaemia/pallor was most strongly associated with leukaemia diagnosis (OR 68.0 [95%CI 31.0 to 149.0]). When the late timeframe for leukaemia was excluded, 7 early symptoms were identified and were fever, tiredness, bruising, cough, weight loss, limb and abdominal pain (Figure 3-4).

Figure 3-4 *Clinical features encountered in primary care associated with CTYA leukaemia in multivariable conditional logistic regression models. Early symptoms denoted with ♦ . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value. Already Known= previously reported in primary-care based studies. New Association= no previous report of association between symptom and cancer in primary care based studies.*

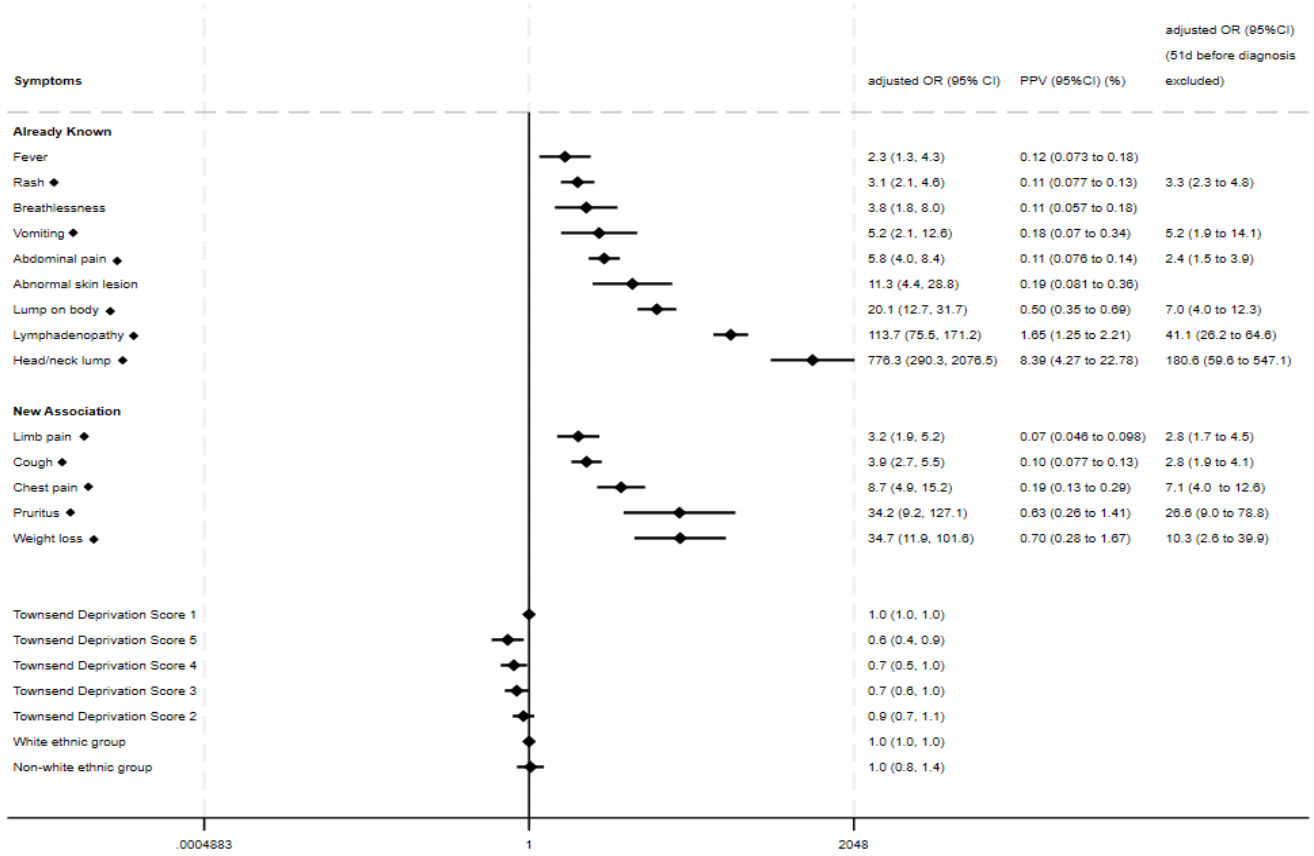


When analysed across different age groups, there were 14 significant clinical features for childhood and 10 for TYA leukaemia cases in the final multivariable models (Appendix Chapter 9, Supplementary Figures 9-3 & 9-4). 6 early symptom/signs were identified for childhood leukaemias and 3 for TYA leukaemias (Appendix Chapter 9, Supplementary Figures 9-3 & 9-4).

For lymphomas, there were 19 symptom/signs associated with lymphoma in univariable models, of which 14 remained in the final multivariable model (Figure 3-5). Head/neck lump was most strongly associated with a lymphoma diagnosis (OR 776.7 [95%CI 290.3 to 2076.5]). 11 early symptoms were identified when the time from median diagnostic interval to diagnosis was excluded. These were lump on body, head/neck lump, lymphadenopathy, cough, limb pain, abdominal pain, chest pain, vomiting, pruritus, weight loss and rash.

Figure 3-5 *Clinical features encountered in primary care associated with CTYA lymphoma in multivariable conditional logistic regression models. Early symptoms denoted with ◆ . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value.*

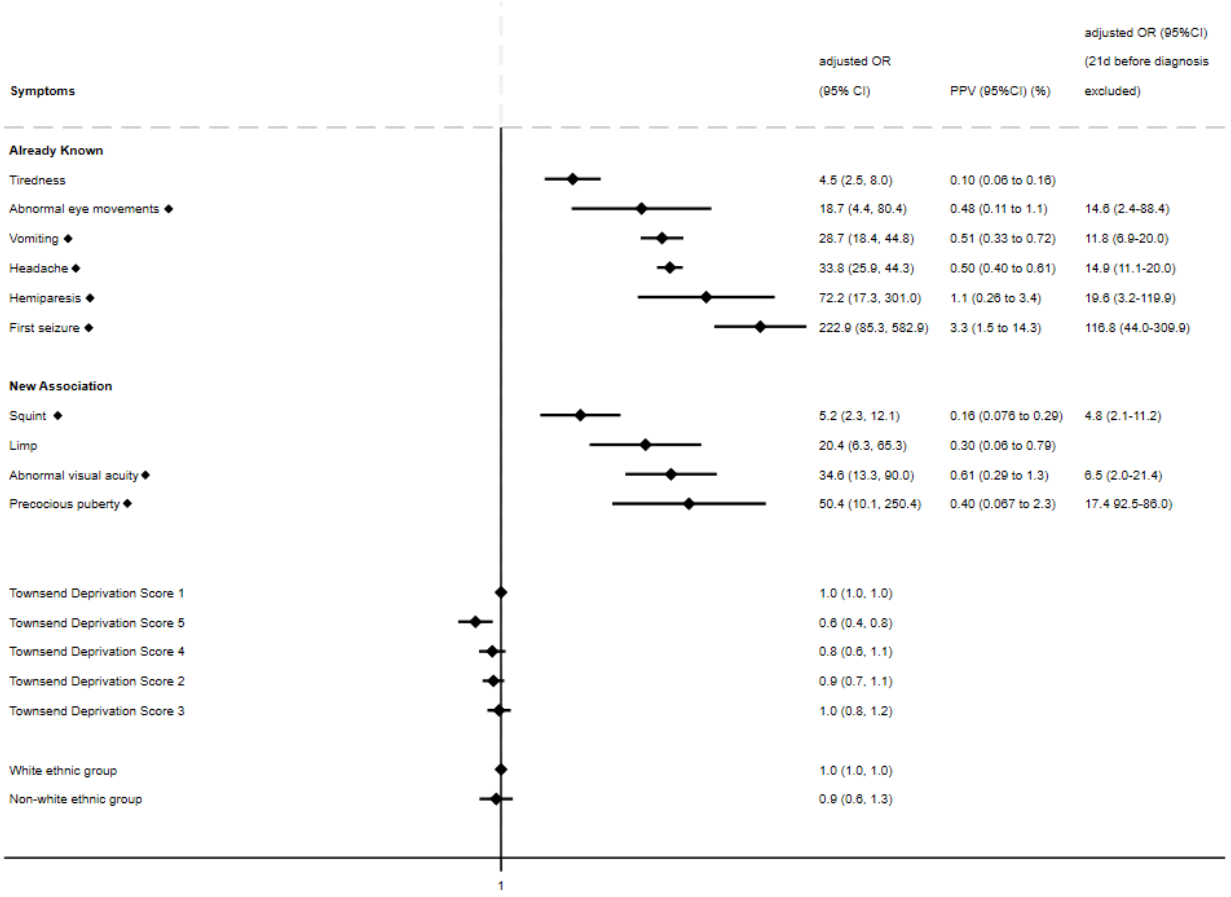
Already Known= previously reported in primary-care based studies. New Association= no previous report of association between symptom and cancer in primary care based studies.



8 symptom/signs were significantly associated with childhood lymphoma diagnoses and 13 to TYA diagnoses (Appendix Chapter 9, Supplementary Figures 9-3 & 9-4). Of these, 3 early clinical features were identified for childhood lymphomas and 10 were observed in TYA lymphomas, where lymphadenopathy and head/neck lump were early features common to both age groups (Appendix Chapter 9, Supplementary Figures 9-3 & 9-4).

For CNS tumours, 15 symptom/signs were associated with a diagnosis of CNS tumours in univariable models and 10 remained in the final multivariable model (Figure 3-6). First seizure was most strongly associated with a CNS tumour diagnosis (OR 222.9 [95%CI 85.3 to 582.9]). For CNS tumours, headache, first seizure, vomiting, hemiparesis, abnormal eye movements and acuity, squint and precocious puberty were categorised as early symptoms.

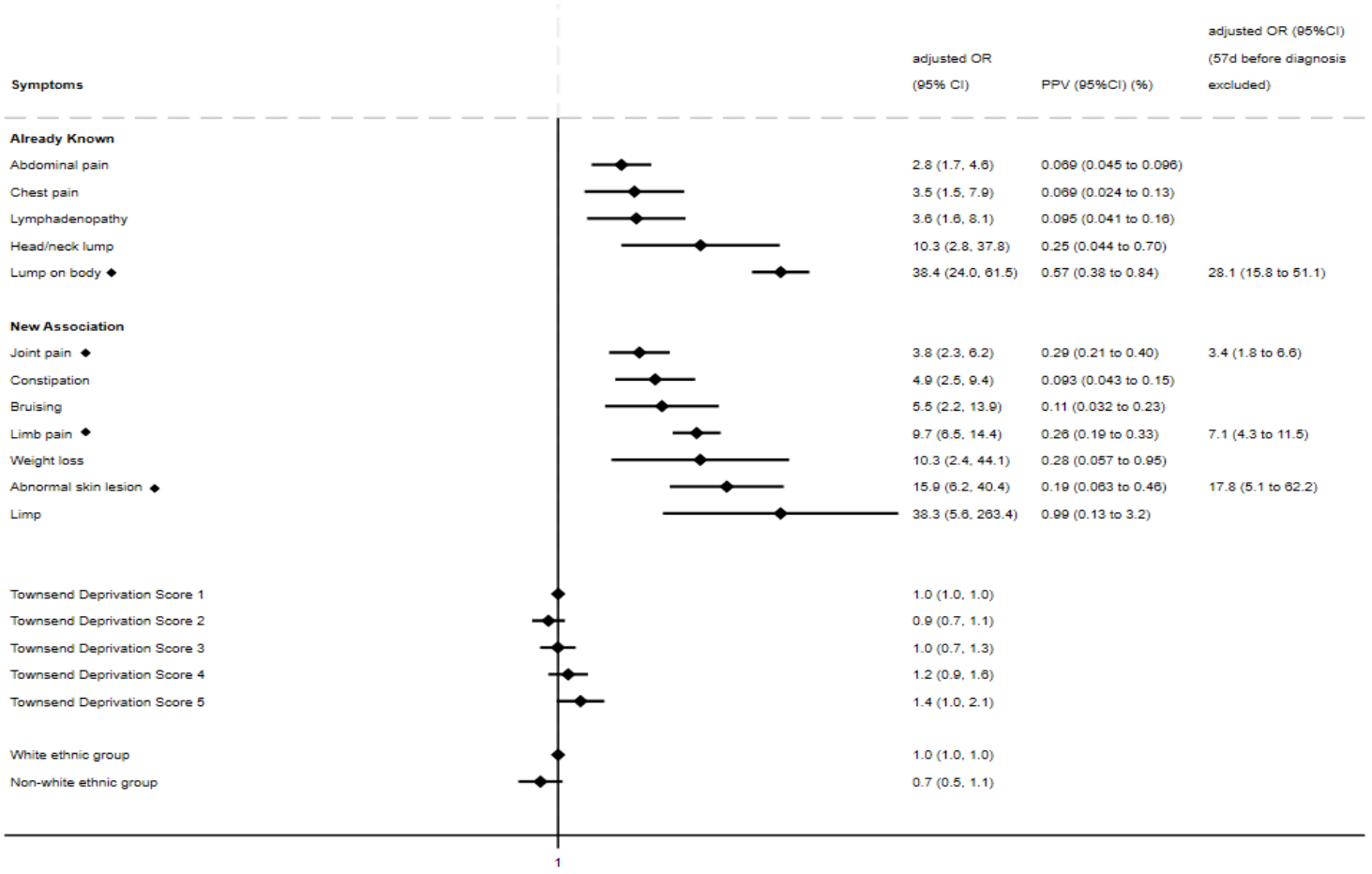
Figure 3-6 *Clinical features encountered in primary care associated with CTYA CNS tumours in multivariable conditional logistic regression models. Early symptoms denoted with ◆ . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value. Already Known= previously reported in primary-care based studies. New Association= no previous report of association between symptom and cancer in primary care based studies.*



Both childhood and TYA CNS tumour diagnosis were associated with headache, first seizure, abnormal visual acuity and vomiting, whilst 6 additional symptom/signs, hemiparesis, abnormal eye movements, limp, precocious puberty, squint and tiredness were linked to childhood CNS tumour diagnoses (Appendix Chapter 9, Supplementary Figures 9-3 & 9-4). 6 early clinical features were associated with childhood CNS tumours, whilst only headache and first seizure were early symptoms for TYA CNS tumours (Appendix Chapter 9, Supplementary Figures 9-3 & 9-4).

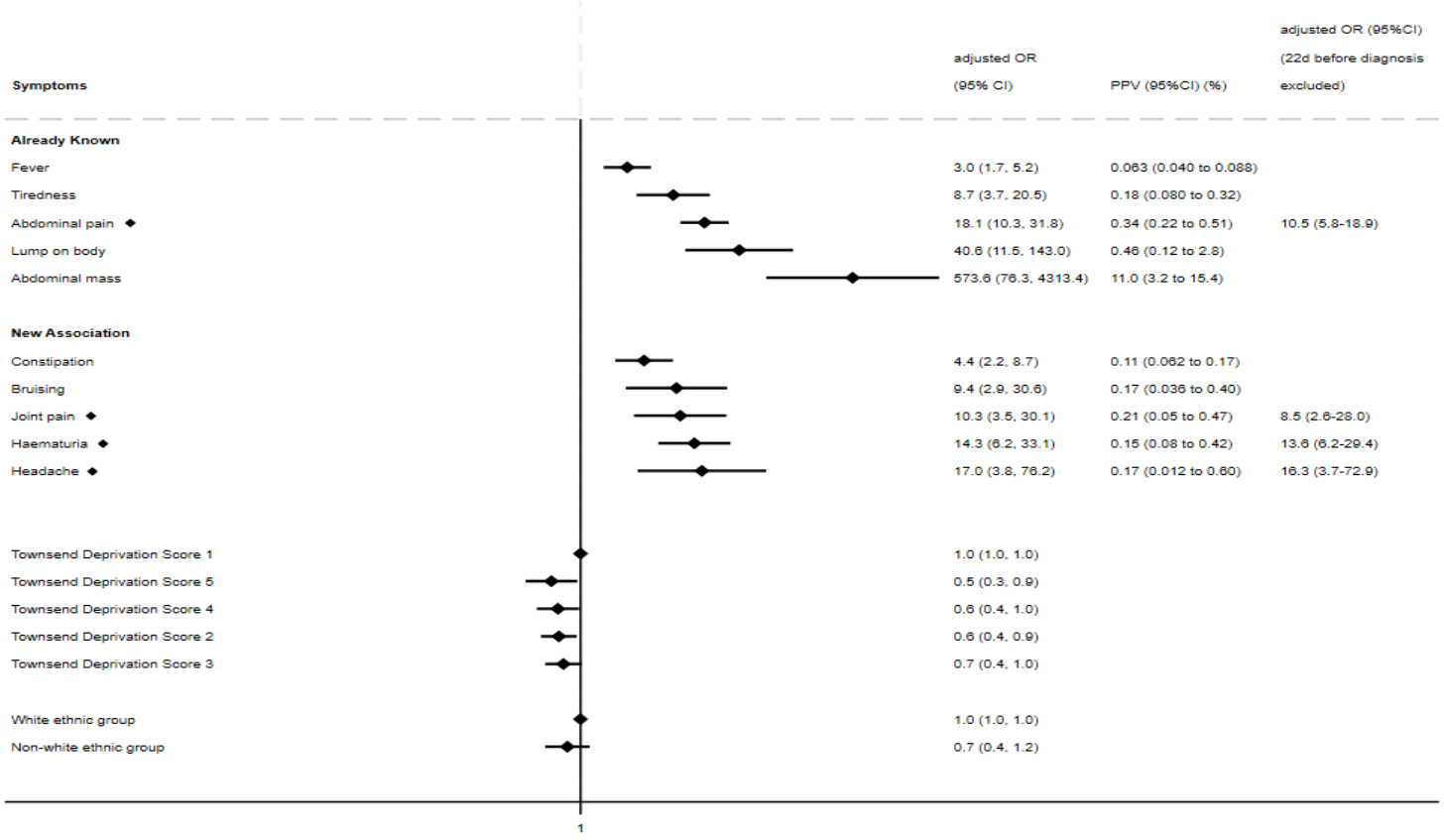
For sarcomas, 13 associated symptoms were identified in univariable models, with 12 remaining in the final multivariable model (Figure 3-7). Lump on body was most strongly associated with sarcoma (OR 38.4 [95%CI 24.0 to 61.5]). 4 symptom/signs (limb pain, joint pain, lump on body and abnormal skin lesion) were early symptom/signs. 8 clinical features significantly associated with both a childhood and TYA sarcoma diagnosis and included lump on body, abnormal skin lesion, bruising, limb, joint abdominal and chest pain (Appendix Chapter 9, Supplementary Figures 9-3 & 9-4). Childhood and TYA sarcoma cases were each also associated with the four early clinical features (Appendix Chapter 9, Supplementary Figures 9-3 & 9-4).

Figure 3-7 *Clinical features encountered in primary care associated with CTYA sarcomas in multivariable conditional logistic regression models. Early symptoms denoted with ♦ . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value. Already Known= previously reported in primary-care based studies. New Association= no previous report of association between symptom and cancer in primary care based studies.*



For abdominal tumours, analyses were restricted to childhood cases only and 12 associated symptom/signs were in univariable models with 10 remaining in the final multivariable model (Figure 3-8). Four early clinical features, including abdominal pain, joint pain, headache and haematuria were identified (Figure 3-8).

Figure 3-8 *Clinical features encountered in primary care associated with CTYA abdominal tumours in multivariable conditional logistic regression models. Early symptoms denoted with ◆ . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value. Already Known= previously reported in primary-care based studies. New Association= no previous report of association between symptom and cancer in primary care based studies.*



For germ cell tumours, 4 associated symptoms were found in univariable models (Chapter 9, Supplementary Table 9-11), with 1 symptom, testicular swelling remaining in the final multivariable model [OR 198.7 (95%CI 46.8 to 841.0)].

3.3.6 Determining combinations of clinical features associated with a cancer diagnosis

3.3.6.1 Clinically-determined paired symptom combinations

When investigating all CTYA cancers together, out of the 17 early clinical features identified in the section above, 4 of these features were identified to have higher odds of being associated with cancer if in paired combination with any of the 28 relevant symptoms than if considered alone (Table 3-7). When analyses were restricted to childhood cancers, out of the 15 early clinical features identified, 3 had higher odds of being associated with cancer if in paired combination with any of the other statistically significant 26 symptoms. For TYA cancers, out of the 13 identified early clinical features, only 1, abdominal pain, had higher odds of being associated with cancer if in paired combination with any of the other statistically significant 21 symptoms. None of the early clinical features identified in individual cancer subtypes had increased odds of being associated with a cancer diagnosis when in paired combination with any of the other statistically significant clinical features identified in multivariable analyses.

Table 3-7 *Strength of association between any cancer diagnosis and early symptoms i) when considered in isolation, ii) when considered in combination with any of the other identified statistically significant symptoms within 6 months of diagnosis. aOR= adjusted Odds Ratio, PPV= Positive predictive value.*

Early Symptom	***Isolated***		***In combination with any of significant symptoms***	
	aOR (95% CI)	PPV (95%CI)	aOR (95%CI)	PPV (95%CI)
All age groups				
Abdominal pain	3.6 (3.1- 4.3)	0.31 (0.27-0.36)	9.6 (7.7-12.1)	0.69 (0.54-0.83)
Chest pain	3.0 (2.1 – 4.2)	0.29 (0.20-0.38)	7.3 (4.8-11.1)	0.54 (0.33-0.78)
Vomiting	6.1 (4.5-8.1)	0.64 (0.49-0.81)	11.8 (8.2-17.0)	0.87 (0.57-1.2)
Limb pain	3.3 (2.6-4.1)	0.36 (0.32-0.47)	6.4 (5.3-7.6)	0.46 (0.38-0.54)
Childhood (0-14 years)				
Abdominal pain	3.6 (3.1-4.3)	0.37 (0.19-0.39)	12.3 (9.2-16.3)	0.84 (0.63-0.98)

Vomiting	6.2 (4.6-8.4)	0.58 (0.32-0.76)	12.6 (8.5-19.1)	0.93 (0.58-1.1)
Limb pain	4.8 (3.57-6.5)	0.38 (0.35-0.42)	8.8 (6.7-11.5)	0.62 (0.46-0.79)
TYA (15-24 years)				
Abdominal pain	3.6 (2.3 – 4.4)	0.26 (0.20-0.32)	6.8 (4.7-9.7)	0.51 (0.34-0.65)

3.3.6.2 Symptom Patterns in CTYA cancer using Principal Component Analysis

Fourteen patterns were retained explaining 50% of the total variance in principal component analysis when the full study population (4,276 cancer cases and 84,071 controls) was used. Individual clinical features providing a significant contribution to each principal component (loading score >0.30) are summarised in Table 3-8. Each component had 1 to 3 original clinical feature which provided a significant contribution.

Table 3-8 *Component Loading Matrix where loading scores indicate the correlation of the original variable with the principal component. Loading scores with significant contribution (above 0.3) shown only. PC= principal component.*

Clinical Feature	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Organomegaly							-0.32		-0.41					
Lymphadenopathy		+0.34			+0.34									
Fever		+0.41												
Limb pain	+0.68													
Joint pain	+0.67													
Limp								-0.38		+0.43				
Anaemic/pallor														
Tiredness														
Abdominal pain				+0.34		+0.34								
Rash														
Abnormal skin lesion											0.48			-0.47
Cough		+0.37												
Chest pain							+0.44							
Breathless														
Headache				+0.32										

Bruising									-0.48					+0.42
Vomiting				+0.31										
Abdominal mass						+0.50								
Head/neck lump					+0.40									+0.37
Lump on body					+0.39									
Weight loss						-0.36		+0.35						
Pruritus														
Papilledema			+0.34							+0.35				
Hemiparesis								+0.34			-0.41	+0.38		
Squint			+0.58											
Abnormal visual acuity			+0.40											
First Seizure									+0.39	-0.32		+0.38		
Precocious puberty												+0.65		
Abnormal eye movements			+0.50	+0.35										
Haematuria							+0.34					+0.33	+0.56	
Constipation						+0.45								

The association between each principal component score and a subsequent cancer diagnosis is summarised in Table 3-9. Whilst 11 principal components had statistically increased odds of a cancer diagnosis, one principal component (chest pain & haematuria pre-dominant) was statistically less likely to be associated with a cancer diagnosis.

Table 3-9 *Association between principal components and CTYA cancers in univariable conditional logistic regression models, adjusted for ethnicity and deprivation level through the Townsend Quintile. PC= principal component, aOR= adjusted Odds Ratio, CI= confidence interval, SD= standard deviation*

Principal Component Description	**PC Score Cases (mean ± SD)**	**PC Score Controls (mean ± SD)**	**aOR (95%CI)**
PC1: limb pain, joint pain predominant	0.82 ± 2.48	-0.042 ± 1.13	1.24 (1.23-1.26)
PC2: lymphadenopathy, fever, cough pre-dominant	1.09 ± 2.50	-0.055 ± 1.01 -	1.63 (1.60-1.66)
PC3: headache, papilledema, abnormal visual acuity pre-dominant	0.04 ± 3.02	-0.002 ± 0.87	1.03 (1.01 – 1.05)
PC4: abdominal pain, headache, vomiting pre-dominant	0.51 ± 2.79	-0.025 ± 0.87	1.35 (1.32-1.38)
PC5: lymphadenopathy, lump on head or body pre-dominant	0.57 ± 2.97	-0.029 ± 0.82	1.34 (1.31-1.36)
PC6: abdominal pain, abdominal mass and constipation pre-dominant	0.12 ± 3.16	-0.006 ± 0.77	1.09 (1.07-1.11)
PC7: chest pain and haematuria pre-dominant	-0.13 ± 2.66	-0.006 ± 0.86	0.89 (0.87-0.91)
PC8: hemiparesis, weight loss pre-dominant	0.14 ± 3.06	-0.0074 ± 0.78	1.12 (1.09-1.15)
PC9: first seizure pre-dominant	0.04 ± 3.12	-0.002 ± 0.76	1.06 (1.01-1.08)
PC10: limp pre-dominant	-0.009 ± 3.04	-0.0048 ± 0.77	0.99 (0.96-1.02)

PC11: abnormal skin lesion pre-dominant	-0.009 ± 2.63	-0.0046 ± 0.84	0.99 (0.96-1.02)
PC12: precocious puberty, hemiparesis, first seizure pre-dominant	0.21 ± 2.87	-0.011 ± 0.80	1.09 (1.07-1.12)
PC13: haematuria pre-dominant	0.048 ± 2.31	-0.024 ± 0.88	1.05 (1.02-1.09)
PC14: head/neck lump and bruising pre-dominant	0.10 ± 2.60	-0.051 ± 0.84	1.1 (1.07-1.13)

When the association between principal components and individual cancer subtypes were investigated, it was found that PC1, PC2, PC4, PC13 and PC14 were statistically significantly associated with higher odds of leukaemia diagnosis; PC1, PC2, PC5, PC8, PC10, PC14 were statistically significantly associated with higher odds of lymphoma diagnosis; PC1, PC2, PC3, PC4, PC8, PC9, PC11, PC12 were statistically significantly associated with higher odds of CNS tumour diagnosis; PC1, PC2, PC4, PC5, PC6, PC7, PC8 were statistically significantly associated with higher odds of sarcoma diagnosis; PC1, PC2, PC4, PC5, PC6, PC10, PC12 were statistically significantly associated with higher odds of abdominal tumour diagnosis; PC2, PC4, PC5, PC6, PC7, PC13 were statistically significantly associated with higher odds of gonadal germ cell tumour diagnosis.

The proportion of cases and controls with the symptom combinations for each principal component is summarised and the corresponding positive predictive values detailed in Table 3-10. As there were no controls with the symptom combinations in principal components 3, 4 and 9, PPVs could not be calculated. Symptom combinations representing principal components 2, 5 and 6 are found to have higher PPVs compared to when symptoms are considered individually (Figure 3-1). The PPVs for symptom combinations representing principal components 1, 12 and 14, were comparable to the PPVs of individual symptoms.

Table 3-10 Proportion of cases and controls with the symptom combination for the principal component present and the corresponding positive predictive values. PC= principal component, PPV= Positive Predictive Value, CI= confidence interval

Principal Component Description	**Number of cases with symptom combination for PC (N=4,276, %)**	**Number of controls with symptom combination for PC (N=84,071, %)**	**PPV (95%CI)**
PC1: limb pain & joint pain present	89 (2.1)	356 (0.4)	0.37 [0.29 to 0.45]
PC2: lymphadenopathy, fever & cough present	6 (0.1)	15 (0.02)	0.59 [95%CI 0.17 to 0.74]
PC3: headache, papilledema & abnormal visual acuity present	<5	<5	-
PC4: abdominal pain, headache & vomiting present	<5	<5	-
PC5: lymphadenopathy & lump on head or body present	41 (0.9)	9 (0.01)	6.3 [3.4 to 15.7]
PC6: abdominal pain, abdominal mass & constipation present	12 (0.2)	6 (0.01)	2.9 [1.1 to 7.8]
PC8: hemiparesis & weight loss	<5	<5	-
PC12: precocious puberty, hemiparesis, & first seizure present	<5	<5	1.9 [95%CI 0.8 to 4.7]
PC14: head/neck lump & bruising present	<5	<5	0.37 [95%CI 0.01 to 0.99]

3.4. Discussion

3.4.1 Summary of Principal Findings

This nested case-control study of 4,276 CTYA cancer cases and 84,071 controls is the largest study to date to investigate the association between clinical features encountered in primary

care and a subsequent diagnosis of cancer in this age group. First, these associations were explored with all cancers grouped together to provide overall and age-specific insight into the most relevant and important symptom/signs that warrant further investigation. Second, the association between clinical features and each cancer subtype was investigated to provide insight into cancer-specific clinical presentations. Overall, this study determined clinical features which are 1) most strongly associated with a cancer diagnosis in primary care, providing support for previously reported associations as well as identifying unreported associations, 2) likely to be relevant to early stages of disease onset and 3) if in combination, more likely to be associated with a cancer diagnosis. It also supports previous findings of significantly increased repeated attendances to primary care prior to a diagnosis [25, 26, 56]. Overall, this information provides novel insight into primary care presentations of CTYA cancers and could have an important role in improving clinical and public awareness of CTYA cancers, narrowing down clinical suspicion of cancer in primary care and ultimately improving earlier diagnosis.

3.4.2 Findings in Context

Overall, 28 statistically significant clinical features were associated with a diagnosis of CTYA cancer. In addition to supporting previously reported clinical features in primary care such as pain, head/neck lumps and seizures [25, 26, 56], this study provides insight into novel associations in primary care, including torticollis, organomegaly, hemiparesis and limping. Furthermore, 17 symptom/signs in this study were likely to be associated with cancer at an earlier stage. Most of these symptom/signs were common and non-specific with low predictive values, including cough, lymphadenopathy, abdominal pain, constipation and headache. However, two symptoms, first seizure, and head/neck lumps were higher-risk symptoms, more infrequently encountered in primary care [56], which appear to be present

at early-disease onset and offer a window of opportunity to be diagnosed earlier. It is possible that investigations such as imaging contribute to longer diagnostic intervals for these two symptoms and expediting investigations may speed diagnosis.

Four early symptoms had higher predictive utility when in paired combination with any of the 28 clinical features associated with a CTYA cancer diagnosis compared to when considered alone. For example, a single presentation of abdominal pain had an increased odds of 3.6 (95%CI 3.1 to 4.3), corresponding to a positive predictive value of 3 in 1000. The odds and corresponding predictive value increased to 9.6 (95%CI 7.7 – 12.1) and 7 in 1000, respectively, if the patient presented with abdominal pain and one of the other 28 significant symptoms at any other time point within 6 months. This can be useful for primary care clinicians to consider when a non-specific symptom may be associated with an underlying cancer diagnosis in this age group.

Principal component analysis identified 14 distinct symptom/sign patterns across the study population. Several patterns described in each component, which were statistically significantly associated with a diagnosis of CTYA cancer, were clinically plausible symptom combinations. For example, headache, papilledema and abnormal visual acuity (symptoms pre-dominantly featuring in principal component 3) are commonly seen in CNS tumours; abdominal pain, abdominal mass, and constipation (symptoms pre-dominantly featuring in principal component 6) in childhood abdominal tumours; and lymphadenopathy, head/neck lump (symptoms pre-dominantly featuring in principal component 5) are commonly occurring symptom/signs in lymphomas. Symptom combinations derived from principal components 5 & 6 also showed higher positive predictive values compared to when the symptoms were considered individually, however, this finding was not consistent across all symptom combinations derived from each principal component. This information can also aid primary

care clinicians to search for other symptoms/signs linked to CTYA cancer when faced with a non-specific symptom such as headache, constipation or lymphadenopathy in young people.

When the association between clinical features and a cancer diagnosis was explored separately across children (0-14 years) and TYA (15-24 years), 71% (17/24) of clinical features were shared across the two age groups. As would be expected, age-specific differences were identified, such as squinting and limping which were only significant in children. The median diagnostic interval for TYA cancer was 4 days longer than childhood cancer (45 [IQR 6-83] vs. 49 [IQR 15-103] days, $p<0.001$ respectively). This is in line with previous evidence and may reflect challenges around young people having their concerns acknowledged by healthcare professionals [95].

When the association between clinical features and a cancer diagnosis was explored across different types of cancers, there were distinct subtype-specific patterns in clinical presentation, which is in line with prior studies [25, 26, 39, 56]. Leukaemias were most strongly associated with anaemia/pallor, organomegaly, bruising and limping; lymphomas with head/neck lumps as well as lymphadenopathy, and CNS tumours with seizures, visual acuity problems, headache and vomiting. Abdominal masses, haematuria and constipation were most frequently associated with abdominal tumours. Finally, lumps and pain in extremities were commonly associated with sarcomas. These cancer subtype-specific clinical presentations closely reflected findings from the BRIGHLIGHT cohort of TYA cancers, where lymphomas were also found to present most frequently with lump/swellings; CNS tumours with headaches, and bone and soft tissue tumours with unexplained pain and lumps/swellings [96]. Cancer subtype-specific clinical features identified in this Chapter are also seen in adulthood cancers, such as anaemia in leukaemia diagnoses, neck lumps in lymphoma diagnoses as well as painless and painful masses in bone and soft tissue tumours [97, 98].

Interestingly, however, there are important age-dependent differences in presentations, with younger patients presenting more frequently with non-specific but faster progressing symptoms, which is likely to be secondary to the difficulties in communicating symptomology in young people as well as cancers occurring in rapidly developing tissues[99].

There were notable differences in median diagnostic intervals, with approximately 3 weeks observed for CNS tumours and 5-8 weeks observed lymphomas and sarcomas. This is in line with previous findings that showed protracted diagnostic intervals for these cancers in TYA [21]. Overall, for each cancer subtype, children and TYA shared similar clinical presentations.

There were also general symptoms that recurred across several cancers and the majority were seen at earlier stages. Examples include abdominal pain and limb pain (seen in leukaemias, lymphomas, sarcomas, abdominal tumours). Limb pain, which has previously also been associated with bacterial meningitis in children, may represent an important marker of serious illness in children who present to primary care [100]. Importantly, both early symptoms had higher predictive values when considered in paired combinations and were associated with all cancer subtypes in principal component analyses. Awareness of these combinations may provide a window of opportunity for clinicians to consider an underlying cancer diagnosis at earlier stages.

Interestingly, a lower overall risk of cancer for CTYA from more deprived backgrounds was observed compared to less deprived backgrounds. Across different cancer subtypes, this trend was also observed across lymphomas and leukaemias. This is in keeping with previous epidemiological studies but should be interpreted with caution as this study was unable to account for important confounders, such as maternal age and birthweight in analyses [82].

3.4.3 Strengths and Limitations

This study has several strengths. Firstly, it is the largest, contemporaneous population-based study exploring the association between clinical features encountered in primary care and a subsequent diagnosis of cancer in young people. Furthermore, it is nationally representative with incidence rates in line with those reported in the UK[1], which suggests results are likely to generalise outside the study population. Secondly, cancer cases were identified using the National Cancer Registry to ensure validity and accuracy of diagnoses, which was not available in previous studies [25, 26]. Finally, this study minimises the recall, observer and selection bias associated with case-control studies through its nested design, incidence density sampling, imputation of missing data and use of routinely collected electronic healthcare records.

This study also has several important limitations. Firstly, despite being the largest study to date, due to the rarity of CTYA cancers, power to detect symptom combinations associated with a diagnosis of individual subtypes of cancer is limited. Furthermore, due to the rarity of these cancers, despite identifying very large odds ratios, the positive predictive values of these symptoms remain low. Secondly, cancer subtypes were categorised broadly according to the International Classification for Childhood Cancers (ICCC-3) and this may limit the interpretation of subtype-specific results. For example, within the lymphomas category we included both Hodgkin's and non-Hodgkin's lymphomas, which are known to present with important differences [101] and this study would not be able to capture this. Thirdly, only clinical features, and not blood test findings were included in this Chapter, as blood test results were not available at the time of data extraction of this study. Fourth, whilst exploring symptom combinations, the date of the symptoms were not restricted (i.e., symptoms could co-occur on same day or across different dates) and this limits overall understanding of how symptom combinations may present temporally together in CTYA cancers. Fifth, there were

important limitations in the chosen clinical feature selection process, which was an initial univariable analysis and subsequent multivariable analysis based on statistical and regression coefficient thresholds. Although there is no consensus in the literature as to which variable selection process performs best [83], in this study, as explained in Section 3.2.6.1, this non-automated approach was deemed the most optimal way to balance clinical relevance and statistical significance. However, this process comes with several disadvantages. These include: 1) the univariable step has the potential to miss weakly associated individual symptoms, which may have a much more significant contribution when combined with other variables, and this increases the risk of selection bias; 2) the a priori statistical and regression coefficient thresholds can also be too constrained, contributing to missing relevant clinical features and, 3) using a non-automated approach can introduce investigator bias[83]. Sixth, the data extracted using SNOMED-CT clinical codes did not allow, due to constraints in the size of the data extract, to have access to all symptoms/signs within the dataset. Thus, it was not possible to explore the presence of potentially relevant clinical features within this dataset (i.e., any features occurring with a frequency of >5% in the dataset). Finally, as this study is an observational study based on routinely collected data, there is the possibility of recording bias of symptoms, although there was no evidence of this bias as the negative control symptom, eczema, was not associated with a cancer diagnosis; ascertainment bias, although the use of the National Cancer Registry minimises this bias; and residual confounding from unaccounted confounders.

3.4.4 Future Directions

This study provides important insight into which clinical features (or feature combinations) are associated with a CTYA cancer diagnosis and at what stage of the diagnostic interval these

features become relevant to a cancer diagnosis. The following research questions arise from the findings of this study:

1. Can these relevant clinical features be utilised to develop statistical models that predict a present but as yet undiagnosed cancer in a primary care setting? This will be addressed in Chapter 4.
2. What other sociodemographic (e.g., deprivation level), personal (e.g., age or sex) and cancer-related (e.g., type of cancer, year of diagnosis, age-at-diagnosis) factors, influence the diagnostic interval of CTYA cancer? This will be addressed in Chapter 5.

Code from Chapter

https://github.com/DefOx2020/Thesis/blob/06e6b1c26357ad03b73a6fc065ef8af9040ef1b4/Chapter%203

Publications from Chapter

Saatci, D., Oke, J., Harnden, A., & Hippisley-Cox, J. (2024). *Identifying early symptoms associated with a diagnosis of childhood, adolescent and young adult cancers: a population-based nested case-control study.* British journal of cancer, 131(6), 1032–1042. https://doi.org/10.1038/s41416-024-02786-5

Chapter 4 Derivation and internal-external validation of clinical prediction models for childhood, teenage and young adult blood and solid cancers

4.1. Background to chapter

Achieving early diagnosis of CTYA cancers has been an important priority worldwide, yet, making advancements have been difficult to date. As early cancer detection strategies in asymptomatic young people, such as screening and biomarker identification, have yielded little success, early diagnosis is mainly reached through the prompt recognition of symptomatic cancer [3]. However, early recognition of symptomatic cancer also remains a significant challenge for both young people and healthcare professionals due to the rarity, diversity and non-specificity of CTYA cancer presentations [23].

In the UK, one successful strategy to improve early recognition of symptomatic CTYA cancer has been through national awareness initiatives, such as the HEADSMART initiative [28]. These initiatives aim to raise public and clinical awareness of symptoms and signs associated with CTYA cancers to facilitate timely recognition and referral. HEADSMART, for example, has been shown to result in a significant reduction of the diagnostic interval for brain tumours [30, 52]. However, these initiatives alone have not been able to tackle all the barriers to early diagnosis in the UK to date. Specifically, delays in the recognition and referral of CTYA cancer from GP remain an ongoing challenge as general practitioners continue to report a high level of diagnostic uncertainty after taking part in the HEADSMART initiative [30, 102]. This is particularly important as CTYA most commonly first present to GP at the earliest stages of their disease in the UK [21]. Thus, there is still a pressing need for novel approaches to be

explored to supplement current of symptomatic CTYA cancer in the UK to improve early diagnosis.

Computer-based clinical decision tools (CDTs) are increasingly being used in clinical settings, supporting medical decision-making where challenges such as diagnostic uncertainty are present [43]. Overall, CDTs have been reported to reduce diagnostic errors, improve clinical practice and patient care [103]. In primary care settings, recent evidence suggests that technology-based CDTs provide the most successful interventions in reducing diagnostic inaccuracies [45].The potential for CDTs in adult cancer diagnosis has been highlighted as an "area of extraordinary opportunity" [46]. A systematic review has shown that these tools for cancer risk prediction have the potential to improve decision-making and clinical service outcomes in adults, as well as one study within the review showing reduction in time-to-diagnosis [104]. Despite these advancements in adult cancers, however, CDTs have not been explored in CTYA cancers.

4.2. Aims & Objectives

Accordingly, the overarching aim of this study is to develop and validate primary care-based risk prediction models to identify symptomatic CTYA who are at increased risk of cancer.

Specific objectives include:

1. Derive risk prediction models for different subgroups of CTYA cancer, including haematological and solid cancers.

2. Evaluate the performance of these models through internal-external cross-validation in order to assess geographical transportability and generalisability.

3. Explore the clinical utility of the models using decision curve analysis.

4.3. Methods

4.3.1 Study Population

The study population generated in Chapter 2 as detailed in Figure 2-6 was used in this Chapter. The full CTYA cohort was utilised, where each CTYA entered the study either at the cohort entry date or the date in which they had a first-recorded cancer-relevant symptom. They were subsequently followed-up for 6 months from this index date.

4.3.2 Outcome of Interest

The main outcome of interest was a cancer diagnosis in CTYA. This was defined in two ways.

First, the cancers were grouped into two broad categories, haematological and solid cancers. These two categories were selected a priori as they are referred to and managed by separate specialities (paediatric haematology and paediatric oncology, respectively) as per previous clinical knowledge. Haematological cancers included 1) Leukaemias and myelodysplastic diseases and 2) lymphomas and reticuloendothelial neoplasms. Solid cancers included 1) central nervous system and intraspinal tumours, 2) soft tissue and bone sarcomas, 3) abdominal tumours (renal tumours, neuroblastomas, hepatic tumours) and 4) gonadal germ cell tumours. All cancer diagnoses were identified and categorised as described in Chapter 2 and date of diagnosis was the earliest date recorded on any dataset.

Second, cancer diagnosis was also defined using the individual cancer subtypes, where sample size was sufficient (as detailed in section 4.3.8 below). These individual subtypes were 1) Leukaemias and myelodysplastic diseases and 2) lymphomas and reticuloendothelial neoplasms and 3) central nervous system and intraspinal tumours.

These two approaches were selected because it was anticipated that general practitioners would want to utilise these models either when they have a concern about a specific cancer

(i.e., individual cancer subtype models) or are overall concerned about an underlying cancer diagnosis (i.e., broad categories of cancer with different referral pathways).

As previous evidence suggests cancer cases in this age group present more frequently up to 12 months before diagnosis and most commonly 3 months before diagnosis, an a priori mid-point was selected at 6 months follow-up as per Chapter 3 [25].

4.3.3 Predictor variable selection

Symptoms, signs, blood tests and risk factors associated with each cancer type were identified through findings from Chapter 3, alongside a literature review, NICE guidelines and patient representative input [35, 38, 78, 81, 82, 105-108] (Table 4-1). Any variable recorded before cohort entry or after cohort exit was excluded.

For blood test values, results determined to be clinically incorrect/implausible/outliers with values above their 99th centile were excluded. Subsequently, UK national laboratory cut-off values were used to categorise blood test results into abnormal and normal [109]. This categorisation was chosen over using continuous blood test values because 1) only 3.1% (154,642/4,861,422) of the cohort had a blood test within their 6 months of follow-up and 2) those without tests were likely to be 'missing' not-at-random. Although categorisation is likely to reduce meaningful information gained from blood test results, given multiple imputation (using predictive mean matching) is likely to introduce bias due to aforementioned reasons [91], categorisation was determined to be the most optimal approach. Furthermore, a binary abnormal/normal categorisation rather than an abnormal/normal/not tested categorisation was selected as 1) clinical experience suggested that those who were not tested were most likely to have a normal blood test result given testing was not clinically indicated and 2) previous evidence has shown potential biases introduced by having a non-tested category, as

those tested with normal results were found to have higher risk of cancer compared to those who were never tested[110].

The date of the blood test had to be after the date of the first-recorded cancer-relevant symptom/study entry date (i.e., index date) and within the 6-month period from this index date, before a cancer diagnosis.

Any continuous candidate variable (e.g., age) was assessed for nonlinearity and handled using fractional polynomials. In this case, categorisation of these non-linear continuous variables was not carried out in order to ensure meaningful information is not lost. Fractional polynomials were chosen over restricted cubic splines in order to balance computational capacity and prior evidence that fractional polynomials capture non-linear associations with the age variable used in QResearch cohorts [111]. Predictor variables were identified through SNOMED-CT and ICD-10 codes on primary care and secondary care records, respectively, as per Chapter 3.

The earliest record, at point of entry to cohort, was used to define sociodemographic risk factors. Ethnicity was defined as self- or parent-reported ethnicity on primary care health records. Ethnic groups were recorded based on the 2021 Census of England and Wales in 2 broad categories (White, Other) to ensure sufficient sample size in each category. Level of deprivation was assessed through the Townsend deprivation score as described in Chapters 2 & 3.

Table 4-1 *Candidate predictor variables explored in model development*

	Candidate variables
Demographic	Age at diagnosis in years
	Sex
	Deprivation level using Townsend Deprivation Score
	Ethnicity in two categories (White/Other)
Clinical Features	Organomegaly/abdominal mass
	Lymphadenopathy
	Fever
	Limb pain

	Joint pain
	Limp/abnormal gait
	Bruising
	Tiredness
	Looks anaemic/pale
	Abdominal pain
	Rash
	Abnormal skin lesion
	Cough
	Chest pain
	Headache
	Vomiting
	Head/neck lump
	Lump on body
	Hemiparesis
	Squint
	Visual acuity problems
	First seizure
	Haematuria
	Constipation
	Dizziness
	Testicular swelling
	Pruritus
	Haemoglobin
Blood Tests	White cell count (including differential count)
	Platelet count
	Mean Cell Volume
	C-reactive protein
	Erythrocyte sedimentation rate
	Ferritin
	Lactate Dehydrogenase
	Alanine transaminase (ALT)
	Bilirubin
	Albumin
	anaemia/pale & low haemoglobin
Interaction terms	anaemia/pale & tiredness
	limb & joint pain
	limb/joint pain & limp
	fever & lymphadenopathy
	fever & cough
	fever & high white cell count
	fever & high CRP
	bruising & low platelets
	lump on body & lymphadenopathy
	head/neck lump & lymphadenopathy
	chest pain & cough
	dizziness and visual acuity problems

4.3.4 Handling Missing Data

Multiple imputation with chained equations was used to impute missing values for ethnicity (using multinomial regression) and Townsend Quintile (using ordered logistic regression) under the missing at random assumption [90]. Due to computational capacity as the full cohort was utilised, 5 imputed datasets were generated to account for % missingness, and this has been previously shown to have high efficiency [112]. The imputation model included all candidate predictors and the outcome (solid/haematological cancers/individual cancers) and statistical models were developed on each of the 5 imputed data sets and the estimates pooled using the Rubin rules [91].

4.3.5 Model Development

Model development consisted of two regression approaches. First, Cox proportional hazards regression modelling was explored and the proportionality assumption tested using Schoenfeld residuals. When the proportionality assumption was violated (Appendix Chapter 9, Supplementary Table 9-12), binary logistic regression was used to develop the model, where the analyses were restricted to the population who had at least 6 month follow up. First, a full model was fitted using all candidate predictors in all imputed datasets and combined using Rubin's rules [91]. Second, the pooled model was used to select any categorical predictor with odds ratios >1.1 or <0.9 (at $p<0.01$) and any continuous predictors with significance of $p<0.01$. Third, these selected predictors were used to refit the final model. This approach was chosen in place of backward/stepwise elimination techniques and penalised regression approaches to ensure that clinical relevance of predictors were fully considered in addition to the statistical magnitude of predictors. Interaction terms, which were identified a priori through available paediatric and general practice expertise, were included within the model development and interactions between variables were considered

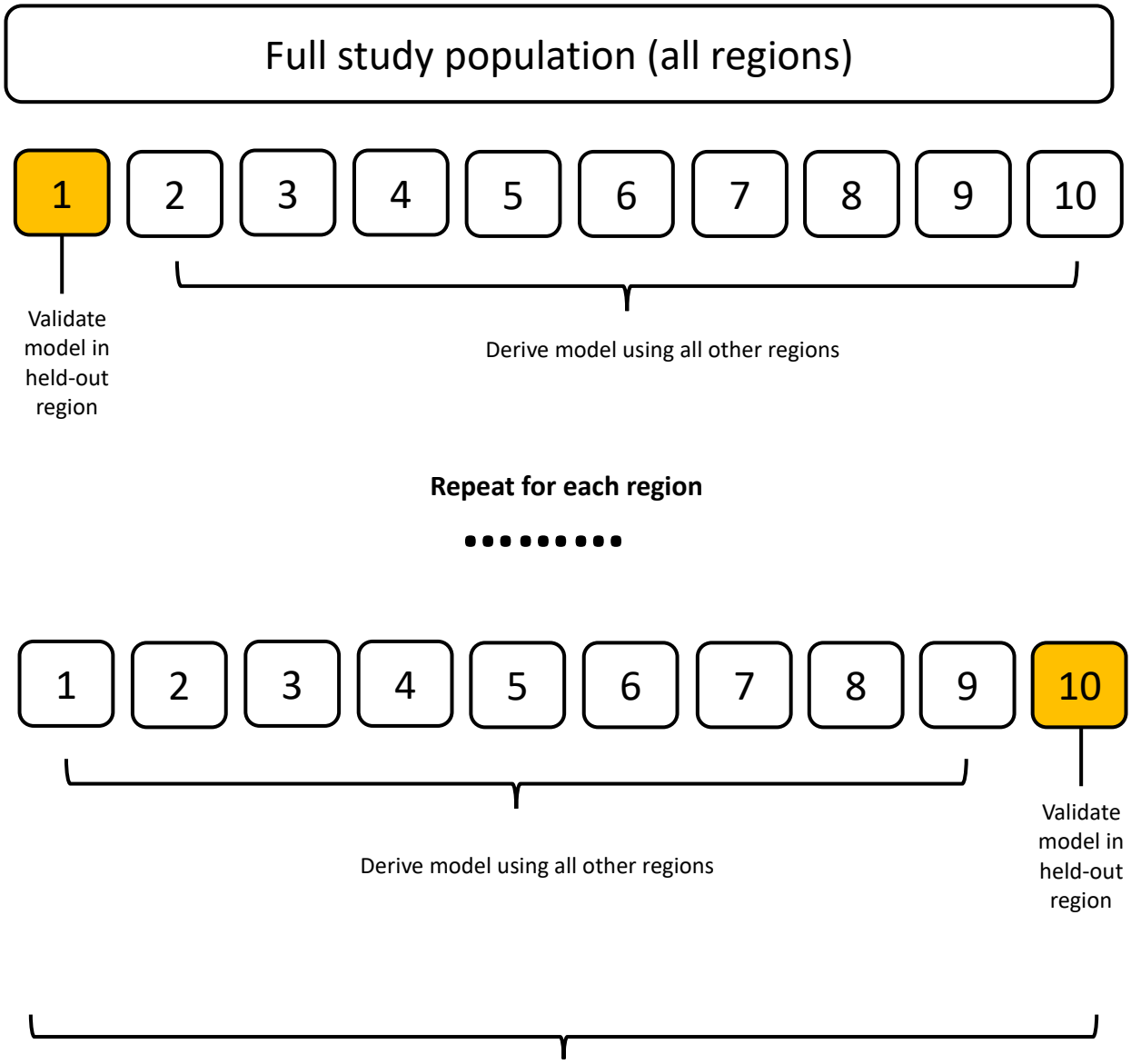
based on clinical plausibility (Table 4-1). To account for clustering of young people within GPs, clustered standard errors were used in regression analyses.

4.3.6 Model Evaluation

Model validation was carried out using the internal-external cross validation approach [113, 114]. This approach was selected as it allows for the assessment of overall model performance as well as potential transportability of the model to a 'distinct' population and takes advantage of the availability of data from different geographical locations within QResearch Database (up to 10 locations across England). Given CTYA cancers are uncommon, this approach is particularly advantageous as it also allows the entire cohort to be used for both derivation and internal-external validation. The summary of this approach is as follows: first, one geographical region is 'selected' to develop the model. Second, data from all other geographical regions are used to assess model performance. These two steps are repeated for each geographical region. Finally, random effects meta-analysis using the Hartung-Knapp-Sidik-Jonkmann method is carried out to pool region-level performance measures and provide a pooled estimate of performance measures [115] (Figure 4-1).

Performance measures used were area-under-the-curve (AUC) for model discrimination, where a good AUC was described for any AUC above 0.8, and calibration-in-the-large, calibration slope and calibration plots for model calibration, where any slope which equals 1 and calibration-in-the-large which equals 0 are defined as perfect calibration [116, 117].

Figure 4-1 *Schematic overview of the internal-external cross-validation process*



Inter-regional heterogeneity of performance measures, measured using I^2, was explored further using meta-regression [118]. This measure assesses the percentage of the total observed variance that is accounted for through between-study variation and is categorised into low (<25%), moderate (25-50%) and high (>50%) [119]. Potential factors which may attribute to heterogeneity were selected a priori through findings in Chapters 2 & 3. These factors were ethnicity, deprivation level and age-at-diagnosis.

Decision curve analysis was used to assess clinical utility. This analysis involves calculating a 'net benefit' (i.e., the trade-off between true positives and false negatives) across a range of

threshold risk probabilities (i.e., cut-offs that prompt an intervention). This net benefit is then compared between the developed models and two default intervention strategies: "intervention for all" and "intervention for none" [120].

To further contextualise clinical utility through decision curve analysis, the developed models were compared to already existent guidance, NICE guidelines for haematological and solid tumours and HEADSMART symptoms/signs for CNS tumours. This was achieved by generating models that only include symptoms/signs listed by NICE for haematological and solid cancers, separately, as well as models that only include symptom/signs listed by HEADSMART. Subsequently each model's clinical utility was evaluated using decision curve analysis.

All analyses were carried out using Stata versions 17 & 18 [92] and adhere to the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) [93], Reporting of studies Conducted using Observational routinely collected health Data (RECORD) [94] and Transparent Reporting of a Multivariable Prediction Model for individual Prognosis or Diagnosis (TRIPOD) guidelines [121].

4.3.7 Sensitivity and Subgroup Analyses

Given that only a small proportion of young people (3.1%) had blood tests carried out, two separate models were developed per outcome. The first model was based on symptom/signs only and the second model was exploratory; derived and evaluated with symptoms/signs as well as blood test results included.

Where sample size was sufficient (see below for sample size calculations), cancer-subtype specific models were derived and internally-externally cross validated. The subtypes with sufficient size included 1) leukaemias, 2) lymphomas, and 3) CNS tumours.

4.3.8 Sample size calculations

Prior to carrying out analyses, sample size calculations were carried out using 'pmsampsize' on Stata [122]. The time point was set at 6 months and used 15% of the maximum permitted

Cox-Snell R-squared (derived from Riley et al., 2020 [122]) as there are no previous risk prediction models for its derivation. CRUK data were used to estimate incidence of cancers. Previous studies using QResearch have identified a cohort of ~5 million children and young adults, with approximately 2 million children and 3 million teenagers and young adults [123]. Altogether, this indicates that our study has sufficient sample size for exploring haematological and solid cancers as well as leukaemias, lymphomas and CNS tumours (Appendix Chapter 9, Supplementary Table 9-13).

4.4. Results

4.4.1 Population Characteristics

There were 4,861,423 children, teenagers and young adults aged 0-24 years registered with a general practitioner between 1st January 1998 and 31st December 2019 with at least 6 months follow-up (excluded due to <6 months follow up n=377,320). Of these, there were 617 cancer diagnoses at 6 months follow-up (0.01%), with 308 haematological and 309 solid cancers. The baseline demographic characteristics, alongside clinical and blood test results are summarised in Table 4-2. Compared to the general population of CTYA, those with cancer were more likely to be male (p=0.028) and otherwise were comparable in terms of ethnicity and deprivation level at baseline.

Table 4-2 *Baseline characteristics of a) haematological cancer, b) solid cancer and c) full population cohorts used in internal-external cross validation. All data except age has been expressed in numbers (% column). Age in years is rounded to the nearest integer.*

Characteristics	Haematological cancers (n=308)	Solid cancers (n=309)	All population (n=4,861,423)
Age			
Median ± IQR (years)	9 [2,18]	8 [1,17]	9 [2,19]
Sex			
Male	177 (57.5)	172 (55.7)	2,491,355 (51.2)
Female	131 (42.5)	137 (44.3)	2,370,068 (48.8)
Ethnicity			
White	80 (26.0)	74 (24.0)	1,201,928 (24.7)
Other	49 (15.9)	48 (15.5)	807,532 (16.6)

Missing	179 (58.1)	187 (60.5)	2,851,963 (58.7)
Townsend Deprivation Score			
1 (most affluent)	65 (21.1)	63 (20.4)	951,952 (19.6)
2	68 (22.1)	69 (22.3)	952,139 (19.6)
3	71 (23.1)	62 (20.1)	966,012 (19.9)
4	59 (19.2)	57 (18.5)	989,351 (20.4)
5 (most deprived)	42 (14.0)	53 (17.1)	976,181 (20.1)
Missing	<5	5 (1.6)	25,788 (0.5)
Clinical Features			
Organomegaly/abdominal mass	<5	<5	974 (0.02)
Lymphadenopathy	65 (21.1)	7 (2.3)	51,544 (1.1)
Fever	42 (13.6)	17 (5.9)	161,056 (3.3)
Limb pain	14 (4.6)	27 (8.7)	106,552 (2.2)
Joint pain	12 (3.9)	15 (4.9)	61,100 (1.3)
Limp/abnormal gait	<5	5 (1.6)	4,959 (0.1)
Bruising	13 (4.2)		21,138 (0.4)
Tiredness	13 (4.2)	<5	59,977 (1.2)
Looks anaemic/pale	7 (2.3)		7,188 (0.1)
Abdominal pain	28 (9.1)	25 (8.1)	122,173 (2.5)
Rash	17 (5.5)	10 (3.2)	312,505 (6.4)
Abnormal skin lesion	<5	<5	20,127 (0.4)
Cough	29 (9.4)	15 (4.9)	305,418 (6.3)
Chest pain	11 (3.6)	<5	33,293 (0.7)
Headache	9 (2.9)	44 (14.2)	84,759 (1.7)
Vomiting	6 (2.0)	13 (4.2)	27,072 (0.5)
Head/neck lump	32 (10.4)	<5	8,909 (0.2)
Lump on body	15 (4.9)	16 (5.2)	32,858 (0.7)
Hemiparesis		6 (1.9)	1,090 (0.02)
Torticollis		<5	6,419 (0.1)
Squint		7 (2.3)	14,213 (0.3)
Visual acuity problems		8 (2.6)	8,464 (0.2)
First seizure		15 (4.9)	5,334 (0.1)
Haematuria	<5	8 (2.6)	29,529 (0.6)
Constipation		<5	6,226 (0.1)
Testicular swelling		6 (1.9)	29 (0.001)
Pruritus	<5	<5	8,535 (0.2)
Weight loss	<5		5,735 (0.1)
Feels unwell	<5		57 (0.001)
Dizziness	<5		377 (0.007)
Blood Test Abnormalities			
Haemoglobin < 12 mg/dl	68 (22.1)	17 (5.5)	20,848 (0.4)
Haemoglobin > 17 mg/dl	<5	<5	162 (0.003)
Neutrophil count $<1.5x10^9/L$	35 (11.4)	6 (1.9)	3,356 (0.07)
Neutrophil count $>8x10^9/L$	21 (6.8)	5 (1.6)	6,404 (0.1)

Lymphocyte count <1.1x10^9/L	11 (3.6)	<5	2,115 (0.04)
Lymphocyte count >4.5x10^9/L	7 (2.3)	<5	8,888 (0.2)
Eosinophil count >0.8 x10^9/L	8 (2.6)	<5	4,094 (0.08)
Platelet count < 150x10^9/L	46 (14.9)	6 (1.9)	2,149 (0.04)
Platelet count > 400x10^9/L	19 (6.2)	7 (2.3)	9,878 (0.2)
Mean Cell Volume <70 fL	<5	<5	4,059 (0.08)
Mean Cell Volume > 96 fL	8 (2.6)	<5	3,237 (0.06)
C-reactive protein > 3 mg/dl	26 (8.4)	<5	3,187 (0.06)
Erythrocyte sedimentation rate > 20 mm/hr	30 (9.7)	6 (1.9)	5,209 (0.1)
Ferritin > 400 ng/ml	8 (2.6)	<5	423 (0.01)
Lactate Dehydrogenase > 300 U/L	11 (3.6)	<5	805 (0.02)
Alanine transaminase (ALT) > 60	6 (2.0)	<5	1,983 (0.04)
Bilirubin > 20 µmol/L	<5	<5	1,100 (0.02)
Albumin ≤ 35 g/L	8 (2.6)	<5	3,961 (0.08)

Furthermore, demographic background of those with and without blood test results are summarised in Table 4-3.

Table 4-3 *Baseline characteristics by blood test results. All data has been expressed in numbers (% row). Age in years is rounded to the nearest integer above. TQ = Townsend Deprivation Quantile 1-5.*

Value	Number tested (N=4,861,423)	Age	Sex - Male	Sex – Female	Ethnic group - White	Ethnic group - Other	Ethnic group - missing	TQ 1	TQ 2	TQ 3	TQ 4	TQ 5	TQ - missing
Haemoglobin (mg/dl)													
Normal (>12 and <17)	98,942	17 [10,20]	40,545 (41.0)	58,397 (59.0)	26,777 (27.0)	17,464 (17.7)	54,701 (55.3)	17,287 (17.4)	18,306 (18.5)	20,425 (20.6)	22,009 (22.2)	20,413 (20.6)	502 (0.5)
>17	151	5 [1,18]	147 (97.4)	<5	37 (24.5)	31 (20.5)	83 (55.0)	32 (21.2)	33 (21.8)	35 (23.2)	28 (18.6)	23 (15.2)	-
<12	20,210	19 [15,21]	7,174 (35.5)	13,036 (64.5)	4,143 (20.5)	6,756 (33.4)	9,311 (46.1)	2,265 (13.0)	3,067 (15.2)	4,063 (20.1)	4,782 (23.7)	5,558 (27.5)	115 (0.6)
Not tested	4,742,119	9 [2,19]	2,443,494 (51.5)	2,298,625 (48.5)	1,170,977 (24.7)	783,282 (16.5)	2,787,860 (58.8)	932,010 (19.7)	930,732 (19.6)	941,498 (19.9)	962,534 (20.3)	950,177 (20.0)	25,168 (0.5)
Neutrophil count (x10^9/L)													
Normal (>1.5 & <8)	105,407	15 [8,20]	42,229 (40.1)	63,178 (59.9)	28,145 (26.7)	21,762 (20.7)	55,500 (52.6)	17,870 (17.0)	19,073 (18.1)	21,816 (20.7)	23,549 (22.3)	22,549 (21.4)	550 (0.5)
<1.5	3,240	9 [2,17]	1,738 (53.6)	1,502 (46.4)	557 (17.2)	1,183 (36.5)	1,500 (46.3)	1,056 (17.3)	1,239 (20.3)	1,257 (20.6)	1,269 (20.8)	1,259 (20.6)	29 (0.4)
>8	6,109	17 [7,20]	2,197 (36.0)	3912 (64.0)	1,682 (27.5)	797 (13.1)	3,630 (59.4)	441 (13.6)	480 (14.8)	595 (18.4)	714 (22.0)	996 (30.7)	14 (0.4)
Not tested	4,746,666	9 [2,19]	2,301,470 (48.5)	2,445,196 (51.5)	1,171,550 (24.7)	783,791 (16.5)	2,791,325 (58.8)	932,587 (19.7)	931,346 (19.6)	942,353 (19.8)	963,821 (20.3)	951,367 (20.0)	25,192 (0.5)
Lymphocyte count (x10^9/L)													
Normal (>1.1 & <4.5)	103,112	16 [9,20]	40,055 (38.8)	63,057 (61.2)	27,907 (27.1)	20,414 (19.8)	54,791 (53.1)	17,495 (17.0)	18,747 (18.2)	21,322 (20.7)	23,074 (22.4)	21,952 (21.3)	522 (0.5)
<1.1	1,989	18 [12, 20]	834 (41.9)	1,155 (58.1)	1,815 (20.7)	2,927 (33.3)	4,039 (46.0)	340 (17.1)	411 (20.7)	404 (20.3)	446 (22.4)	379 (19.1)	9 (0.4)

>4.5	8,781	1 [1,3]	4,853 (55.3)	3,298 (44.7)	497 (25.0)	323 (16.2)	1,169 (58.8)	1,330 (15.2)	1,431 (16.3)	1,774 (20.2)	1,976 (22.5)	2,218 (25.3)	52 (0.6)
Not tested	4,747,540	9 [2,19]	2,445,618 (51.5)	2,301,922 (48.5)	1,171,715 (24.7)	783,869 (16.5)	2,791,956 (58.8)	932,789 (19.7)	931,549 (19.6)	942,521 (19.9)	963,857 (20.3)	951,622 (20.0)	25.202 (0.5)
Eosinophil count (x10^9/L)													
Normal (<0.8)	109,835	16 [8,20]	43,728 (39.8)	66,107 (60.2)	29,183 (26.6)	22,608 (20.6)	58,044 (52.8)	18,441 (16.8)	19,819 (18.0)	22,587 (20.6)	24,576 (22.4)	23,847 (21.7)	565 (0.5)
Eosinophil count >0.8	3,964	9 [4,16]	1,938 (48.9)	2,026 (51.1)	946 (23.9)	1,012 (25.5)	2,006 (50.6)	689 (17.4)	718 (18.1)	837 (21.1)	865 (21.8)	834 (21.0)	21 (0.5)
Not tested	4,747,623	9 [2,19]	2,445,694 (51.5)	2,301,929 (48.5)	1,717,805 (24.7)	783,913 (16.5)	2,791,905 (58.8)	932,824 (19.7)	931,601 (19.6)	942,597 (19.9)	963,912 (20.3)	951,490(20.0)	25,199 (0.5)
Platelet count (x10^9/L)													
Normal (> 150 & < 400)	107,184	16 [9,20]	42,574 (39.7)	64,610 (60.3)	28,213 (26.3)	20,970 (19.6)	58,001 (54.1)	17,936 (16.7)	19,354 (18.1)	22,018 (20.5)	24,139 (22.5)	23,187 (21.6)	550 (0.5)
< 150	2,019	17 [9, 21]	1,059 (52.5)	960 (47.5)	473 (23.4)	406 (20.1)	1,140 (56.5)	384 (19.0)	346 (17.1)	399 (19.8)	446 (22.1)	437 (21.6)	7 (0.5)
> 400	9,649	4 [1,13]	4,204 (43.6)	5,445 (56.4)	2,225 (23.1)	2,713 (28.1)	4,711 (48.8)	1,512 (15.7)	1,708 (17.7)	2,010 (20.8)	2,173 (22.5)	2,186 (22.7)	60 (0.6)
Not tested	4,742,570	9 [2,19]	2,443,523 (51.5)	2,299,047 (48.5)	1,171,023 (24.7)	783,444 (16.5)	2,788,103 (58.8)	932,122 (19.7)	930,730 (19.6)	941, 594 (19.9)	962,595 (20.3)	950,361 (20.0)	25,168 (0.5)
Mean Cell Volume (fl)													
Normal (>70 & <96)	117,587	16 [7,20]	44,660 (40.4)	65,942 (59.6)	29,367 (26.6)	21,642 (19.6)	59,593 (53.9)	18,923 (17.1)	20,233 (19.3)	22,784 (20.6)	24,553 (22.2)	23,536 (21.3)	573 (0.5)
< 70	3,984	3 [1, 15]	1883 (47.3)	2101 (52.7)	496 (12.5)	2,050 (51.5)	1,438 (36.1)	2379 (9.5)	494 (12.4)	804 (20.2)	1,068 (26.8)	1,219 (30.6)	20 (0.5)
> 96	3,001	20 [17,22]	762 (25.4)	2,239 (74.6)	942 (31.4)	302 (10.1)	1,757 (58.6)	433 (14.4)	489 (16.3)	673 (22.4)	735 (24.5)	652 (21.7)	19 (0.6)

Not tested	4,744,601	9 [2,19]	2,444,289 (51.5)	2,300,312 (49.5)	1,171,334 (24.7)	783,630 (16.5)	2,789,637 (58.8)	932,362 (19.7)	931,082 (19.6)	941,922 (19.8)	963,162 (20.3)	950,890 (20.4)	25,183 (0.5)
C-reactive protein level (mg/dl)													
Normal (<3)	17,816	15 [7,20]	7,659 (43.0)	10,157 (57.0)	5,412 (30.4)	3,565 (20.0)	8,839 (49.6)	3,545 (19.9)	3,649 (20.5)	3,941 (22.1)	3,797 (21.3)	2,755 (15.5)	129 (0.7)
> 3	19,738	17 [8,20]	8,017 (40.6)	11,721 (59.4)	5,895 (29.9)	3,811 (19.3)	10,032 (50.8)	3,343 (16.9)	3,495 (17.7)	3,971 (20.1)	4,381 (22.2)	4,447 (22.5)	101 (0.5)
Not tested	4,824,634	9 [2,19]	2,475,918 (51.3)	2,348,716 (48.7)	1,190,832 (24.7)	800,248 (19.6)	2,833,554 (58.7)	945,209 (19.6)	945,154 (19.6)	958,271 (19.9)	981,340 (20.3)	969,095 (20.1)	25,565 (0.5)
Erythrocyte sedimentation rate (mm/hr)													
Normal (< 20)	41,813	16 [9,20]	17,830 (42.6)	23,983 (57.4)	11,641 (27.8)	7,890 (18.9)	22,282 (53.3)	7,512 (17.9)	7,721 (18.5)	8,509 (20.4)	8.834 (21.1)	9,010 (21.6)	225 (0.5)
> 20 mm/hr	4,973	15 [6, 20]	1,533 (30.8)	3,440 (69.2)	1,165 (23.4)	1,493 (30.0)	2,315 (46.6)	767 (15.4)	860 (17.3)	942 (19.0)	1,166 (23.4)	1,216 (24.4)	22 (0.4)
Not tested	4,815,402	9 [2,19]	2,472,231 (51.3)	2,343,171 (48.7)	1,189,333 (24.7)	798,241 (16.6)	2,827,828 (58.7)	943,817 (19.6)	943,717 (19.6)	956,732 (19.8)	979,518 (20.3	966,071 (20.1)	25,547 (0.5)
Ferritin (ng/ml)													
Normal (<400)	42,439	16 [7,20]	13,382 (31.5)	29,048 (68.5)	11,351 (26.8)	12,731 (30.0)	18,348 (43.2)	5,597 (13.2)	6,293 (14.8)	8,603 (20.2)	10,265 (14.2)	11,446 (27.0)	226 (0.5)
> 400	118	19 [15,22]	88 (74.6)	30 (25.4)	35 (29.7)	37 (31.4)	46 (38.9)	21 (17.8)	20 (16.9)	21 (17.8)	26 (22.0)	30 (25.4)	-
Not tested	4,819,640	9 [2,19]	2,478,124 (51.4)	2,341,516 (48.6)	1,190,753 (24.7)	794,856 (16.5)	2,834,031 (58.8)	946,479 (19.6)	945,985 (19.6)	957,559 (19.8)	979,227 (20.3)	964,821 (20.0)	25,569 (0.5)
Lactate Dehydrogenase (U/L)													
Normal (< 300)	402	19 [13,20]	164 (40.8)	238 (59.2)	116 (28.9)	56 (13.9)	230 (57.2)	70 (17.4)	99 (24.6)	88 (21.9)	70 (17.4)	74 (18.4)	-
> 300	16	8 [4,19]	<5	12 (75.0)	<5	6 (27.5)	8 (50.0)	<5	<5	<5	<5	<5	-

Not tested	4,861,770	9 [2,19]	2,492,418 (51.3)	2,370,352 (48.7)	1,202,021 (24.7)	807,562 (16.6)	2,852,187 (58.7)	952,026 (19.6)	952,196 (19.6)	966,093 (19.9)	989,444 (20.4)	976,217 (20.1)	25,794 (0.5)
Alanine transaminase (U/L)													
Normal (< 60)	70,502	16 [9,20]	28,239 (40.1)	42,263 (59.9)	19,939 (28.3)	15,379 (21.8)	35,184 (49.9)	12,287 (17.4)	12,716 (18.0)	14,386 (20.4)	15,666 (22.2)	15,082 (21.4)	365 (0.5)
> 60	1,876	19 [13, 21]	691 (36.8)	1,185 (63.2)	497 (26.5)	350 (18.7)	1,029 (54.8)	332 (17.7)	344 (18.4)	395 (21.1)	393 (20.9)	402 (21.4)	10 (0.5)
Not tested	4,789,810	9 [2,19]	2,462,170 (51.4)	2,327,640 (48.6)	1,181,703 (24.7)	791,895 (16.5)	2,816,212 (58.8)	939,478 (19.6)	939,238 (19.6)	951,402 (19.9)	973,459 (20.3)	960,812 (20.1)	25,420 (0.5)
Bilirubin (µmol/L)													
Normal (< 20)	66,686	16 [8,20]	26,267 (39.4)	40,419 (60.6)	19,106 (28.7)	14,459 (21.7)	33,121 (49.7)	11,599 (17.4)	11,924 (17.9)	13,643 (20.5)	14,809 (22.2)	14,355 (21.5)	356 (0.5)
> 20	3,745	18 [13,21]	2,247 (60.0)	1,498 (40.0)	1,045 (27.9)	693 (18.5)	2,007 (53.6)	683 (18.2)	693 (18.5)	779 (20.6)	794 (21.2)	790 (21.1)	15 (0.4)
Not tested	4,791,757	9 [2,19]	2,463,080 (51.4)	2,328,677 (48.6)	1,181,988 (24.7)	792,472 (16.5)	2,817,297 (58.8)	939,815 (19.6)	939,681 (19.6)	951,770 (19.9)	973,915 (20.3)	961,152 (20.1)	25,424 (0.5)
Albumin (g/L)													
Normal (> 35)	77,653	16 [9,20]	32,036 (41.3)	45,617 (58.7)	21,457 (27.6)	17,237 (22.2)	38,959 (50.2)	12,056 (16.8)	13,667 (17.6)	15,852 (20.4)	17,411 (22.4)	17,271 (22.2)	396 (0.5)
≤ 35	1,034	18 [5,21]	267 (25.8)	767 (74.2)	262 (25.3)	258 (25.0)	514 (49.7)	162 (15.7)	181 (17.5)	207 (20.0)	235 (22.7)	240 (23.2)	9 (0.8)
Not tested	4,783,501	9 [2,19]	2,459,291 (51.4)	2,324,210 (48.6)	1,180,420 (24.7)	790.129 (16.5)	2,812,952 (58.8)	938,879 (19.6)	938,450 (19.6)	950,124 (19.9)	971,872 (20.3)	958,786 (20.0)	25,390 (0.5)

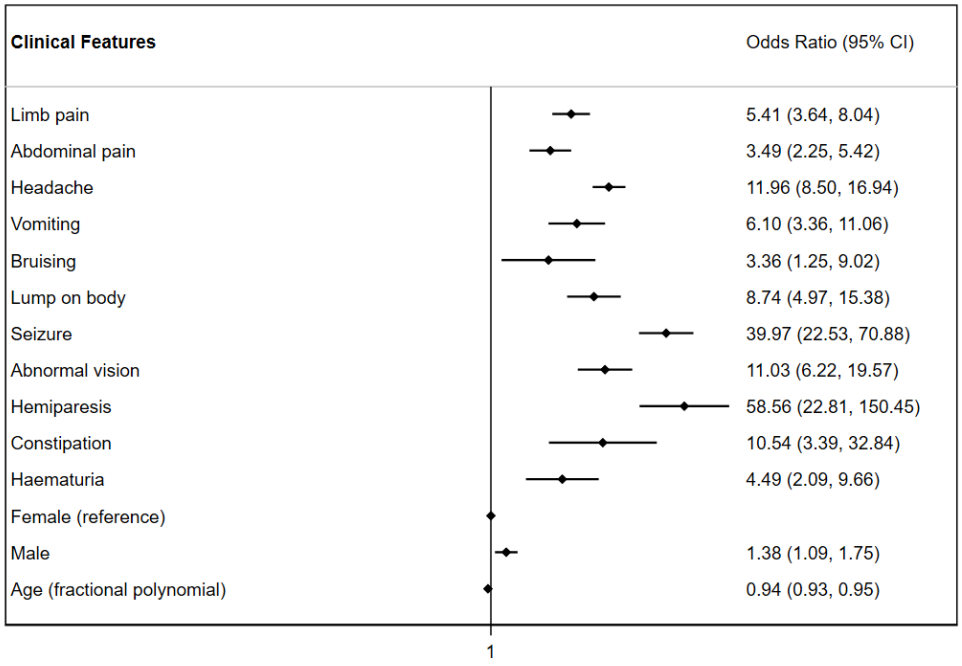
4.4.2 Models for Combined Solid Cancers

4.4.2.1 Model Derivation

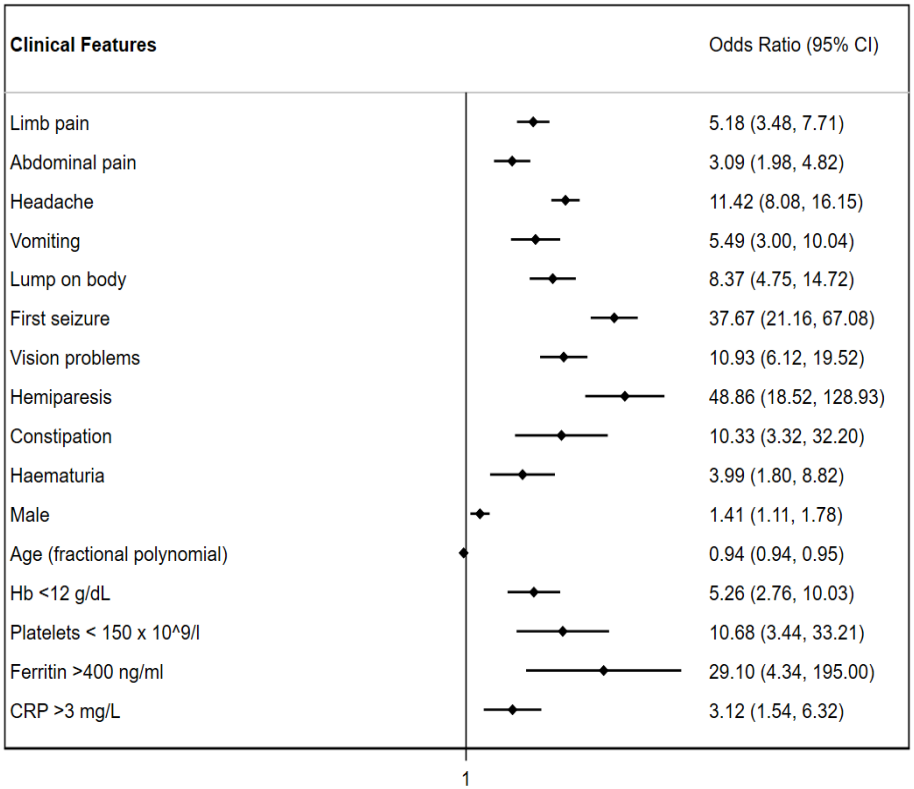
The predictor variables in the final model included sociodemographic factors such as sex and non-linear fractional polynomial terms for age (-2, -2) (Appendix Chapter 9, Supplementary Figure 9-4); 10 symptoms/signs (limb pain, abdominal pain, headache, vomiting, lump on body, first seizure, vision problems, hemiparesis, constipation, haematuria); and in the exploratory symptom + test model, an additional 4 abnormal blood test results (low haemoglobin (<12 mg/dl), low platelets (<150 x 10^9/L), high ferritin (>400 ng/ml) and high CRP (>3mg/dl)) (Figure 4-2).

Figure 4-2 *Final logistic regression model predicting odds of solid cancer 6 months following study entry/symptom onset in a) symptom-only and b) symptom + blood test model. Models contain fractional polynomials for age (-2,-2). CI= Confidence interval, Hb= haemoglobin, CRP=C-reactive protein.*

a)



b)

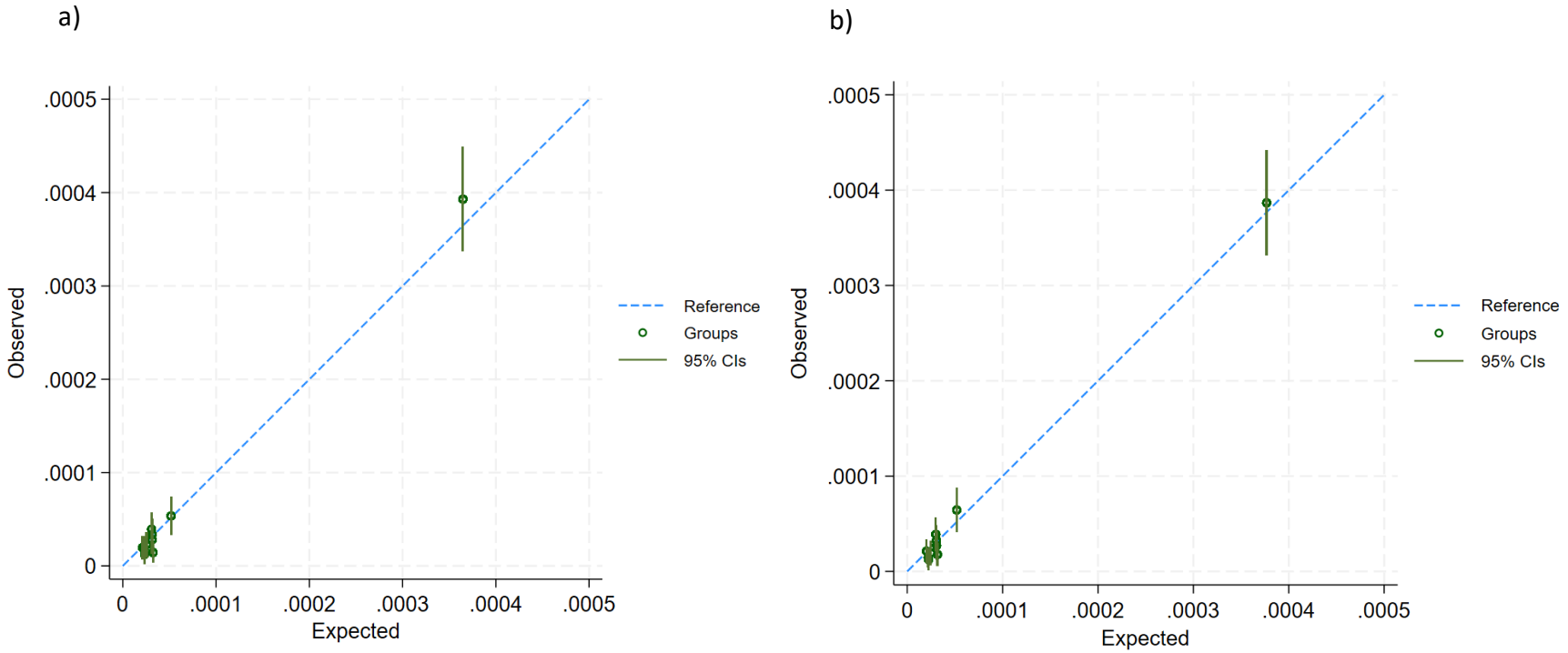


4.4.2.2 Model Performance

The symptom-only model was found to have good discrimination and was well-calibrated following internal-external validation (AUC=0.80 [95%CI 0.77 to 0.83], slope=0.98 [0.92 to 1.04], calibration-in-the-large=0.01 [-0.11 to 0.12]). Developing the solid cancer model with blood test results did not greatly alter performance measures (AUC=0.81 [95%CI 0.78 to 0.84], slope=0.98 [0.91 to 1.04], calibration-in-the-large=-0.02 [-0.09 to 0.04]).

Calibration plots for both solid cancer models (symptom only and symptom + test) demonstrated that models tended towards overestimation at the lower predicted risk ranges (Figure 4-3).

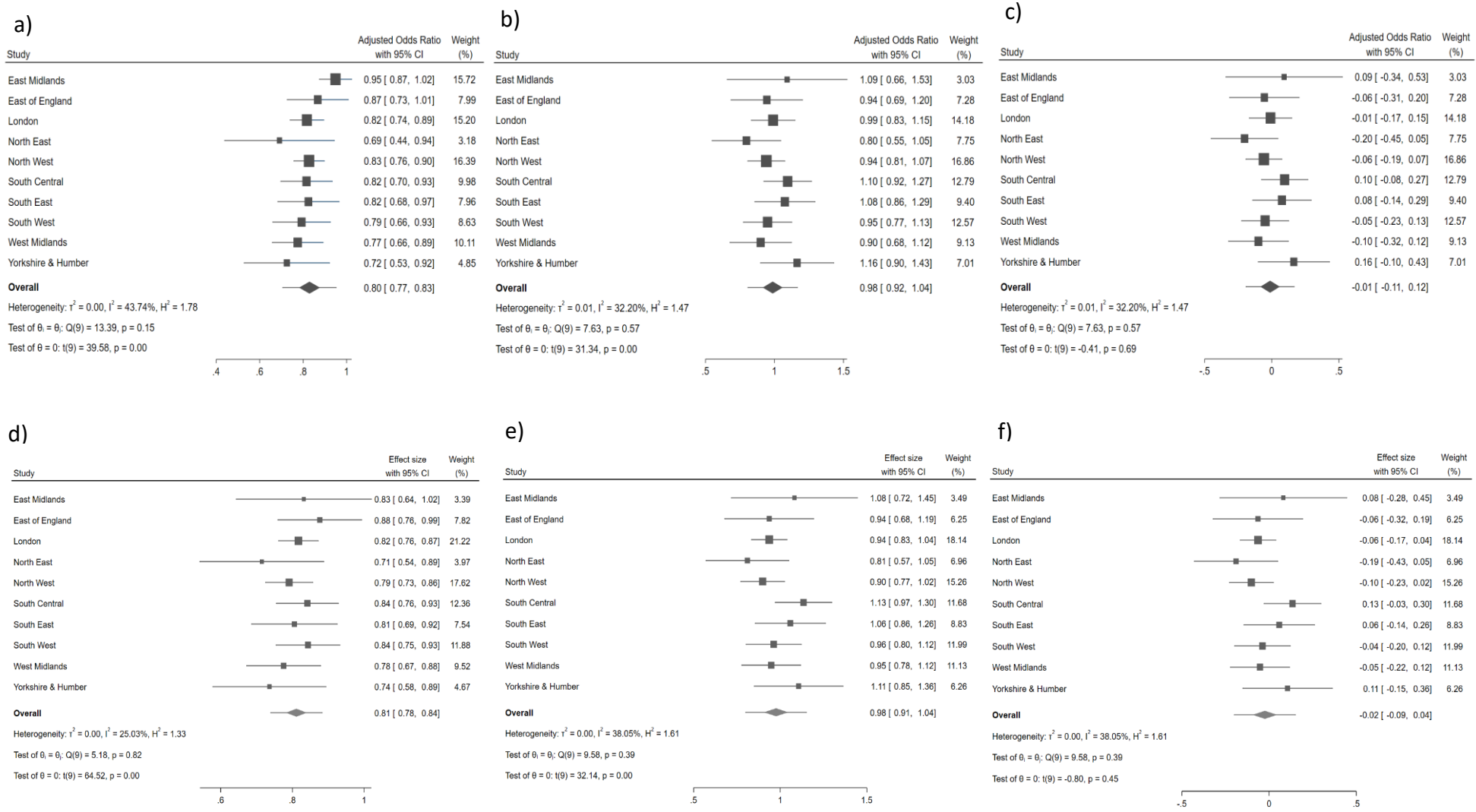
Figure 4-3 *Calibration of the a) symptom-only and b) symptom + blood test solid cancer models tested demonstrated through calibration plots which show the alignment between predicted and observed risks across 10 risk deciles/groups.*



4.4.2.3 Heterogeneity

Meta-analysis of performance measures across the 10 geographical regions demonstrated that there was moderate level of heterogeneity for AUC (I^2=43.5%), calibration slope (I^2=32.2%) and calibration-in-the large (I^2=32.2%) (Figure 4-4). Age, ethnicity and deprivation level did not account for the observed heterogeneity in meta-regression analyses (Appendix Chapter, Supplementary Table 9-14). The heterogeneity of performance measures for the symptom + blood test model was similar, with slightly lower heterogeneity observed for discrimination measures [AUC (I^2=25.0%), and higher heterogeneity for calibration slope and calibration-in-the-large (I^2=38.0%)] (Figure 4-4).

Figure 4-4 *Region-level and pooled performance estimates following internal-external cross validation for a) & d) area-under-the-curve (AUC), b) & e) calibration slope and, c) & f) calibration-in-the-large for symptom-only (a-c) and symptom + blood test solid cancer models (d-f). Effect Size=adjusted Odds ratio, CI= Confidence interval.*



4.4.2.4 Model fairness

When the model performance was assessed across different age groups (child 0-14 years & TYA 15-24 years), sexes (male & female) and ethnic groups (White & non-White ethnic groups), models had lower discriminatory ability for TYA, females and non-white ethnic groups but with overlapping confidence intervals (Table 4-4).

Table 4-4 *Performance measures per age, sex and ethnic subgroups for a) symptom-only and b) symptom + blood test solid cancer model. AUC= area under the curve, 95% CI = 95% confidence interval*

	AUC (95% CI)	**Calibration Slope (95% CI)**	**Calibration-in-the-large (95% CI)**
Symptom only model			
Age			
Children	0.83 (0.80 to 0.87)	1.02 (0.94 to 1.09)	0.2 (0.04 to 0.35)
TYA	0.75 (0.70 to 0.80)	0.91 (0.80 to 1.01)	-0.21 (-0.38 to -0.03)
Sex			
Male	0.81 (0.76 to 0.86)	0.98 (0.89 to 1.06)	0.004 (-0.17 to 0.18)
Female	0.77 (0.67 to 0.88)	0.92 (0.77 to 1.07)	-0.4 (-0.69 to -0.12)
Ethnic Group			
White	0.77 (0.71 to 0.81)	0.90 (0.79 to 1.02)	-0.26 (-0.45 to -0.06)
Other	0.72 (0.61 to 0.82)	0.94 (0.83 to 1.05)	-0.14 (-0.33 to 0.05)
Symptom + test model			
Age			
Children	0.83 (0.80 to 0.87)	0.99 (0.92 to 1.07)	0.19 (0.03 to 0.35)
TYA	0.76 (0.71 to 0.81)	0.90 (0.81 to 0.99)	-0.22 (-0.40 to -0.05)
Sex			
Male	0.81 (0.77 to 0.86)	0.96 (0.88 to 1.04)	-0.013 (-0.19 to 0.16)
Female	0.75 (0.70 to 0.80)	0.97 (0.89 to 1.04)	-0.009 (-0.16 to 0.14)
Ethnic Group			
White	0.81 (0.74 to 0.86)	0.95 (0.83 to 1.06)	-0.05 (-0.29 to 0.19)
Other	0.74 (0.65 to 0.84)	0.91 (0.79 to 1.03)	0.16 (-0.46 to 0.13)

4.4.2.5 Measures of Accuracy

Table 4-5 shows the measures of accuracy of the models, following validation, based on the top 0.1%, 0.5%, 1%, 2%, 5%, 10%, 20%, 40%, 60%, 80%, 90% and 95% at highest risk. Specifically, when the 90th centile threshold of risk (cut-off 0.03%) was used to define young

people at “high risk” of solid cancer, both models showed 61% sensitivity and 90% specificity, with a positive predictive value of 0.03% and negative predictive value of 99.9%.

Table 4-5 *Measures of accuracy of each model, following validation, based on the top 0.1%, 0.5%, 1%, 2%, 5%, 10%, 20%, 40%, 60%, 80%, 90% and 95% at highest risk in symptom-only and symptom + blood test solid cancer models*

	Risk Threshold (%)	True positive	True negative	False positive	False negative	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
0.1%									
Symptom only model	0.36	45	4,851,437	9,677	264	14.6	99.8	0.46	99.9
Symptom + test model	0.45	54	4,851,468	9,646	255	17.4	99.8	0.55	99.9
0.5%									
Symptom only model	0.20	67	4,836,874	24,240	242	21.7	99.5	0.28	99.9
Symptom + test model	0.24	71	4,836,900	24,214	238	23.0	99.5	0.29	99.9
1%									
Symptom only model	0.12	82	4,812,582	48,532	227	26.5	99.0	0.17	99.9
Symptom + test model	0.28	93	4,812,880	48,520	216	30.1	99.0	0.20	99.9
2%									
Symptom only model	0.10	96	4,763,982	97,132	213	31.1	98.0	0.10	99.9
Symptom + test model	0.18	106	4,764,280	97,120	203	34.4	98.0	0.11	99.9
5%									
Symptom only model	0.049	140	4,618,183	242,931	169	45.1	95.0	0.060	99.9
Symptom + test model	0.061	143	4,618,478	242,922	166	43.1	95.0	0.060	99.9
10%									
Symptom only model	0.032	190	4,374,802	485,952	119	61.5	90.0	0.039	99.9
Symptom + test model	0.037	187	4,375,457	485,943	122	61.0	90.0	0.038	99.9
20%									
Symptom only model	0.018	216	3,889,046	972,068	93	70.0	80.0	0.022	99.9
Symptom + test model	0.021	219	3,889,359	972,041	90	70.9	80.0	0.022	99.9
40%									
Symptom only model	0.010	239	2,916,784	1,944,330	70	77.3	60.0	0.012	99.9
Symptom + test model	0.011	240	2,916,807	1,944,307	69	77.7	60.0	0.012	99.9

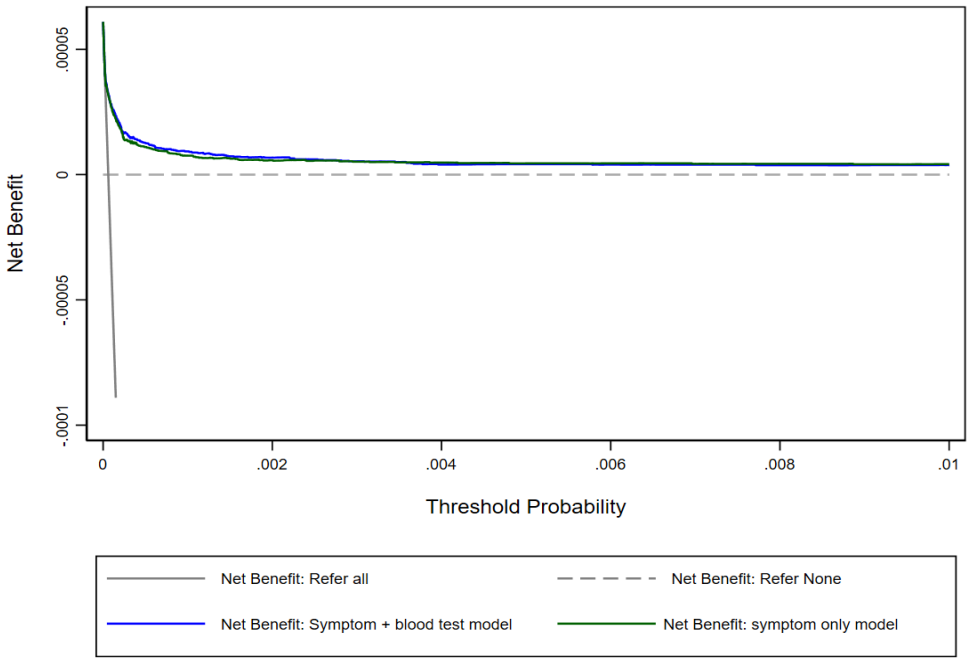
60%									
Symptom only model	0.0082	272	1,944,533	2,916,581	37	88.0	40.0	0.0093	99.9
Symptom + test model	0.0084	275	1,944,558	2,916,556	34	89.0	40.0	0.0094	99.9
80%									
Symptom only model	0.0068	289	972,265	3,888,849	20	93.5	20.0	0.0074	99.9
Symptom + test model	0.0069	289	972,287	3,888,827	20	93.5	20.0	0.0074	99.9
90%									
Symptom only model	0.0063	297	486,132	4,374,982	12	96.1	10.0	0.0068	99.9
Symptom + test model	0.0064	298	486,154	4,374,960	11	96.4	10.0	0.0068	99.9
95%									
Symptom only model	0.0060	305	243,069	4,618,045	4	98.7	5.0	0.0066	99.9
Symptom + test model	0.0061	306	243,091	4,618,023	3	99.0	5.0	0.0066	99.9

4.4.2.6 Decision Curve Analysis

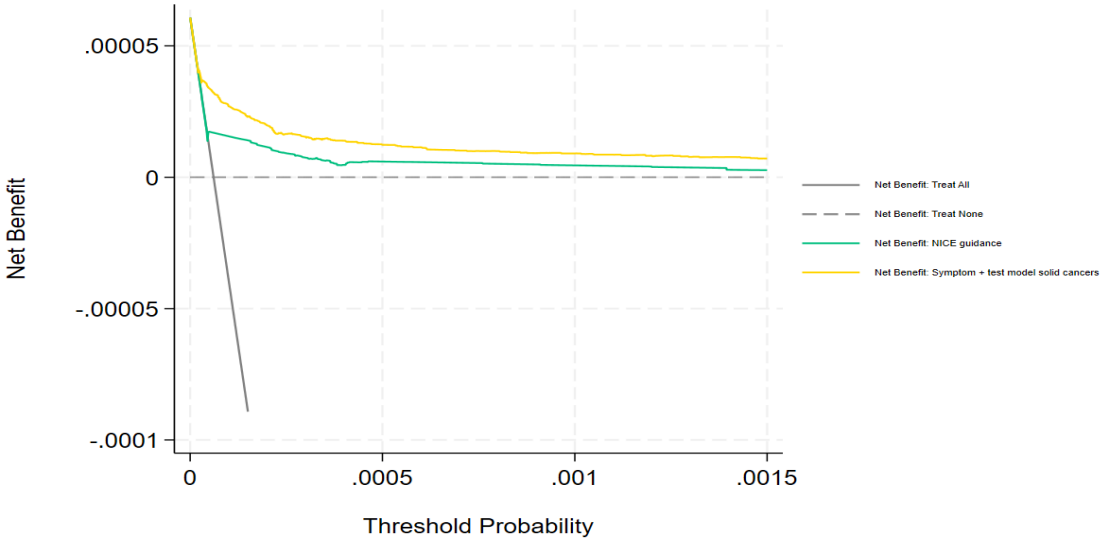
Decision curve analysis demonstrated that the models (symptom only and symptom + blood test) had similar net benefit across all threshold probabilities compared to refer-none or refer-all strategies (Figure 4-5). The clinical utility of the developed symptom + test model was higher than that of symptoms/signs used for suspected cancer referral in NICE guidelines [78].

Figure 4-5 *a) Decision curve to assess net benefit of symptom only and symptom + blood test models across a range of threshold probabilities, b) Decision curve to assess net benefit of symptom+ test model compared to NICE guidance across a range of threshold probabilities*

a)



b)



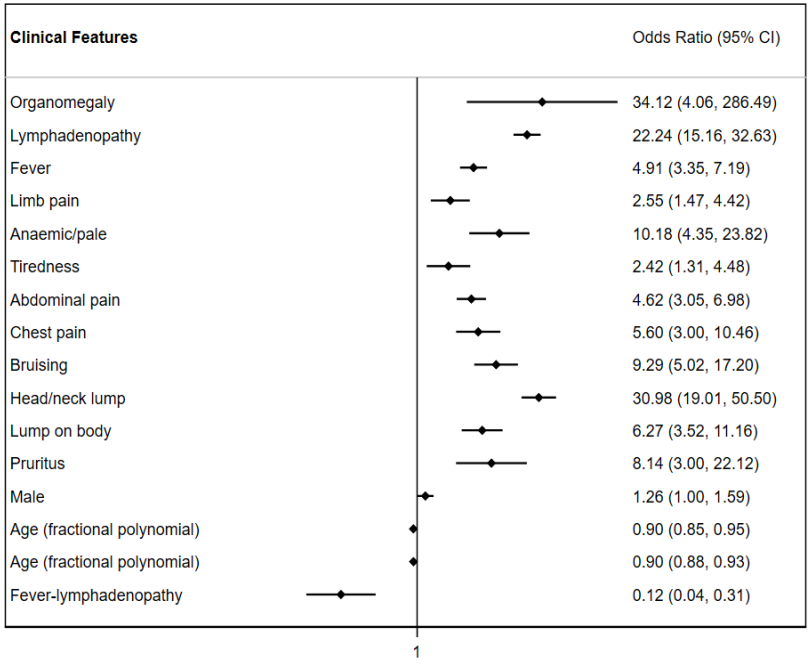
4.4.3 Models for Combined Haematological Cancers

4.4.3.1 Model Derivation

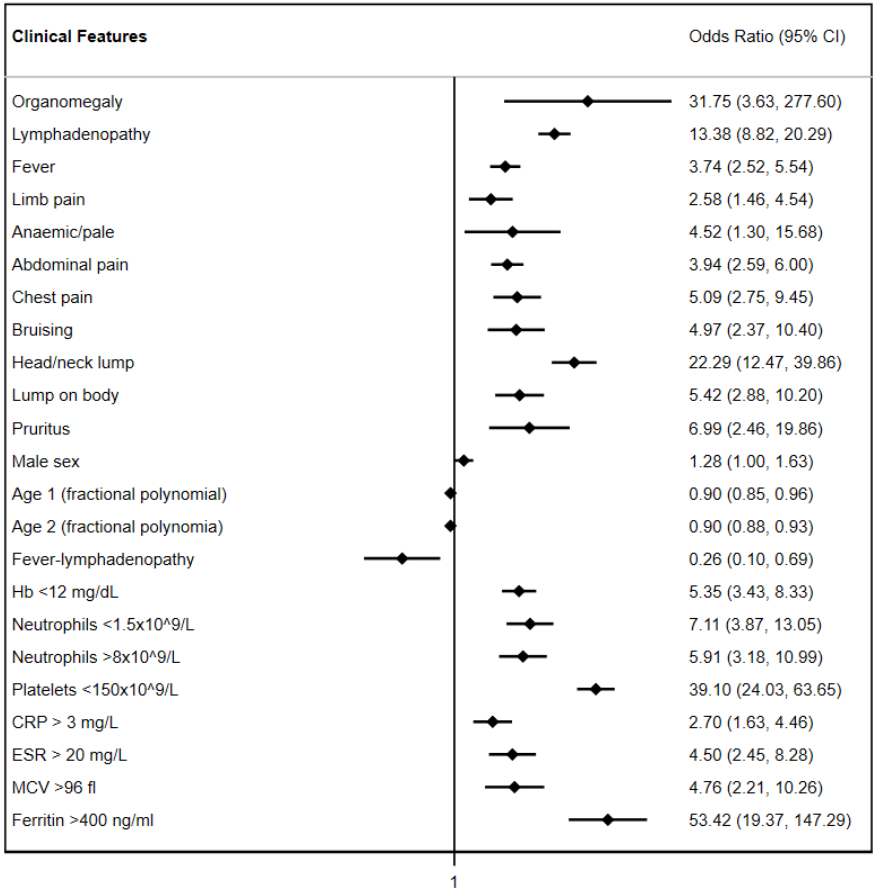
The predictor variables in symptom only and symptom + blood test model included sociodemographic factors sex and non-linear fractional polynomial terms for age (-2, -2) (Appendix Chapter 9, Supplementary Figure 9-4). There were 11 symptoms/signs (organomegaly, lymphadenopathy, fever, limb pain, anaemia/pallor, abdominal pain, chest pain, bruising, head/neck lump, lump on body, pruritus) in both models (Figure 4-6); and a further 8 abnormal blood test results (low haemoglobin (<12 mg/dl), low neutrophils (<1.5 x 10^9/L), low platelets (<150 x 10^9/L), high neutrophils (>8 x 10^9/L), high ESR (>20 mm/hr), high MCV (>96 fl), high ferritin (>400 ng/ml) and high CRP (>3 mg/dl)) in the exploratory symptom + blood test model (Figure 4-6).

Figure 4-6 *Final logistic regression model predicting odds of haematological cancer 6 months following study entry/symptom onset in a) symptom-only and b) symptom + blood test model. Models contain fractional polynomials for age (-2,-2). Adjusted OR= Adjusted Odds ratio, CI= Confidence interval, Hb= haemoglobin, MCV= mean cell volume, ESR= Erythrocyte sedimentation rate, CRP=C-reactive protein.*

a)



b)

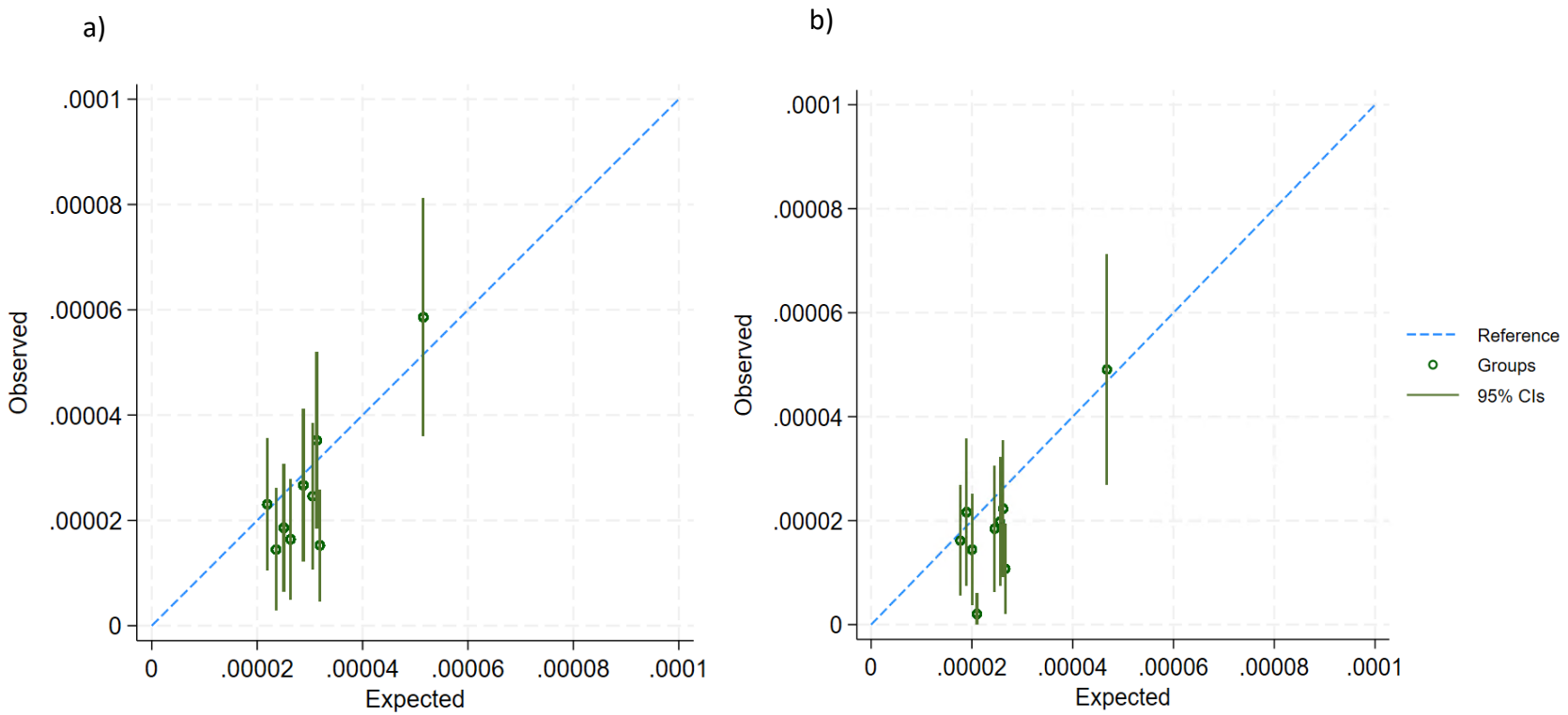


4.4.3.2 Model performance

The symptom only model had good discrimination and was well-calibrated following internal-external validation (AUC=0.83 [95%CI 0.80 to 0.87], slope=0.98 [95%CI 0.93 to 1.03], calibration-in-the-large=0.007 [95%CI -0.11 to 0.12]). The symptom + blood test model was found to have higher discrimination (AUC=0.88 [95%CI 0.84 to 0.92]) and was also well-calibrated (slope=0.98 [95%CI 0.91 to 1.04], calibration-in-the-large=-0.02 [95%CI -0.09 to 0.04]).

Calibration plots for both haematological cancer models (symptom only and symptom + test) demonstrated that models tended towards overestimation at the lower predicted risk ranges (Figure 4-7).

Figure 4-7 *Calibration of the a) symptom-only and b) symptom + blood test haematological cancer models tested demonstrated through calibration plots which show the alignment between predicted and observed risks across 10 risk deciles/groups.*

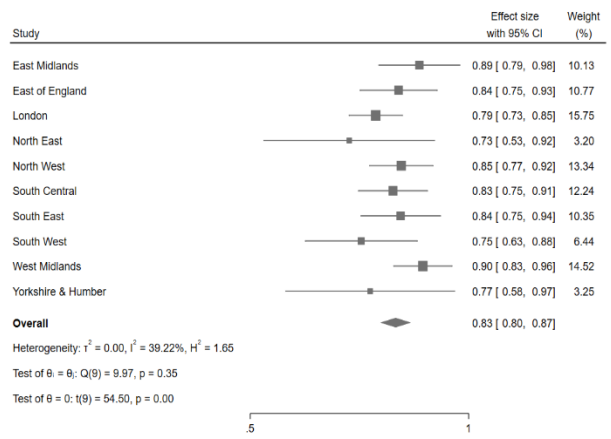


4.4.3.3 Heterogeneity

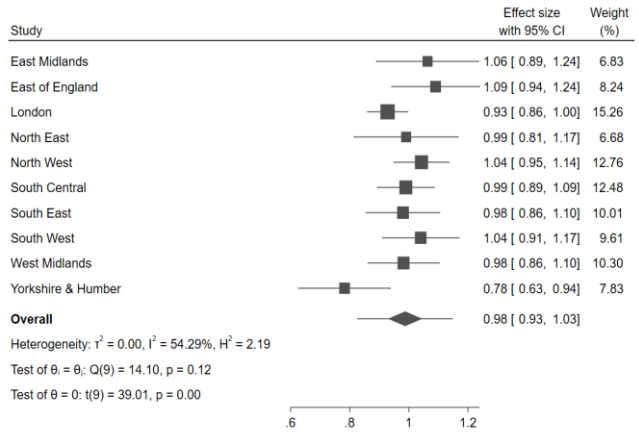
For the symptom-only model, meta-analysis of performance measures across the 10 geographical regions demonstrated that there was moderate level of heterogeneity for AUC (I^2=39.2%), and high for calibration slope (I^2=54.3%) and calibration-in-the-large (I^2=54.3%) (Figure 4-8). The heterogeneity for all performance measures were comparable in the symptom + blood test model (I^2=61.8%, I^2=38.7%, I^2=29.0% for AUC, slope and calibration-in-the-large, respectively) (Figure 4-8). Age, ethnicity and deprivation level did not account for the observed heterogeneity in meta-regression analyses (Appendix Chapter 9, Supplementary Table 9-14).

Figure 4-8 *Region-level and pooled performance estimates following internal-external cross validation for a) & d) area-under-the-curve (AUC), b) & e) calibration slope and, c) & f) calibration-in-the-large for symptom-only (a-c) and symptom + blood test haematological cancer models (d-f). Effect Size=adjusted Odds ratio, CI= Confidence interval.*

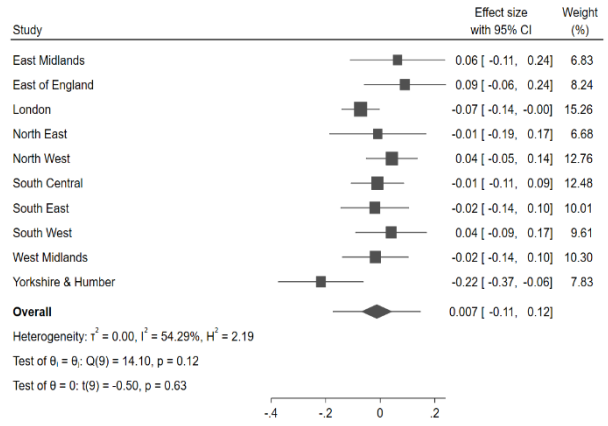
a)



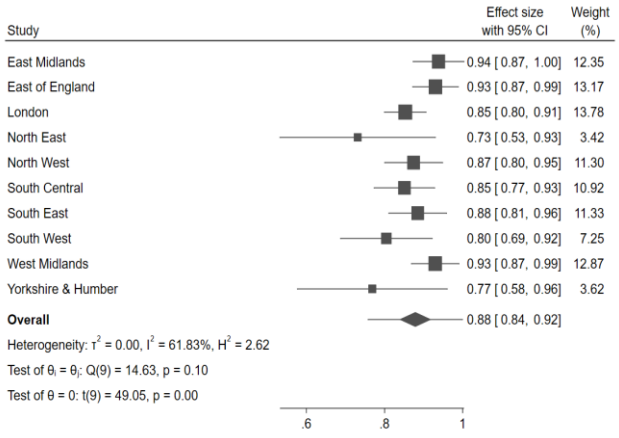
b)



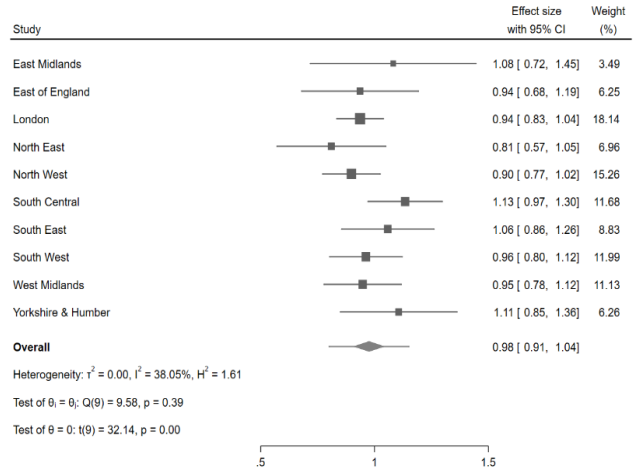
c)



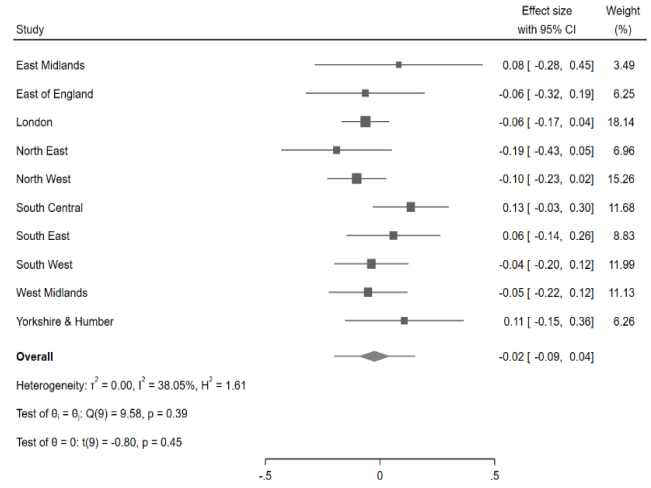
d)



e)



f)



4.4.3.4 Model fairness

The model performance was comparable across different age groups (child 0-14 years & TYA 15-24 years), sexes (male & female) and ethnic groups (white & non-white) for both symptom-only and symptom + blood test models (Table 4-6).

Table 4-6 *Performance measures per age, sex and ethnic subgroups for a) symptom-only and b) symptom + blood test haematological cancer models. AUC= area under the curve, 95% CI = 95% confidence interval*

	AUC (95% CI)	Calibration Slope (95% CI)	Calibration-in-the-large (95% CI)
Symptom only			
Age			
Children	0.83 (0.80 to 0.87)	0.93 (0.86 to 0.99)	-0.014 (-0.18 to 0.16)
TYA	0.80 (0.75 to 0.84)	1.04 (0.97 to 1.11)	-0.0016 (-0.16 to 0.15)
Sex			
Male	0.78 (0.73 to 0.82)	0.97 (0.91 to 1.04)	-0.016 (-0.17 to 0.14)
Female	0.81 (0.76 to 0.86)	0.99 (0.91 to 1.07)	0.0037 (-0.17 to 0.18)
Ethnic Group			
White	0.81 (0.77 to 0.84)	0.98 (0.93 to 1.03)	-0.006 (-0.13 to +0.12)
Other	0.82 (0.78 to 0.85)	0.99 (0.94 to 1.05)	-0.024 (-0.16 to 0.11)
Symptom + test			
Age			
Children	0.89 (0.85 to 0.91)	0.97 (0.92 to 1.02)	-0.0058 (-0.18 to 0.17)
TYA	0.85 (0.79 to 0.88)	0.99 (0.94 to 1.043)	-0.038 (-0.2 to 0.12)
Sex			
Male	0.85 (0.80 to 0.89)	0.97 (0.92-1.03)	-0.15 (-0.19 to 0.17)
Female	0.86 (0.81 to 0.89)	0.98 (0.94-1.03)	-0.03 (-0.19 to 0.13)
Ethnic Group			
White	0.83 (0.77-0.89)	0.96 (0.89-1.03)	0.078 (-0.15 to 0.31)
Other	0.93 (0.89-0.97)	0.95 (0.86 to 1.02)	-0.29 (-0.58 to 0.017)

4.4.3.5 Measures of Accuracy

The model's measures of accuracy based on the top 1%, 2%, 5%, 10%, and 20% at highest risk was assessed following validation. Specifically, when 90th centile threshold of risk (cut-off 0.04%) was used to define young people at high-risk of haematological cancer, the symptom + blood test model showed 75% sensitivity and 90% specificity, with a positive predictive value

of 0.05% and negative predictive value of 99.9%, whist for the symptom only model sensitivity was lower at 63% (Table 4-7).

Table 4-7 *Measures of accuracy of each model, following validation, based on the top 0.1%, 0.5%, 1%, 2%, 5%, 10%, 20%, 40%, 60%, 80%, 90%, 95% at highest risk in symptom-only and symptom + blood test haematological cancer models*

	Risk Threshold (%)	True positive	True negative	False positive	False negative	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
0.1%									
Symptom only model	0.33	23	4,851,902	9,213	285	7.4	99.8	0.25	99.0
Symptom + test model	0.47	51	4,851,930	9,185	257	16.5	99.8	0.55	99.0
0.5%									
Symptom only model	0.26	67	4,836,875	24,240	241	21.7	99.7	0.28	99.0
Symptom + test model	0.37	130	4,836,938	24,177	178	42.2	99.7	0.53	99.0
1%									
Symptom only	0.16	98	4,812,599	48,516	210	31.8	99.7	0.20	99.0
Symptom + test	0.30	149	4,812,650	48,465	159	48.4	99.7	0.30	99.0
2%									
Symptom only	0.10	135	4,764,022	97,093	173	43.8	98.0	0.14	99.9
Symptom + test	0.20	178	4,764,065	97,050	130	57.8	98.0	0.18	99.9
5%									
Symptom only	0.053	160	4,618,204	242,911	148	51.9	95.0	0.065	99.9
Symptom + test	0.074	198	4,618,242	242,873	110	64.3	95.0	0.090	99.9
10%									
Symptom only	0.033	195	4,375,168	485,947	113	63.3	90.0	0.040	99.9
Symptom + test	0.042	226	4,375,199	485,916	82	73.4	90.0	0.045	99.9
20%									
Symptom only	0.019	221	3,889,052	972,063	87	71.8	80.0	0.023	99.9

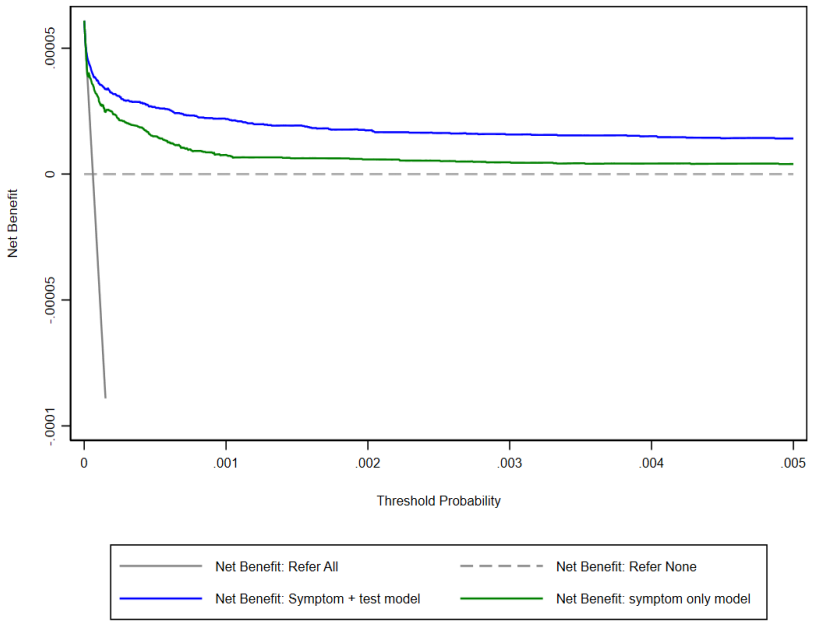
Symptom + test	0.023	245	3,889,076	972,039	63	80.0	80.0	0.025	99.9
40%									
Symptom only model	0.011	247	2,916,793	1,944,322	61	80.2	60.0	0.012	99.0
Symptom + test model	0.011	262	2,916,809	1,944,306	46	85.0	60.0	0.013	99.0
60%									
Symptom only model	0.0084	272	1,944,534	2,916,581	36	88.3	40.0	0.0093	99.0
Symptom + test model	0.0085	281	1,944,544	2,916,571	27	91.2	40.0	0.0096	99.0
80%									
Symptom only model	0.0069	289	972,266	3,888,849	19	93.8	20.0	0.0074	99.0
Symptom + test model	0.0069	289	971,168	3,889,947	19	93.8	20.0	0.0074	99.0
90%									
Symptom only model	0.0064	296	486,131	4,374,984	12	96.1	10.0	0.0068	99.0
Symptom + test model	0.0063	299	486,136	4,374,979	9	97.0	10.0	0.0068	99.0
95%									
Symptom only model	0.0061	304	243,068	4,618,047	4	98.7	5.0	0.0065	99.0
Symptom + test model	0.0061	306	243,071	4,618,043	2	99.3	5.0	0.0065	99.0

4.4.3.6 Decision Curve Analysis

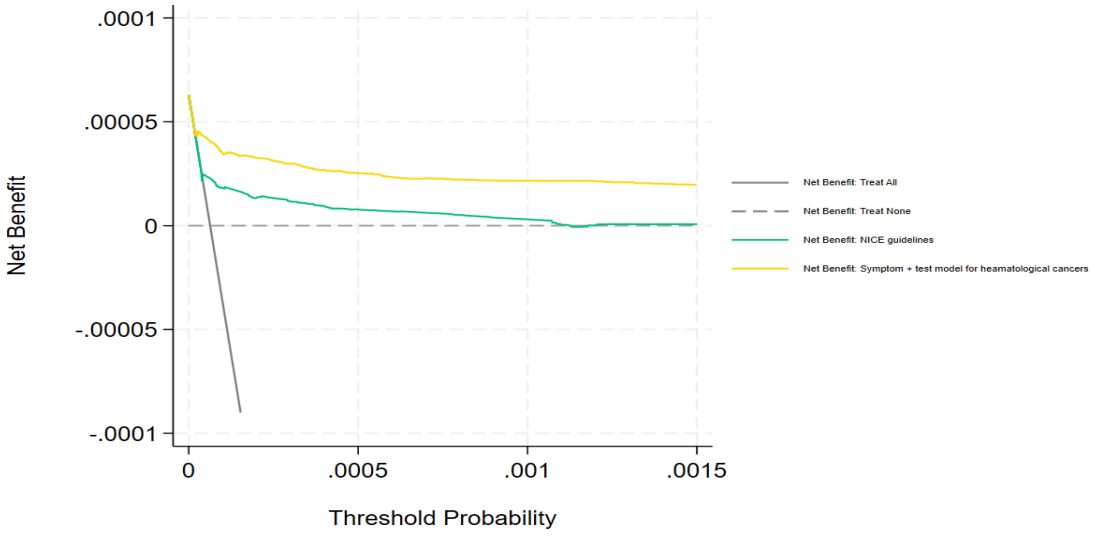
Decision curve analysis demonstrated that the models (symptom only and symptom + test) had higher net benefit across all risk threshold probabilities compared to refer-none or refer-all strategies (Figure 4-9), with the symptom + blood test model showing highest net benefit. The clinical utility of the developed symptom + test model was higher than that of symptoms/signs used for suspected cancer referral in NICE guidelines [78].

Figure 4-9 *a) Decision curve to assess net benefit of symptom only and symptom + blood test models across a range of threshold probabilities, b) Decision curve to assess net benefit of symptom+ test model compared to NICE guidance across a range of threshold probabilities*

a)



b)



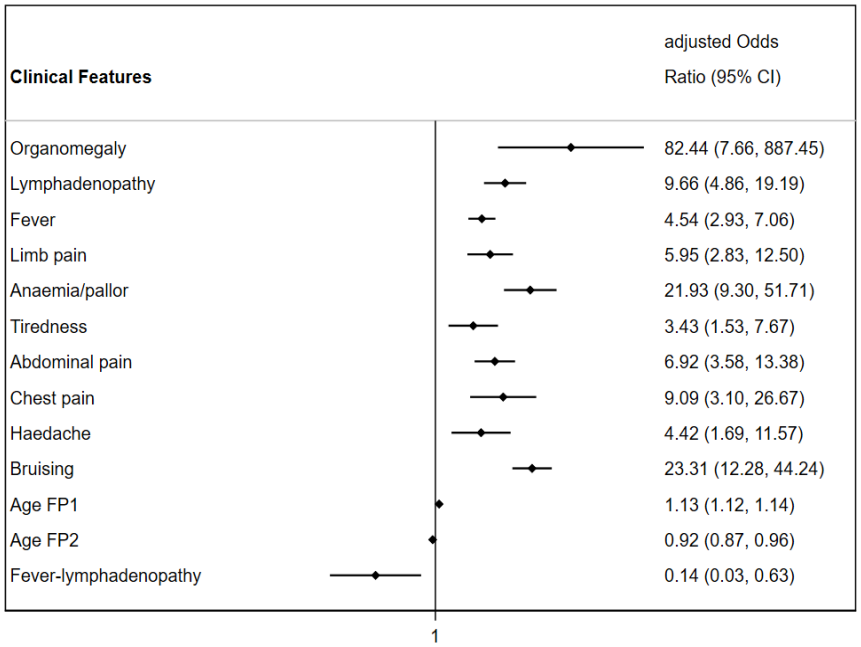
4.4.4 Models for Leukaemia

4.4.4.1 Model Derivation

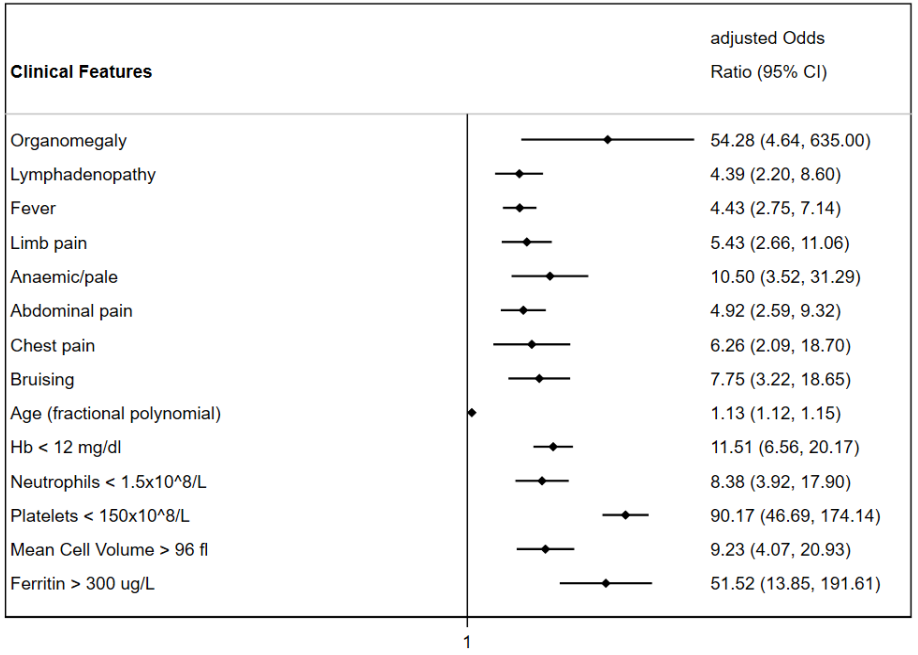
The predictor variables in symptom only and symptom + test model included non-linear fractional polynomial terms for age (-2, -2) (Appendix Chapter 9, Supplementary Figure 9-4). There were 11 symptoms/signs (organomegaly, lymphadenopathy, fever, limb pain, anaemia/pallor, abdominal pain, tiredness, bruising, chest pain, headache) in both models (Figure 4-10); and a further 5 abnormal blood test results (low haemoglobin (<12 mg/dl), low neutrophils (<1.5 x 10^9/L), low platelets (<150 x 10^9/L), high ESR (>20 mm/hr) and high ferritin (>400 ng/ml)) in the exploratory symptom + blood test model (Figure 4-10).

Figure 4-10 *Final logistic regression model predicting odds of leukaemia 6 months following study entry/symptom onset in a) symptom-only and b) symptom + blood test model. Models contain fractional polynomials for age (-2,-2). Adjusted OR= Adjusted Odds ratio, CI= Confidence interval, Hb= haemoglobin, FP= fractional polynomial.*

a)



b)

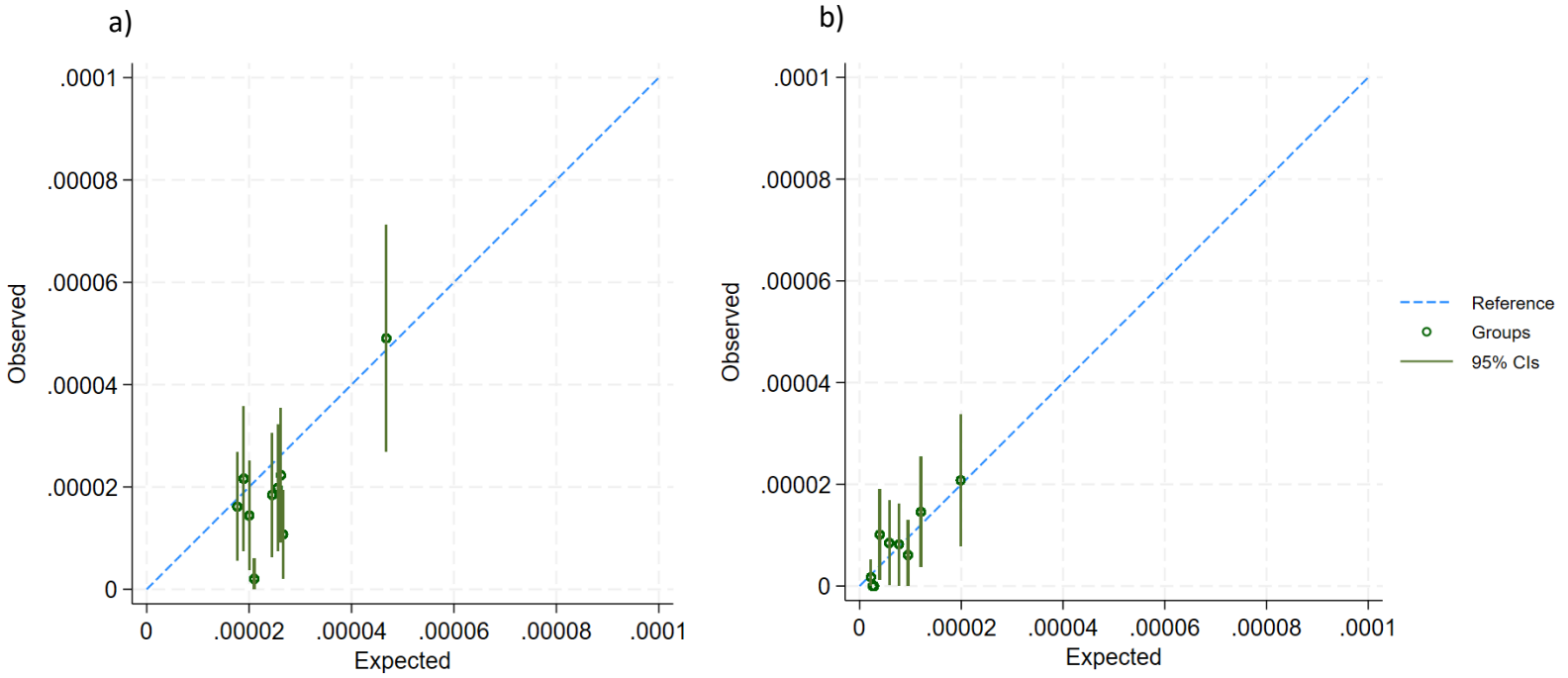


4.4.4.2 Model performance

Following internal-external cross validation, the symptom-only model was found to discriminate well (AUC=0.86 [95%CI 0.83 to 0.89]), and was well-calibrated (slope=0.98 [95%CI 0.92 to 1.04], calibration-in-the-large=-0.01 [95%CI -0.17 to 0.16]). The symptom + blood test model showed similar performance, with higher discrimination (AUC=0.90 [95%CI 0.87 to 0.93], slope=0.97 [95%CI 0.93 to 1.02], calibration-in-the-large=-0.05 [95%CI -0.22 to 0.13]).

Calibration plots demonstrated that the symptom-only models tended towards overestimation at the lower predicted risk ranges whilst the symptom + blood test model appeared to be well-calibrated (Figure 4-11).

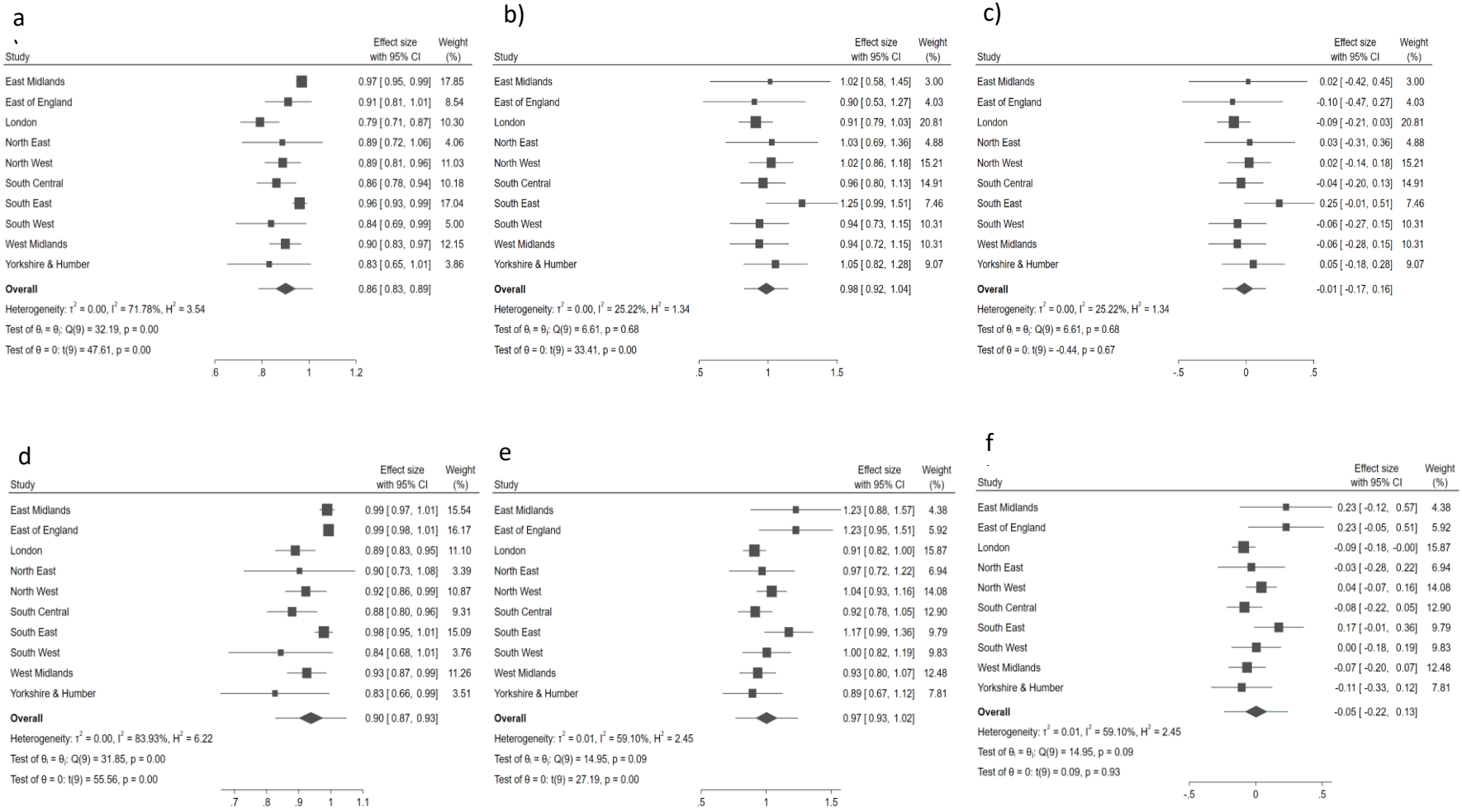
Figure 4-11 *Calibration of the a) symptom-only and b) symptom + blood test leukaemia models, demonstrated through calibration plots which show the alignment between predicted and observed risks across 10 risk deciles/groups.*



4.4.4.3 Heterogeneity

For the symptom only model, heterogeneity for AUC was high (I^2=71.8%), but heterogeneity was moderate for calibration slope and calibration-in-the large (I^2=25.2%) (Figure 4-12). Age, ethnicity and deprivation level did account for 39% of the observed heterogeneity in the symptom only model (Appendix Chapter 9, Supplementary Table 9-14). For the symptom + blood test model, meta-analysis of performance measures across the 10 geographical regions demonstrated that there was high level of heterogeneity for AUC (I^2=83.9%), calibration slope (I^2=59.1%) and calibration-in-the large (I^2=59.1%) (Figure 4-12). Age, ethnicity and deprivation level did account for approximately 13% of observed heterogeneity for the AUC in meta-regression analyses (Appendix Chapter 9, Supplementary Table 9-14).

Figure 4-12 *Region-level and pooled performance estimates following internal-external cross validation for a) & d) area-under-the-curve (AUC), b) & e) calibration slope and, c) & f) calibration-in-the-large for symptom-only (a-c) and symptom + blood test leukaemia models (d-f). Effect Size=adjusted Odds ratio, CI= Confidence interval.*



4.4.4.4 Model Fairness

The model performance was comparable across different age groups (child 0-14 years & TYA 15-24 years), sexes (male & female) and ethnic groups (white & non-white) for both symptom-only and symptom + test models (Table 4-8).

Table 4-8 *Performance measures per age, sex and ethnic subgroups for i) symptom-only and ii) symptom + blood test leukaemia models. AUC= area under the curve, 95% CI = 95% confidence interval*

	AUC (95% CI)	Calibration Slope (95% CI)	Calibration-in-the-large (95% CI)
Symptom only			
Age			
Children	0.85 (0.81 to 0.89)	1.02 (0.94 to 1.09)	-0.069 (-0.26 to 0.12)
TYA	0.80 (0.72 to 0.88)	0.87 (0.70 to 1.07)	0.17 (-0.15 to 0.50)
Sex			
Male	0.85 (0.80 to 0.89)	0.97 (0.89 to 1.05)	0.12 (-0.11 to 0.33)
Female	0.88 (0.82 to 0.93)	0.99 (0.89 to 1.10)	-0.16 (-0.42 to 0.10)
Ethnic Group			
White	0.86 (0.82 to 0.89)	0.97 (0.90 to 1.04)	-0.048 (-0.24 to 0.14)
Other	0.87 (0.83 to 0.90)	0.99 (0.92 to 1.06)	-0.012 (-0.21 to 0.18)
Symptom + test			
Age			
Children	0.89 (0.86 to 0.93)	0.97 (0.92 to 1.03)	-0.12 (-0.33 to 0.09)
TYA	0.88 (0.92 to 0.94)	0.98 (0.90 to 1.07)	0.16 (-0.17 to 0.49)
Sex			
Male	0.90 (0.86 to 0.94)	0.95 (0.89 to 1.01)	0.03 (-0.22 to 0.26)
Female	0.91 (0.86 to 0.96)	0.99 (0.92 to 1.07)	-0.14 (-0.41 to 0.13)
Ethnic Group			
White	0.89 (0.86 to 0.92)	0.99 (0.94 to 1.04)	0.015 (-0.18 to 0.21)
Other	0.91 (0.88 to 0.95)	0.99 (0.94 to 1.04)	-0.08 (-0.29 to 0.12)

4.4.4.5 Measures of Accuracy

The model's measures of accuracy based on the top 1%, 2%, 5%, 10%, and 20% at highest risk was assessed following validation. Specifically, when 90th centile threshold of risk (cut-off 0.02%) was used to define young people at high-risk of leukaemia, the symptom + test model showed 77% sensitivity and 90% specificity, with a positive predictive value of 0.02% and

negative predictive value of 99.9%, whist for the symptom only model sensitivity was lower at 62% (Table 4-9).

Table 4-9 *Measures of accuracy of each model, following validation, based on the top 0.1%, 0.5%, 1%, 2%, 5%, 10%, 20%, 40%, 60%, 80%, 90%, 95% at highest risk in symptom-only and symptom + blood test leukaemia models*

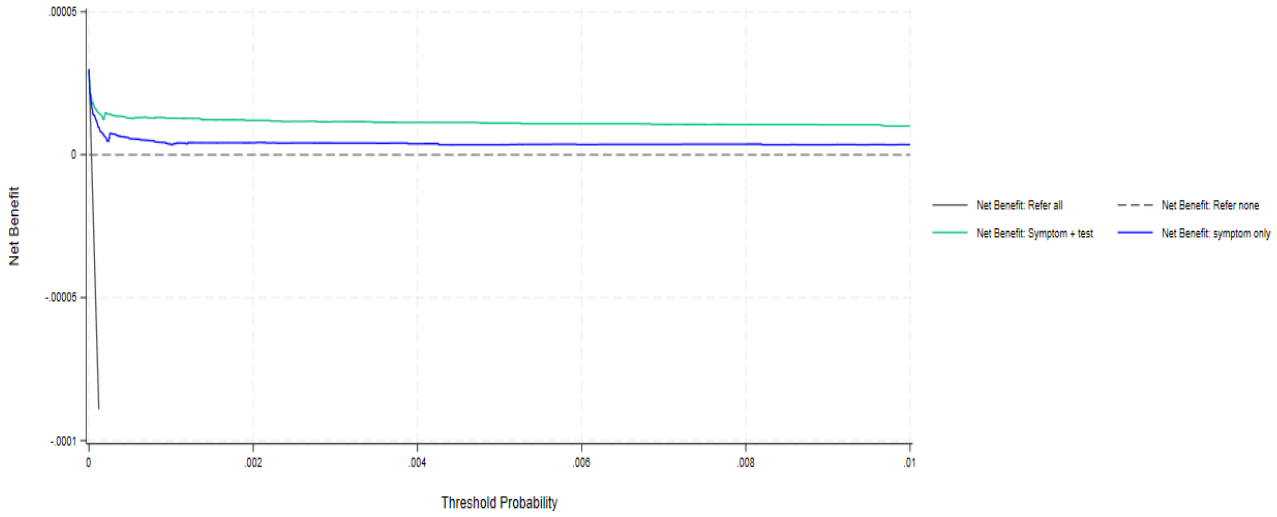
	Risk Threshold (%)	True positive	True negative	False positive	False negative	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
0.1%									
Symptom only	0.25	30	4,851,747	9,530	116	20.5	99.8	0.31	99.0
Symptom + test	0.53	70	4,851,787	9,490	76	47.9	99.8	0.73	99.0
0.5%									
Symptom only	0.14	40	4,837,172	24,105	106	27.3	99.5	0.17	99.0
Symptom + test	0.24	75	4,837,207	24,070	71	51.3	99.5	0.31	99.0
1%									
Symptom only	0.084	47	4,812,871	48,406	99	32.2	99.0	0.09	99.0
Symptom + test	0.13	77	4,812,901	48,376	69	52.7	99.0	0.16	99.0
2%									
Symptom only	0.054	48	4,764,258	97,019	98	32.9	98.0	0.05	99.9
Symptom + test	0.073	81	4,764,291	96,986	65	55.5	98.0	0.083	99.9
5%									
Symptom only	0.031	79	4,618,446	242,831	67	54.1	95.0	0.032	99.9
Symptom + test	0.036	100	4,618,467	242,810	46	68.5	95.0	0.041	99.9
10%									
Symptom only	0.018	91	4,375,387	485,890	55	62.3	90.0	0.019	99.9
Symptom + test	0.020	112	4,375,408	485,869	34	76.7	90.0	0.022	99.9
20%									
Symptom only	0.011	117	3,889,271	972,006	29	80.1	80.0	0.012	99.9
Symptom + test	0.011	122	3,889,276	972,001	24	83.6	80.0	0.013	99.9
40%									
Symptom only	0.0061	126	2,916,996	1,944,281	20	86.3	60.0	0.0065	99.9
Symptom + test	0.0060	132	2,917,002	1,944,275	14	90.4	60.0	0.0068	99.9
60%									
Symptom only	0.0043	137	1,944,723	2,916,554	9	93.8	40.0	0.0047	99.9

Symptom + test	0.0042	140	1,944,726	2,916,551	6	95.8	40.0	0.0048	99.9
80%									
Symptom only	0.0034	144	972,445	3,888,832	2	98.6	20.0	0.0037	99.9
Symptom + test	0.0032	145	972,446	3,888,831	1	99.3	20.0	0.0037	99.9
90%									
Symptom only	0.0030	145	486,304	4,374,973	1	99.3	10.0	0.0033	99.9
Symptom + test	0.0029	146	486,305	4,374,972	0	100.0	10.0	0.0033	99.9
95%									
Symptom only	0.0029	146	243,234	4,618,043	0	100.0	5.0	0.0031	99.9
Symptom + test	0.0028	146	243,234	4,618,043	0	100.0	5.0	0.0031	99.9

4.4.4.6 Decision Curve Analysis

Decision curve analysis demonstrated that the models (symptom only and symptom + test) had higher net benefit across all risk threshold probabilities compared to refer-none or refer-all strategies (Figure 4-13), with the symptom + test model once again showing highest net benefit.

Figure 4-13 *Decision curve to assess net benefit of symptom-only and symptom + blood test models across a range of threshold probabilities*



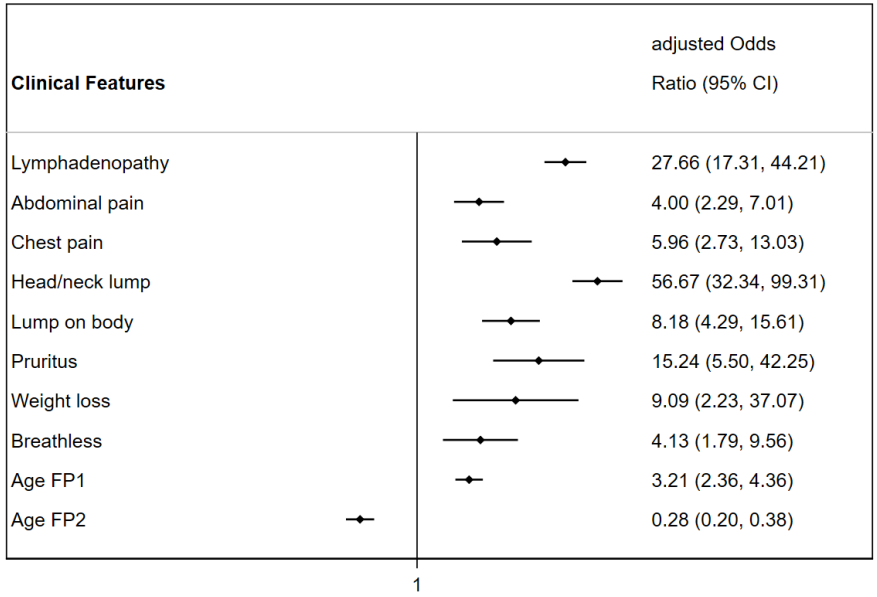
4.4.5 Models for Lymphoma

4.4.5.1 Model Derivation

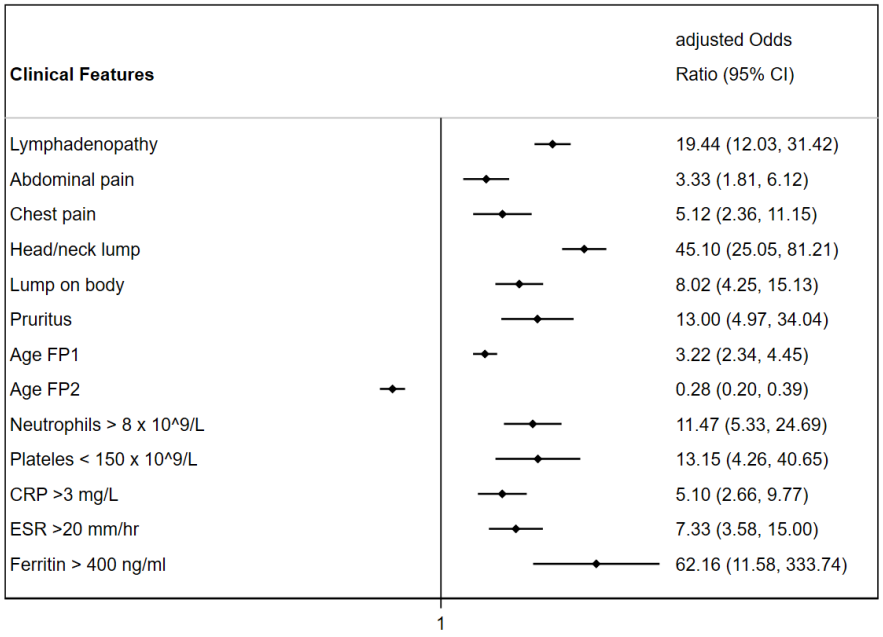
The final model included non-linear fractional polynomial terms for age (-2, -2) (Chapter Appendix, Supplementary Figure 9-4), as well as 8 symptoms/signs (lymphadenopathy, abdominal pain, chest pain, head/neck lump, lump on body, pruritus, weight loss and breathlessness) in the symptom only model (Figure 4-14). The symptom + blood test model included a further 5 abnormal blood test results (high neutrophils (>8 x 10^9/L), low platelets (<150 x 10^9/L), high CRP (>3 mg/L) high ESR (>20 mm/hr) and high ferritin (>400 ng/ml)).

Figure 4-14 *Final logistic regression model predicting odds of lymphoma 6 months following study entry/symptom onset in a) symptom-only and b) symptom + blood test model. Models contain fractional polynomials (FP) for age (-2,-2), CI= Confidence interval, Hb= haemoglobin, CRP = C-reactive protein, ESR = Erythrocyte Sedimentation Rate.*

a)



b)

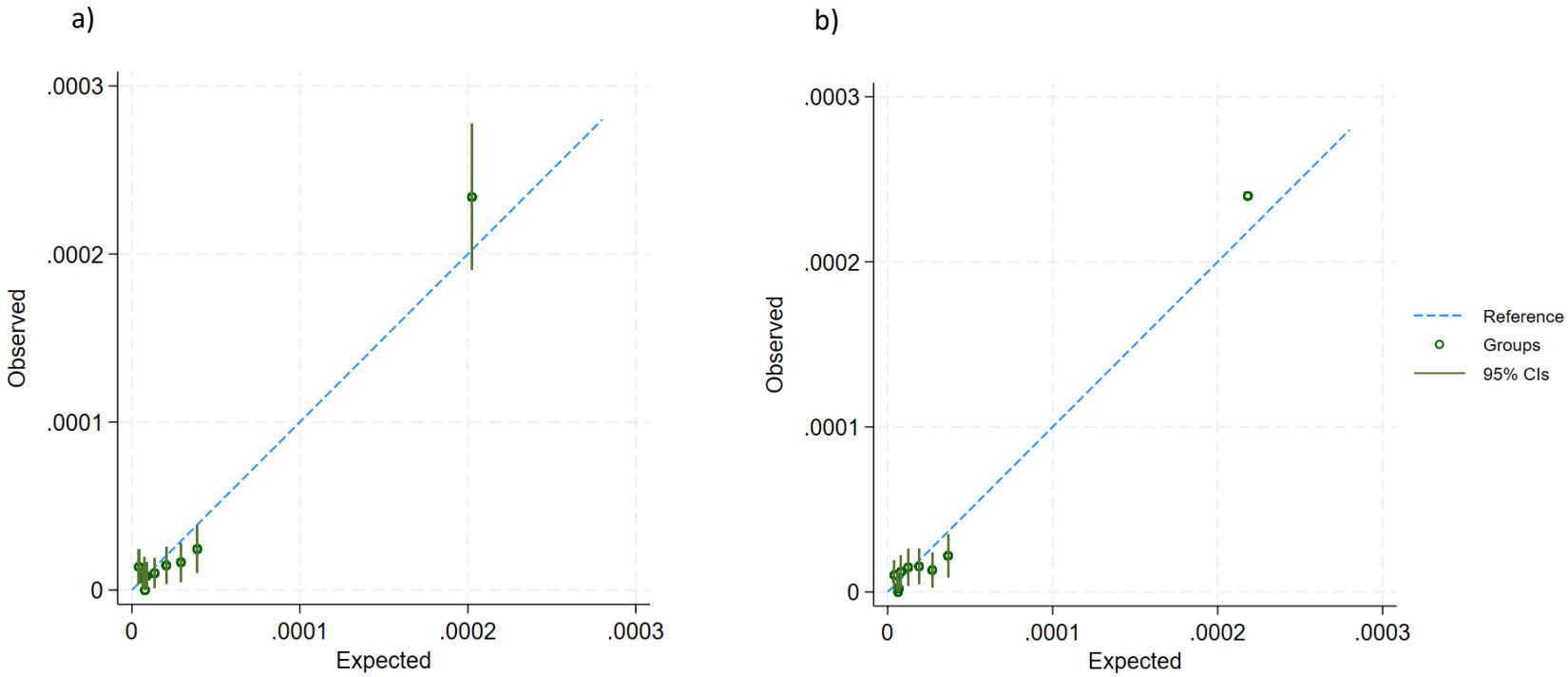


4.4.5.2 Model performance

The symptom-only model showed good discrimination and was well-calibrated, following internal-external validation ((AUC=0.86 [95%CI 0.82 to 0.89], slope=0.99 [95%CI 0.93 to 1.04], calibration-in-the-large=-0.01 [95%CI -0.16 to 0.15]). The symptom + blood test model was also found to discriminate well (AUC=0.86 [95%CI 0.82 to 0.90]) and was well-calibrated (slope=0.98 [95%CI 0.94 to 1.03], calibration-in-the-large=-0.01 [95%CI -0.17 to 0.15]).

Calibration plots demonstrated that models overestimated at lower risk thresholds and underestimated in higher risk thresholds (Figure 4-15).

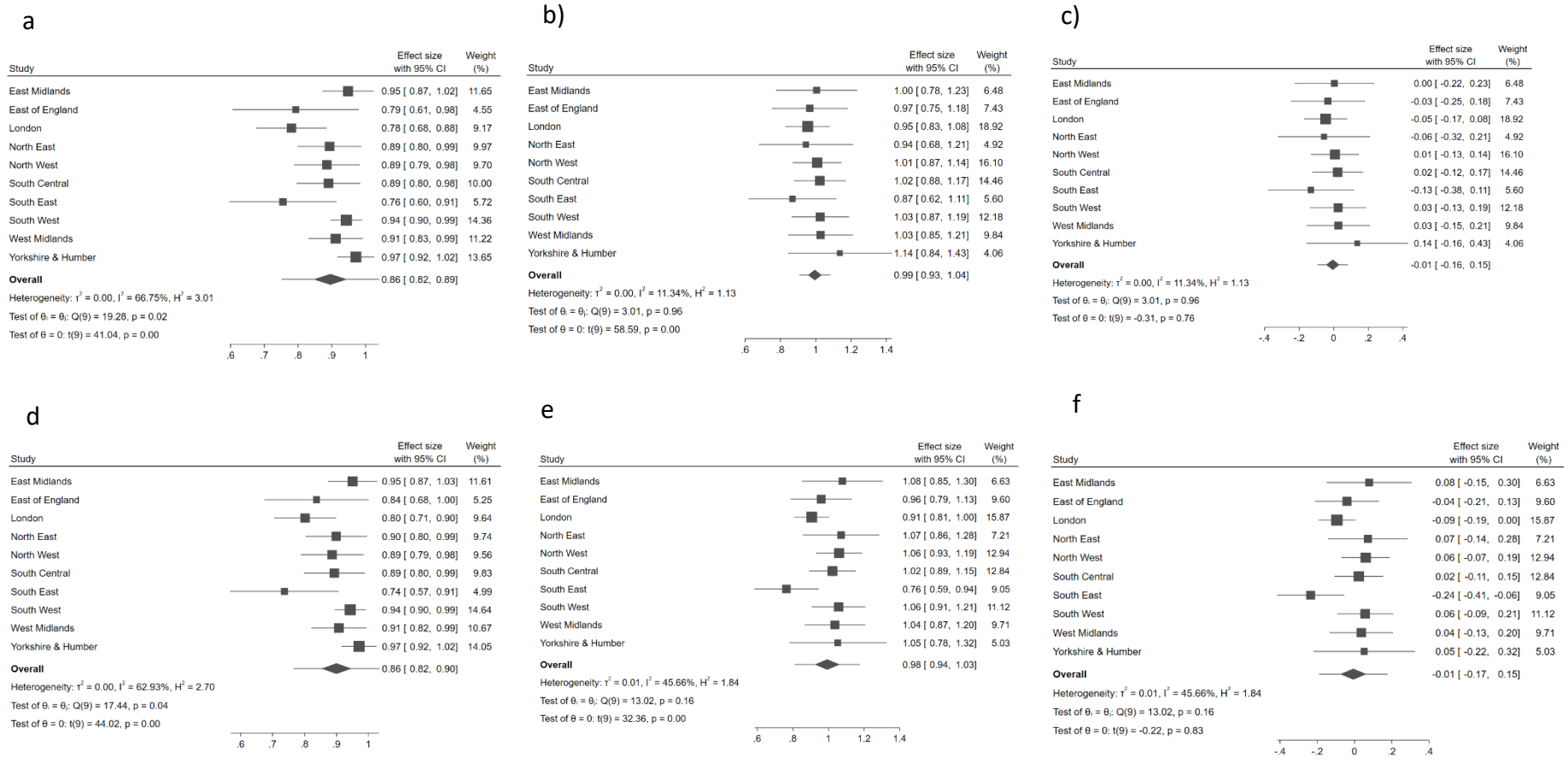
Figure 4-15 *Calibration of the a) symptom-only and b) symptom + blood test lymphoma models, demonstrated through calibration plots which show the alignment between predicted and observed risks across 10 risk deciles/groups.*



4.4.5.3 Heterogeneity

For the symptom + blood test model, meta-analysis of performance measures across the 10 geographical regions demonstrated that there was high level of heterogeneity for AUC (I^2=62.9%), for calibration slope (I^2=45.7%) and calibration-in-the large (I^2=45.7%) (Figure 4-16). For the symptom only model, heterogeneity for AUC was high (I^2=66.7%), but heterogeneity was low for calibration slope and calibration-in-the large (I^2=11.3%). Age, ethnicity and deprivation level did account for 17% of observed heterogeneity for the AUC in meta-regression analyses (Appendix Chapter 9, Supplementary Table 9-14).

Figure 4-16 *Region-level and pooled performance estimates following internal-external cross validation for a) & d) area-under-the-curve (AUC), b) & e) calibration slope and, c) & f) calibration-in-the-large for symptom-only (a-c) and symptom + blood test lymphoma models (d-f). Effect Size=adjusted Odds ratio, CI= Confidence interval.*



4.4.5.4 Model Fairness

When the model performance was assessed across different age groups (child 0-14 years & TYA 15-24 years), sexes (male & female) and ethnic groups (white & non-white), it was found that the models had lower discriminatory ability for children and non-white ethnic groups (Table 4-10)

Table 4-10 *Performance measures per age, sex and ethnic subgroups for i) symptom-only and ii) symptom + blood test lymphoma models. AUC= area under the curve, 95% CI = 95% confidence interval*

	AUC (95% CI)	Calibration Slope (95% CI)	Calibration-in-the-large (95% CI)
Symptom only			
Age			
Children	0.72 (0.61 to 0.82)	0.83 (0.70 to 0.96)	-0.055 (-0.36 to 0.24)
TYA	0.90 (0.86 to 0.93)	1.04 (0.97 to 1.08)	0.012 (-0.17 to 0.19)
Sex			
Male	0.84 (0.79 to 0.90)	0.96 (0.89 to 1.04)	0.09 (-0.12 to 0.29)
Female	0.86 (0.80 to 0.92)	1.02 (0.93 to 1.10)	-0.11 (-0.34 to 0.12)
Ethnic Group			
White	0.88 (0.92 to 0.94)	0.95 (0.85 to 1.06)	0.088 (-0.21 to 0.39)
Other	0.78 (0.63 to 0.92)	1.00 (0.84 to 1.16)	-0.18 (-0.61 to 0.24)
Symptom + test model			
Age			
Children	0.74 (0.65 to 0.84)	0.80 (0.68 to 0.92)	-0.065 (-0.37 to 0.24)
TYA	0.90 (0.87 to 0.94)	1.04 (0.99 to 1.10)	0.0070 (-0.18 to 0.19)
Sex			
Male	0.87 (0.81 to 0.92)	1.01 (0.94 to 1.08)	-0.10 (-0.34 to 0.13)
Female	0.86 (0.81 to 0.91)	0.96 (0.90 to 1.02)	0.068 (-0.14 to 0.28)
Ethnic Group			
White	0.88 (0.82 to 0.95)	0.96 (0.87 to 1.05)	0.078 (-0.22 to 0.38)
Other	0.82 (0.69 to 0.95)	0.88 (0.76 to 1.01)	-0.38 (-0.84 to 0.074)

4.4.5.5 Measures of Accuracy

The model's measures of accuracy based on the top 1%, 2%, 5%, 10%, 20% at highest risk was assessed following validation. Specifically, when 90th centile threshold of risk (cut-off 0.02%) was used to define young people at high-risk of lymphoma, the symptom + test model showed 70% sensitivity and 90% specificity, with a positive predictive value of 0.02% and negative

predictive value of 99.9%, and the symptom only model sensitivity was similar at 69% (Table 4-11).

Table 4-11 *Measures of accuracy of each model, following validation, based on the top 0.1%, 0.5%, 1%, 2%, 5%, 10%, 20%, 40%, 60%, 80%, 90%, 95% at highest risk in symptom-only and symptom + blood test lymphoma models*

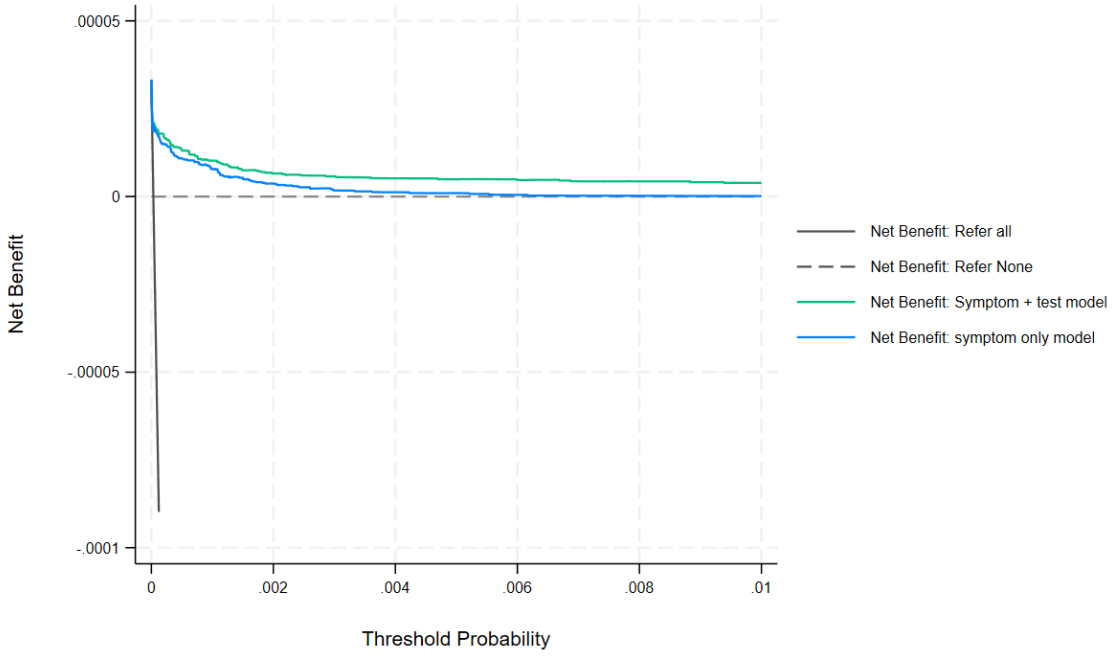
	Risk Threshold (%)	True positive	True negative	False positive	False negative	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
0.1%									
Symptom only	0.45	41	4,860,161	9,536	121	25.3	99.8	0.43	99.0
Symptom + test	0.57	59	4,851,596	9,518	193	36.4	99.8	0.61	99.0
0.5%									
Symptom only	0.23	65	4,837,017	24,097	97	40.1	99.5	0.27	99.0
Symptom + test	0.27	79	4,837,031	24,083	83	48.7	99.5	0.32	99.0
1%									
Symptom only	0.13	82	4,812,727	48,387	80	50.6	99.0	0.17	99.0
Symptom + test	0.15	92	4,812,737	48,377	70	56.7	99.0	0.19	99.0
2%									
Symptom only	0.076	90	4,764,121	96,993	72	55.6	98.0	0.094	99.9
Symptom + test	0.082	96	4,764,113	96,985	66	59.2	98.0	0.10	99.9
5%									
Symptom only	0.035	106	4,618,294	242,820	56	65.4	95.0	0.043	99.9
Symptom + test	0.035	110	4,618,300	242,816	52	67.9	95.0	0.045	99.9
10%									
Symptom only	0.020	112	4,375,229	485,885	50	69.1	90.0	0.023	99.9
Symptom + test	0.020	114	4,375,231	485,883	48	70.3	90.0	0.024	99.9
20%									
Symptom only	0.012	124	3,889,099	972,015	38	76.5	80.0	0.013	99.9
Symptom + test	0.012	124	3,889,099	972,015	38	76.5	80.0	0.013	99.9
40%									
Symptom only	0.0072	139	2,916,830	1,944,284	23	85.8	60.0	0.0071	99.9
Symptom + test	0.0072	138	2,916,829	1,944,285	24	85.1	60.0	0.0071	99.9
60%									
Symptom only	0.0052	148	1,944,555	2,916,559	14	91.3	40.0	0.0051	99.9

Symptom + test	0.0051	151	1,944,558	2,916,556	11	93.2	40.0	0.0051	99.9
80%									
Symptom only	0.0037	153	972,276	3,888,838	9	94.4	20.0	0.0039	99.9
Symptom + test	0.0040	155	972,277	3,888,837	7	95.6	20.0	0.0040	99.9
90%									
Symptom only	0.0035	155	486,136	4,374,978	7	95.6	10.0	0.0035	99.9
Symptom + test	0.0036	157	486,137	4,374,977	5	96.9	10.0	0.0036	99.9
95%									
Symptom only	0.0034	157	243,067	4,618,047	5	96.9	5.0	0.0036	99.9
Symptom + test	0.0034	158	243,067	4,618,047	4	97.5	5.0	0.0036	99.9

4.4.5.6 Decision Curve Analysis

Decision curve analysis demonstrated that the model (with and without abnormal blood test results) had higher net benefit across all threshold probabilities compared to refer-none or refer-all strategies (Figure 4-17).

Figure 4-17 *Decision curve to assess net benefit of models across a range of threshold probabilities for lymphoma models*

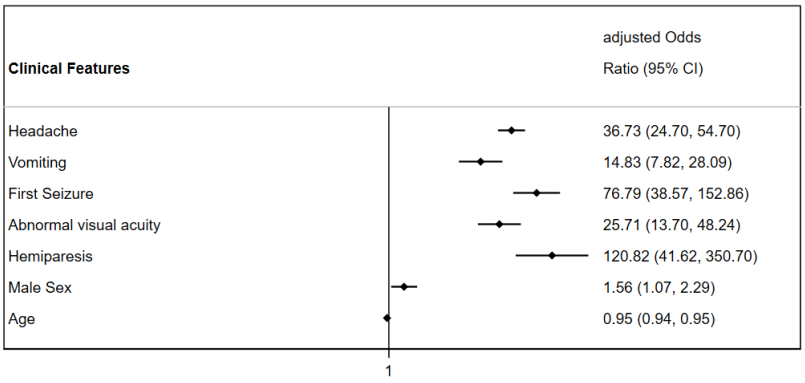


4.4.6 Model for Central Nervous System Tumours

4.4.6.1 Model Derivation

The final model was inclusive of sociodemographic factor sex, and non-linear fractional polynomial terms for age (-2, -2) (Appendix Chapter, Supplementary Figure 9-4), as well as 5 symptoms/signs (headache, vomiting, first seizure, abnormal visual acuity, hemiparesis) (Figure 4-18). There were no blood tests which could predict the risk of central nervous system tumours and thus the final model only included symptoms.

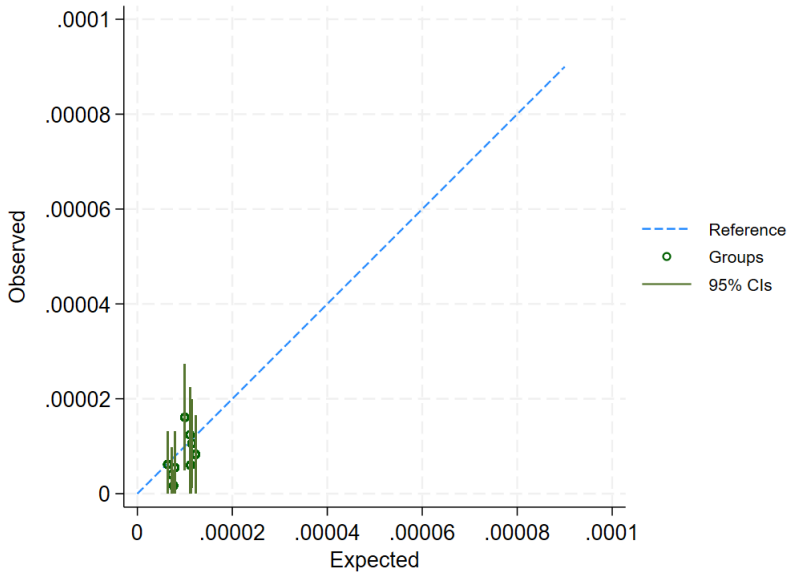
Figure 4-18 *Final logistic regression model predicting odds of CNS tumours 6 months following study entry/symptom onset. Models contain fractional polynomials (FP) for age (-2,-2), CI= Confidence interval*



4.4.6.2 Model performance

Following internal-external cross validation, the model was found to discriminate well (AUC=0.85 [95%CI 0.81 to 0.90]) and was well-calibrated (slope=0.99 [95%CI 0.92 to 1.05], calibration-in-the-large=-0.01 [95%CI -0.06 to 0.03]). Calibration plots demonstrated that there was overestimation at lower risk thresholds (Figure 4-19).

Figure 4-19 *Calibration of CNS tumour model, demonstrated through a calibration plot which shows the alignment between predicted and observed risks across 10 risk deciles/groups.*

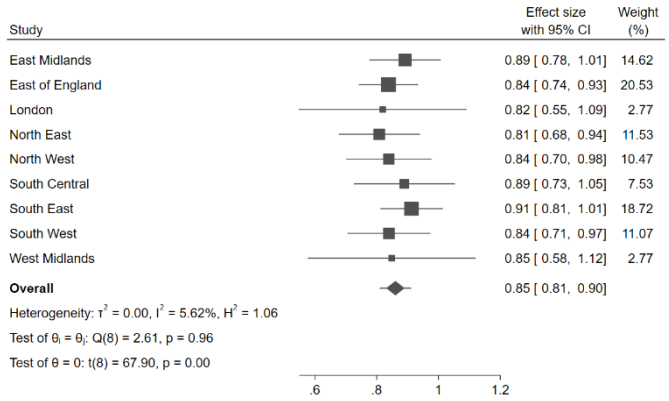


4.4.6.3 Heterogeneity

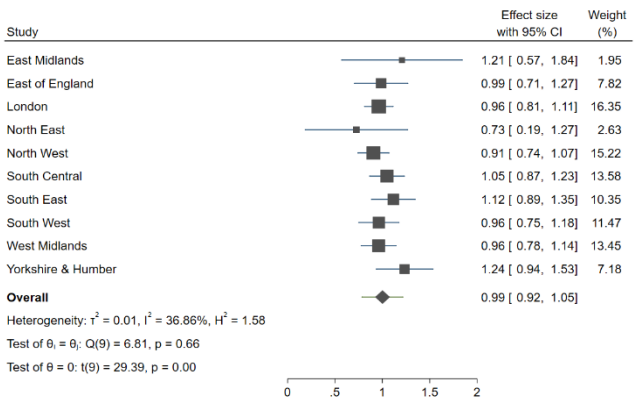
Meta-analysis of performance measures across the 10 geographical regions demonstrated that there was low level of heterogeneity for AUC (I^2=5.6%), and moderate for calibration slope (I^2=35.6%) and calibration-in-the large (I^2=36.8%) (Figure 4-20). Age, ethnicity and deprivation level did not account for the observed heterogeneity in meta-regression analyses (Appendix Chapter 9, Supplementary Table 9-14).

Figure 4-20 *Region-level and pooled performance estimates following internal-external cross validation for a) area-under-the-curve (AUC), b) calibration slope and, c) calibration-in-the-large for symptom-only CNS tumour models. Effect Size=adjusted Odds ratio, CI= Confidence interval.*

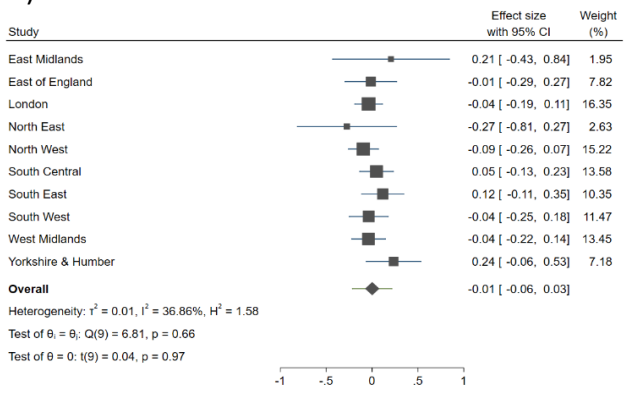
a)



b)



c)



4.4.6.4 Model Fairness

When the model performance was assessed across different age groups (child 0-14 years & TYA 15-24 years), sexes (male & female) and ethnic groups (white & non-white), the models had lower discriminatory ability for TYA and the male sex, but, confidence intervals were overlapping (Table 4-12).

Table 4-12 *Performance measures per age, sex and ethnic subgroups CNS tumour model. AUC= area under the curve, 95% CI = 95% confidence interval*

	AUC (95% CI)	**Calibration Slope (95% CI)**	**Calibration-in-the-large (95% CI)**
Age			
Children	0.88 (0.83 to 0.93)	1.04 (0.96 to 1.13)	0.3 (0.05 to 0.54)
TYA	0.80 (0.72 to 0.88)	0.92 (0.81 to 1.03)	-0.3 (-0.5 to -0.02)
Sex			
Male	0.81 (0.74 to 0.88)	0.99 (0.91 to 1.09)	0.003 (-0.24 to +0.24)
Female	0.88 (0.81 to 0.94)	0.97 (0.87 to 1.07)	-0.007 (-0.29 to 0.27)
Ethnic Group			
White	0.86 (0.81 to 0.90)	0.99 (0.91 to 1.06)	0.04 (-0.16 to 0.23)
Other	0.85 (0.79 to 0.90)	0.98 (0.91 to 1.06)	0.03 (-0.17 to 0.24)

4.4.6.5 Measures of Accuracy

The model's measures of accuracy based on the top 1%, 2%, 5%, 10%, and 20% at highest risk was assessed following validation.

Table 4-13 *Measures of accuracy, following validation, based on the top 0.1%, 0.5%, 1%, 2%, 5%, 10%, 20%, 40%, 60%, 80%, 90%, 95% at highest risk in CNS tumour model*

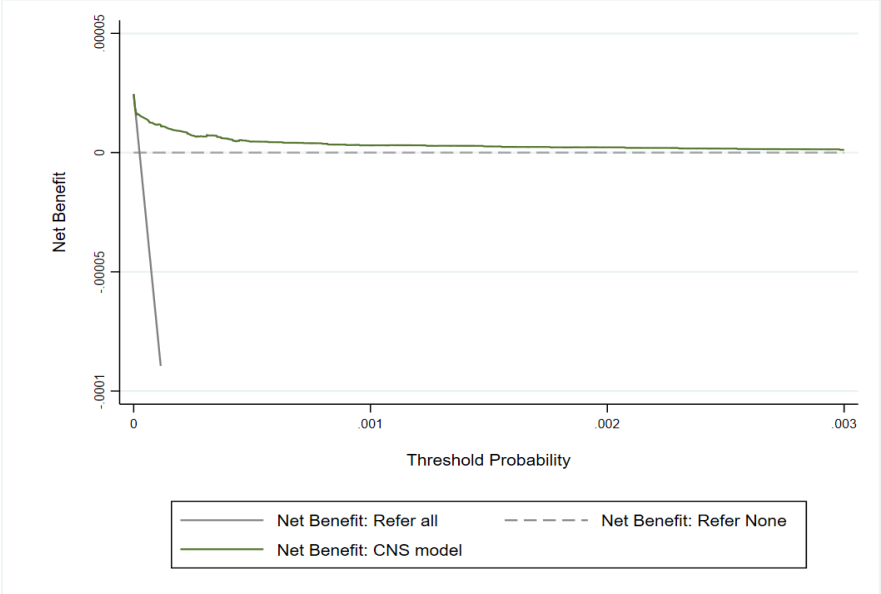
	Risk Threshold (%)	True positive	True negative	False positive	False negative	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
0.1%									
	0.36	30	4,812,926	9,505	69	42.5	99.0	0.1	99.0
0.5%									
	0.33	34	4,812,926	24,086	69	42.5	99.0	0.1	99.0
1%									
	0.28	51	4,812,926	48,376	69	42.5	99.0	0.1	99.0
2%									
	0.18	59	4,764,320	96,982	61	49.2	98.0	0.06	99.9
5%									
	0.061	81	4,618,499	242,803	39	67.5	95.0	0.03	99.9
10%									
	0.015	86	4,375,433	485,869	34	71.7	90.0	0.018	99.9
20%									
	0.008	90	3,889,295	972,007	30	75.0	80.0	0.009	99.9
40%									
	0.0047	98	3,889,295	1,944,283	30	75.0	80.0	0.009	99.9
60%									
	0.0035	112	3,889,295	2,916,553	30	75.0	80.0	0.009	99.9
80%									
	0.0028	115	3,889,295	3,888,834	30	75.0	80.0	0.009	99.9
90%									
	0.0026	117	3,889,295	4,374,974	30	75.0	80.0	0.009	99.9
95%									
	0.0025	117	3,889,295	4,618,045	30	75.0	80.0	0.009	99.9

4.4.6.6 Decision Curve Analysis

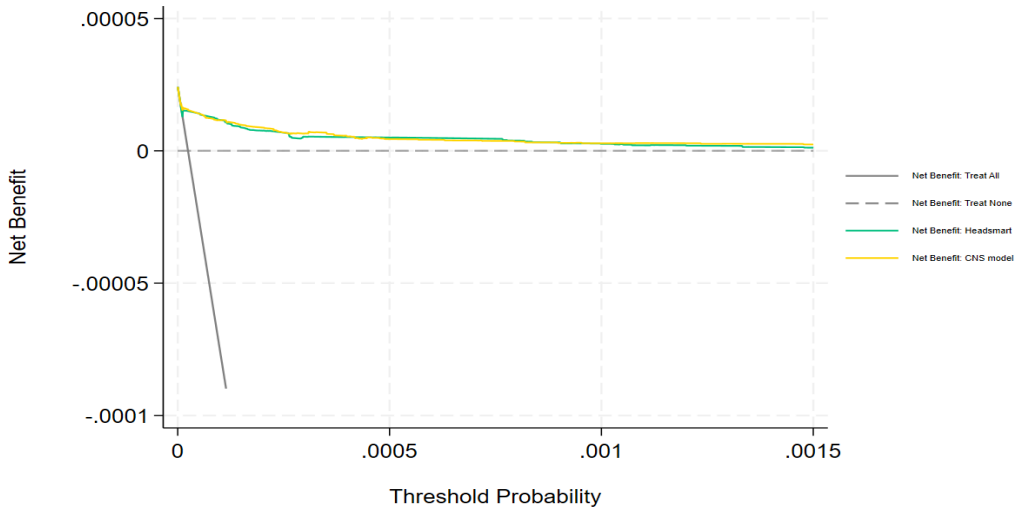
Decision curve analysis demonstrated that the model (with and without abnormal blood test results) had higher net benefit across all threshold probabilities compared to refer-none or refer-all strategies (Figure 4-21). The clinical utility of the developed model was comparable to that of the symptoms/signs used in the HEADSMART initiative.

Figure 4-21 *a) Decision curve to assess net benefit of the CNS tumour model across a range of threshold probabilities, b) Decision curve to assess net benefit of the developed CNS tumour model compared to symptoms/signs listed by HEADSMART initiative, across a range of threshold probabilities.*

a)



b)



4.4.7 Clinical Scenarios

Table 4-14 provides six examples of clinical scenarios from a range of ages and different clinical presentations and the obtained predicted risks probabilities using the models.

Table 4-14 Clinical Scenarios and predicted risks using developed models for a) haematological (leukaemia and lymphoma individually), b) solid cancers (CNS tumour individually).

a)	**Predicted Absolute Risk of Haematological Cancer (%)**	**Relative Risk of Haematological Cancer compared to baseline**	**Predicted Absolute Risk of Leukaemia (%)**	**Relative Risk of Leukaemia compared to baseline**	**Predicted Absolute Risk of Lymphoma (%)**	**Relative Risk of Lymphoma compared to baseline**
17 year old female, no other clinical features (baseline)	0.002	-	0.0006	-	0.004	-
17 year old female with lump on the neck, pruritus, CRP > 3mg/L	0.88	440	0.0006	1	2.2	550
2 year old male, no other clinical features (baseline)	0.012	-	0.01	-	0.00002	-
2 year old male with fever, lymphadenopathy & limb pain	0.43	35.8	1.1	110	0.00003	1.5

b)	**Predicted Absolute Risk of Solid Cancer (%)**	**Relative Risk of Solid Cancer compared to baseline**	**Predicted Absolute Risk of CNS tumour (%)**	**Relative Risk of CNS Tumour compared to baseline**
12 year old male, no other clinical features (baseline)	0.003	-	0.001	-
12 year old male with headache, vomiting and new onset vision problems	2.0	667	1.4	1,400
3 year old female, no clinical features (baseline)	0.005	-	-	-
3 year old female with abdominal pain and haematuria	0.058	11.6	-	-
9 year old female, no other clinical features (baseline)	0.002	-	-	-
9 year old female with lump on body, limb pain, CRP > 3 mg/L	0.29	145	-	-

4.5. Discussion

4.5.1 Summary of Principal Findings

Overall, both the combined haematological and solid cancer models, as well as the individual cancer subtype models, have good discriminative ability and are well-calibrated, and further show potential for clinical utility. Overall, this offers opportunities to improve early recognition and referral of young people at risk of cancer in primary care, where they are likely to present at the earliest stages of their disease.

4.5.2 Findings in Context

Although no previous risk prediction models have been developed to detect young people at risk of cancer in primary care, the benefits of integrating such models have been explored in adult cancers. QCancer, for example, was evaluated at a national level in the UK through the Accelerate, Coordinate, Evaluate (ACE) programme and it was mainly found to 1) improve awareness of cancer symptoms/signs in primary care and 2) support GPs to refer a suspected cancer case even when uncertainty exits [124]. In CTYA cancers, where diagnosis in primary care is often challenging, raising awareness and providing support for GPs have the potential to improve earlier detection and timely referrals. It is important to note however that the association between QCancer scores and cancer diagnosis could not be established in the ACE programme, and accordingly there was still limited understanding of the risk prediction model's ability to identify high-risk patients in primary care accurately. Further research is still required to better understand how risk prediction models can aid earlier cancer diagnosis.

The combined and individual haematological and solid cancer models were inclusive of demographic factors, including age and sex, as well as clinical factors, including symptoms, clinical signs and abnormal blood test results. Both combined models, as well as lymphoma and CNS tumour models, identified male sex as a risk factor, and this is consistent with the marginally higher number of CTYA cancers seen in males [6].

For haematological cancers, clinical features included previously well-characterised presentations, such as lymphadenopathy, head/neck lumps, organomegaly and bruising [36, 38]. For solid cancers, clinical features differed as expected and consisted of headache, vomiting, seizures, which is also included in the central nervous system tumour model developed, as well as lumps on the body, abdominal pain and haematuria (as previously reported for sarcomas and abdominal tumours) [35, 37]. Both combined haematological and solid models were inclusive of limb pain, which is a less well-established but potentially important clinical feature in cancers in this age group, as previously highlighted in Chapter 3.

Several abnormal blood test results were common to both combined models, including a low haemoglobin count (<12 mg/dL) and thrombocytopenia (platelets <150x10^9/L), likely due to bone marrow pathology in haematological cancers and some solid cancers (e.g., neuroblastomas)[125], and elevated inflammatory markers (ferritin>300 ng/ml, ESR>20 mm/hr and CRP>20 mg/L), probably representing a cancer-driven acute inflammatory response [126]. Interestingly, both neutropenia (<1.5x10^9/L) and neutrophilia (>8x10^9/L) were included in the model for haematological cancers. This is most likely due to the model being inclusive of both leukaemias and lymphomas, which have different effects on the haematopoietic system, as can also be seen from the models for each haematological cancer, where the leukaemia models is inclusive of only neutropenia and lymphoma model of only neutrophilia. The CNS tumour model was not inclusive of any abnormal blood test result and highlights the importance of a tailored approach to blood test requests in the diagnostic workup.

In combination, inclusion of these demographic and clinical predictors within the models provides two important opportunities to aid earlier diagnosis of cancers within this age group. Firstly, it allows clinicians to contextualise clinical features that are relevant to a subsequent cancer diagnosis given a young person's demographic background. This can be particularly

useful when young people present with symptoms/signs, such as lymphadenopathy, that are also common to self-limiting childhood illnesses. Secondly, it allows clinicians to consider undertaking targeted blood test investigations if there is an underlying suspicion of cancer and this information can help aid earlier discussions with haemato-oncological services.

Following internal-external validation, both combined models, as well as the three individual cancer models, showed good discriminative ability. Additionally, all models were well-calibrated overall, i.e. were able to accurately estimate the absolute cancer risk. With good discrimination and calibration, the models appear to have higher net clinical benefit than refer-all or refer-none strategies and current NICE guidance, as seen following decision curve analysis. Altogether, this has several important implications for improving clinical decision-making and considering risk thresholds for referral from primary care. Determining risk thresholds however would need to be in parallel with cost-effectiveness considerations and clinical guidance.

Reassuringly, the model performances were large comparable across different age groups (childhood vs. TYA), sexes and ethnic groups (White vs. non-White ethnic groups). There was marginally worse performance, with large confidence intervals, in the solid cancer model for TYA compared to children and in the lymphoma model, for children compared to TYA. This finding may be due to lower incidence of TYA solid cancers and childhood lymphomas [6].

Interestingly, there was a moderate level of heterogeneity of performance measures across the different models. Age, ethnicity and deprivation level were able to account for a small proportion of heterogeneity seen in leukaemia/lymphoma models only. Other factors, which could not be measured, such as variable access to healthcare across different regions in England[127], educational and employment levels of young people and their families, may contribute to this heterogeneity.

4.5.3 Strengths and Limitations

This study had several strengths. Firstly, it used large, representative population-level data, linked and validated for CTYA cancer. This is likely to reduce the risk of selection, ascertainment and recall bias. Secondly, this study benefitted from the availability of both symptomology and blood test results, which offered opportunities to develop more comprehensive models. Thirdly, this study was able to develop models in both broad categories, haematological and solid cancers, as well as individual cancer subtypes, and they all appear to provide good overall performance. Thus, GPs can utilise these models either when they have a concern about a specific cancer or are overall concerned about an underlying cancer diagnosis (i.e., in a non-specific presentation where the exact subtype presentation may not be clear). Finally, an internal-external cross-validation approach was taken, which allowed to explore heterogeneity and generalisability of the models, whilst also demonstrating no evidence of overfitting.

This study also had several important limitations. First, certain important predictors, such as repeated attendances [25, 26] and genetic predisposition markers, could not be included in the models and would need to be considered by clinicians in parallel to using these models. Second, inclusion of more detailed predictor information that could improve model performance (e.g., duration of fever) was not possible. However, we were able to explore certain interactions between clinical features within our models (e.g., fever and lymphadenopathy). Third, we did not externally validate the models using an independent dataset due to large sample size requirements and this will need to be undertaken in the future to assess generalisability of the models further. Fourth, blood test results were dichotomised into normal and abnormal, resulting in several limitations. This included, the reduction of meaningful information that arose from categorisation, but it was overall decided to be a better approach than using multiple imputation (using predictive mean matching) of

continuous blood test results, as it was likely to introduce bias given the small proportion of those tested [91]. Furthermore, those who were not tested were grouped into the "normal" blood test category because it was assumed that this untested group would most likely have a normal blood test result given testing was not clinically indicated. Additionally, having an untested category may introduce bias, as a previous study that investigated inflammatory marker testing and the risk of an adult cancer diagnosis found that untested patients had even lower risk of cancer than those who were tested but with normal inflammatory marker results [110]. On the other hand, this dichotomised categorisation approach may also introduce bias by underestimating the risk of cancer in those categorised as "normal" and also significantly limits insight into the potentially important differences in the tested and untested CTYA populations. Fifth, the overall clinical utility of the models are limited by the rarity of CTYA cancers. This can be seen through the clinical scenarios highlighted in Table 4-14, where even though the relative risk of cancer in those with clinical features is significantly higher than those without clinical features, the absolute risk of cancer in symptomatic CTYA, and corresponding PPVs, remain low. In one example clinical scenario in Table 4-14, the absolute risk of haematological cancer in a 2 year old boy with fever, lymphadenopathy and limb pain was 0.43%, which corresponds to a PPV of 0.6%. This PPV is significantly lower than the 3% PPV threshold for referral set by NICE guidance for adult cancers [128]. This threshold for adult cancers may however be too high for CTYA cancers and accordingly recent NICE guidance has suggested that CTYA cancers should be an exception to the use of this 3% threshold [128]. Despite this limitation, it is important to note that the models show improved clinical utility compared to current guidance. Finally, as this study is an observational study based on routinely collected data, there is the possibility of recording bias of clinical features (which can influence variable selection during model development) and residual confounding from unaccounted confounders.

4.5.4 Future Directions

This study developed and evaluated primary care-based risk prediction models for CTYA cancers, demonstrating good overall performance. Utilising these models have the important potential to improve timely referral and ultimately earlier recognition of symptomatic CTYA cancers in primary care in the UK. Next steps will need to evaluate external validation and cost effectiveness, as well as integration of these models in primary care settings to assess their acceptability and impact.

Code from Chapter

https://github.com/DefOx2020/Thesis/blob/06e6b1c26357ad03b73a6fc065ef8af9040ef1b4/Chapter%204

Publications in progress from Chapter

Saatci D., Harnden A., Hippisley-Cox J. *Development and internal-external validation of risk prediction models for childhood, teenage and young adult cancers: a population-level study of 4.8 million young people (in submission)*

Chapter 5 Identifying factors that affect diagnostic pathways in childhood, teenage and young adult cancer

5.1. Background to chapter

The underlying causes resulting in delays to cancer diagnosis are multi-factorial. These factors have been largely studied across adult cancers and have been found to include 1) person-specific factors, such as age [129], sex [21], presence of comorbidities [130] and ethnicity [131, 132], 2) cancer-specific factors, such as the type of tumour [23, 133], its biology [23] and associated symptoms/signs [134], 3) health system-specific factors, such as differences in access to services across regions [135], and 4) societal factors, such as economic and health crises (e.g., the coronavirus-19 pandemic) [50, 136].

In the UK, where diagnostic intervals of CTYA cancers are some of the longest in Europe [137], factors associated with diagnostic delays are not fully characterised across all cancers in this age group. Most evidence is available for cancers in 15-to-24 year-olds, through the BRIGHTLIGHT cohort, and longer diagnostic intervals have been associated with age, sex and tumour type [21]. Delays in central nervous system tumours have also been studied and have shown to be linked to age and tumour location [133]. Knowledge of factors that are associated with other cancer subtypes, particularly in children (0-14 year olds), is lacking. Factors that are associated with emergency cancer presentations in CTYA, which are closely linked to diagnostic delays [138], also have not been studied extensively in this age group, particularly at a population-level.

5.2. Aims and Objectives

The overarching aim is to identify factors that influence diagnostic pathways. Specific objectives are:

1. To explore how systemic diagnostic intervals are associated with person-specific, cancer-specific, health system-specific and societal factors.

2. To explore how route-to-diagnosis of cancer is associated with person-specific, cancer-specific and health system-specific factors.

5.3. Methods

5.3.1 Data Source & Study Population

This Chapter utilises the full cohort of selected cancer cases identified through the open cohort of CTYA generated in Chapter 2, as detailed in Figure 2-6. In brief, this is 4,389 young people with a cancer diagnosis from a cohort of 5,238,743 young people between the ages of 0-24 years. The cancers were categorised into the commonest subtypes in this age group, which are defined by the National Cancer Registration and Analysis Service and in accordance with the ICCC-3 [2]: 1) Leukaemias, myeloproliferative and myelodysplastic diseases (n=1,127), 2) lymphomas and reticuloendothelial neoplasms (n=919), 3) central nervous system and intraspinal tumours (n=933), 4) soft tissue and bone sarcomas (n=637), 5) abdominal tumours (renal tumours, neuroblastomas, hepatic tumours) (n=369) and 6) gonadal germ cell tumours (n=404).

5.3.2 Outcomes of Interest

5.3.2.1 Diagnostic interval

Diagnostic interval was defined as “from the date of the first presentation to GP to diagnosis date”. This represents the systemic diagnostic interval, combining the primary care (i.e., from date of first presentation to GP to referral) and secondary care (i.e., from date of referral by

GP to diagnosis date) intervals. This definition was chosen as it is widely reflective of the literature exploring diagnostic intervals in cancer using EHR [32]. It will be referred to as "diagnostic interval" in this Chapter. This systemic diagnostic interval was chosen to be studied in combination instead of separating primary care and secondary care intervals individually, as only 16.3% (n=854,358/5,238,743) of the cohort had recorded referral data.

First presentation was defined as the first recorded cancer-relevant symptom/sign for each cancer subtype, as identified in Chapter 3. For example, leukaemia diagnoses included time from first recording of "anaemia/pallor" to diagnosis date; whilst central nervous system tumour diagnoses included time from "visual acuity problems" to diagnosis date. Then each subtype specific diagnostic interval was combined together to give an overall diagnostic interval for cancers in this age group.

5.3.2.2 Route-to-diagnosis

Cancer route-to-diagnosis is recorded within the National Cancer Registry and is defined by the following eight routes-to-diagnosis, as listed in Table 5-1. As this is the most comprehensive information available on route-to-diagnosis on linked datasets, it was used to define the route-to-diagnosis outcome.

Table 5-1 *Route to cancer diagnosis (adapted National Cancer Registry[34])*

Route	Description
Urgent Suspected Cancer/ two week wait (TWW)	Urgent referrals from GP for suspected cancer
GP referral	Routine and urgent GP referrals (non-TWW)
Emergency presentation	Emergency defined as attendance to accident and emergency (A&E), emergency GP referral, emergency transfer or admission up to 30 days before diagnosis
Other outpatient	Outpatient elective route
Screen-detected	Identified through screening programme.

Inpatient elective	Inpatient elective route
DCO	Death certificate only.
Unknown	No identified data in HES, screening or National Cancer Waiting Times.

The routes were re-categorised into two broad groups; those reflective of emergency presentations (i.e. emergency presentation category) or non-emergency presentations (i.e., GP referral, other outpatient, screen-detected, inpatient elective). Broadening the categories were deemed to be the most relevant for young people with cancer due to prior clinical experience and evidence that two-week wait, death certificate only and screening routes are rare in young people [3, 139].

5.3.3 Exposures of Interest

The literature across adult and CTYA cancers was reviewed using EMBASE and MEDLINE databases to identify factors previously shown to be associated with diagnostic delay or emergency presentations. The search strategy for this can be found in the Supplementary Appendix (Appendix Chapter 9, Supplementary Table 9-15). Identified factors were subsequently categorised into person-specific, cancer-specific, health system-specific and societal factors and are listed in Table 5-2. These identified factors, where available in the data provided by the QResearch Database and linked datasets, were included in analyses and the equivalent variables in the datasets are listed in Table 5-2. Variable outcomes were taken at the point of study entry, prior to cancer diagnosis date.

Table 5-2 *Factors associated with diagnostic delays and emergency presentations*

Factors	**Details of factors**	**Variables available on the QResearch Database**
Person-specific	Age at presentation [129] Sex [21] Ethnicity [131, 132] Presence of comorbidities [130]	Age at presentation Sex Ethnicity Presence of comorbidities

	Socioeconomic status	Socioeconomic status
Cancer-specific	Tumour type [23] Tumour biology [23] Tumour location [23] Symptoms/signs [134]	Cancer subtype Aggressiveness Symptoms/signs
Health system-specific	Test request (blood or imaging) [23] Clinician seen (generalist vs. specialist) [23] Availability of diagnostic tests [23] Referral type [23]	Previous presentation to GP within 6 months of diagnosis Test request Region
Global/Societal	COVID-19 pandemic 2008 economic crisis [50, 136]	COVID-19 pandemic

In further detail these included:

1. Age was categorised into children (0-14 years) and TYA (15-24 years), as previously described, to reflect the categorisation available on the National Cancer Registration and Analysis Service [1].

2. As per previous Chapters, socioeconomic status was explored using the Townsend Deprivation Score on the QResearch Database [65].

3. Self-assigned ethnicity was categorised into White and Other ethnic groups, following adaptation from the 2021 Census [64] and to ensure sufficient sample in each group.

4. Presence of comorbidities: mental health disorders (depression, anxiety, and psychosis) and autism were selected due to their relevance in young people and previous evidence that they influence diagnostic intervals in adult cancers [130]. Only diagnoses prior to cancer diagnosis date were included and only secondary care data (ICD-10 codes) were available at the time of analysis.

5. Cancer biology was categorised into a binary outcome, aggressive vs. non-aggressive. Aggressiveness was defined through histology codes (Appendix Chapter 9, Supplementary Table 9-16).

6. Symptoms and signs relevant to cancer within 6 months of diagnosis were identified in Chapter 3 per age group and cancer subtype. Median diagnostic interval of each clinical feature was characterised. In addition, symptoms and signs were grouped into “red flag” and not “red flag”. Red flag symptoms were identified through NICE [78], Grace Kelly Trust [80] and HEADSMART [30] guidelines.

7. Test request was either a cancer relevant blood test request identified in Chapter 4 or relevant imaging request (based on availability on datasets and clinical knowledge of tests used to diagnose cancer in young people) within 6 months prior to diagnosis. These tests were ultrasound (any body part), computed tomography (CT, any body part), positron emission tomography (any body part) and magnetic resonance imaging (MRI) of the central nervous system.

8. Previous presentation to GP in the 6 months prior to diagnosis was used to identify CTYA who had sought medical advice prior to a potential emergency presentation. 6 months was used as a cut-off to balance sufficient sample size and as evidence has shown that young people are most likely to present to their GP within a 6 month period before cancer diagnosis [21, 25].

9. Region was classified into 3 groups: North (North West England, North East England, and Yorkshire & Humber), Midlands (East and West Midlands) and South (London, South West England, South England, South East England, and East of England). The 10 regions available on QResearch Database were broadened to achieve larger samples per region for comparisons.

10. COVID-19 pandemic: Pre-pandemic periods 1st January- 1st May 2017-2019 and 2nd May-31st December 2017-2019 were compared to the first wave (1st January 2020- 1st May 2020, period 1), the second wave (1st January 2021- 1st May 2021, period 2), and pre-vaccination rollout (2nd May 2020-31st December 2020, period 3), respectively.

5.3.3.1 The COVID-19 Cohort

In order to study diagnostic pathways across the COVID-19 pandemic, a more contemporary CTYA cohort was required than the main cohort described in Chapter 2. Accordingly, the cohort of CTYA (n= 5,099,095, 0-24 year olds) between 2017 to December 2021 were extracted from the QResearch Database and the baseline demographics of this cohort is detailed in the Supplementary Appendix (Appendix Chapter 9, Supplementary Tables 9-17 & 9-18).

This COVID-19 cohort had some important differences from the main cohort. First, as there was no National Cancer Registry information available (last update available on QResearch Database is 2019), the CTYA cancers were grouped into the ICCC-3 framework described in Chapter 2 using primary and secondary care datasets only. This was deemed sufficient given results of Chapter 2 showed high sensitivity when both datasets were used. Second, as cancer route information is only available from the National Cancer Registry, only diagnostic interval outcomes were explored. Third, the study was carried out between 2020 and 2021 to provide real-time information that was clinically relevant to the pandemic. This meant that it was carried out prior to the nested case-control study described in Chapter 3 and thus diagnostic intervals were instead calculated using only a list of established pre-diagnostic symptoms (as detailed in the NICE Guideline for Suspected Cancer [78]) within 6 months from diagnosis. The symptoms set by NICE are available within the Supplementary Appendix (Appendix Chapter 9, Supplementary Table 9-19). Furthermore, only 5 CTYA cancer subtypes, i.e., leukaemias,

lymphomas, CNS tumours, sarcomas and renal tumours, were requested at the time of the study.

5.3.4 Statistical Analysis

For descriptive analyses, continuous variables were presented as means (medians if non-normal) and categorical variables were presented in frequencies and comparisons between cancers and the general population were carried out using t-tests (Mann-Whitney U if non-normal) and $\chi 2$ tests, respectively.

5.3.4.1 Outcome 1: Diagnostic Interval

For the outcome-of-interest, diagnostic interval, quantile regression with bootstrap method for standard errors was used to explore the association between the diagnostic interval and exposures-of-interest. Exposures-of-interest were investigated following adjustment for age, sex, cancer subtypes and cancer biology, given prior clinical and epidemiological evidence of their association with diagnostic time intervals [23]. Quantile regression was chosen over other regression approaches as 1) it provides a more comprehensive and clinically interpretable result (e.g., compared to a hazard ratio, which does not provide information on the number of days), and 2) it does not assume a normal distribution (diagnostic intervals are not normally distributed) [140]. Furthermore, bootstrapped standard errors were calculated as this approach also does not make any assumptions about the distribution [141]. To balance the large number of comparisons and clinical relevance as per Chapters 3 and 4, a significance threshold of 0.01 was considered to be statistically significant.

5.3.4.2 Outcome 2: Route-to-diagnosis

For the outcome-of-interest, route-to-diagnosis, logistic regression was used to explore the association between the route and exposures-of-interest. Exposures-of-interest were investigated following adjustment for age, sex, deprivation level, cancer subtypes and cancer

histological aggressiveness, given prior evidence of their association [23]. To account for clustering of young people within GPs, clustered standard errors were used.

5.3.4.3 Handling Missing Data

There was missing data both in the exposures-of-interest (ethnicity and Townsend Deprivation Score) and outcomes (diagnostic interval and route-to-diagnosis).

For exposures-of-interest, ethnicity and Townsend Deprivation Score, as per Chapter 3, the data was deemed to be missing under the missing at random assumption (i.e., any systematic differences between the missing and observed data can be explained through associations within the observed data) [142]. 31% of the cohort (1,359/4,389) had a recorded diagnostic interval and 64.2% of the cohort (2,814/4,389) had a recorded route-to-diagnosis. Both were also considered to be missing at random, given their method of collection, described above in Table 5-2. Accordingly, both complete case analyses and sensitivity analyses following multiple imputation were carried out for this outcome.

Multiple imputation with chained equations was used to replace missing data. 50 imputations were performed, given % missingness of ethnicity and diagnostic intervals and no computational restrictions [91]. Imputation models included all exposure variables, and the primary outcome variables; statistical models were developed on each of the 50 imputed data sets and the estimates pooled using the Rubin rules [91]. All continuous variables were assessed for normality [91]. In this Chapter, the only continuous variable included was the diagnostic interval, which was not normally distributed and was transformed with its natural logarithm to follow a normal distribution.

5.3.5 Sample Size and Power Calculations

For diagnostic intervals, prior to carrying out analyses, a simulated dataset with a normal distribution where a median can be specified was generated to calculate sample sizes across

different delta values (differences in medians) at 80% power and significance threshold of 0.05 and found that with a sample size of 1,359 in this study, a difference of 20% was detectable.

For emergency admissions, given the total sample size of 2,814 for emergency admissions, at 80% power and significance threshold of 0.05, a minimum OR of 1.16 would be detectable.

5.4. Results

5.4.1 Outcome 1: Diagnostic Intervals

There was a total of 1,359 (31% of full cohort) young people with a recorded diagnostic interval. The exposures-of-interest in those with and without recorded diagnostic intervals are presented in Table 5-3. There were important differences across those with and without recorded diagnostic intervals. First, more males without a recorded diagnostic interval were observed compared to females (p=0.01). This was driven by TYA cancers (diagnostic interval not recorded, males n=926, 64.7%, p=0.001) rather than childhood cancers (diagnostic interval not recorded, males n=906, 56.0%, p=0.9). Second, lymphomas had a higher proportion with a recorded diagnostic interval, whilst gonadal germ cell tumours had a much smaller proportion with a recorded diagnostic interval, compared to those who had no record of a diagnostic interval (<0.001). Third, among those who did have a recorded diagnostic interval, there was a higher proportion of blood test requests recorded (<0.001). Other exposures of interest were comparable. Of note, there was a very small proportion of young people with selected comorbidities (n<5) and accordingly, were not included in further analyses.

Table 5-3 *Exposures-of-interest in those with and without a recorded diagnostic interval. All percentages are column percentages (%). Age is in integers and rounded up to the nearest integer. IQR= interquartile range.*

Exposure of interest	Outcome recorded (n=1,359)	Outcome not recorded (n=3,030)	P-value
Age at cohort entry			
Median ± IQR (years)	5 [1,12]	6 [1,14]	0.4
Sex			
Male	772 (56.8)	1,835 (60.6)	0.01
Female	587 (43.2)	1,195 (39.4)	
Ethnicity			
White	352 (26.0)	793 (26.2)	0.2
Other	211 (15.5)	417 (13.8)	
Missing	796 (58.5)	1,820 (60.0)	
Townsend Deprivation Score			

1 (most affluent)	325 (23.9)	756 (25.0)	0.2
2	317 (23.3)	633 (20.9)	
3	285 (21.0)	598 (19.7)	
4	225 (16.6)	556 (18.4)	
5 (most deprived)	193 (14.2)	459 (15.1)	
Missing	14 (1.0)	28 (0.9)	
Comorbidities			
Present	<5	<5	0.9
Absent	-	-	
Cancer groups			
Leukaemia	358 (26.3)	769 (25.4)	<0.001
Lymphoma	437 (32.2)	482 (15.9)	
CNS tumours	290 (21.3)	643 (21.2)	
Sarcomas	165 (12.1)	472 (15.6)	
Abdominal tumours	84 (6.2)	285 (9.4)	
Gonadal germ cell tumours	25 (1.8)	379 (12.5)	
Cancer Biology			
Aggressive	310 (22.8)	662 (21.9)	0.5
Non-aggressive	1,049 (77.2)	2,368 (78.1)	
Test Request			
Bloods requested	511 (37.6)	544 (17.9)	<0.001
Bloods not requested	848 (62.4)	2,486 (82.1)	
Imaging requested	74 (5.5)	120 (4.0)	0.03
Imaging not requested	1,285 (94.5)	2,910 (96.0)	
Region			
North	395 (29.1)	884 (29.2)	0.9
Midlands	181 (13.3)	395 (13.0)	
South	783 (57.6)	1,751 (57.8)	

5.4.2 The association between exposures and the cancer diagnostic interval

The 25th centile, median and 75th centile diagnostic intervals for each exposure variable unadjusted and adjusted for age, sex, cancer subtype and cancer biology are provided below. All days are rounded to the integer above. Findings following multiple imputation of the diagnostic interval is available in the Supplementary Appendix (Appendix Chapter 9, Supplementary Table 9-20) and have comparable results to the analyses detailed below.

1. *Age-at-diagnosis*

Compared to children (0-14 year olds), in TYA (15-24 year olds), the 25th, median and 75th centile time-to-diagnosis was 8, 17 and 23 days longer in unadjusted models, respectively and all were statistically significant. Following adjustment, the 25th, median and 75th time-to-diagnosis was found to be 2, 6 and 7 days longer in TYA, respectively, however this was not statistically significant (Table 5-4).

Table 5-4 *The association between age-at-diagnosis and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for sex, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Percentile	Children (0-14 years, n=741) (days, 95% CI)	TYA (15-24 years, n=618) (days, 95%CI)	Unadjusted (days, 95% CI)	p-value	Adjusted (days, 95%CI)	p-value
25th	3 [2,5]	11 [8,13]	+8 [+5,+11]	<0.001	+2 [-1, +4]	0.08
50th	18 [15,22]	34.5 [29,40]	+17 [+11,+24]	<0.001	+6 [0,+12]	0.04
75th	60 [52,67]	83 [74,90]	+23 [+13,+33]	<0.001	+7 [-4,+18]	0.2

2. *Sex*

Males and females had similar diagnostic intervals in unadjusted and adjusted models (Table 5-5).

Table 5-5 *The association between sex and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Percentile	Female (n=587) (days, 95% CI)	Male (n=772) (days, 95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	6 [5,8]	6 [4,7]	0 [-2, +1]	0.9	0 [-2, +1]	0.9
50th	26 [22,30]	24 [21,29]	-2 [-9, +4]	0.5	-1 [-6, +4]	0.7
75th	73 [63,84]	67 [60,77]	-6 [-15,3]	0.2	-9 [-18, 0]	0.06

3. Ethnicity

Diagnostic intervals were comparable across White and Other ethnic groups (Table 5-6).

Table 5-6 *The association between ethnic group and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, sex, cancer subtype and cancer aggressiveness. 95% CI = 95% confidence interval.*

Percentile	White ethnic group (n=352) (days, 95% CI)	Other ethnic groups (n=211) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	6 [4, 11]	7 [4, 9]	1 [-4, +6]	-	0 [-2, +2]	0.9
50th	30 [25, 37]	26 [18, 35]	+4 [-11, +13]	-	-2 [-10,+6]	0.6
75th	73 [63, 83]	82 [63, 104]	+9 [-20, +38]	-	+1 [-17, +21]	0.8

4. Deprivation level

Although the 25th, 50th and 75th centile diagnostic intervals were shorter in the least deprived (Quintile 1) compared to the more deprived (Quintile 4/5), the difference was not statistically significant in adjusted analyses (Table 5-7).

Table 5-7 *The association between deprivation level, as assessed by the Townsend Deprivation Score, and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, sex, cancer subtype and cancer aggressiveness. Baseline quintile is Quintile 1 (i.e., least deprived). 95%CI = 95% confidence interval.*

Percentile	Quintile 1 (n=325) (days, 95% CI)	Quintile 2 (n=317) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	5 [3,7]	6 [3,9]	+1 [-3, +4]	0.6	0 [-2, +2]	0.9
50th	23 [20,29]	22 [18,27]	-1 [-7, +6]	0.8	-1 [-8, +5]	0.8
75th	64 [58,76]	60 [51,73]	-4 [-17, +8]	0.5	0 [-14, +14]	0.9

Percentile	Quintile 1 (n=325) (days, 95% CI)	Quintile 3 (n=285) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	5 [3,7]	5 [2,6]	0 [-3, +3]	0.9	0 [-2, +2]	0.9
50th	23 [20,29]	24 [18,31]	+1 [-8, +10]	0.8	0 [-7, +8]	0.9
75th	64 [58,76]	71 [58,89]	+7 [-12, +24]	0.4	+7 [-10, +23]	0.4

Percentile	Quintile 1 (n=325) (days, 95% CI)	Quintile 4 (n=225) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	5 [3,7]	9 [7,13]	+4 [0, +8]	0.04	+1 [-1, +4]	0.4
50th	23 [20,29]	32 [26,41]	+9 [-0.2, +18]	0.05	+3 [-8, +13]	0.6
75th	64 [58,76]	83 [66,98]	+19 [+2, +36]	0.02	+12 [-6, +31]	0.2

Percentile	Quintile 1 (n=325) (days, 95% CI)	Quintile 5 (n=193) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	5 [3,7]	7 [2,9]	+2 [-3, +7]	0.5	+1 [-1, +3]	0.3
50th	23 [20,29]	27 [20,36]	+4 [-5, +13]	0.4	-2 [-10, +5]	0.6
75th	64 [58,76]	83 [63, 104]	+19 [-2, +40]	0.08	+17 [-7,+40]	0.1

5. Type of CTYA Cancer

Diagnostic intervals of cancer subtypes varied substantially. Compared to leukaemia diagnoses, who had the shortest median diagnostic interval of 9 days [95% CI 6 to 15], lymphomas, sarcomas and abdominal tumours had statistically significantly longer median diagnostic intervals in unadjusted and adjusted models (Table 5-8). This difference was also observed at the 25th centile for these cancer subtypes and at the 75th centile, for lymphomas and sarcomas only.

Table 5-8 *The association between cancer types and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Baseline cancer type is leukaemia. Adjusted for age and sex. 95%CI = 95% confidence interval.*

Percentile	Leukaemias (n=358) (days, 95% CI)	Lymphomas (n=437) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	1 [1,1]	14 [12,16]	+13 [+11,+15]	<0.001	+12 [+9,+15]	<0.001
50th	9 [6,15]	38 [32,44]	+29 [+21,+36]	<0.001	+26 [+18,+34]	<0.001
75th	44 [33,56]	83 [75,90]	+39 [+26,+52]	<0.001	+37 [+23,+51]	<0.001

Percentile	Leukaemias (n=358) (days, 95% CI)	CNS tumours (n=290) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	1 [1,1]	5 [3,7]	+4 [+2,+6]	0.002	+3 [+1,+5]	0.1
50th	9 [6,15]	19 [14,24]	+10 [+3,+17]	0.003	+7 [+1,+13]	0.03
75th	44 [33,56]	58 [49,69]	+14 [-1,+29]	0.07	+12 [-5,28]	0.2

Percentile	Leukaemias (n=358) (days, 95% CI)	Sarcomas (n=165) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	1 [1,1]	18 [9,24]	+17 [+10,+24]	<0.001	+17 [+11,+23]	<0.001
50th	9 [6,15]	49 [36,60]	+40 [+27,+53]	<0.001	+42 [+29,+55]	<0.001
75th	44 [33,56]	103 [90,115]	+59 [40,78]	<0.001	+58 [+40, +75]	<0.001

Percentile	Leukaemias (n=358) (days, 95% CI)	Abdominal tumours (n=84) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	1 [1,1]	7 [4,12]	+6 [+3,+11]	<0.001	+6 [+2,+10]	0.003
50th	9 [6,15]	27 [15,38]	+18 [+8,+32]	0.001	+22 [+8,+34]	<0.001
75th	44 [33,56]	67 [47,95]	+23 [-4, +50]	0.1	+29 [-3,+61]	0.08

Percentile	Leukaemias (n=358) (days, 95% CI)	Gonadal germ cell tumours (n=25) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	1 [1,1]	3 [0,16]	+2 [-7, +11]	0.7	0 [-9, +9]	0.9
50th	9 [6,15]	20 [5,31]	+11 [-1, +23]	0.07	+5 [-8,+18]	0.4
75th	44 [33,56]	38 [21,120]	-7 [-39, +25]	0.6	-15 [-54, +24]	0.4

6. Aggressiveness of Cancer

Aggressive cancers were found to have statistically significant shorter diagnostic intervals compared to other cancers, following adjustment for age, sex and cancer type (Table 5-9).

Table 5-9 *The association between cancer biology (as defined by histological aggressiveness) and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, sex and cancer subtype. 95%CI = 95% confidence interval.*

Percentile	Aggressive (n=310) (days, 95% CI)	Other (n=1,049) (days, 95% CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	7 [5,9]	6 [4,7]	+1 [-2, +4]	0.5	-2 [-4, +1]	0.07
50th	24 [20,30]	25 [22,29]	-1 [-8, +6]	0.8	-9 [-14,-4]	0.001
75th	71 [60,82]	70 [63,79]	+1 [-14, +16]	0.9	-16 [-30, -2]	0.02

7. Symptoms/signs

The presence of red flag symptoms and signs were not statistically significantly associated with diagnostic intervals, following adjustment for age, sex, cancer type and cancer biology (Table 5-10).

Table 5-10 *The association between presence of red flag symptoms (as defined by NICE, Grace Kelly and HEADSMART guidance) and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, sex, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Percentile	Red flags (n=1,194) (days, 95% CI)	No red flags (n=165) (days, 95% CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	6 [5,7]	6 [3, 9]	0 [-3, +3]	1.0	+1 [-2, +3]	0.6
50th	26 [23, 28]	24 [14,31]	+2 [-6, +10]	0.4	+2 [-6, +10]	0.4
75th	71 [64, 79]	58 [45, 81]	+14 [-2, +30]	0.09	+13 [-2, +28]	0.09

For each individual cancer subtype, after adjusting for age and sex, only leukaemias, sarcomas and abdominal tumours were found to have symptom/signs that were associated with variations in diagnostic intervals (Table 5-11, only showing statistically significant results).

Table 5-11 *The association between symptoms/signs and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models for leukaemias, sarcomas and abdominal tumours. Adjusted for age and sex. 95%CI = 95% confidence interval*

Symptom/signs	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
Leukaemias				
25th Centile				
Anaemic/pale	-1 [-2, 0]	0.001	-1 [-2, 0]	0.009
Bruising	-1 [-2, -0]	0.001	-1 [-2, 0]	0.003
50th Centile				
Limb pain	+19 [+10, +28]	<0.001	+20 [+9, +31]	<0.001
Abdominal pain	+16 [+8, +24]	0.001	+15 [+7, +23]	0.001
75th Centile				
Limb pain	+47 [+21, +73]	0.001	+46 [+21, +71]	0.001
Abdominal pain	+28 [+11, +45]	0.002	+30 [+7, +53]	0.01
Sarcomas				
50th Centile				
Gait abnormalities	-28 [-47, -8]	0.005	-25 [-41, -7]	0.005
Abdominal Tumours				
25th Centile				
Constipation	+33 [+29, +37]	0.001	+33 [+27, +39]	0.001

8. Relevant test requests

Cancer patients who had a blood test requested had longer 25th centile and median diagnostic intervals compared to those who had no blood test requests, in adjusted and unadjusted analyses (Table 5-12). Those who had imaging requests had similar diagnostic intervals to those who did not (Table 5-13).

Table 5-12 *The association between blood test request and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, sex, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Percentile	Test Requested (n=511) (days, 95% CI)	Test not requested (n=848) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	10 [8,14]	4 [3,6]	+6 [+2,+10]	0.002	+3 [+1,+5]	0.01
50th	34 [28,41]	21 [17,24]	+13 [+6,+20]	0.001	+8 [+1,+15]	0.02
75th	79 [68,85]	64 [57,73]	+16 [+5,+27]	0.004	+6 [-5,+17]	0.3

Table 5-13 *The association between imaging test request and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, sex, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Percentile	Test Requested (n=74) (days, 95% CI)	Test not requested (n=1,285) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	6 [3,14]	6 [5,7]	0 [-8, +8]	0.2	+7 [-11, +25]	0.4
50th	35 [21,57]	25 [22,27]	+13 [-8,+34]	0.9	+2 [-4,+8]	0.5
75th	72 [52,125]	69 [63,78]	+1 [-33, +36]	0.9	+16 [-16, +48]	0.3

9. Region

There were no statistically significant regional differences in diagnostic intervals (Table 5-14).

Table 5-14 *The association between regions and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Baseline region selected was South of England. Adjusted for age, sex, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval*

Percentile	South (n=783) (days, 95% CI)	North (n=395) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	6 [6,8]	5 [3,8]	-1 [-4, +2]	0.4	+1 [-1, +3]	0.4
50th	24 [21,28]	26 [21,30]	+2 [-4,+8]	0.5	0 [-7, +7]	1.0
75th	71 [63, 82]	67 [58,87]	-3 [-18, +12]	0.7	-5 [-15, +7]	0.4

Percentile	South (n=783) (days, 95% CI)	Midlands (n=181) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	6 [6,8]	6 [3,10]	0 [-4, +4]	1.0	+1 [-1, +2]	0.2
50th	24 [21,28]	28 [20,42]	+4 [-6,+14]	0.4	+1 [-8, +10]	0.8
75th	71 [63, 82]	68 [57,80]	-4 [-22, +15]	0.7	+1 [-14, +16]	0.9

10. COVID-19 Pandemic

The baseline characteristics of the COVID-19 cohort can be found in the Supplementary Appendix (Chapter 9, Supplementary Table 9-17). Compared to the pre-pandemic period, the first and second wave of the pandemic was observed to have similar diagnostic intervals (Table 5-15). Diagnosis of cancer during the pre-vaccination role out time-period between the first and second wave (Period 3), however, was found to be statistically significantly faster compared to the pre-pandemic period.

Table 5-15 *The association between the three COVID-19 pandemic periods and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Baseline period was selected pre-pandemic period (2017-2019). Adjusted for age, sex, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Percentile	Pre-pandemic (Jan-May 2017-2019, n=249) (days, 95% CI)	Period 1 (Jan-May 2020) (n=49) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	13 [10,17]	8 [4,15]	-5 [-10,0]	0.05	-4 [-9, +2]	0.2
50th	35 [28,53]	33 [13,49]	-2 [-14, +11]	0.8	-2 [-20, +14]	0.7
75th	85 [66,101]	76 [49,99]	-9 [-40, +22]	0.6	-14 [-41, +13]	0.3

Percentile	Pre-pandemic (2017-2019, n=249) (days, 95% CI)	Period 2 (Jan-May 2021) (n=63) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	14 [10,18]	15 [7,22]	+1 [-8, +10]	0.8	+1 [-5, +7]	0.8
50th	36 [29,46]	34 [23, 55]	-2 [-21, +17]	0.8	-2 [-15, +10]	0.7
75th	82 [62,102]	91 [57,122]	+4 [-25, +32]	0.8	+4 [-26, +34]	0.8

Percentile	Pre-pandemic (2017-2019, n=309) (days, 95% CI)	Period 3 (May-December 2020) (n=124) (95%CI)	Unadjusted (days, 95% CI)	p	Adjusted (days, 95%CI)	p
25th	15 [12,20]	6 [4,9]	-9 [-14, -4]	<0.001	-6 [-11,-1]	0.01
50th	39 [32,44]	28 [18,41]	-10 [-21, +1]	0.07	-12 [-22, -2]	0.01
75th	96 [78,112]	63 [49,96]	-24 [-66, -3]	0.04	-28 [-61, +5]	0.09

5.4.3 Outcome 3: Route-to-Diagnosis

There was a total of 2,814 (64% of full cohort) young people with a recorded route-to-diagnosis. The exposures-of-interest in those who had an emergency presentation, non-emergency presentation and no recorded route-to-diagnosis are summarised in Table 5-16. There were important differences across those with and without emergency presentations.

First, median age at cohort entry was younger in those who presented as an emergency compared to those who had not (p=0.01). Second, a higher proportion of leukaemias and CNS tumours had emergency presentations, whilst lymphomas and gonadal germ cell tumours had higher proportions as non-emergency presentations. There was a very small proportion of young people with selected comorbidities (n<5) and accordingly, was not included in further analyses.

Table 5-16 *Exposures-of-interest in those with and emergency, non-emergency and those without a recorded presentation. All percentages are column percentages (%). Age is in integers and rounded up to the nearest integer. IQR= interquartile range. Note: p-value for analyses between those with and without a presentation (i.e., column 1&2 vs column 3).*

Characteristics	Emergency presentation (n=1,292)	Non-emergency presentation (n=1,522)	Not recorded (n=1,575)	P-value
Age at cohort entry				
Median ± IQR (years)	2 [1,9]	7 [2,14]	8 [2,16]	0.3
Sex				
Male	751 (58.1)	905 (59.5)	951 (60.4)	0.4
Female	541 (41.9)	617 (40.5)	624 (39.6)	
Ethnicity				
White	362 (28.0)	475 (31.2)	308 (19.6)	0.7
Other	226 (17.5)	238 (15.6)	164 (10.4)	
Missing	704 (54.5)	809 (53.2)	1,103 (70.0)	
Townsend Deprivation Score				
1 (most affluent)	321 (24.9)	370 (24.3)	390 (24.8)	0.6
2	279 (21.6)	341 (22.4)	330 (20.9)	
3	256 (19.8)	292 (19.2)	335 (21.3)	
4	221 (17.1)	288 (18.9)	273 (17.3)	
5 (most deprived)	204 (15.8)	219 (14.4)	229 (14.5)	
Missing	11 (0.9)	12 (0.8)	18 (1.1)	
Comorbidities				
Present	<5	<5	<5	0.1
Absent	-	-	-	
Cancer groups				
Leukaemia	535 (41.4)	203 (13.3)	389 (24.7)	0.008
Lymphoma	189 (14.6)	433 (28.5)	297 (18.9)	
CNS tumours	310 (24.0)	280 (18.4)	343 (21.8)	
Sarcomas	102 (7.9)	307 (20.2)	228 (14.5)	

Abdominal tumours	97 (7.5)	118 (7.8)	154 (9.8)	
Gonadal germ cell tumours	59 (4.6)	181 (11.9)	164 (10.4)	
Cancer Biology				
Aggressive	319 (24.7)	333 (21.9)	320 (20.3)	0.03
Non-aggressive	973 (75.3)	1189 (78.1)	1255 (79.7)	
Test Request				
Bloods requested	316 (24.5)	428 (28.1)	311 (19.8)	<0.001
Bloods not requested	976 (75.5)	1094 (71.9)	1264 (80.2)	
Imaging requested	88 (6.8)	69 (4.5)	37 (2.4)	<0.001
Imaging not requested	1204 (93.2)	1453 (95.5)	1538 (97.5)	
Region				
North	445 (34.4)	406 (26.7)	428 (27.2)	0.01
Midlands	157 (12.2)	227 (14.9)	192 (12.2)	
South	690 (53.4)	889 (58.4)	955 (60.6)	

5.4.4 The association between exposures and route-to-diagnosis

The association between exposures-of-interest and route-to-diagnosis was explored using both complete case analyses and analyses following multiple imputation, which yielded comparable results. Accordingly, the complete-case analysis results are presented below (Figures 5-1 & 5-2), whilst the results following multiple imputation can be found in the Supplementary Appendix (Supplementary Appendix Figures 9-21 & 9-22).

1. *Age-at-diagnosis*

TYA (15-24 year olds) were less likely to present via an emergency presentation compared to children (0-14 year olds) in adjusted logistic regression models.

2. *Sex*

Emergency presentations were similar between males and females in adjusted models.

3. *Ethnicity*

Emergency presentations were similar between White and other ethnic minority groups in adjusted models.

4. *Deprivation level*

There were no differences in the odds between emergency presentations across different deprivation levels in adjusted models.

5. *Type of CTYA Cancer*

Compared to leukaemias, all other cancer subtypes were less likely to present as an emergency in adjusted models.

6. *Cancer Biology/Aggressiveness of Cancer*

Cancers with aggressive histology were found to be more likely to present as an emergency in adjusted models.

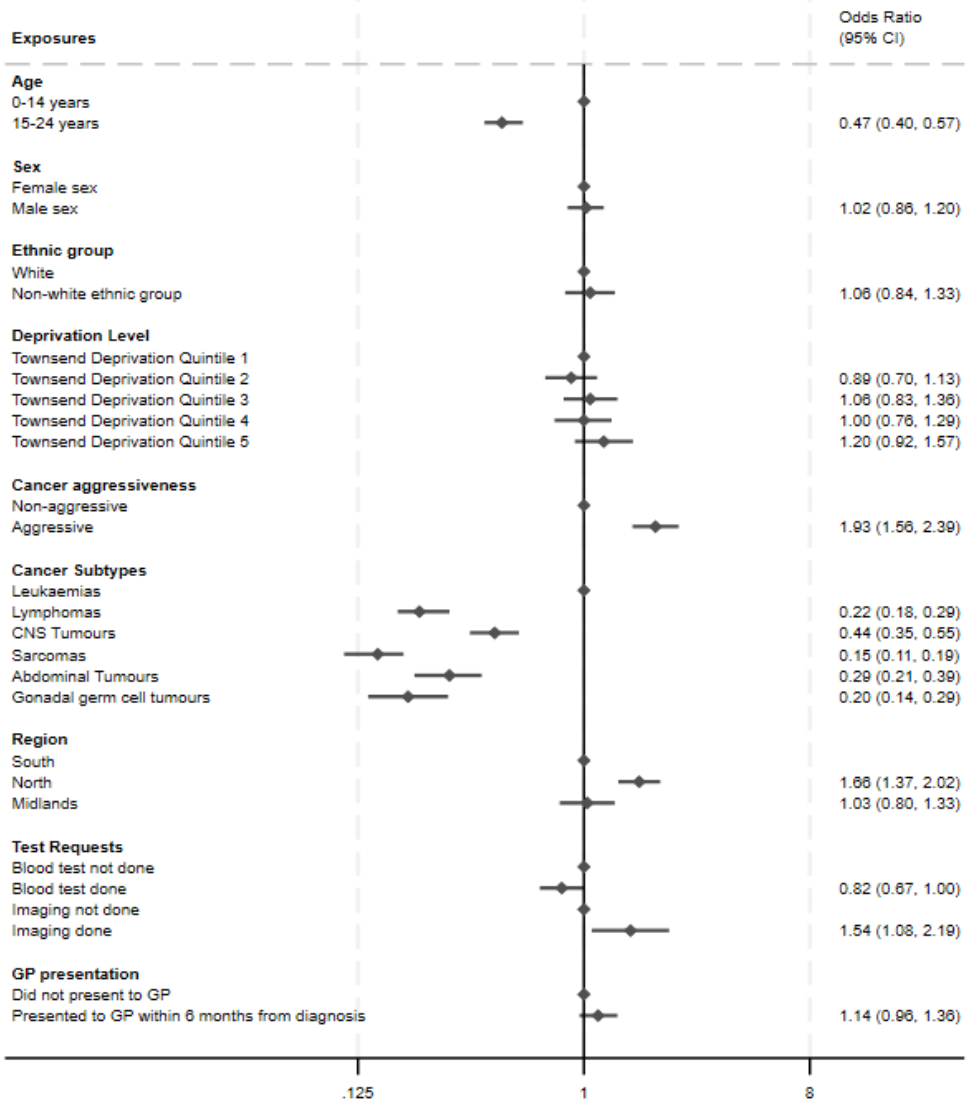
7. *Relevant test requests*

There were no statistically significant differences in odds of emergency presentation between those who had a blood or imaging test requested and those who did not (cut-off p-value set at 0.01). Similarly, there was no difference in the odds of emergency presentation between young people who presented to their GP within 6 months before diagnosis and those who did not.

8. *Region*

Compared to the South, CTYA cancer presentations in the North were more likely to be as emergencies in adjusted models.

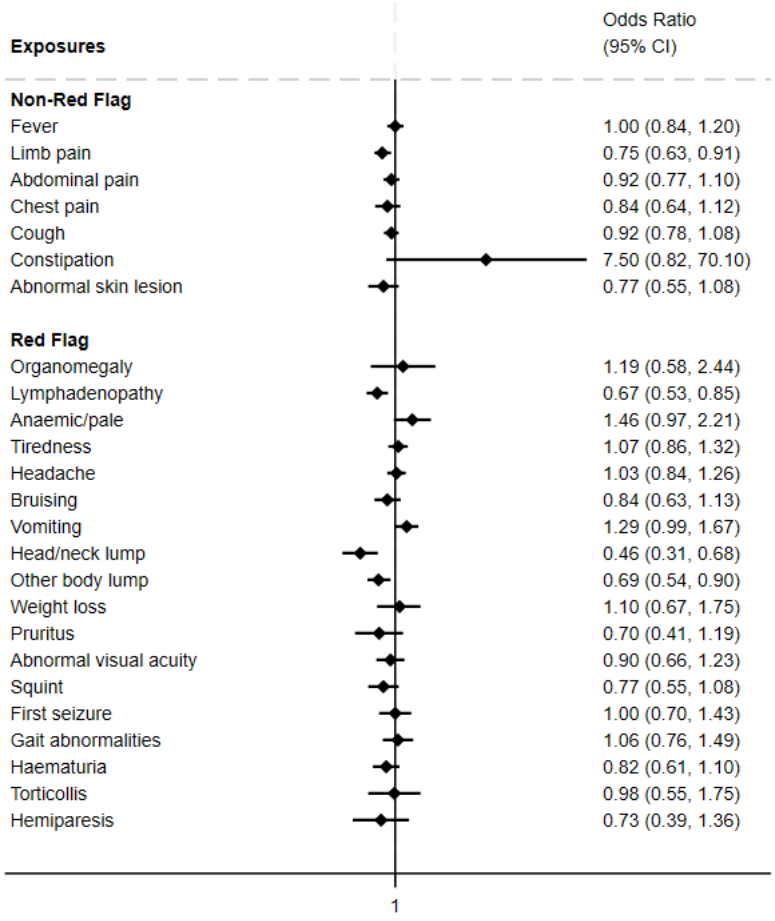
Figure 5-1 *Forest plot demonstrating adjusted odds of emergency presentation compared to non-emergency presentation in CTYA cancers for each factor/exposure. 95% CI = 95% confidence interval.*



9. Symptoms/signs

When the association between individual symptom/signs and emergency presentations were explored following adjustment for age, sex, cancer aggressiveness and cancer subtype, lymphadenopathy [adjusted (aOR) 0.67 (95%CI 0.53-0.85)]; limb pain [aOR 0.75 (95%CI 0.63-0.91)]; head/neck lump [aOR 0.46 (95%CI 0.31-0.68)]; and other lump on body [aOR 0.67 (95%CI 0.54-0.90)] were found to be statistically significantly less likely to present as emergencies (Figure 5-2). There were no symptom/signs associated with higher odds of emergency presentations.

Figure 5-2 *Forest plot demonstrating adjusted odds of emergency presentation compared to non-emergency presentation in CTYA cancers for symptoms only.*



5.5. Discussion

5.5.1 Summary of Principal Findings

This study comprehensively evaluated factors that may influence diagnostic pathways and highlights the complexity of these pathways in CTYA cancers. Several factors, including person-level (i.e., age-at-diagnosis), cancer-specific (i.e., cancer subtype, presenting features and cancer biology) and health-system specific (i.e., presence of investigative test requests as well as regional differences) factors were found to be associated with variations in diagnostic pathways.

The majority of factors that influence diagnostic pathways were common for the two outcomes, systemic diagnostic intervals and route-to-diagnosis. For example, emergency presentations and shorter diagnostic intervals were found to be influenced by a leukaemia diagnosis and aggressive cancer, whilst non-emergency presentations and long diagnostic intervals were associated with a TYA cancer diagnosis. These outcomes also had key differences in associated factors, where presence of a blood test request was found to be associated with longer diagnostic intervals but not with route-to-diagnosis; whilst the North of England showed higher emergency presentations compared to the South of England but this was not seen for diagnostic intervals.

In this study, it was also possible to explore systemic diagnostic intervals during different stages of the COVID-19 pandemic and findings suggests that time-to-diagnosis was faster in the period following the first wave of the pandemic.

5.5.2 Findings in Context

There were several factors found to be associated with longer diagnostic intervals in this study. This included older age-at-diagnosis, in unadjusted analyses, and cancer subtype, which have both been consistently shown across previous studies [21-23, 143]. Similar to previous

findings, this study also identified that compared to leukaemias, other cancer subtypes have longer diagnostic intervals [21, 23, 139].

In this study, it was also possible to investigate the influence of presenting symptom/signs on diagnostic intervals for each cancer subtype. Certain symptoms, such as limb and abdominal pain in leukaemia patients as well as constipation in patients with abdominal tumours, were found to lengthen time-to-diagnosis, whilst anaemia/pallor and bruising in leukaemia patients were found to shorten it. This is consistent with previous reports that young people presenting with common symptoms, such as pain, were more likely to have a delayed diagnosis compared to those that present with rarer symptoms [23].

Additionally, this study identified an association between longer diagnostic intervals and the presence of a blood test request, which is also in line with adult cancer studies, where longer diagnostic intervals were found in those who had blood test requests prior to diagnosis [41].

Finally, cancers classified as “histologically aggressive” were found to have shorter diagnostic intervals, reflecting previous reports that rapidly advancing disease can have shorter lag times [144].

There were several factors also found to be associated with emergency presentations, including 1) cancers diagnosed in childhood compared to during teenage and young adult years, 2) cancers categorised as “histologically aggressive” compared to those that are not, and 3) cancers diagnosed as leukaemias. Interestingly, although longer diagnostic intervals are thought to play a role in increased risk of emergency presentations, these factors identified in this study did not show this; in fact they were associated with shorter diagnostic intervals. This highlights the complexity of cancer presentations and may reflect a category of young people with rapidly advancing disease for which emergency presentations are expected and unlikely to be modifiable [138].

Importantly, however, the finding that cancers in the North of England compared to the South of England being identified to have a higher risk of emergency presentations, which is also seen in adult cancers [145], is likely to represent a potentially modifiable factor and may be linked to diagnostic delays. This study was unable to provide evidence for this, as diagnostic intervals were comparable across different regions, however, this may be due to the small sample sizes per region available for diagnostic intervals.

This study was also able to comprehensively explore diagnostic intervals during different stages of the COVID-19 pandemic and interestingly identified shorter times-to-diagnosis in young people presenting in between the two biggest waves of the COVID-19 pandemic in the UK. Related work I had carried out had also shown that the number of incident cancer diagnoses were significantly lower in the first wave of the pandemic and this subsequent shorter interval may represent young people presenting post lock-down and with more advanced disease [123].

5.5.3 Strengths and Limitations

The two biggest strengths of this study are 1) the use of a nationally-representative, population-based cohort of CTYA cancers to draw conclusions from and 2) the availability of linked data which has allowed to explore a significant proportion of the patient's diagnostic journey, from primary care to secondary care and ultimately to diagnosis.

However, there are some important limitations in this study. First, diagnostic intervals were constructed using symptoms/signs identified to be relevant per cancer subtype. Although this approach has been used in primary care-based cancer diagnostic interval studies [41], one cannot be certain, without asking the patient, what the actual presenting symptoms for their cancer were. Second, there is the possibility of recording bias, where the clinician reviewing the patient may either be more likely to record symptoms in someone they suspect to have cancer (although there was no evidence for this in Chapter 3) or less likely to record

symptoms in those with atypical symptoms. Recording bias may either overestimate or underestimate the association between the exposure and diagnostic interval. The third limitation is the presence of missing data in this study. For both outcomes, the missingness was deemed to be missing at random given the method of collection of EHR and thus was analysed through both complete case and multiple imputation approaches, and reassuringly, both showed comparable results. The fourth limitation arises due to small sample sizes; even though this is a population-based cohort, given the rarity of CTYA cancers, exploring diagnostic pathways individually across important subgroups, such as cancer subtype, was not possible. Therefore, subgroups were accounted for in multivariable regression analyses. Finally, given this study relies on routinely collected electronic healthcare data, it was not possible to calculate the time interval from symptom-onset to first presentation to healthcare (i.e., the patient interval), which is also likely to provide important information on health-seeking behaviours and subsequent link to diagnostic delays.

5.5.4 Implications and Future Directions

Findings of this study have important implications for public health, clinical and patient decision-making.

There are four main public health implications that arise from this study. First, findings provide further evidence for older age and cancer subtype as factors that can be specifically targeted by public health interventions (e.g., awareness campaigns or risk stratification specific to these factors) to aid shorter diagnostic intervals. Second, this study also identified a novel association between region and emergency CTYA cancer presentations. This factor could also be useful for targeted public health interventions but importantly highlights the need to develop more tailored diagnostic pathways for CTYA cancers across the country. Third, findings relating to the COVID-19 pandemic in this study also provide useful evidence for future pandemic preparedness. Finally, findings of this study demonstrate the complex and

heterogeneous relationship between CTYA cancer and diagnostic pathways, with a potential central role for cancer biology (assessed through histological aggressiveness) in this relationship. These findings highlight the further need for future studies to explore this relationship between cancer biology and diagnostic pathways in CTYA cancer with more granularity.

Findings of this study have important implications for clinicians and patients too. Knowledge that certain cancer subtypes, such as lymphomas and sarcomas, or certain presenting symptoms, such as abdominal pain and constipation, have longer diagnostic intervals, for example, may prompt clinicians to think of cancer earlier and support their clinical decision-making. Furthermore, knowledge of these factors may also help raise awareness of cancer presentations in young people, particularly around identifying non-specific symptoms such as limb pain. Additionally, other factors that may prolong diagnostic intervals, such as a test request, may result in careful consideration of its clinical need and aid more targeted requests before referral.

Finally, factors identified to influence diagnostic pathways may ultimately also be associated with long-term outcomes. Accordingly, this DPhil will investigate this in Chapters 6 and 7.

Code from Chapter

https://github.com/DefOx2020/Thesis/blob/06e6b1c26357ad03b73a6fc065ef8af9040ef1b4/Chapter%205

Publication from Chapter

Saatci, D., Oke, J., Harnden, A., & Hippisley-Cox, J. (2022). *Childhood, teenage and young adult cancer diagnosis during the first wave of the COVID-19 pandemic: a population-based observational cohort study in England.* Archives of disease in childhood, 107(8), 740–746. https://doi.org/10.1136/archdischild-2021-322644

Chapter 6 The association between diagnostic pathways and 1- and 5-year mortality in childhood, teenage and young adult cancers

6.1. Background to chapter

Despite vast treatment-related improvements, cancer remains one of the commonest causes of mortality-by-disease among young people globally and ranks as the 6th leading cause of cancer burden worldwide [8]. This global trend is strongly mirrored in the UK, where 10-year survival has now increased to approximately 80%, yet mortality due to cancer remains the leading cause of death in young people [9]. Furthermore, cancer survival in the UK is one of the lowest in Europe [5]. Although the underlying reasons for this is likely to be multifactorial and complex, factors such as diagnostic/treatment delay and emergency presentations may play a contributory role, as has been shown in adult cancers [12, 138].

The association between diagnostic pathways and cancer mortality in young people is not well understood. Two systematic reviews summarising the evidence for the link between diagnostic intervals and cancer mortality conclude that the results are conflicting. One of the reviews, which reviewed 98 studies, found that there was some evidence for increased mortality risk following diagnostic delay in cancer subtypes such as leukaemias and nephroblastomas, whilst no or paradoxically, lower mortality risk in CNS tumours and sarcomas [15]. The second review, which reviewed 32 studies based in Europe and the United States, found similar paradoxical findings [146]. However individual studies in these systematic reviews were substantially limited by 1) small sample sizes, 2) non-population based designs subject to selection and recall biases, heterogeneity in cancer subtype and age, and 3) did not account for statistical challenges such as the “waiting-time paradox” [12]

commonly seen in this field of research or 4) did not account for all relevant confounding factors (i.e., those explored in Chapter 5 in this thesis).

6.2. Aims and Objectives

The overall aim of this Chapter is to explore whether CTYA cancer diagnostic pathways affect mortality outcomes. Specific objectives are:

1. To explore the association between systemic diagnostic intervals and 1- and 5- year all-cause mortality, following adjustment for confounders identified in Chapter 5.

2. To explore the association between emergency presentations and 1- and 5- year all-cause mortality, following adjustment for confounders identified in Chapter 5.

6.3. Methods

6.3.1 Study Population

For 1-year mortality, this was the full cohort of 4,389 young people with a cancer diagnosis from a cohort of 5,238,743 young people between the ages of 0-24 years, generated from QResearch Database in Chapter 2. For 5-year mortality, to ensure all young people have a minimum of 5 years follow-up, the cohort was restricted to an exit date of 31st January 2017, as there was follow up available only until 31st January 2022. Accordingly, this was a cohort of 3,883,032 young people between the ages of 0-24 years and 3,214 cancers. Using the same ICCC-3 framework described in Chapter 2, the subtypes were: 1) Leukaemias, myeloproliferative and myelodysplastic diseases (n=813), 2) lymphomas and reticuloendothelial neoplasms (n=656), 3) central nervous system and intraspinal tumours (n=722), 4) soft tissue and bone sarcomas (n=462), 5) abdominal tumours (renal tumours, neuroblastomas, hepatic tumours) (n=269) and 6) gonadal germ cell tumours (n=292). The incidence rates of each cancer subtype in this restricted population were similar to the full cohort and national estimates (Chapter 9, Supplementary Table 23). The mortality rate for 1-

9 year olds in this cohort was 8.9 [95% CI 7.7-10.0] per 100,000, which was comparable to national estimates provided for 1-9 year olds (10.0 per 100,000) [9].

The study population entry was the date at which the young person received their cancer diagnosis. The study population exit for the 1-year mortality study was the earliest of study end date (31st Jan 2022) or 1 year (365.25 days) from diagnosis. The study population exit for the 5-year mortality study was earliest of study end date (31st Jan 2022) or 5 years (1826.25 days) from diagnosis.

Exclusion criteria included any young person who had a death recorded before or on their date of diagnosis. Any young person who had a death recorded after the study end date was censored.

6.3.2 Exposures of Interest

The two primary exposures-of-interest are systemic diagnostic intervals as defined in Chapter 5 (i.e., time from first presentation to primary care to date of diagnosis, with maximum of 6 months) and emergency presentation at diagnosis as defined in Chapter 5, using the National Cancer Registry records of route-to-diagnosis. Presentations and diagnostic intervals were only recorded if the date of presentation and dates of symptoms/signs used to define the diagnostic interval were prior to the cancer diagnosis (i.e., cohort entry date).

Diagnostic intervals were explored in three ways detailed below. Several ways were explored because of previous findings of the complex and non-linear relationship between diagnostic intervals and CTYA cancer mortality as well as reports of the “waiting-time paradox” bias [12].

1. First, diagnostic intervals derived from Chapter 5 were directly included in analyses as a continuous variable (referred in the Chapter as Approach 1).

2. Second, they were explored using restricted cubic splines (i.e., natural splines) to model the non-linear nature of the relationship (referred in the Chapter as Approach

2) [13]. The number of knots was specifically based on recommended percentiles by Harrell et al., (2001) [147].

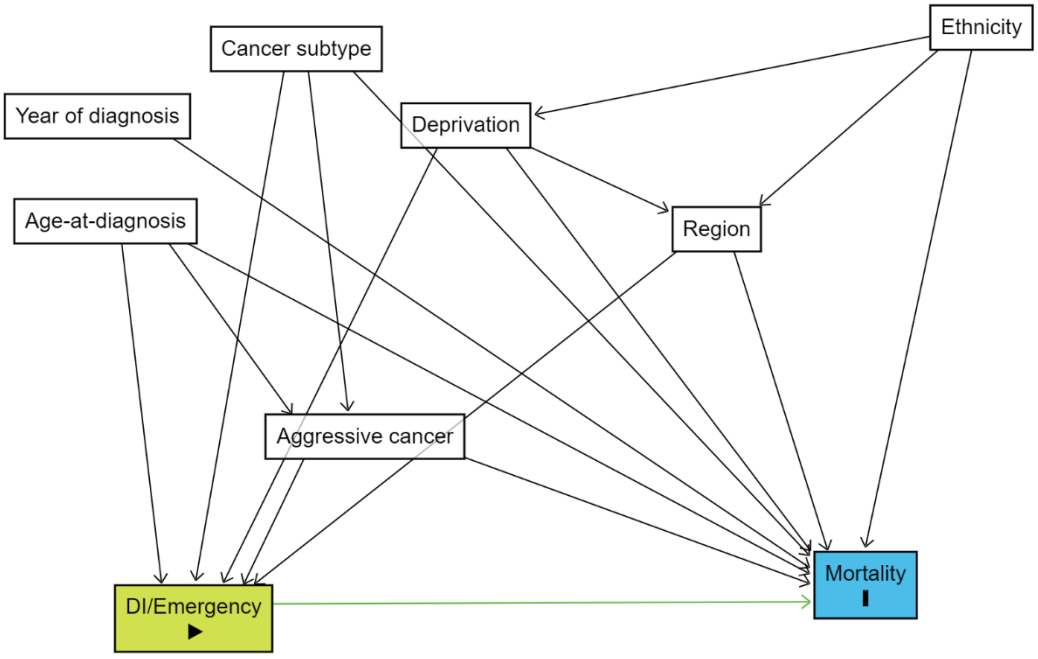
3. Third, an exploratory approach was taken where diagnostic intervals were stratified by 0-25th percentiles, 25th-50th percentiles, and 50th-75th percentiles. The association between the interval (still a continuous variable) and mortality within each length was explored separately (referred in the Chapter as Approach 3).

Another approach to explore the relationship between mortality and diagnostic intervals identified in the literature was the removal of red flag symptoms [12], however, an a priori decision not to carry this out was taken, given the non-specific symptoms/signs associated with cancers in this age group as well as findings from Chapter 5 suggesting no significant association between red flags and the diagnostic interval.

6.3.3 Potential Confounders

In Chapter 5, several factors were found to be associated with diagnostic intervals and emergency presentations. These have been summarised in the direct acyclic graph in Figure 6-1 below. Furthermore, as detailed in Figure 6-1, sociodemographic factors such as ethnicity, deprivation level have previously been shown to be associated with mortality in this age group [9]. Collectively, all listed factors in Figure 6-1 may represent confounders and were therefore accounted for within analyses. Age-at-diagnosis was defined into two categories (childhood, 0-14 years and TYA, 15-24 years). Year of diagnosis was defined into 2 categories (1998-2008 and >2008). Cancer subtype was grouped into two categories (blood and solid cancers). Deprivation level was grouped into 3 categories (1- Townsend I & II, 2- Townsend III, 3- Townsend IV & V). Region was grouped into three (North, South and Midlands). Ethnicity was grouped into two (White and non-White ethnic groups). Categories were broad to generate larger samples per category for comparisons to achieve sufficient power.

Figure 6-1 *Direct acyclic graph demonstrating causal pathways between diagnostic pathway (DI/Emergency) and mortality, with biasing paths through age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressive cancer, deprivation level, region and ethnic group. DI= diagnostic interval.*



6.3.4 Outcome of interest

The outcomes-of-interest were 1-year and 5-year all-cause mortality. This was identified through both ONS death records and National Cancer Registry records. All-cause mortality was chosen as it avoids limitations around ascertainment of cause-of-death [138].

6.3.5 Statistical Analysis

For descriptive analyses, continuous variables were presented as means (medians if non-normal) and categorical variables were presented in frequencies and comparisons between cancers and the general population were carried out using t-tests (Mann-Whitney U if non-normal) and $\chi 2$ tests, respectively.

To explore the association between exposures-of-interest and all-cause mortality outcomes, multivariable Cox proportional hazard regression modelling was used. Cox models summarise how an exposure affects the hazard (i.e., event rate) over a set follow-up time through the

hazard ratio. This regression approach makes the assumption that the hazards are proportional over time (i.e, the effect of the exposure is constant over time). Whether this assumption was met in this study was tested using the Schoenfeld residuals. The models accounted for the correlation because of clustering of children within practices through a robust variance estimator.

The relationship between 5-year mortality and the diagnostic interval was visually explored through calculating the predicted probability of 5-year mortality ($Predicted\ probability = 1 - baseline\ survival\ function^{\exp(linear\ predictor)}$) and plotting it against the diagnostic interval in days, where the baseline survival function and linear predictor were calculated following regression analysis.

6.3.6 Handling Missing Data

There was missing data in four exposures-of-interest: diagnostic interval, emergency presentation, ethnicity and Townsend Deprivation Score. Exposures with missing data were deemed to be missing under the missing at random assumption (i.e., any systematic differences between the missing and observed data can be explained through associations within the observed data). This decision was made because of the method in which these exposures are collected (i.e., through routinely collected electronic healthcare records) [142]. Accordingly, both complete case analyses and analyses following multiple imputation were carried out. For complete case analyses, ethnicity was excluded from analyses due to the significant proportion of missingness.

Restricted cubic splines generated in the complete case cohort were included as variables to be imputed as this has been shown to perform better than passive imputation [148].

Multiple imputation with chained equations was used to replace missing data. 50 imputations were performed, given % missingness of ethnicity and diagnostic intervals [91]. Imputation models included all exposure variables, and the primary outcome variables;

statistical models were developed on each of the 50 imputed data sets and the estimates pooled using the Rubin rules [91]. Any continuous variable was assessed for normality [91]. As the diagnostic interval did not follow a normal distribution and was right-skewed, it was transformed using its natural logarithm prior to imputation analyses.

6.3.7 Sample Size Calculations

Stata's *stpower log-rank test* comparing two survival rates method was carry out a priori for sample size calculations [149]. The survival probability in the "control" population was set at 0.95 (i.e, 95%) for 1-year survivors and 0.8 (i.e., 80%) for 5-year survivors. The only studies which specifically explore the association between diagnostic interval/cancer route and mortality are in adults and they report effect sizes between 1.5 and 3 [16]. Accordingly for the lowest reported effect size of 1.5, 80% power and a significant threshold of 0.05, the total sample size required for this study was 1,045. As the cohorts have 4,389 and 3,241 young people for 1-year and 5-year mortality studies, respectively, the study was deemed to have sufficient sample size to detect this effect size.

6.4. Results

In the cohort of 4,389 young people with cancer, 11 met the exclusion criteria and there were 163 deaths recorded within 1 year from diagnosis (3.7%). Within the restricted cohort of 3,241 young people with cancer, 34 met the exclusion criteria and there were 309 deaths within 5 years from diagnosis (9.5%). The baseline characteristics of those with and without a mortality outcome are detailed in Table 6-1.

Table 6-1 *Baseline characteristics of young people in the cohorts (cohort 1 for 1-year follow-up, cohort 2 for 5-year follow-up) who had or did not have death recorded. All percentages are column percentages (%). Age is in integers and rounded up to the nearest integer. IQR= interquartile range.*

Exposure of interest	**Cohort 1 (n=4,378)**		**p-value**	**Cohort 2 (n=3,207)**		**P-value**
	Outcome recorded (n=163)	**Outcome not recorded (n=4,215)**		**Outcome recorded (n=309) (%)**	**Outcome not recorded (n=2,898) (%)**	
Age at cohort entry						
Median ± IQR (years)	8 [2,14]	5 [1,13]	0.2	9 [3,14]	6 [1,14]	0.2
Sex						
Male	97 (59.5)	2,503 (59.4)	0.9	191 (61.8)	1,739 (60.0)	0.01
Female	66 (40.5)	1,712 (40.6)		118 (38.2)	1,159 (40.0)	
Ethnicity						
White	14 (8.6)	1,130 (26.8)	0.004	30 (9.7)	735 (25.4)	<0.001
Other	20 (12.3)	607 (14.4)		41 (13.3)	388 (13.4)	
Missing	129 (79.1)	2,478 (58.8)		238 (77.0)	1,775 (61.2)	
Townsend Deprivation Score						
1 (most affluent)	35 (21.5)	1,042 (24.7)	0.5	64 (20.7)	732 (25.3)	0.2
2	30 (18.4)	919 (21.8)		61 (19.7)	643 (22.2)	
3	35 (21.5)	846 (20.1)		67 (21.7)	578 (20.0)	
4	35 (21.5)	743 (17.6)		63 (20.4)	491 (16.9)	
5 (most deprived)	28 (17.2)	624 (14.8)		53 (17.2)	434 (15.0)	
Missing	-	41 (1.0)		<5	20 (0.6)	
Cancer groups						
Leukaemia	48 (29.5)	1,075 (25.5)	<0.001	83 (26.9)	729 (25.2)	<0.001
Lymphoma	19 (11.7)	896 (21.3)		31 (10.0)	622 (21.5)	
CNS tumours	57 (35.0)	874 (20.7)		87 (28.2)	633 (21.8)	
Sarcomas	25 (15.3)	612 (14.5)		78 (25.2)	384 (13.3)	
Abdominal tumours	12 (7.4)	356 (8.5)		24 (7.8)	244 (8.4)	
Gonadal germ cell tumours	<5	402 (9.5)		6 (1.9)	286 (9.9)	
Cancer Biology						
Aggressive	62 (38.0)	908 (21.5)	<0.001	176 (57.0)	573 (19.8)	<0.001
Non-aggressive	101 (62.0)	3,307 (78.4)		133 (43.0)	2,325 (80.2)	
Region						
North	43 (26.4)	1,233 (29.3)	0.5	81 (26.2)	889 (30.4)	0.2

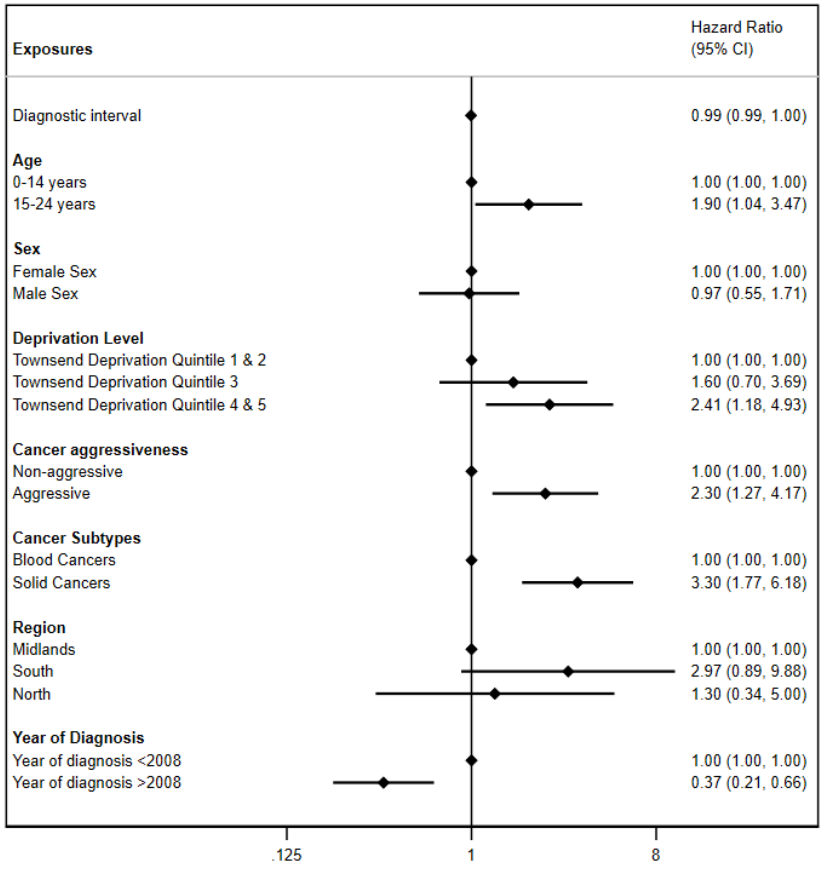
Midlands	19 (11.7)	555 (13.2)		35 (11.3)	368 (12.7)	
South	101 (61.9)	2,427 (57.6)		193 (62.5)	1,650 (56.9)	
Diagnostic intervals						
0-7 days	16 (9.8)	366 (8.7)	0.6	24 (7.8)	242 (8.4)	<0.001
7-25 days	12 (7.4)	287 (6.8)		22 (7.1)	200 (6.9)	
25-67 days	11 (6.8)	318 (7.5)		19 (6.2)	218 (7.5)	
>58 days	8 (4.9)	334 (7.9)		24 (7.8)	211 (7.3)	
Missing	116 (71.7)	2,906 (68.9)		220 (71.2)	2,024 (69.8)	
Route-to-diagnosis						
Emergency	66 (40.5)	1,223 (29.0)	<0.001	96 (31.1)	834 (28.8)	0.001
Non-emergency	28 (17.2)	1,490 (35.4)		67 (21.7)	1,002 (34.6)	
Missing	69 (42.3)	1,502 (35.6)		146 (47.2)	1,062 (37.7)	
Diagnostic time period						
1998-2008	93 (57.1)	1,407 (33.4)	<0.001	199 (64.4)	1,244 (42.9)	0.001
>2008	70 (42.9)	2,808 (66.6)		110 (35.6)	1,654 (57.1)	

6.4.1 Exposure 1: Diagnostic Intervals

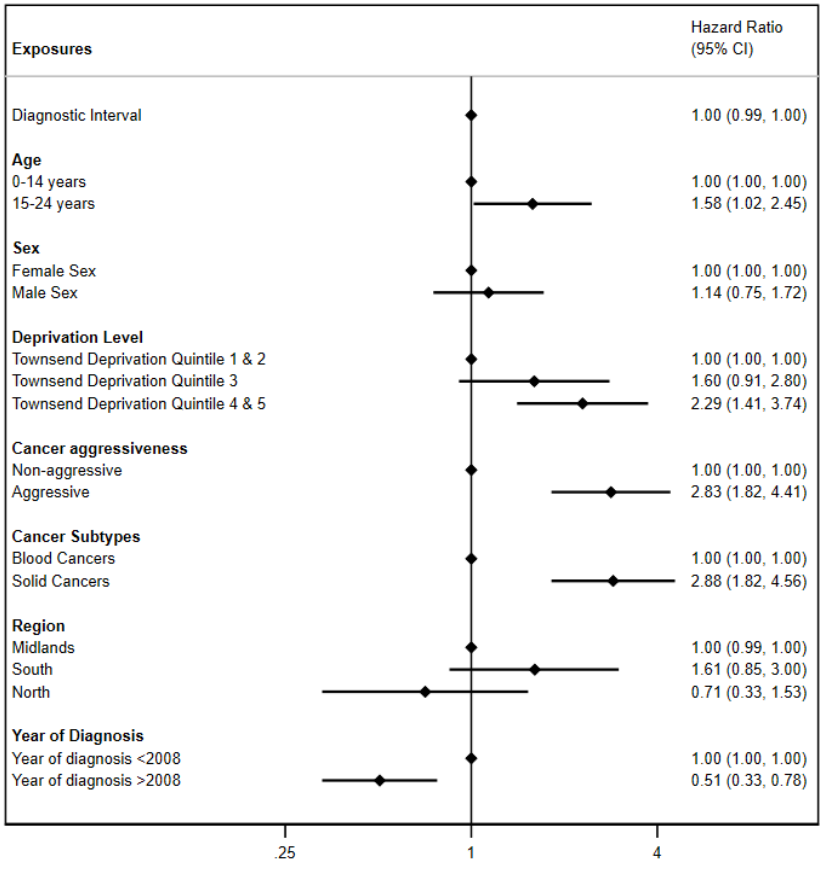
Using Approach 1 (i.e., no modifications to diagnostic interval), there was no significant association between the diagnostic interval and risk of mortality at 1-year or 5-years following adjustment for sociodemographic (sex, deprivation level) and cancer-specific factors (age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressiveness) in complete-case analyses (Figure 6-2).

Figure 6-2 *Forest plots demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and diagnostic intervals following Cox proportional regression using Approach 1.*

a)



b)



Approach 2 fitted restricted cubic splines with 4 knots according to percentiles defined by Harrell et al., 2001 (153). Accordingly, the knots corresponded to day 0, day 13, day 47, day 158; generating 3 separate terms. The relationship between diagnostic interval and 1-year mortality was found to be non-linear, with an initial increase in risk followed by a plateau and a decrease in risk (Figure 6-3). The relationship between diagnostic interval and 5-year mortality was also found to be non-linear, with an initial decrease in risk followed by an increase then plateau. There was no significant association between each of the 3 intervals and risk of mortality at 1-year or 5-years, following adjustment for sociodemographic (sex, deprivation level, region) and cancer-specific factors (age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressiveness) in complete case analyses (Figure 6-4).

Figure 6-3 *Predicted probability of a) 1- and b) 5-year mortality across the diagnostic interval modelled using restricted cubic splines with 4 knots.*

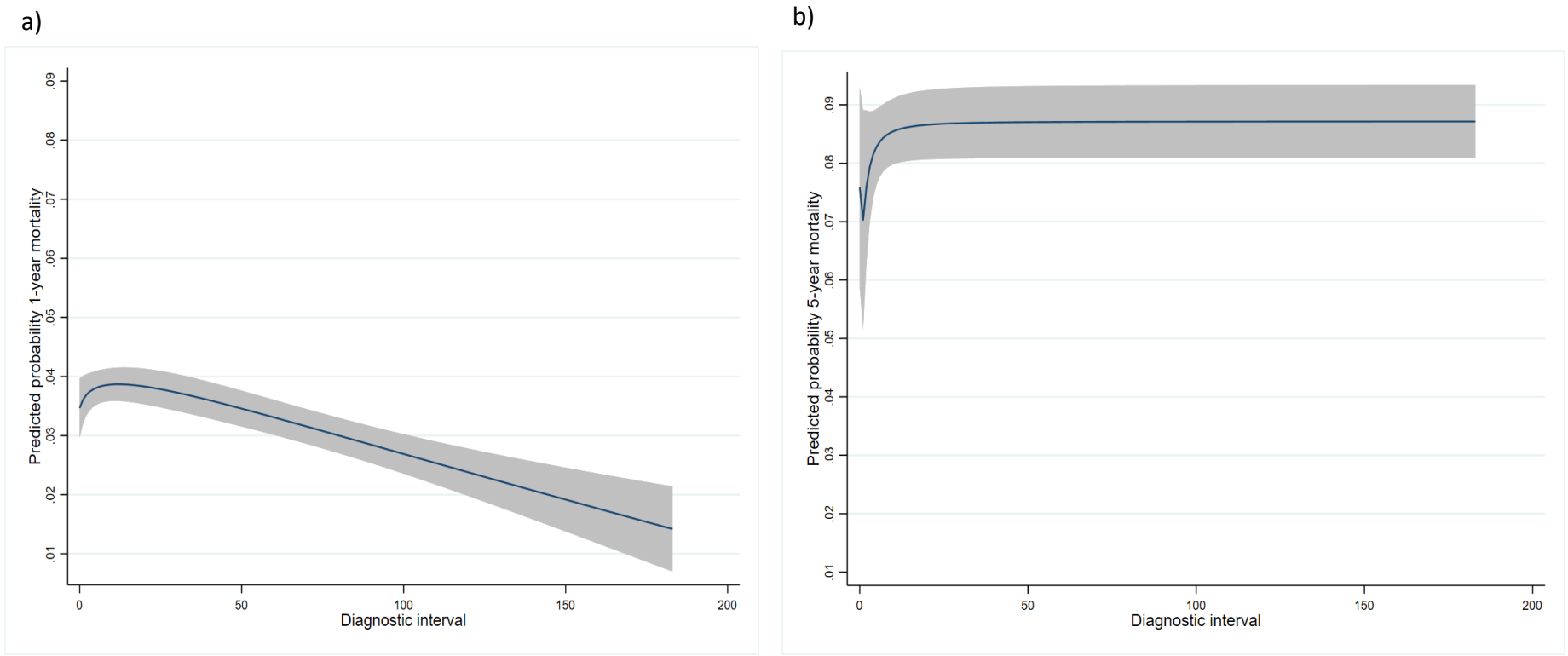
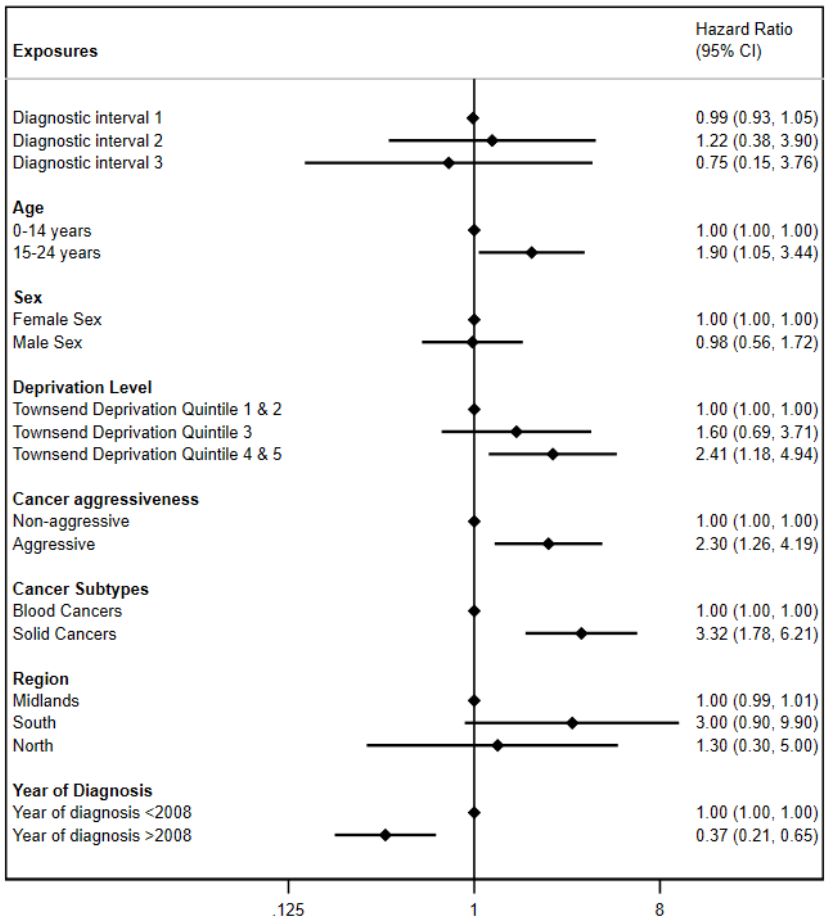
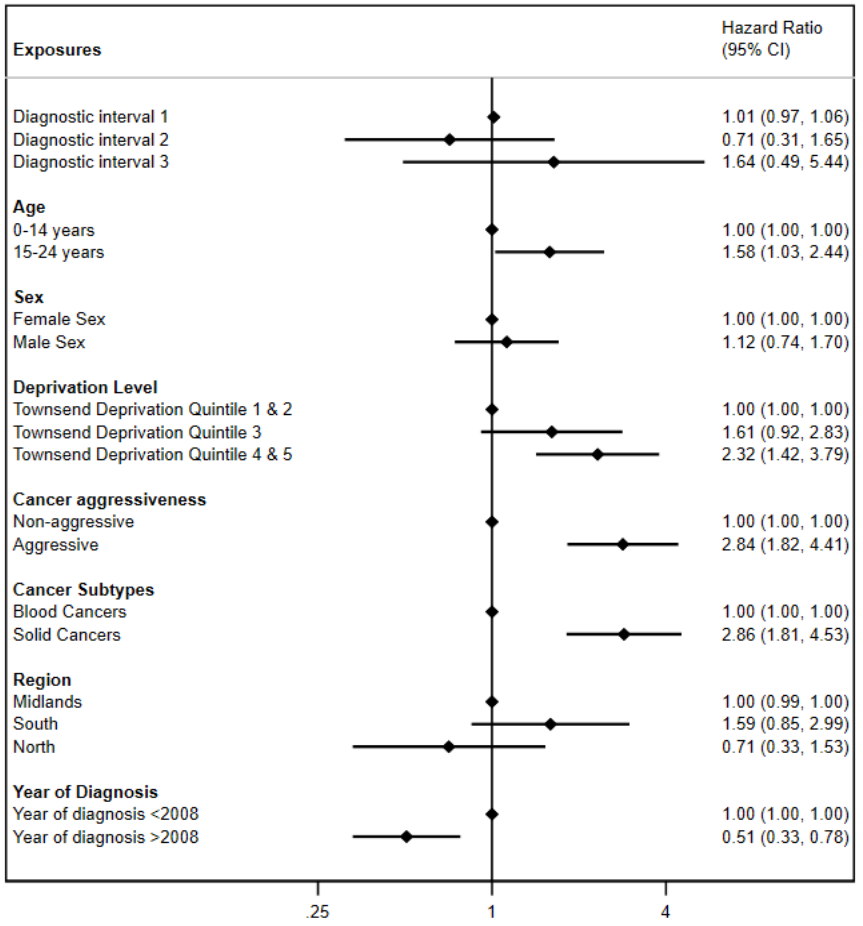


Figure 6-4 *Forest plots demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and diagnostic intervals following Cox proportional regression using Approach 2.*

a)

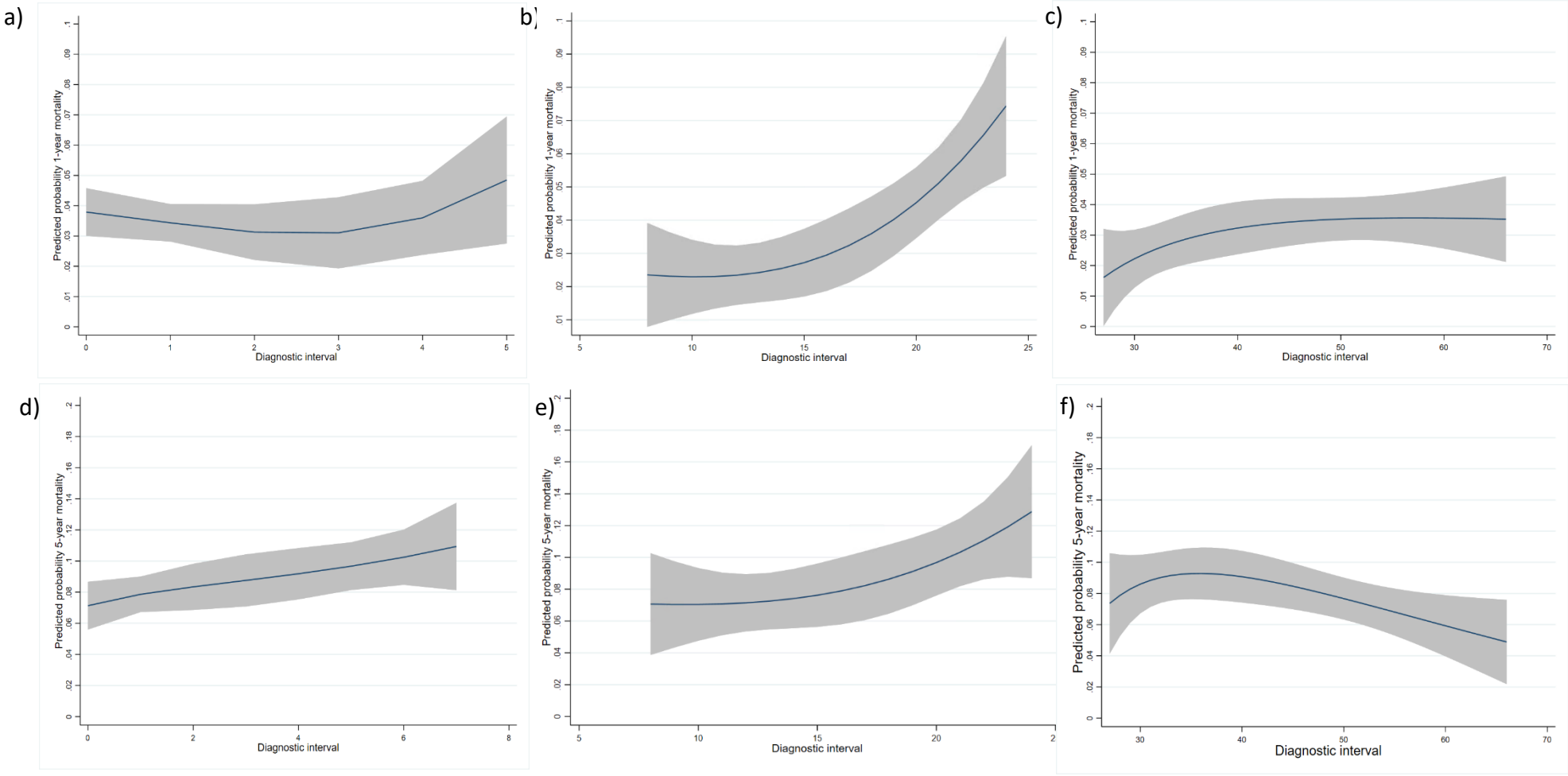


b)



Approach 3 was stratifying according to a priori set diagnostic interval lengths. These lengths were defined by the 25th, median and 75th centile of the diagnostic interval and were accordingly: 0-6 days (0-25th percentile, n=221), 7 to 25 days (25th-50th percentile, n=215) and 26-67 days (50th-75th percentile, n=221) (Figure 6-5). When the first 6 days were considered, there was no significant increased risk in mortality at 1- or 5 years [aHR 0.95 (95%CI 0.68 - 1.44), p=0.8 and aHR 1.05 (95%CI 0.77-1.42), p=0.7, respectively]. When 7-25 days were considered, there was a significant increase in mortality risk at 1-year and 5-years [aHR 1.13 (95%CI 1.01-1.26), p=0.03 and aHR 1.09 (95%CI 1.01-1.18), p=0.02]. When 26-67 days were considered, there was no significant increased risk in mortality at 1- or 5 years [aHR 1.01 (95%CI 0.97-1.07), p=0.6 and aHR 0.98 (95%CI 0.94-1.02), p=0.4, respectively].

Figure 6-5 *Predicted probability of 1-year mortality across the restricted diagnostic intervals, a) 0-6 days; b) 7-25 days; c) 26-67 days (top panel). Predicted probability of 5-year mortality across the restricted diagnostic intervals, d) 0-6 days; e) 7-25 days; f) 26-67 days (bottom panel).*



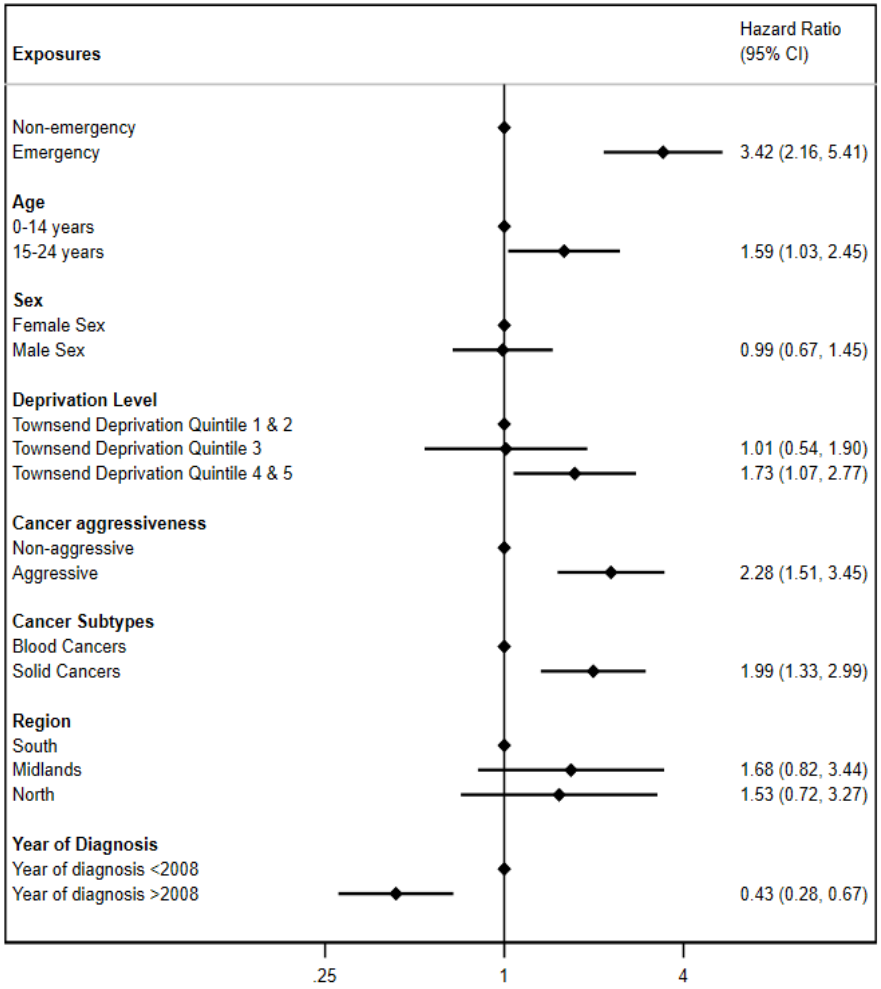
Following multiple imputation, for Approach 1 and 2, similar results to the complete-case analysis was found; there was no significant association between the diagnostic interval and risk of mortality at 1- or 5-years following adjustment for sociodemographic (sex, ethnicity, deprivation level) and cancer-specific factors (age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressiveness) (Appendix Chapter 9, Supplementary Figures 9-5 & 9-6).

6.4.2 Exposure 2: Route-to-Diagnosis

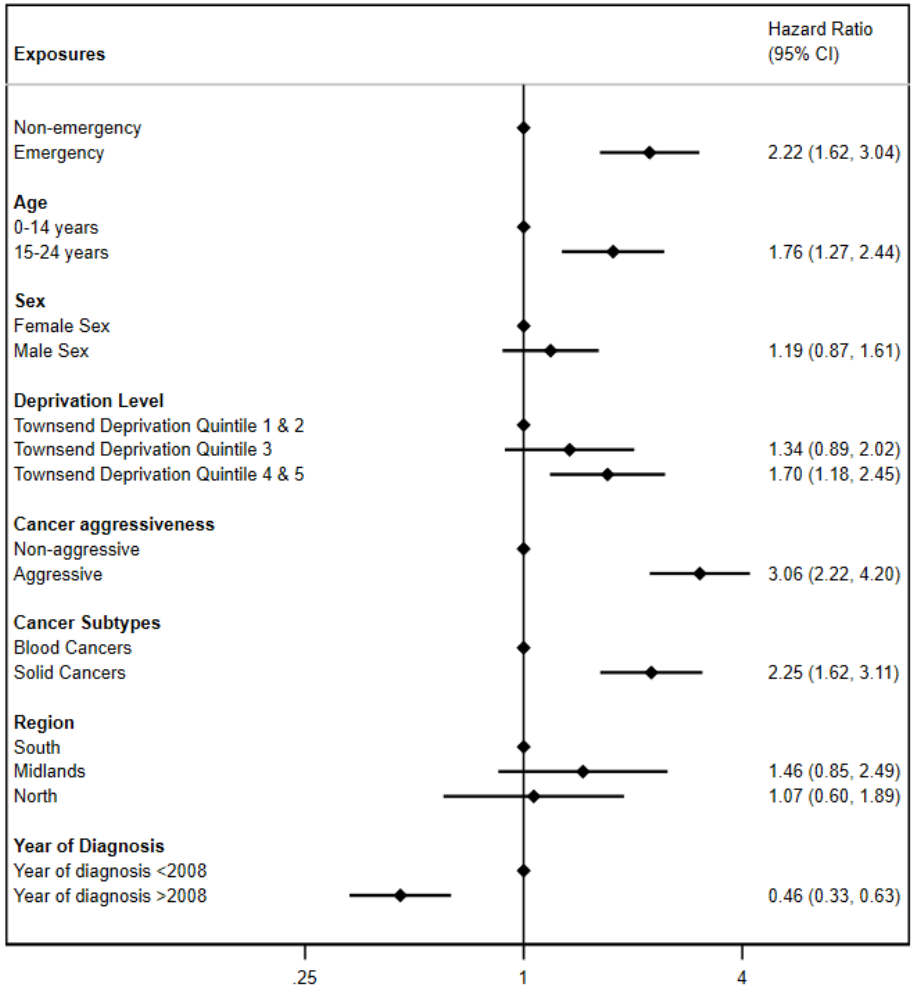
Emergency presentations were associated with significantly higher risk of mortality at 1- and 5 years, following adjustment for sociodemographic (sex, ethnicity, deprivation level, region) and cancer-specific factors (age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressiveness) (Figure 6-6). Findings were similar following multiple imputation (Appendix Chapter 9, Supplementary Figure 9-7).

Figure 6-6 *Forest plot demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and emergency presentations following Cox proportional regression.*

a)



b)



6.5. Discussion

6.5.1 Summary of Principal Findings

This Chapter identified a significant association between emergency cancer presentations at diagnosis in young people and a subsequent higher risk of mortality at 1 years and 5 years. Additionally, in exploratory analyses, there was a statistically significant higher risk of 1- and 5- year mortality in young people with longer diagnostic intervals, when the diagnostic intervals was restricted to between 7 and 25 days (i.e., between the 25th and 50th centiles). However, when the whole diagnostic interval, from 0 to 183 days was considered, this association was not observed.

6.5.2 Findings in Context

The association between diagnostic delay and mortality outcomes following a cancer diagnosis is a complex non-linear relationship, which has been observed in both young people and adults in previous studies [12, 15]. This Chapter also provides evidence for this non-linear relationship. Restricted cubic splines have previously been shown to better model this relationship [13], and accordingly, this approach was utilised in this Chapter but no statistically significant association was found between diagnostic intervals and mortality at 1- and 5- years. In exploratory analyses of 221 young people with cancer, higher mortality risk at 1- and 5- years was observed in those who had longer diagnostic intervals, when the interval was restricted to 7-25 days. This may reflect the interval in which both the advanced/aggressive and the indolent cancer presentations are excluded. This offers insight into a potential role of delay in mortality for cancers with similar progression rates. However, larger studies are warranted before drawing conclusions.

Although the underlying causes of emergency presentations of cancer diagnoses are multifactorial, diagnostic delay is known to play a contributory role [138]. Emergency presentations of adult cancer diagnoses have been shown to be an independent prognostic

factor, with worse survival outcomes in those who present as an emergency compared to those who do not, independent of stage of cancer [138]. This Chapter identified that emergency presentations are associated with higher mortality risk at 1- and 5- years in young people with cancer as well, following adjustment for sociodemographic factors, year-of-diagnosis, cancer type and aggressiveness. Overall, these findings suggest that emergency presentation may also be an important independent prognostic factor in young people with cancer.

In line with previous studies [150], this Chapter also identified that other sociodemographic and cancer-specific factors are associated with higher mortality in CTYA with cancer. These included higher mortality risk seen in 1) older age-at-diagnosis, 2) more deprived compared to less deprived young people, 3) leukaemia and CNS tumour diagnosis compared to other diagnoses, 4) more aggressive compared to less aggressive cancers and 5) earlier year-of-diagnosis (<2008) compared to more contemporary time periods (>2008).

6.5.3 Strengths and Limitations

The main strength of this study is the representativeness and size of the population-based cohort used. Additionally, linked information from cancer and civil registry data allows to accurately ascertain cancer cases and subsequent mortality outcomes.

However, there are also several important limitations in this study. Firstly, there is limited information on relevant prognostic factors, such as the stage of diagnosis (and for leukaemias the risk status at diagnosis), and stage-at-diagnosis in adult cancers have been shown to confound the association between diagnostic delay and mortality outcomes [138]. Secondly, as highlighted in Chapter 5, there are limitations in the way in which systemic diagnostic intervals are constructed using EHR, even though this approach has been used in primary care-based cancer diagnostic interval studies [41]. Thirdly, there was missing data in several exposures-of-interest, including diagnostic interval and emergency presentations. Although

the missingness was deemed to be missing at random, analyses were carried out using both complete case and multiple imputation approaches, and both showed comparable results. Finally, given the rarity of CTYA cancers, exploring diagnostic pathways individually across important subgroups, such as cancer subtype, was not possible due to small sample sizes.

6.5.4 Implications and Future Directions

This Chapter provides a comprehensive analysis of the relationship between diagnostic pathways and mortality risk 1- and 5-years following a cancer diagnosis in young people and 1) identifies, for the first time in young people with cancer that emergency presentations are associated with higher long-term mortality risk and, 2) provides further insight into the complex, non-linear relationship between diagnostic intervals and mortality. Altogether, these findings have important research, public health and clinical implications.

From a public health perspective, findings call for better strategies to identify potentially preventable emergency presentations through targeted public health interventions. From a clinical perspective, there may be a role for integrating emergency presentations during prognostication and risk assessment of the young person during their treatment, alongside their age, cancer type and biology, to improve long-term outcomes. From a research perspective, findings firstly emphasise the need to better investigate the underlying causal pathways that lead to emergency presentations in CTYA cancer, which are currently not well-studied. Secondly, findings call for future more personalised mortality risk prediction research. Finally, findings warrant future research to better delineate the underlying complexity and heterogeneity of the relationship between diagnostic intervals and long-term mortality.

Code from Chapter

https://github.com/DefOx2020/Thesis/blob/06e6b1c26357ad03b73a6fc065ef8af9040ef1b4/Chapter%206

Chapter 7 The association between diagnostic pathways and mental health outcomes in 1- and 5- year survivors of childhood, teenage and young adult cancers

7.1. Background to chapter

Survival following CTYA cancers has improved worldwide with 10-year survival now estimated around 80% in the UK [1]. This improvement was also observed in Chapter 6, where 1- and 5- year mortality risk was significantly lower in young people diagnosed after 2008 compared to before 2008. Unfortunately, however, it is now becoming apparent that survival is associated with a wide range of debilitating physical and psychosocial complications, with survivors experiencing more chronic health problems compared to the general population [151]. Early recognition, better management and prevention of these complications remain important research priorities for young people with cancer [11].

Risk factors associated with developing these complications are not well understood. Sociodemographic factors such as age-at-diagnosis and sex as well as cancer-specific factors such as cancer subtype and treatment modality are known to play possible roles [152]. Specifically, there is strong evidence that chemotherapy and radiotherapy are linked to complications in a dose-dependent manner [153-155]. As treatment regimens intensify when the cancer is diagnosed at more advanced stages, early diagnosis may therefore play a potentially important role in preventing complications.

Currently, the association between diagnostic pathways and complications following cancer survival are poorly understood. One important complication, which has been

highlighted as an area of high-priority by young people with cancer, is the impact of early diagnosis on mental health disorders [11]. There is strong evidence that survivors of CTYA cancer are at higher risk of mental health disorders, including anxiety, depression and psychosis [156] but the role of early diagnosis is not well understood. There has been some insight of worse mental health outcomes and quality of life following a delayed diagnosis but this has been limited to studies focusing only on teenagers and young adults [21]. Furthermore, these studies have relied on questionnaire-based study designs, which may be subject to selection and recall biases. Importantly, even though mental health disorders are most likely to first present and be managed in primary care [157], there are no primary care-based studies exploring the burden and risk factors associated with mental health outcomes in CTYA cancer survivors. Given the availability of a population-based cohort of CTYA and linked primary care-based and cancer treatment data in the QResearch Database, there is a unique opportunity to study the association between diagnostic pathways and complications following cancer survival.

7.2. Aims & Objectives

The overarching aim of this Chapter is to identify if there is an increased risk of mental health disorders in 1-year and 5-year survivors of CTYA cancer compared to the general population and if so, to investigate the association between diagnostic pathways and these disorders. The specific objectives were:

1. To ascertain the incidence of mental health disorders (depression, anxiety and psychotic disorders) in 1- and 5- year cancer survivors and compare it to 10 age- and sex- matched general population controls.

2. To explore if diagnostic pathways (systemic diagnostic intervals and emergency presentations as defined in Chapters 5 and 6) are associated with the risk of developing these disorders.

7.3. Methods

7.3.1 Data Source & Source Population

For 1 year survivors, the source population was the open cohort of CTYA generated in Chapter 2. For 5 year survivors, the source population was open cohort of CTYA registered with a GP within QResearch Database between 1st January 1998 and 31st January 2017 generated in Chapter 6. These time periods were selected in order to ensure a minimum of 1-year or 5-year follow-up following diagnosis was available for all CTYA.

7.3.2 Study Population

Any young person (0-24 years) diagnosed with selected cancers defined in Chapter 2 and alive 1- and 5- years following diagnosis (irrespective of disease and treatment status) were used to generate two survivor cohorts. Cohort entry was the date in which the young person was alive and either 1- or 5-years post cancer diagnosis date (index date for each survivor). Cohort entry was restricted to >8 years as mental health disorders prior to that age are very rare in CTYA [158]. Any young person with an outcome that was recorded *before* cohort entry was excluded.

For each case, up to 10 year-of-birth and sex- matched controls were selected, who also entered at the index date and had at least 1-year or 5-year follow-up, depending on the survivor cohort. Year-of-birth instead of age was used as a matching variable to ensure cases and controls were followed-up in the same time period.

Both survivor and control cohorts were followed up until i) the date of the outcome of interest, ii) leaving a QResearch practice or iii) study end date (31st January 2022). Any young person who had an outcome after the study end date was censored.

7.3.3 Outcomes of interest

The outcomes-of-interest were incident anxiety-depression or psychotic disorders. Each disorder was defined in a composite category as follows:

- Diagnosis-based: any condition that matched the SNOMED-CT or ICD-10 diagnostic code for the relevant outcome.
- Prescription-based: any prescription of medication relevant to the outcome

Both diagnosis and prescription-based definitions were subsequently combined to make a composite code for outcomes: anxiety-depression as the first composite outcome and psychosis as the second composite outcome. Table 7-1 provides an overview of the diagnoses and drug groups used to define the outcomes and detailed SNOMED-CT/ICD-10 codes per outcome is available in the Supplementary Appendix (Appendix Chapter 9, Supplementary Table 9-24). The first record of the disorder within the study period was used as the date of diagnosis of the outcome.

Table 7-1 *Description of diagnoses and drug names used to define each outcome in this Chapter.*

	Diagnoses	**Prescriptions**
Mental health disorders	• Anxiety (phobias and panic disorders) • Depressive disorders (including self-harm)	• Serotonin re-uptake inhibitors (SSRIs), tricyclic antidepressants (TCI), monoamine oxidase inhibitors • Anxiolytics (e.g., diazepam)
	• Psychosis (affective and non-affective)	• Antipsychotics (e.g., clozapine), lithium

7.3.4 Exposures of Interest

For Objective 1, where the incidence of mental health disorders were compared between cancer survivors and matched controls representing the general CTYA population, the exposure of interest was a diagnosis of cancer. Potential confounders were identified to be deprivation level, ethnicity and family history of the disorder [156].

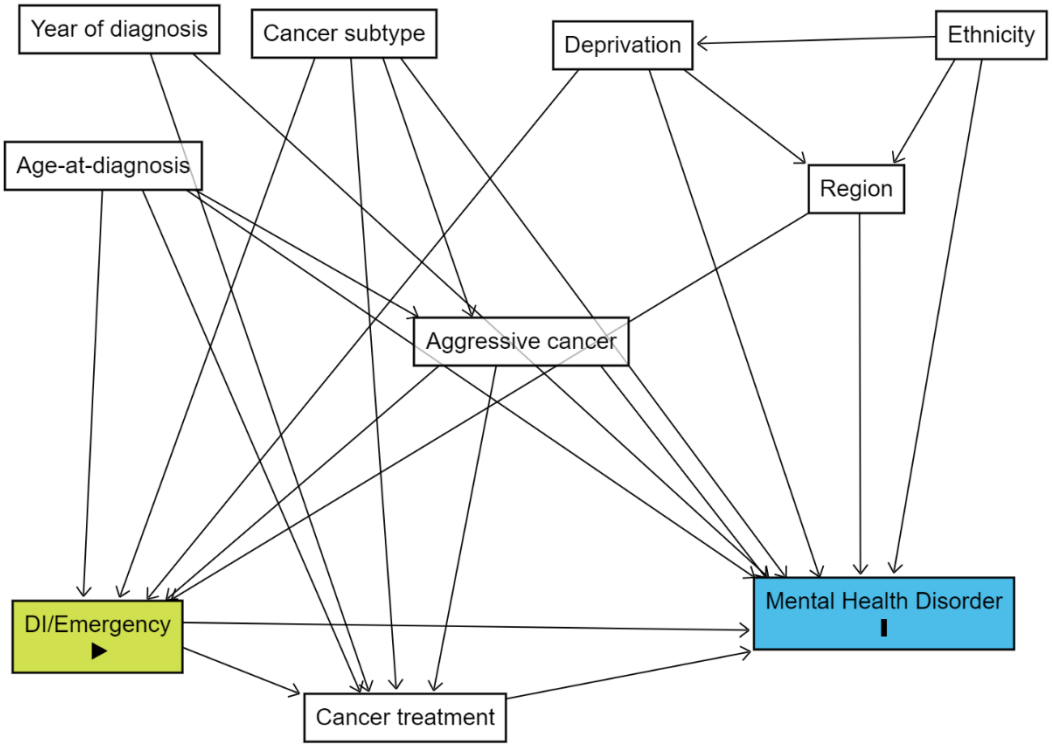
For objective 2, the two primary exposures-of-interest are systemic diagnostic intervals and emergency presentation at diagnosis as defined in Chapter 5. However, as the timing of the data extract for this Chapter was prior to the execution of Chapter 3, the diagnostic intervals were defined solely through symptoms/signs previously reported to be relevant in CTYA by NICE guidance [78] (as described in Chapter 5 for the COVID-19 related analyses and available through Chapter 9, Supplementary Appendix Table 9-19).

Diagnostic intervals were explored using the same approach described in Chapter 6: Approach 1 defined diagnostic interval as a continuous, linear variable; Approach 2 defined diagnostic interval as a continuous variable, fitted with restricted cubic splines; Approach 3 stratified diagnostic intervals according to a priori set diagnostic interval lengths, defined by the 25th, median and 75th centile of the diagnostic interval. In addition to the potential confounding exposures identified in Chapter 6, complications following survival also require adjustment for cancer treatment (chemotherapy and radiotherapy). This is because diagnostic pathways may influence cancer treatment, which in turn may be associated with the risk of developing these selected complications. Confounding exposures are summarised in Figure 7-1.

Cancer treatment data was available through Classification of Interventions and Procedures Dataset (OPCS-4) for surgery, chemotherapy and radiotherapy from 1st January 1998 [159]. OPCS-4 codes were categorised into two broad groups: chemotherapy or radiotherapy. The

young person was categorised as having cancer treatment only if the treatment was given after the cancer diagnosis (i.e., cohort entry) and before the date of the outcome.

Figure 7-1 *Updated direct acyclic graph demonstrating causal pathways between diagnostic pathway (DI/Emergency) and outcomes-of-interest, with biasing paths through age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressive cancer, cancer treatment, deprivation level, region and ethnic group. DI= diagnostic interval.*



7.3.5 Statistical Analysis

For descriptive analyses, continuous variables were presented as means (medians if non-normal) and categorical variables were presented in frequencies and comparisons between cancers and the general population were carried out using t-tests (Mann-Whitney U if non-normal) and $\chi 2$ tests, respectively.

For both objectives, to explore the association between developing the selected disorders and primary exposures-of-interest, multivariable Cox proportional hazard regression modelling was used. The proportional hazards assumption was tested using the Schoenfeld residuals. The models accounted for the correlation generated by clustering of children within practices through a robust variance estimator.

7.3.6 Handling Missing Data

There was missing data in exposures-of-interest (diagnostic interval, emergency presentation, ethnicity and Townsend Deprivation Score) and all missing data were deemed to be missing under the missing at random assumption due to the mechanism in which they are collected in electronic healthcare data [142]. Both complete case analyses and analyses following multiple imputation were carried out.

Multiple imputation with chained equations was used to replace missing data. 50 imputations were performed, given % missingness of ethnicity and diagnostic intervals[91]. Imputation models included all exposure variables, and the primary outcome variables; statistical models were developed on each of the 50 imputed data sets and the estimates pooled using the Rubin rules [91]. Any continuous variable was assessed for normality [91]. As the diagnostic interval did not follow a normal distribution and was right-skewed, it was transformed using its natural logarithm prior to imputation analyses.

7.3.7 Power Calculations

For objective 1, Stata's *stpower log-rank test* comparing two survival rates method was carried out for power calculations [149]. The incidence/hazard rate of each outcome in the general CTYA population was estimated through previous studies (Table 7-2) [156]. As seen in Table 7-2 there is sufficient sample size to detect these risk estimates at 80% power and a significance threshold of 0.05, given this study has sample sizes of 4,389 (up to 2019) and 3,214 (up to 2017) young people with cancer as determined in Chapter 6.

Table 7-2 *Minimum detectable hazard ratios in this study at 80% power and significance threshold of 0.05, given sample sizes of 4,389 (up to 2019) and 3,214 (up to 2017) young people with cancer. Comparison with previously reported hazard ratios [156].*

Outcome of interest	Reported prevalence or incidence rate in CTYA	Minimum detectable hazard ratio in study	Previous reports of risk of disorder in 5- year cancer survivors
Anxiety	5.9%	1.23	1.29
Depression	3.3%	1.33	1.57
Acute Psychosis	1.2%	1.47	1.56

For objective 2, Stata's *stpower log-rank test* comparing two survival rates method was carried out for power calculations. Following completion of analysis of Objective 1 and before starting analysis for Objective 2, the survival probability in the "control" population was determined, using incidence rate findings from Objective 1. The only study exploring diagnostic intervals and depression/anxiety outcomes reported an increased risk (unadjusted hazard ratio of 1.9) with longer diagnostic intervals [16]. Accordingly, the minimum sample size required to detect this given effect size, at 80% power and significance threshold of 0.05 was calculated to be 300. Given findings in Chapter 6, where 30% and 60% of young people with cancer had a diagnostic interval and cancer route recording, respectively, the smallest sample sizes available were estimated to be 964 (i.e., 30% of cancer cohort n=3,214) for diagnostic interval studies and 1,928 (i.e., 60% of cancer cohort n=3,214) for cancer route studies, which provide evidence that sufficient sample size is available to study depression/anxiety outcomes.

Although no studies have explored the association between diagnostic pathways and psychotic disorders in cancer survivors, the minimum sample size required to detect the same hazard ratio of 1.9 at 80% power and significance threshold of 0.05 was 1,422. For the study investigating diagnostic interval as an exposure, sample size was deemed to be insufficient (estimated 964 young people with recorded diagnostic intervals) and thus was not carried out.

For the study investigating cancer route as an exposure, sample size was likely to be sufficient (estimated 1,928 young people with recorded cancer route).

7.4. Results

7.4.1 Objective 1: Risk of mental health disorders in CTYA cancer survivors and controls

Overall, there were 3,267 1-year cancer survivors and 27,197 year-of-birth and sex-matched controls and 2,035 5-year cancer survivors and 17,475 year-of-birth and sex-matched controls. In line with the full cohort, cancer cases (both 1- and 5-year survivors) were more likely to be White compared to other ethnic groups ($p<0.001$) and more likely to come from lower deprivation levels (i.e., higher Townsend Deprivation Scores) compared to higher deprivation levels ($p<0.001$) (Table 7-3).

Table 7-3 *Baseline characteristics of cancer cases and matched controls. All percentages are column percentages (%). Age is in integers and rounded up to the nearest integer. IQR= interquartile range.*

Characteristics	**1-year survivor**		**P-value**	**5-year survivor**		**P-value**
	Cases (n=3,267)	**Controls (n=27,197)**		**Cases (n=2,035)**	**Controls (n=17,475)**	
Age at entry						
Median ± IQR (years)	13 [5, 19]	13 [5,20]	0.9	11 [5, 19]	12 [5,19]	0.8
Sex						
Male	1,927 (59.0)	16,176 (59.3)	0.8	1,202 (59.1)	10,364 (59.3)	0.9
Female	1,340 (41.0)	11,021 (40.7)		833 (40.9)	7,111 (40.7)	
Ethnic Group						
White	1,494 (45.7)	12,287 (45.2)	<0.001	1,025 (50.4)	8,176 (46.8)	<0.001
Other	440 (13.5)	4,635 (17.0)		268 (13.2)	2,879 (16.5)	
Missing	1,333 (40.8)	10,275 (37.8)		742 (36.4)	6,420 (36.7)	
Deprivation Level						
1 (most affluent)	792 (24.2)	5,903 (21.7)	<0.001	527 (25.9)	3,999 (22.8)	<0.001
2	733 (22.4)	5,480 (20.2)		465 (22.9)	3,689 (21.1)	
3	671 (20.5)	5,251 (19.3)		430 (21.1)	3,591 (20.6)	
4	573 (17.5)	5,326 (19.6)		337 (16.6)	3,197 (18.3)	
5 (most deprived)	477 (14.6)	5,134 (18.9)		269 (13.2)	2,925 (16.7)	
Missing	21 (0.6)	103 (0.4)		7 (0.3)	74 (0.4)	
Family history of mental health disorder						
Yes	10 (0.3)	95 (0.4)	0.7	6 (0.3)	64 (0.4)	0.6
No	3,257 (99.7)	27,102 (99.6)		2,029 (99.7)	17,411 (99.6)	
Anxiety and/or depressive disorders						
Yes	1,065 (32.6)	5,955 (21.9)	<0.001	752 (37.0)	4,744 (27.2)	<0.001
No	2,202 (67.4)	21,242(78.1)		1,283 (63.0)	12,731 (72.8)	
Psychotic disorders						
Yes	91 (2.8)	450 (1.7)	<0.001	213 (10.5)	1,101 (6.3)	<0.001
No	3,176 (97.2)	26,747 (98.3)		1,822 (89.5)	16,374 (93.7)	

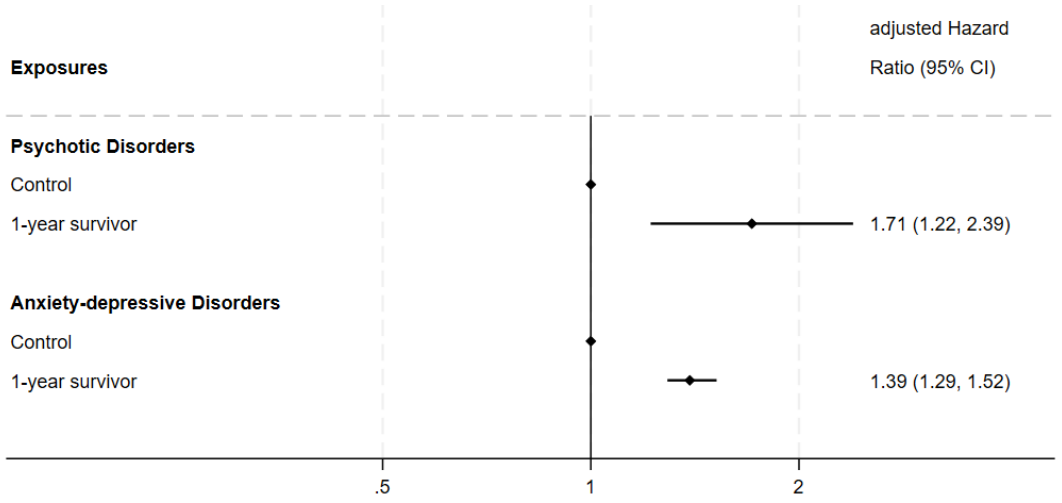
Out of 3,267 1-year cancer survivors, 525 had a mental health outcome recorded before the index date and were thus excluded. Similarly, out of 27,197 controls, 2,116 had a mental health outcome recorded before the index date and were excluded. The number of mental health disorders recorded in cancer survivors and controls, and their corresponding incidence rates, are detailed in Table 7-4.

Table 7-4 *Mental health disorders in 1-year cancer survivors and controls.*

Disorder	Cases (n=2,742) (n, %)	Incidence Rate (per 10,000 population) (95%CI)	Controls (n=25,081) (n, %)	Incidence Rate (per 1000 population)
Anxiety and/or depression	570 (20.8)	42.6 (39.2-46.2)	4026 (16.1)	27.4 (26.6-28.3)
Psychosis	40 (1.4)	3.0 (2.2-4.1)	234 (0.9)	1.6 (1.4-1.8)

Compared to year-of-birth and sex matched controls, 1-year cancer survivors were at higher risk of both anxiety-depressive disorders and psychotic disorders, following adjustment for deprivation level, ethnicity and family history of mental health disorders in complete case analyses (Figure 7-2). Findings were comparable following multiple imputation (Appendix Chapter 9, Supplementary Figure 9-8).

Figure 7-2 *Forest plot showing risk (measured using adjusted hazard ratio) of psychotic and anxiety-depressive disorders in 1 year CTYA cancer survivors. 95% CI = 95% confidence interval*



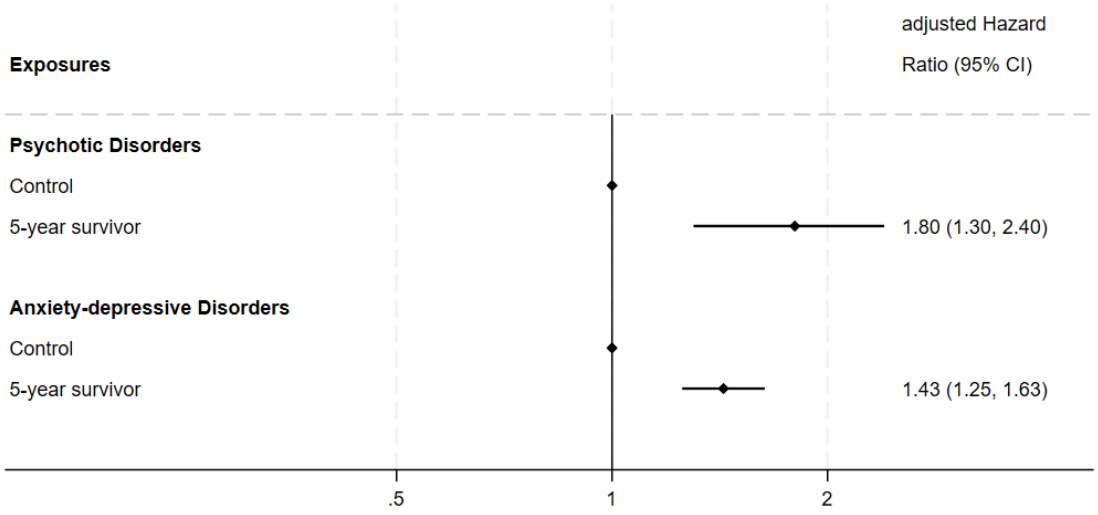
Out of 2,035 5-year cancer survivors, 515 had a mental health disorder recorded before the index date and were thus excluded. Similarly, out 17,475 controls, 2,489 had a mental health disorder recorded before the index date and were excluded. The number of mental health disorders recorded in 5-year cancer survivors and the controls, and their corresponding incidence rates, are detailed in Table 7-5.

Table 7-5 *Mental health disorder in 5-year cancer survivors and controls*

Disorder	Cases (n=1,520) (n, %)	Incidence Rate (10,000 population) (95%CI)	Controls (n=14,986) (n, %)	Incidence Rate (10,000 population)
Anxiety-depressive disorders	275 (18.1)	30.4 (27.1-34.3)	2,471 (16.5)	25.3 (24.3-26.3)
Psychotic disorders	55 (3.6)	6.1 (4.7-7.9)	406 (2.7)	4.2 (3.8-4.6)

Compared to year-of-birth and sex matched controls, 5-year cancer survivors were at higher risk of both anxiety-depressive disorders and psychotic disorders, following adjustment for deprivation level, and family history in complete case analyses (Figure 7-3). Findings were comparable following multiple imputation (Appendix Chapter 9, Supplementary Figure 9-8)

Figure 7-3 *Forest plot showing risk (measured using adjusted hazard ratio) of psychotic and anxiety-depressive disorders in 5- year CTYA cancer survivors. 95% CI = 95% confidence interval*



7.4.2 Objective 2: Diagnostic pathways and their association with the risk of mental health disorders

There were 350 out of 2,742 (12.7%) 1-year and 222 out of 1,520 (14.7%) 5-year cancer survivors with a recorded diagnostic interval. Given the small proportion of those with recorded diagnostic intervals, only complete case analyses were carried out. Furthermore, of the 350 1-year survivors with a recorded diagnostic interval, 95 had an anxiety-depressive disorder and only 3 had psychotic disorders, whilst of the 222 5-year survivors, 54 had an anxiety-depressive disorder and 12 had psychotic disorders. Given the low number of psychotic disorder diagnoses and insufficient power, regression analyses exploring the association between diagnostic intervals and psychotic disorders were not carried out.

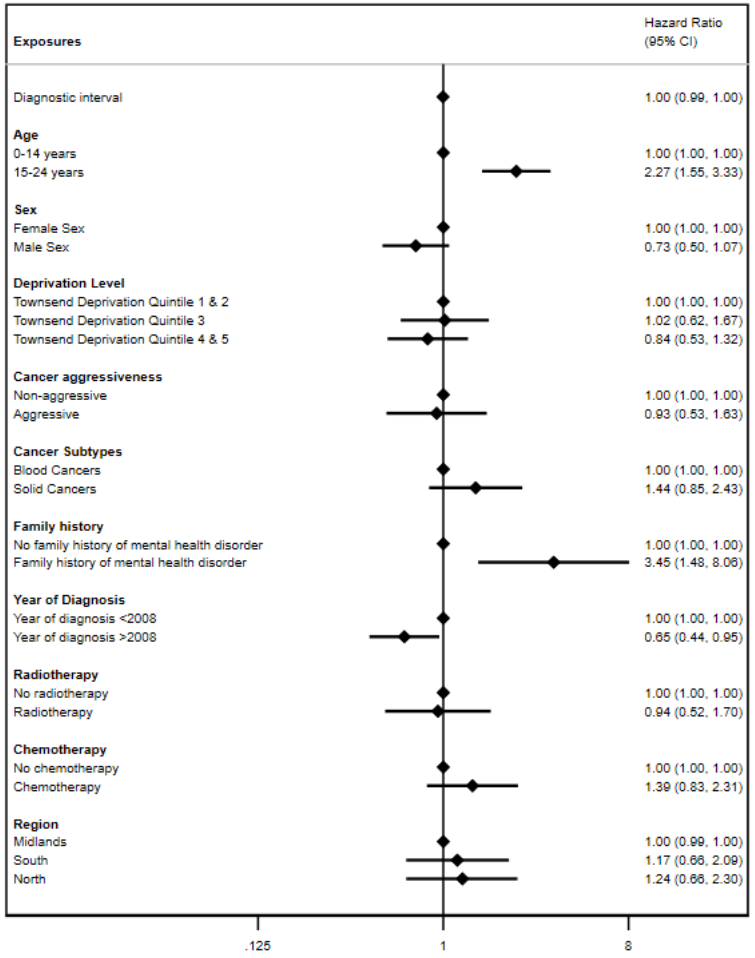
There was 1,381 out of 2,742 (50.3%) 1-year cancer survivors with a recorded route-to-diagnosis, of which 309 had a diagnosis of anxiety-depressive disorders and 22 had psychotic disorders. There were 904 out of 1,520 (59.5%) 5-year cancer survivors with a recorded route-to-diagnosis, of which 275 had an anxiety-depressive disorder diagnosis and 19 had a psychotic disorder diagnosis.

7.4.3 Diagnostic Intervals and anxiety-depressive disorders in 1-year and 5-year survivors

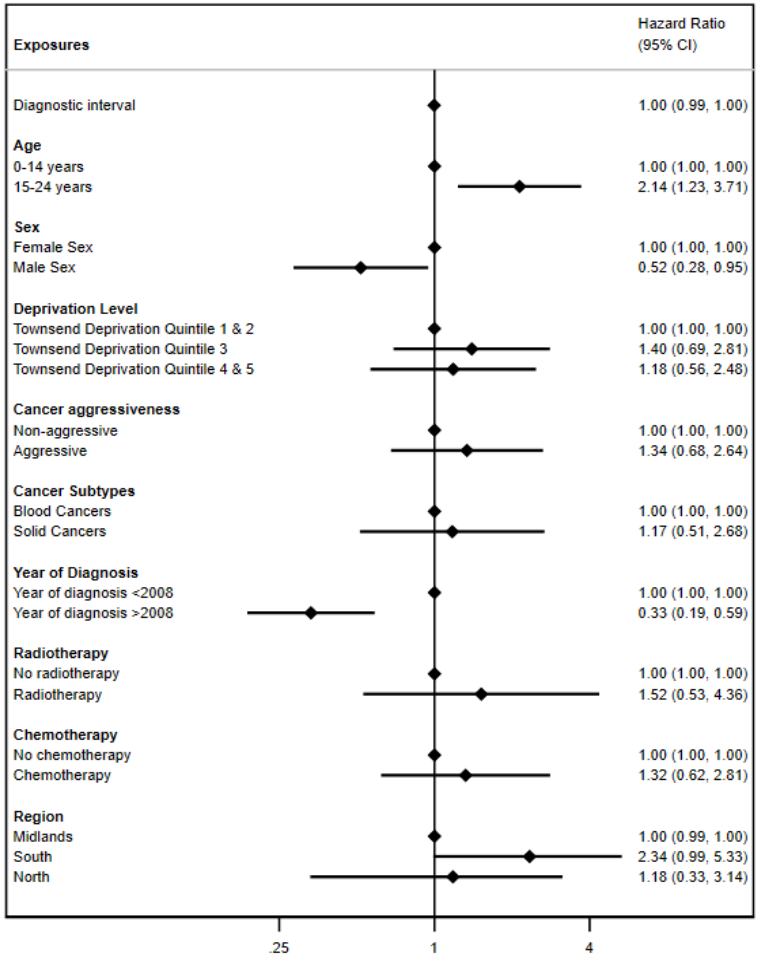
Using Approach 1 described in Chapter 6 (i.e., no modifications to diagnostic interval), there was no significant association between the diagnostic interval and risk of anxiety-depressive disorders in 1-year or 5-year survivors following adjustment for sociodemographic (sex, deprivation level, region) and cancer-specific factors (age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressiveness, cancer treatment) (Figure 7-4). Family history of psychiatric disease was only included in the 1-year survivor analysis as there were a small number with a recorded family history in 5-year survivors.

Figure 7-4 *Forest plot demonstrating hazard ratios (with 95% confidence intervals) for diagnostic interval and adjusted factors following Cox proportional regression using Approach 1 in a) 1-year survivors and b) 5-year survivors.*

a)

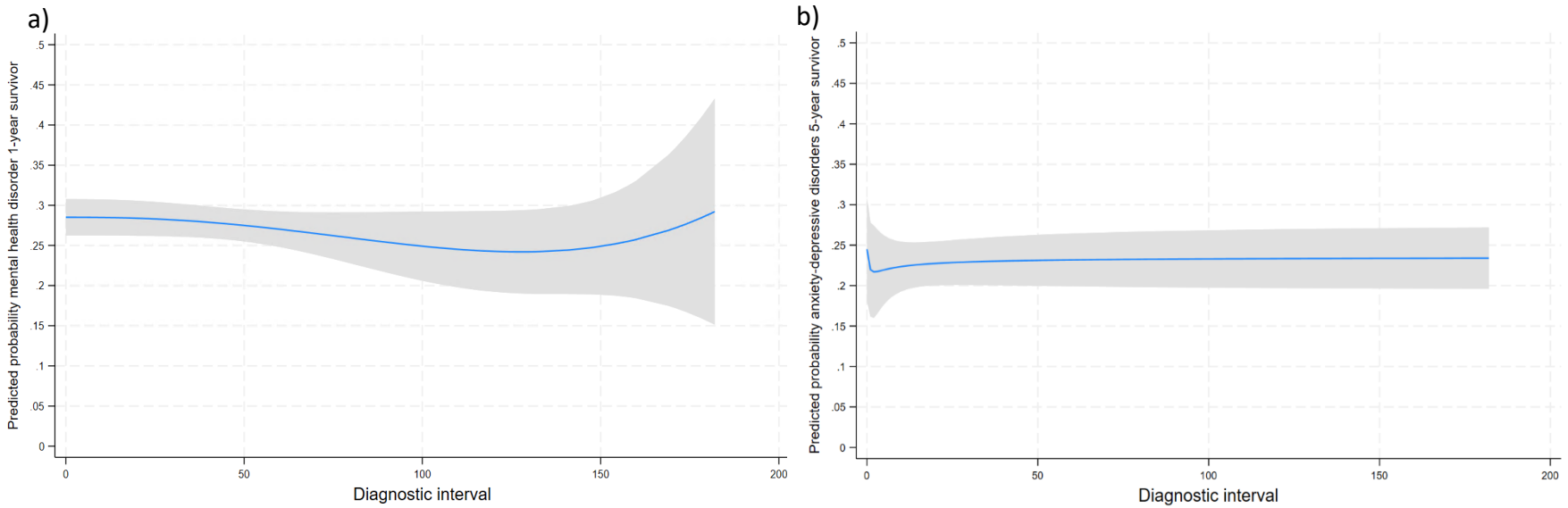


b)



Approach 2 fitted restricted cubic splines with 4 knots for the diagnostic interval according to percentiles defined by Harrell et al., 2001 [147]. Accordingly, the knots corresponded to day 0, day 11, day 44, day 132; generating 3 separate terms. The relationship was found to be non-linear, with an initial increase followed by a decrease in risk in both 1- and 5-year survivors (Figure 7-5).

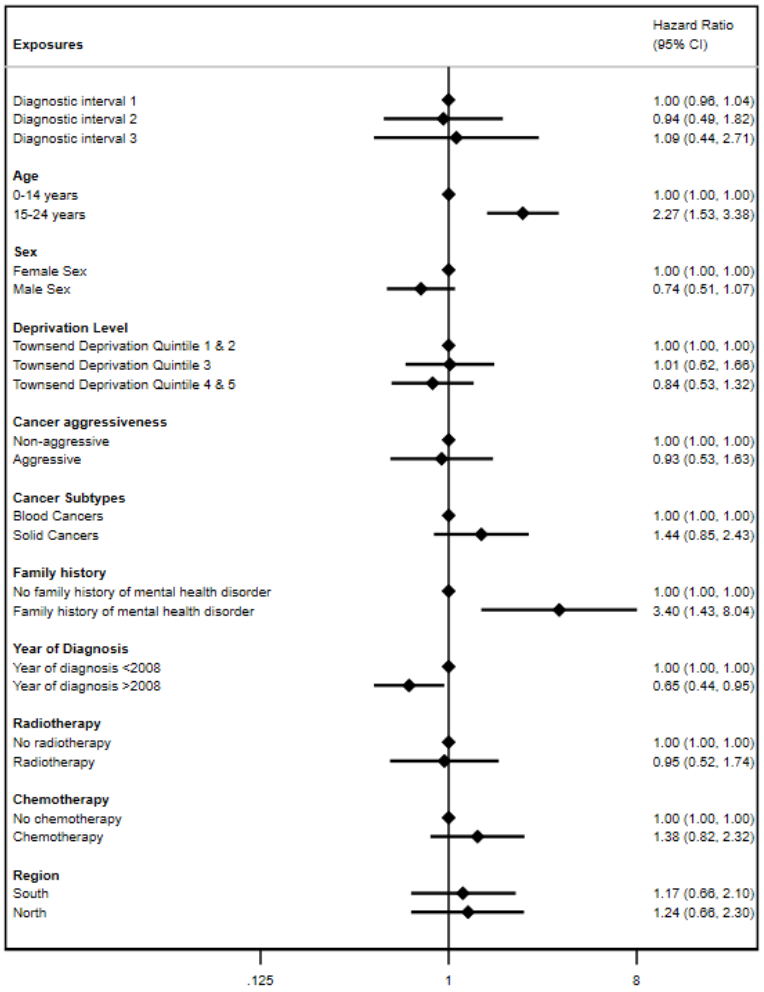
Figure 7-5 *Predicted probability of mental health disorder in 1-year and 5-year survivors of cancer across full systemic diagnostic interval*



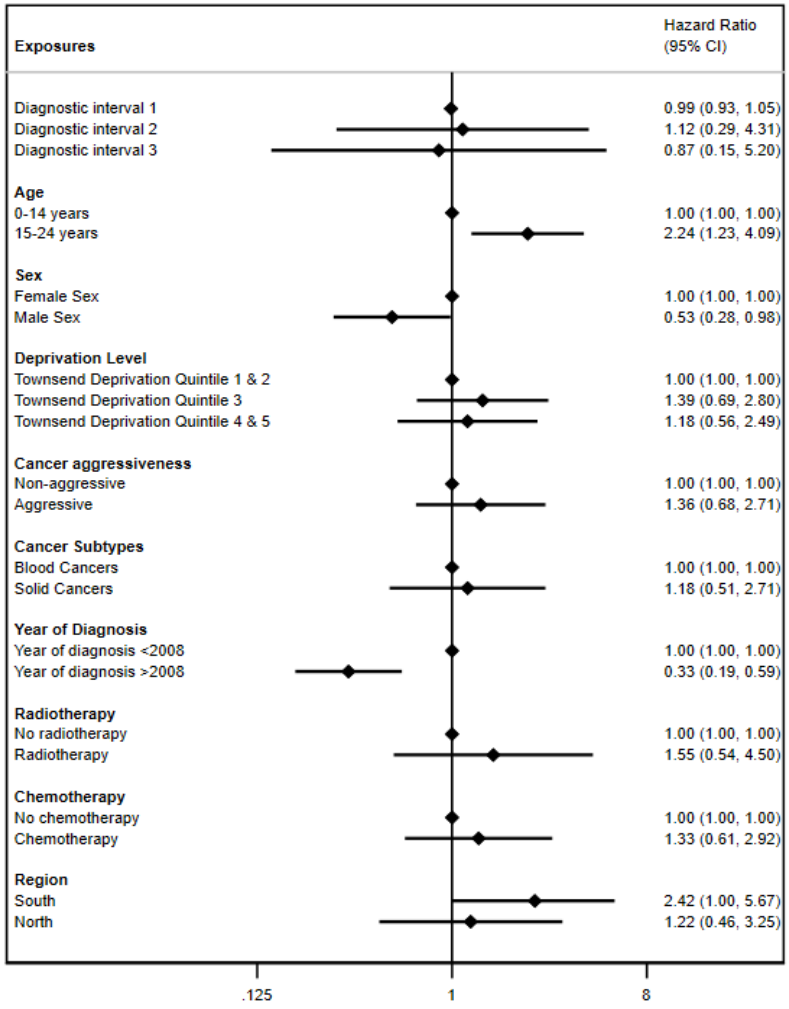
There was no significant association between each interval and risk of anxiety-depressive disorders in 1- or 5-year survivors following adjustment for sociodemographic (sex, deprivation level, region) and cancer-specific factors (age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressiveness, cancer treatment) (Figure 7-6).

Figure 7-6 *Forest plot demonstrating hazard ratios (with 95% confidence intervals) for diagnostic interval and adjusted factors following Cox proportional regression using Approach 2 in a) 1-year survivors and b) 5-year survivors.*

a)

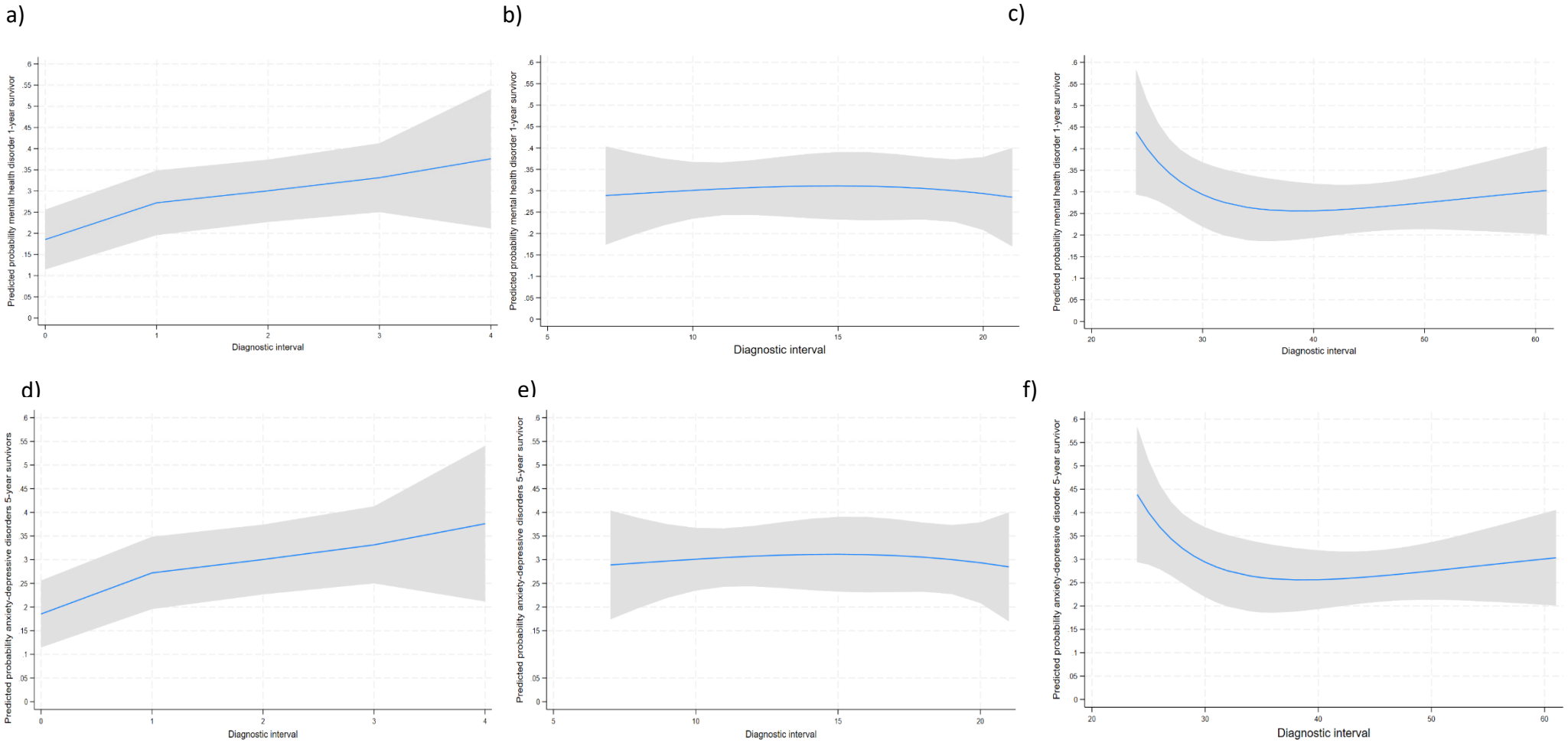


b)



Approach 3 was exploratory and stratified according to set diagnostic interval lengths as per Chapter 6. These lengths were defined by the 25th, median and 75th centile of the diagnostic interval and were 0-5 days (0-25th percentile, n=87), 6 to 22 days (25th-50th percentile, n=76) and 23-62 days (50th-75th percentile, n=83) (Figure 7-7). When the first 5 days were considered, there was a significant increase in the risk of anxiety-depressive disorders in 1-year survivors who had longer diagnostic intervals [HR 1.94 (95%CI 1.39 -2.71), p<0.001]. This persisted when explored in 5-year survivors [HR 2.82 (95%CI 1.54 -5.13), p<0.001, n=62]. When 6-22 days were considered, there was no significant increased risk in 1-year survivors with longer diagnostic intervals [HR 0.99 (95%CI 0.89-1.09), p=0.8]. When 23-62 days were considered, there was no significant increased risk in 1-year survivors [HR 0.98 (95%CI 0.96-1.01), p=0.2]. Neither was changed when explored in 5-year survivors [HR 1.03 (95%CI 0.82-1.28), p=0.8, n=52 and HR 0.98 (95%CI 0.94-1.03), p=0.4, n=46].

Figure 7-7 *Predicted probability of 1-year anxiety-depressive disorders across the restricted diagnostic intervals, a) 0-5 days; b) 6-22 days; c) 23-62 days (top panel). Predicted probability of 5-year anxiety-depressive disorders across the restricted diagnostic intervals, d) 0-5 days; e) 6-22 days; f) 23-62 days (bottom panel).*

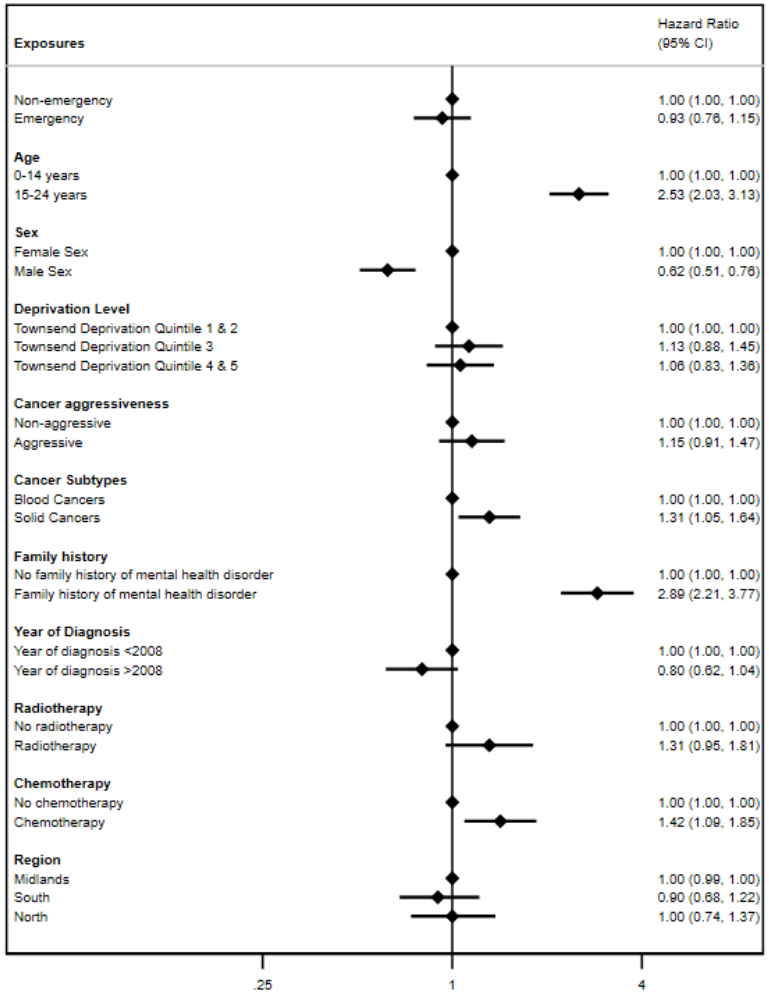


7.4.4 Route-to-diagnosis and anxiety-depressive disorders in 1 year and 5-year survivors

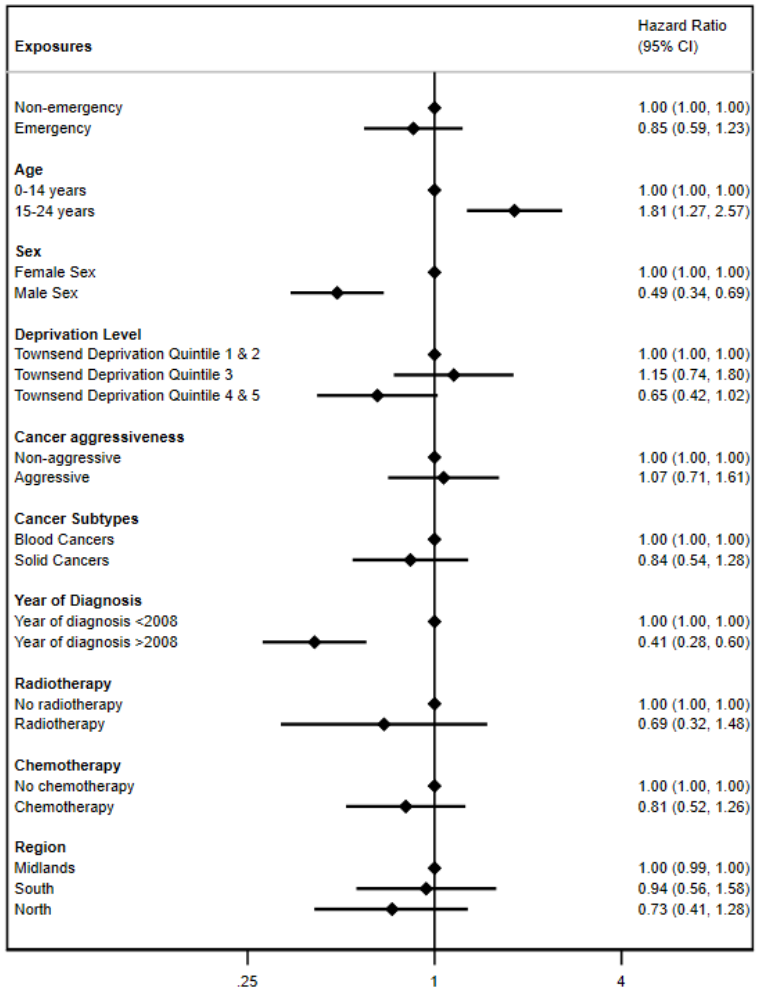
Emergency presentations were not associated with a higher risk of anxiety-depressive disorders in 1-year or 5-year survivors, following adjustment for sociodemographic (sex, ethnicity, deprivation level, region) and cancer-specific factors (age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressiveness) (Figure 7-8). Findings were similar following multiple imputation (Appendix Chapter 9, Supplementary Figure 9-9).

Figure 7-8 *Forest plot demonstrating hazard ratios (with 95% confidence intervals) of anxiety-depressive disorders for cancer-route and adjusted factors following Cox proportional regression in a) 1-year survivors and b) 5-year survivors.*

a)



b)

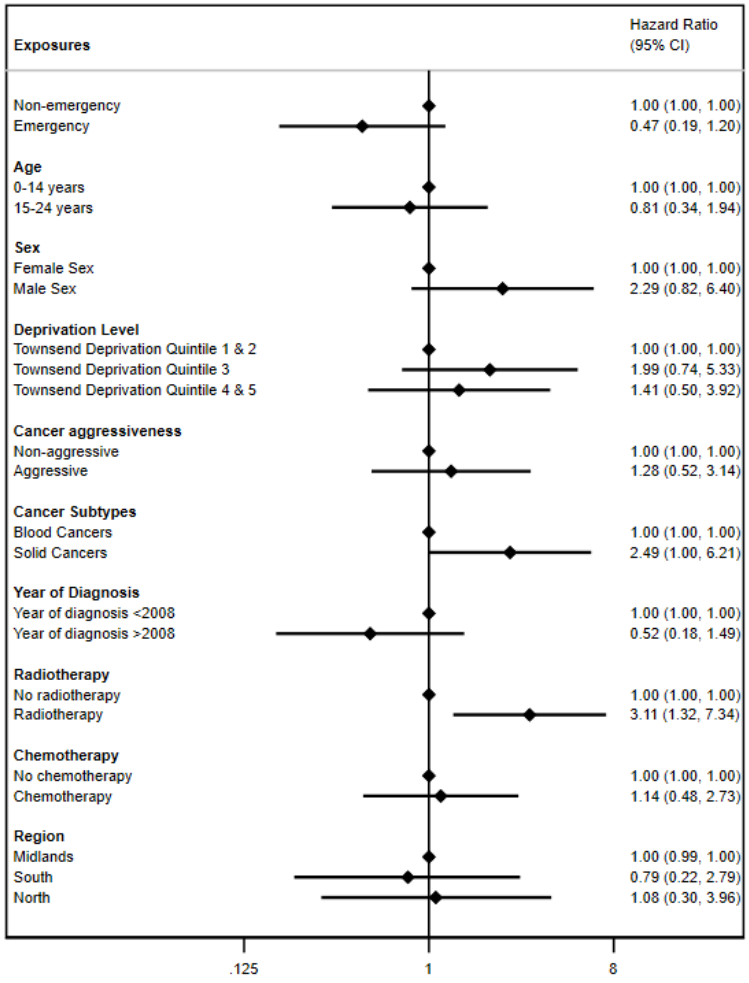


7.4.5 Route-to-diagnosis and psychotic disorders in 1- and 5-year survivors

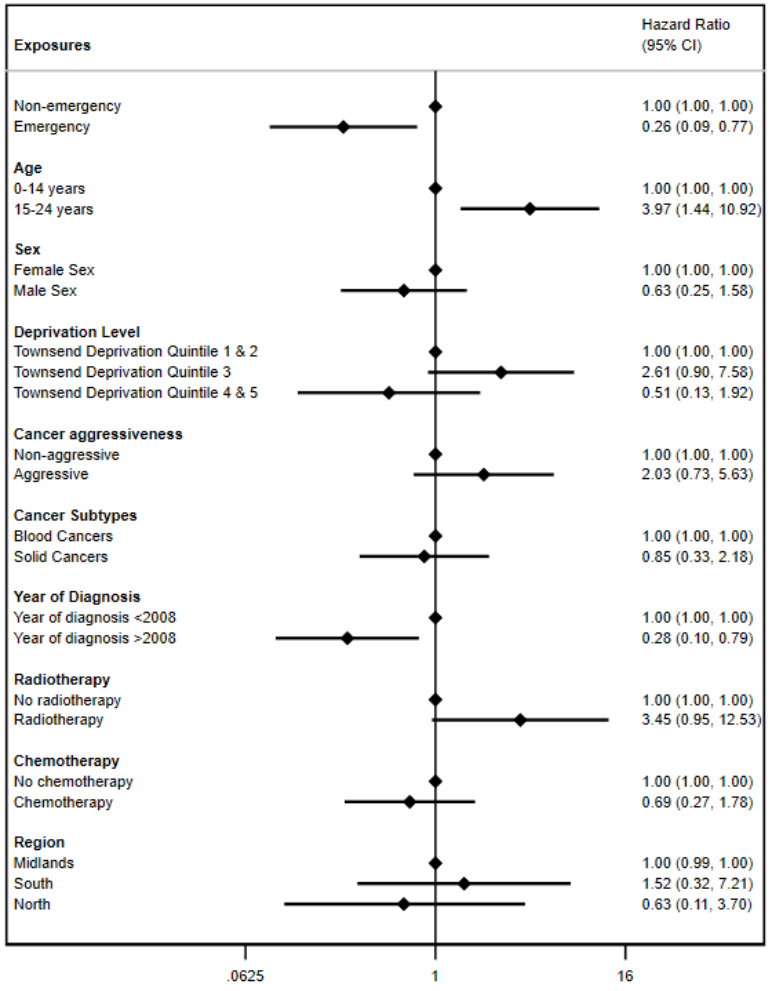
Emergency presentations were not associated with a higher risk of psychotic disorders in 1-year survivors and were associated with lower risk in 5-year survivors, however this was not statistically significant (p=0.02), following adjustment for sociodemographic (sex, ethnicity, deprivation level, region) and cancer-specific factors (age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressiveness) (Figure 7-9). Findings were similar following multiple imputation (Appendix Chapter 9, Supplementary Figure 9-10)

Figure 7-9 *Forest plot demonstrating hazard ratios (with 95% confidence intervals) of psychotic disorders for cancer-route and its association with psychotic disorders following Cox proportional regression in a) 1-year survivors and b) 5-year survivors.*

a)



b)



7.5. Discussion

7.5.1 Summary of Principal Findings

This Chapter explored mental health outcomes of long-term CTYA cancer survivors and identified that there was an increased risk of anxiety-depressive as well as psychotic disorders in cancer survivors compared to the control population. Whilst diagnostic pathways did not have a clear association with mental health outcomes in CTYA cancer survivors, in exploratory analyses, those who have longer diagnostic intervals when the interval is restricted to up to 5 days, were at increased risk. Other factors found to be associated with increased risk of mental health outcomes were cancer subtype, cancer treatment and year-of-diagnosis, as well as older age-at-diagnosis compared to younger age-at-diagnosis, female sex compared to male sex, and higher deprivation level compared to lower deprivation level,.

7.5.2 Findings in Context

The mental health of CTYA cancer survivors is a growing area of survivorship research and is being studied across population-based survivor cohorts internationally [156]. A recent meta-analysis summarising this evidence has shown a higher risk of anxiety, depression and psychosis in CTYA cancer survivors compared to both sibling and general population controls, as early as 1-year after diagnosis [156]. Findings of this study strongly reflect reports of the meta-analysis.

Given CTYA cancer survivors are at increased risk of mental health outcomes, another important factor that requires addressing is identifying which of these survivors are at increased risk. Current understanding of those at risk of poor mental health is limited, with factors such as female sex, having radiotherapy or a brain tumour thought to play a role [156]. Only one UK-based study, using the BRIGHLIGHT cohort, has explored diagnostic delays as a potential risk factor, and identified higher risk of both anxiety and depression in those who

had multiple presentations to GPs but not those with overall longer diagnostic intervals, but these intervals were defined in broad categories (<91 days vs >91 days) only [16].

Findings from this Chapter, similarly to mortality outcomes studied in Chapter 6, however, suggest that the relationship between diagnostic intervals and mental health outcomes is complex and non-linear. For example, this Chapter identified a potential subset of young people who present with a very short diagnostic interval (0-5 days) who may have a higher risk of anxiety-depressive disorders as their diagnostic interval lengthens, after adjusting for other relevant factors such as cancer subtype, biology and treatment. This may suggest that in these group of young people, perhaps with more advanced disease at diagnosis, delay plays a more significant impact on their long-term anxiety/depression outcomes. However, findings remain exploratory and larger studies are warranted before drawing conclusions.

There have been no studies to date that explore the association between emergency presentations and poorer mental health outcomes in adult or CTYA cancer survivors. Findings of this Chapter did not identify a statistically significant link between emergency presentations and anxiety-depressive or psychotic disorder risk in 1- or 5-year survivors.

7.5.3 Strengths and Limitations

The main strength of the study in this Chapter is the availability of linked data from primary care, which allows to 1) better define mental health outcomes (as they most commonly present to and are managed in primary care) and 2) integrate information on family history, which is strongly associated with mental health disorders [157], whilst also providing information on the pre-diagnosis period for the young person. Furthermore, having a representative population-based cohort of CTYA cancer survivors provides further opportunities to comprehensively study the relationship between diagnostic pathways and mental health outcomes.

However, this study has several limitations that need to be highlighted. Firstly, given the rarity of CTYA cancers, exploring the association between mental health outcomes and diagnostic pathways was limited by small sample sizes. Secondly, as highlighted in Chapters 5 and 6, there are limitations in the way in which systemic diagnostic intervals are constructed using EHR, even though this approach has been used in primary care-based cancer diagnostic interval studies [41]. Thirdly, there was missing data in several exposures-of-interest, including diagnostic interval and emergency presentations. This was accounted for with multiple imputation. Finally, there was a lack of availability of other important potential confounders, particularly accurate information on coping mechanisms, relationship status or employment, which are known to have strong associations with mental health disorders as well as survival from cancer in young people [160, 161].

7.5.4 Implications and Future Directions

Findings of this Chapter suggest that CTYA cancer survivors are at higher risk of mental health disorders compared to their peers. This provides further evidence for the need to provide emotional support to young people following their cancer treatment and highlights the need to identify those at highest risk. Although diagnostic pathways did not appear to be a definitive risk factor for mental health disorders in survivors, this could be limited by sample size and further studies are required to explore this. However, this Chapter identified several other risk factors, such as radiotherapy and cancer subtype, which can be used to risk stratify young people at a more personalised level and provide them with tailored long-term support.

Code from Chapter

https://github.com/DefOx2020/Thesis/blob/06e6b1c26357ad03b73a6fc065ef8af9040ef1b4/Chapter%207

Chapter 8 Discussion

Diagnostic delay of cancer is a common problem in CTYA and has potentially serious consequences [7]. In order to achieve earlier CTYA cancer diagnosis, a better understanding of their diagnostic pathway, from the earliest presentation to the confirmation of diagnosis, is required to overcome challenges that arise from the rarity, diversity and non-specificity of cancer in this age group. Through generating a nationally-representative and population-level EHR cohort of young people between the ages of 0 and 24, this DPhil addresses key gaps in the understanding of this diagnostic pathway. It advances knowledge of presenting clinical features of cancer; identifies person-, cancer- and healthcare-dependent factors that influence diagnostic pathways, and better characterises risk of cancer in primary care through the development of statistical risk prediction models. It also provides insight into how delays may be linked to poorer outcomes, in order to emphasise the importance of achieving early diagnosis.

8.1. Summary of Findings and Implications

8.1.1 Developing a platform to study CTYA cancer diagnostic pathways using EHR

One research gap that has limited early diagnosis research in CTYA cancer is the absence of large, population-level, representative cohorts of CTYA with comprehensive follow up. Chapter 2 of this DPhil addresses this gap by generating a CTYA cancer cohort that is representative of the general population in England, through leveraging routinely-collected, linked EHR available on the QResearch Database. Furthermore, it provides evidence that the clinical codes available on the three linked datasets (the primary care, the secondary care and the National Cancer Registry datasets) can comprehensively and accurately capture cancer diagnoses within this age group. Overall, this representative cohort provides opportunities to

draw generalizable conclusions about diagnostic pathways of CTYA cancer, which can support future decision-making in clinical practice and in policy. Future research using this cohort will also be feasible and given it is EHR-based, can be easily updated to provide the most contemporary information.

8.1.2 Advancing knowledge of presenting clinical features to better recognise CTYA cancer

Given the relative rarity of CTYA cancers, clinical presentations, particularly encountered at the earliest stages, are not well-characterised and this restricts early recognition [17]. To address this gap in Chapter 3, a nested case-control study was carried out to identify clinical features and feature-combinations encountered in primary care up to 6 months prior to a subsequent cancer diagnosis. Clinical features were studied initially across all cancers together, then across childhood and TYA age groups and finally in individual cancer types. This study was able to advance knowledge of clinical presentations of CTYA cancer in primary care, in further depth than previous studies. First, cancer-relevant clinical features in primary care were ascertained, a significant proportion of which had not previously been reported to be relevant to primary care [25, 26, 56]. Second, using the median diagnostic interval as cut-off, clinical features relevant to the earliest time-points of the diagnostic interval were identified. Third, relevant clinical feature combinations were identified using either a clinically-driven or an unsupervised statistical method (principal component analysis) approach. Overall, this information provides more comprehensive insight into primary care presentations of CTYA cancers, which can be integrated into healthcare guidance, such as the currently limited NICE guidance [78], or awareness initiatives, such as ChildrenCancerSmart, which aims to extend the HEADSMART initiative from just brain tumours to all childhood cancers [162]. Ultimately,

findings could have an important role in improving clinical and public awareness of CTYA cancers to aid early diagnosis.

8.1.3 Developing a statistical model to ascertain the risk of CTYA cancer in primary care

In addition to being uncommon, CTYA cancers are also diverse and as identified in Chapter 3, present with clinical features that are non-specific and of low predictive value. Overall, this makes assessment of cancer risk very challenging and awareness initiatives such as HEADSMART have not been able to fully overcome this challenge [28]. Accordingly, Chapter 4 focuses on identifying an alternative way to ascertain and stratify cancer risk in CTYA, through the development of statistical risk prediction models. Utilising information about the sociodemographic background of CTYA and relevant clinical features identified in Chapter 3, Chapter 4 derives and internally-externally validates statistical models for combined cancers (haematological and solid cancers) as well as, where sample size is sufficient, models for individual cancers (leukaemias, lymphomas and CNS tumours).

Overall, models have good discriminative ability, are well-calibrated, and show potential for clinical utility. They also provide insight into model fairness and potential sources of heterogeneity in the model. Thus, these models provide a promising starting point for the use of risk prediction tools in CTYA cancers and have the important potential for better stratifying cancer risk in primary care in order to aid earlier detection and referral.

8.1.4 Factors that influence diagnostic pathways

Diagnostic pathways of CTYA cancer have been reported to include repeated attendances, and prolonged diagnostic intervals, all pointing towards protracted delays [21, 25, 26, 56, 102]. Factors that contribute to these delays have been shown to include older age, female sex, cancer subtype and medical speciality consulted, however, comprehensive exploration of all relevant factors in CTYA cancer diagnostic pathways is lacking. Chapter 5 addressed this

research gap by initially identifying potentially relevant factors through a literature review and subsequently investigating how these factors affect diagnostic pathways. The two quantifiable components of the diagnostic pathway explored in this Chapter were systemic diagnostic intervals, using quantile regression analyses, and diagnostic routes, through logistic regression analyses.

Several factors, including person-level (i.e., age-at-diagnosis), cancer-specific (i.e., cancer subtype, presenting features and cancer biology) and health-system specific (i.e., presence of investigative test requests as well as regional differences) were associated with variations in diagnostic pathways. These results provide further support for findings of previous studies, such as older age being associated with longer diagnostic intervals [23], but also provide more comprehensive understanding of how tumour biology, such as histological aggressiveness, and other healthcare factors, such as blood test requests, can affect diagnostic pathways. Specifically, analyses during the COVID-19 pandemic provide an interesting insight into how reduced access to healthcare services can influence diagnostic pathways in subsequent time periods.

Findings of Chapter 5 also highlight the complexity of diagnostic pathways in CTYA cancers. Specifically, delayed diagnosis does not necessarily equate to the longest diagnostic intervals, given the heterogeneity of tumour subtype and biology, and that using the length of the diagnostic interval as a sole indicator for delay could result in inferring inaccurate conclusions. A more refined approach, considering the characteristics of cancer (e.g., histology and subtype) is required.

Importantly, Chapter 5 provides the first study to explore factors associated with diagnostic routes in CTYA cancer. Findings offer novel insight, in particular, regional differences identified in this study, may play an important role in future healthcare planning and policymaking decisions.

8.1.5 Diagnostic pathways and subsequent cancer burden

Delays in the diagnosis of symptomatic cancer have been shown to worsen cancer burden in adults [12]. The impact of delays on cancer burden in young people is less well-established and less well-studied. This lack of evidence has important clinical and policy-making implications, and limits support for further early cancer diagnosis research. Chapters 6 and 7 address this gap using time-to-event analyses. Chapter 6 explores the association between diagnostic pathways and mortality at 1- and 5- year following diagnosis. Chapter 7 explores how diagnostic pathways are linked to one morbidity outcome, mental health disorders, in 1- and 5-year survivors.

Both Chapters suggest in exploratory analyses, where the diagnostic interval is split into centile-based categories, that diagnostic delay may be associated with higher risk of mortality and morbidity. Although exploratory in nature, and thus limited in its interpretation, these analyses highlight the complexity of defining diagnostic delay through diagnostic intervals and provide evidence for why some prior studies, which use the length of the diagnostic interval as their sole indicator for delay without considering the heterogeneity of these intervals, may get conflicting results about associations and report the "waiting time paradox" [133].

Interestingly, Chapter 6 shows a statistically significant association between emergency presentations and subsequent mortality in CTYA cancer. This is the first population-based study to explore this association in this age group and findings have important healthcare planning indications and the following clinical implications: First, should reduction of emergency presentations in CTYA cancer, like adult cancers, be a top priority and if so, how can this be achieved? Second, should emergency presentation be included in prognostication and contribute to decisions around choice and length of treatment?

8.2. Strengths

This DPhil has several strengths, which provide opportunities to contribute to the advancement of early diagnosis research in CTYA. First, as detailed above, a large, population-based, nationally representative cohort of CTYA cancer with comparable incidence rates to national estimates was generated as part of this thesis. Drawing conclusions that are generalizable to the whole population of CTYA in England is therefore possible. This, in turn, can be used to develop national healthcare and policy strategies relevant to early cancer diagnosis research in young people. Another strength of this cohort is that it is derived from routinely collected EHR with high quality linked datasets. This reduces the risk of several sources of bias as follows: 1) selection bias that may arise from recruitment to prospective cohort studies or loss-to-follow up [163], 2) misclassification bias that may arise from under-ascertainment and 3) information bias that may occur from different recall of events and differential recording of events by investigators [164, 165], specifically in questionnaire-based designs. The third strength of this thesis is the use and availability of data representative of a large spectrum of the patient journey (from first medical presentation to long-term outcomes) through the availability of high-quality linkages of EHR from primary care, secondary care and registry-based datasets. This allowed to comprehensively explore the relationship between clinical presentations at the start of the patient journey (available via the primary care dataset) and diagnostic pathways (available via the cancer registry dataset) as well as long-term outcomes (available via the secondary care and civil registry datasets), which has historically been difficult to achieve in this field. It also allowed for better ascertainment of outcomes across different datasets. The final strength of this thesis is that it addresses research gaps that are clinically-driven. Thus, findings may be impactful to clinical practice and improve support to young people who are receiving a cancer diagnosis.

8.3. Limitations

Alongside these strengths, this DPhil also has several notable limitations that need to be highlighted.

8.3.1 The limitations of clinical coding and EHR

Although the use of routinely collected EHR has its advantages, it also has several disadvantages.

In the context of this thesis, one of the biggest limitations was on clinical coding for several reasons. First, as discussed in the validation study of Chapter 2, coding across different systems (e.g., between SNOMED-CT and ICD-O-3) may not be fully harmonised, and result in inconsistencies. This was seen, for example, with sarcoma diagnoses in Chapter 2, where "fibrohistiocytic neoplasm" on SNOMED-CT coding did not fully match to a benign or malignant ICD-O-3 code. Second, as clinical codes used in this study are collected as part of everyday clinical practice, there may be possibilities of non-differential misclassification through non-recording/missing data (e.g., clinician does not code a symptom picked up in consultation or ethnicity not recorded on records), or differential misclassification through over-recording (e.g., clinician worried about cancer and thus asks and records relevant symptoms/signs more often)[166]. These possibilities may bias results in different directions, with non-recording underestimating associations, and over-recording overestimating associations [166]. The possibility of over-recording/recording bias was explored in Chapter 3 and reassuringly was not detected in the nested case-control study. Non-differential misclassification and missing data were accounted for using multiple imputation throughout the thesis. Third, clinical codes used in EHR are standardised for comparability but are limited in granularity. For example, useful information such as duration and severity of a presenting feature, which may play an important role in discriminating cancer cases from non-cases and strengthen results of thesis, is not available.

A second drawback of routinely collected EHR is the limitations associated with demographic indices within datasets. An example of this is the Townsend Deprivation Score was used to define deprivation level. This score is limited due to its use of only a few markers of deprivation, such as car and house ownership, and is calculated at the area-level, potentially masking pockets of deprivation within neighbourhoods[65]. Additionally, the Townsend Score used is based on the 2011 Census, which limits its applicability to the more contemporary cohort (2010-2019) in this Thesis. Although other indices such as the Index of Multiple Deprivation more accurately capture deprivation level due to availability of more markers of deprivation, in this Thesis, the Townsend Deprivation Score was used due to being the only available deprivation index on QResearch Database.

Another limitation of routinely collected EHR is the absence of the patient's voice and perspective. In order to overcome this limitation, a patient representative panel was set up in this DPhil, where priorities of young people with cancer were discussed and wherever feasible, the thesis was tailored to address patient-relevant priorities. An example of this was in Chapters 3 and 4, where patient representatives were consulted about presenting features that could be included in studies and drew from their broader experiences.

A cancer-specific limitation that arose from using EHR was the lack of cancer stage data available for young people with cancer on the National Cancer Registry. Cancer stage provides information into how far the cancer has progressed and thus provides key insight into delay and prognosis [167]. This is further complicated by the fact that not all CTYA cancers use the "traditional" staging systems available for adults and on the Cancer Registry. An example of this is with leukaemia which quantifies the degree of advancement of cancer through a risk-status (a score based on blood test results and age) rather than a staging system available on the registry [168]. Therefore, there is a pressing need for further information at diagnosis on

the molecular/genetic status and blood results needed on the Cancer Registry datasets, in order to further advance early cancer diagnosis research in CTYA.

8.3.2 Limitations of study size

Another significant limitation of this thesis is the size of the CTYA cancer cohort. Although it represents one of the largest cohorts of CTYA cancer available, with 4,389 identified cancer cases, given the relative rarity of cancer in this age group, the size of the cohort makes it difficult to have enough power to study early diagnosis at a more granular level, e.g., across different histological subtypes or smaller age-bands. This limitation was highlighted in Chapter 4, where risk prediction models for rarer cancers, was not possible, as well as in Chapters 5, 6 and 7, where diagnostic intervals could not be explored with confidence in smaller cancer subgroups. This limitation is likely to resolve as more EHR data becomes available for research.

8.4. Future Directions

Findings from this thesis open up several future avenues of research in early diagnosis and beyond for young people with cancer.

Firstly, the CTYA cancer cohort generated in Chapter 2 offers a reliable study population that can be used by different researchers to ask many important questions relevant to young people with cancer. Its strength as a cohort for cancer research is only expected to increase given every year more data will be available through EHR and the cohort's power to answer research questions will improve. One future aim, for example, is to utilise this cohort to expand the study presented in Chapter 7 and to explore how elements of the diagnostic pathway are associated with other serious long-term comorbidities in cancer survivors.

Secondly, the next step for findings of Chapter 3 is to explore the value of integrating these relevant clinical features into awareness initiatives and national healthcare guidance. Especially findings of early clinical features prior to a cancer diagnosis raises important

research avenues about identifying the most tailored, least invasive approaches to utilise this information for early cancer diagnosis.

Thirdly, the risk prediction models developed in Chapter 4 represent a starting point for better risk stratification of CTYA cancer. Future research is required to undertake independent external validation to provide further evidence of the generalisability of the models alongside acceptability, cost-effectiveness and impact analyses in primary care.

Finally, findings of Chapter 5, 6, and 7 demonstrate the complexity of using the diagnostic interval as an indicator of diagnostic delay in CTYA cancers and highlights the future research need to refine the diagnostic interval to account for the heterogeneity/confounding that arises from cancer subtypes, cancer biology and possibly host-tumour interactions. As the EHR-derived CTYA cohort gets larger over time, it will be more and more possible to achieve this.

8.5. Concluding remarks

As highlighted in my motivation to study this topic, I started my DPhil journey with the two aims: identify ways to improve diagnosis in young people with cancer and understand the consequences of delayed diagnosis. Through this thesis, I believe I have taken important strides in addressing these aims. First, this work advances knowledge of how CTYA cancers present and how we can better differentiate them in clinical settings. Second, it sheds light into modifiable factors that may contribute to diagnostic delays. Finally, it provides useful insight into how diagnostic pathways may ultimately affect the long-term outcomes of young people with cancer. I hope that the evidence generated in this thesis will help young people, their families, clinicians and policymakers with their decision-making and drive forward further important research in the field.

Key Professional Development and Skills Acquisition

During my DPhil, I have developed several important professional and research skills to progress my clinical academic career forward through formal training, conference attendances and publications.

General Clinical Research Training: I completed 1-day courses on "Research ethics and research integrity", "Planning, designing and critically appraising clinical research" and "Experimental design" offered in-house to DPhil students at the University of Oxford.

Statistical Training: As part of my DPhil, I wanted to advance my skills in medical statistics and learn to use statistical packages and develop code. To achieve this, I initially attended an 8-week course on "Essential Medical Statistics" offered by the Department of Continuing Education, University of Oxford and subsequently completed a further 8-week course on "Advanced Statistical Computing with R" in the same department. I supplemented my skills further with 1 and 2-day courses offered in-house at the University of Oxford, specific to developing skills in the statistical software Stata ("Stata: A beginner's tour, Stata: Data manipulation and descriptive statistics, Stata: Statistical analysis and graphs") and data simulation ("Simulating data to improve your research").

Communication Skills: I developed important oral presentation and written communication skills during my DPhil. This was through a successful Cancer Research UK project grant application as a co-investigator in 2021; 15 peer-reviewed publications (4 as first-author) between 2021-2024; oral presentations in regional (Oxford Cancer Centre Annual Symposium 2023) and national conferences (National Cancer Research Institute Annual Conference, 2021; Cancer Research UK – Children and Young People's Cancer Conference 2023; Royal College of Paediatrics and Child Health Annual Conference 2024; Cambridge Centre for Data-Driven Discovery (C2D3) Workshop 2024).

Teaching Skills: I attended the teaching course offered by the University of Oxford "Preparing for learning and teaching at Oxford".

Leadership and Management Skills: In November 2023, I chaired a session in the national conference, Cancer Research UK – Children and Young People's Cancer Conference. This helped me develop important skills in time management and leadership.

Chapter 9 Supplementary Appendix

Supplementary Table 9-1 *ICD-10 codes used to define CTYA cancers in HES Dataset*

Cancer Type	**ICD-10 codes**
CNS Tumours	C70, C70.0, C70.1, C70.9, C71, C71.0, C71.1, C71.2, C71.3, C71.4, C71.5, C71.6, C71.7, C71.8, C71.9, C72, C72.0, C72.1, C72.2, C72.3, C72.4, C72.5, C72.6, C72.7, C72.8, C72.9, C79.3, C79.4
Lymphoma	C81,C81.0, C81.1, C81.2, C81.3, C81.4, C81.7,C81.9, C82, C82.0, C82.1, C82.2, C82.3, C82.4, C82.5, C82.6, C82.7,C82.9 C83, C83.0, C83.1, C83.3, C83.5, C83.7,C83.9, C84, C84.0, C84.1, C84.44, C81.3, C81.4, C81.7,C81.9 C85, C85.1, C85.2, C85.7, C85.9
Leukaemia	C91, C91.0, C91.1, C91.3, C91.4, C91.5, C91.6, C91.7, C91.8, C91.9, C92, C92.0, C92.1, C92.2, C92.3, C92.4, C92.5, C92.6, C92.7, C92.8, C92.9 C93, C93.0, C93.1, C93.3, C93.7, C93.9 C94, C94.0, C94.2, C94.3, C94.4, C94.6, C94.7 C95, C95.0, C95.1, C95.7, C95.9
Sarcoma	C40, C40.0, C40.1, C40.2, C40.3, C40.8, C40.9 C41, C41.0, C41.1, C41.2, C41.3, C41.4, C41.8, C41.9 C46, C46.1, C46.2, C46.3, C46.7, C46.8, C46.9 C49, C49.0, C49.1, C49.2, C49.3, C49.4, C49.5, C49.6, C49.8 C49.9
Abdominal Tumours	C64, C64.X, C65, C65.X, C74. 9, C22. 2
Gonadal germ cell tumours	C56, C62

Supplementary Table 9-2 SNOMED-CT codes used to define CTYA cancers in GP Dataset

Cancer Type	SNOMED ConceptID	SNOMED concept Fully Specified Name
Leukemias, myeloproliferative disease and myelodysplastic diseases	109995007	Myelodysplastic syndrome (disorder)
	110004001	Acute promyelocytic leukemia, FAB M3 (disorder)
	110005000	Acute myelomonocytic leukemia, FAB M4 (disorder)
	110006004	Prolymphocytic leukemia (disorder)
	110007008	Adult T-cell leukemia/lymphoma (disorder)
	127225006	Chronic myelomonocytic leukemia (disorder)
	128623006	Myelodysplastic neoplasm (morphologic abnormality)
	128831004	Chronic myelomonocytic leukemia (morphologic abnormality)
	128832006	Juvenile myelomonocytic leukemia (morphologic abnormality)
	128931003	Leukemia - category (morphologic abnormality)
	128934006	Myeloid leukemia - category (morphologic abnormality)
	17788007	Acute myeloid leukemia, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	188725004	Lymphoid leukemia (disorder)
	188726003	Subacute lymphoid leukemia (disorder)
	188728002	Aleukemic lymphoid leukemia (disorder)
	188729005	Adult T-cell leukemia (disorder)
	188732008	Myeloid leukemia (disorder)
	188733003	Chronic eosinophilic leukemia (disorder)
	188734009	Chronic neutrophilic leukemia (disorder)
	188736006	Subacute myeloid leukemia (disorder)
	188737002	Chloroma (disorder)
	188738007	Granulocytic sarcoma (disorder)
	188741003	Aleukemic myeloid leukemia (disorder)
	188768003	Myelomonocytic leukemia (disorder)
	188770007	Subacute myelomonocytic leukemia (disorder)
	255102004	Angioendotheliomatosis (disorder)
	277473004	B-cell chronic lymphocytic leukemia (disorder)
	277567002	T-cell prolymphocytic leukemia (disorder)
	277571004	B-cell acute lymphoblastic leukemia (disorder)
	277597005	Myelodysplastic syndrome with isolated del(5q) (disorder)
	277619001	B-cell prolymphocytic leukemia (disorder)
	28950004	Acute promyelocytic leukemia with PML::RARA fusion (morphologic abnormality)
	30962008	Acute myelomonocytic leukemia (morphologic abnormality)
	35287006	Myeloid sarcoma (morphologic abnormality)
	37810007	Myeloid leukemia, no International Classification of Diseases for Oncology subtype (morphologic abnormality)

	445227008	Juvenile myelomonocytic leukemia (disorder)
	445448008	Acute myeloid leukemia with myelodysplasia-related changes (disorder)
	445738007	Myelodysplastic/myeloproliferative disease (disorder)
	449108003	Philadelphia chromosome positive chronic myelogenous leukemia (disorder)
	449386007	Philadelphia chromosome negative chronic myelogenous leukemia (disorder)
	63364005	Chronic myelogenous leukemia, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	815361000000107	Acute myeloid leukaemia with 11q23 abnormality (disorder)
	863741000000108	Clinical stage A chronic lymphocytic leukaemia (disorder)
	863761000000109	Clinical stage B chronic lymphocytic leukaemia (disorder)
	863781000000100	Clinical stage C chronic lymphocytic leukaemia (disorder)
	91857003	Acute lymphoid leukemia, disease (disorder)
	91861009	Acute myeloid leukemia (disorder)
	92814006	Chronic lymphoid leukemia, disease (disorder)
	92818009	Chronic myeloid leukemia (disorder)
	94719007	Myeloid sarcoma, disease (disorder)

Cancer Type	**SNOMED Concept ID**	**SNOMED concept Fully Specified Name**
Lymphomas and Reticuloendothelial neoplasms	1090241000000100	Angioimmunoblastic T-cell lymphoma with dysproteinaemia (disorder)
	1091861000000100	Follicular lymphoma, cutaneous follicle centre (disorder)
	1091921000000100	B-cell non-Hodgkin's lymphoma (disorder)
	109962001	Diffuse non-Hodgkin's lymphoma (disorder)
	109965004	Diffuse non-Hodgkin's lymphoma, lymphoblastic (disorder)
	109966003	Diffuse non-Hodgkin's lymphoma, immunoblastic (disorder)
	109969005	Diffuse large B-cell malignant lymphoma (disorder)
	109975001	T-zone lymphoma (disorder)
	109976000	Lymphoepithelioid lymphoma (disorder)
	109977009	Peripheral T-cell lymphoma (disorder)
	109978004	T-cell lymphoma (disorder)
	109988003	Histiocytic sarcoma (disorder)
	112687003	Hodgkin lymphoma, lymphocyte depletion (morphologic abnormality)
	118599009	Hodgkin's disease (disorder)

Lymphomas and Reticuloendothelial neoplasms	118600007	Malignant lymphoma (disorder)
	118601006	Non-Hodgkin's lymphoma (disorder)
	118602004	Hodgkin's granuloma (disorder)
	118605002	Hodgkin lymphoma, nodular lymphocyte predominance (disorder)
	118606001	Hodgkin's sarcoma (disorder)
	118607005	Hodgkin lymphoma, lymphocyte-rich (disorder)
	118608000	Hodgkin's disease, nodular sclerosis (disorder)
	118609008	Hodgkin's disease, mixed cellularity (disorder)
	118610003	Hodgkin's disease, lymphocytic depletion (disorder)
	118611004	Sezary's disease (disorder)
	118612006	Malignant histiocytosis (disorder)
	118614007	Langerhans cell histiocytosis, disseminated (disorder)
	118615008	Malignant mast cell tumor (disorder)
	118617000	Burkitt's lymphoma (disorder)
	118618005	Mycosis fungoides (disorder)
	128799007	Hodgkin lymphoma, lymphocyte-rich (morphologic abnormality)
	128875000	Primary cutaneous CD30 antigen positive large T-cell lymphoma (disorder)
	128930002	Hodgkin lymphoma - category (morphologic abnormality)
	129000002	Eosinophilic granuloma (disorder)
	14537002	Hodgkin lymphoma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	16893006	Hodgkin lymphoma, lymphocyte depletion, diffuse fibrosis (morphologic abnormality)
	188510001	Burkitt's lymphoma of lymph nodes of head, face and neck (disorder)
	188511002	Burkitt's lymphoma of intrathoracic lymph nodes (disorder)
	188512009	Burkitt's lymphoma of intra-abdominal lymph nodes (disorder)
	188513004	Burkitt lymphoma of lymph nodes of axilla and upper limb (disorder)
	188514005	Burkitt lymphoma of lymph nodes of inguinal region and lower limb (disorder)
	188515006	Burkitt's lymphoma of intrapelvic lymph nodes (disorder)
	188516007	Burkitt's lymphoma of spleen (disorder)
	188517003	Burkitt's lymphoma of lymph nodes of multiple sites (disorder)
	188524002	Hodgkin's paragranuloma of intrathoracic lymph nodes (disorder)

Lymphomas and Reticuloendothelial neoplasms	188529007	Hodgkin's paragranuloma of intrapelvic lymph nodes (disorder)
	188531003	Hodgkin's paragranuloma of lymph nodes of multiple sites (disorder)
	188534006	Hodgkin's granuloma of lymph nodes of head, face and neck (disorder)
	188536008	Hodgkin's granuloma of intra-abdominal lymph nodes (disorder)
	188537004	Hodgkin's granuloma of lymph nodes of axilla and upper limb (disorder)
	188538009	Hodgkin's granuloma of lymph nodes of inguinal region and lower limb (disorder)
	188541000	Hodgkin's granuloma of lymph nodes of multiple sites (disorder)
	188544008	Hodgkin's sarcoma of lymph nodes of head, face and neck (disorder)
	188547001	Hodgkin's sarcoma of lymph nodes of axilla and upper limb (disorder)
	188548006	Hodgkin's sarcoma of lymph nodes of inguinal region and lower limb (disorder)
	188551004	Hodgkin's sarcoma of lymph nodes of multiple sites (disorder)
	188554007	Hodgkin's disease, lymphocytic-histiocytic predominance of lymph nodes of head, face and neck (disorder)
	188558005	Hodgkin's disease, lymphocytic-histiocytic predominance of lymph nodes of axilla and upper limb (disorder)
	188559002	Hodgkin's disease, lymphocytic-histiocytic predominance of lymph nodes of inguinal region and lower limb (disorder)
	188562004	Hodgkin's disease, lymphocytic-histiocytic predominance of lymph nodes of multiple sites (disorder)
	188565002	Hodgkin's disease, nodular sclerosis of lymph nodes of head, face and neck (disorder)
	188566001	Hodgkin's disease, nodular sclerosis of intrathoracic lymph nodes (disorder)
	188567005	Hodgkin's disease, nodular sclerosis of intra-abdominal lymph nodes (disorder)
	188568000	Hodgkin's disease, nodular sclerosis of lymph nodes of axilla and upper limb (disorder)
	188569008	Hodgkin's disease, nodular sclerosis of lymph nodes of inguinal region and lower limb (disorder)
	188570009	Hodgkin's disease, nodular sclerosis of intrapelvic lymph nodes (disorder)
	188572001	Hodgkin's disease, nodular sclerosis of lymph nodes of multiple sites (disorder)

Lymphomas and Reticuloendothelial neoplasms	188575004	Hodgkin's disease, mixed cellularity of lymph nodes of head, face and neck (disorder)
	188576003	Hodgkin's disease, mixed cellularity of intrathoracic lymph nodes (disorder)
	188577007	Hodgkin's disease, mixed cellularity of intra-abdominal lymph nodes (disorder)
	188578002	Hodgkin's disease, mixed cellularity of lymph nodes of axilla and upper limb (disorder)
	188579005	Hodgkin's disease, mixed cellularity of lymph nodes of inguinal region and lower limb (disorder)
	188580008	Hodgkin's disease, mixed cellularity of intrapelvic lymph nodes (disorder)
	188582000	Hodgkin's disease, mixed cellularity of lymph nodes of multiple sites (disorder)
	188585003	Hodgkin's disease, lymphocytic depletion of lymph nodes of head, face and neck (disorder)
	188586002	Hodgkin's disease, lymphocytic depletion of intrathoracic lymph nodes (disorder)
	188587006	Hodgkin's disease, lymphocytic depletion of intra-abdominal lymph nodes (disorder)
	188589009	Hodgkin's disease, lymphocytic depletion of lymph nodes of axilla and upper limb (disorder)
	188590000	Hodgkin's disease, lymphocytic depletion of lymph nodes of inguinal region and lower limb (disorder)
	188591001	Hodgkin's disease, lymphocytic depletion of intrapelvic lymph nodes (disorder)
	188592008	Hodgkin's disease, lymphocytic depletion of spleen (disorder)
	188593003	Hodgkin's disease, lymphocytic depletion of lymph nodes of multiple sites (disorder)
	188609000	Nodular lymphoma of lymph nodes of head, face and neck (disorder)
	188612002	Nodular lymphoma of lymph nodes of axilla and upper limb (disorder)
	188613007	Nodular lymphoma of lymph nodes of inguinal region and lower limb (disorder)
	188627002	Mycosis fungoides of lymph nodes of multiple sites (disorder)
	188630009	Sezary's disease of lymph nodes of head, face and neck (disorder)
	188631008	Sezary's disease of intrathoracic lymph nodes (disorder)
	188632001	Sezary's disease of intra-abdominal lymph nodes (disorder)
	188633006	Sezary's disease of lymph nodes of axilla and upper limb (disorder)

Lymphomas and Reticuloendothelial neoplasms	188634000	Sezary's disease of lymph nodes of inguinal region and lower limb (disorder)
	188635004	Sezary's disease of intrapelvic lymph nodes (disorder)
	188637007	Sezary's disease of lymph nodes of multiple sites (disorder)
	188640007	Malignant histiocytosis of lymph nodes of head, face and neck (disorder)
	188641006	Malignant histiocytosis of lymph nodes of axilla and upper limb (disorder)
	188642004	Malignant histiocytosis of lymph nodes of inguinal region and lower limb (disorder)
	188645002	Leukemic reticuloendotheliosis of lymph nodes of head, face and neck (disorder)
	188648000	Leukemic reticuloendotheliosis of lymph nodes of axilla and upper limb (disorder)
	188649008	Leukemic reticuloendotheliosis of lymph nodes of inguinal region and lower limb (disorder)
	188660004	Malignant mast cell tumors (disorder)
	188662007	Mast cell malignancy of lymph nodes of head and neck (disorder)
	188663002	Mast cell malignancy of intrathoracic lymph nodes (disorder)
	188664008	Mast cell malignancy of intra-abdominal lymph nodes (disorder)
	188665009	Mast cell malignancy of lymph nodes of axilla and upper limb (disorder)
	188666005	Mast cell malignancy of lymph nodes of inguinal region and lower limb (disorder)
	188667001	Mast cell malignancy of intrapelvic lymph nodes (disorder)
	188668006	Malignant mast cell sarcoma of spleen (disorder)
	188669003	Mast cell malignancy of lymph nodes of multiple sites (disorder)
	188672005	Follicular non-Hodgkin's mixed small cleaved and large cell lymphoma (disorder)
	188674006	Diffuse malignant lymphoma - small non-cleaved cell (disorder)
	188675007	Malignant lymphoma - small cleaved cell (disorder)
	188676008	Malignant lymphoma - mixed small and large cell (disorder)
	188679001	Diffuse non-Hodgkin's lymphoma undifferentiated (diffuse) (disorder)
	188691005	Malignant immunoproliferative small intestinal disease (disorder)
	1929004	Non-Hodgkin lymphoma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)

Lymphomas and Reticuloendothelial neoplasms	269475001	Malignant tumor of lymphoid, hemopoietic AND/OR related tissue (disorder)
	269476000	Nodular lymphoma (disorder)
	275524009	Immunoproliferative neoplasm (disorder)
	277606000	Lymphoreticular tumor (disorder)
	277613000	Cutaneous/peripheral T-cell lymphoma (disorder)
	277615007	Low grade B-cell lymphoma (disorder)
	277622004	Mucosa-associated lymphoma (disorder)
	277625002	Follicular malignant lymphoma - small cleaved cell (disorder)
	277641001	Follicular malignant lymphoma - large cell (disorder)
	277654008	Enteropathy-associated T-cell lymphoma (disorder)
	302842009	Diffuse malignant lymphoma - centroblastic (disorder)
	308121000	Follicular lymphoma (disorder)
	39086001	Hodgkin lymphoma, nodular sclerosis, cellular phase (morphologic abnormality)
	397009000	Mast cell malignancy (disorder)
	400122007	Primary cutaneous T-cell lymphoma (disorder)
	404133000	Subcutaneous panniculitic cutaneous T-cell lymphoma (disorder)
	414166008	Extranodal natural killer/T-cell lymphoma, nasal type (disorder)
	41529000	Hodgkin lymphoma, mixed cellularity (morphologic abnormality)
	43985008	Hodgkin lymphoma, nodular sclerosis, grade 2 (morphologic abnormality)
	443487006	Mantle cell lymphoma (disorder)
	444910004	Primary mediastinal (thymic) large B-cell lymphoma (disorder)
	445105005	Blastic plasmacytoid dendritic cell neoplasm (disorder)
	445269007	Extranodal marginal zone B-cell lymphoma of mucosa-associated lymphoid tissue (disorder)
	445406001	Hepatosplenic T-cell lymphoma (disorder)
	446643000	Sarcoma of dendritic cells (accessory cells) (disorder)
	448212009	Anaplastic lymphoma kinase negative anaplastic large cell lymphoma (disorder)
	449220000	Diffuse follicle center lymphoma (disorder)
	45572000	Hodgkin lymphoma, nodular sclerosis, grade 1 (morphologic abnormality)
	46923007	Hodgkin sarcoma [obs] (morphologic abnormality)
	52248008	Hodgkin lymphoma, nodular sclerosis (morphologic abnormality)

Lymphomas and Reticuloendothelial neoplasms	539691000000104	Other malignant neoplasm of lymphoid and histiocytic tissue (disorder)
	65399007	Langerhans cell histiocytosis (disorder)
	70600005	Hodgkin lymphoma, nodular lymphocyte predominance (morphologic abnormality)
	71109004	Hodgkin lymphoma, lymphocyte depletion, reticular (morphologic abnormality)
	738770003	Anaplastic lymphoma kinase positive anaplastic large cell lymphoma (disorder)
	74189002	Hodgkin granuloma [obs] (morphologic abnormality)
	77381001	Burkitt lymphoma (morphologic abnormality)
	847481000000109	Follicular lymphoma grade 1 (disorder)
	847631000000107	Follicular lymphoma grade 2 (disorder)
	847651000000100	Follicular lymphoma grade 3 (disorder)
	847691000000108	Follicular lymphoma grade 3a (disorder)
	847701000000108	Follicular lymphoma grade 3b (disorder)
	847741000000106	Diffuse large B-cell lymphoma (disorder)
	853831000006109	Non hodgkin lymphoma (disorder)
	89796001	Mastocytoma (morphologic abnormality)
	907291000006103	[RFC] Hodgkin's disease (finding)
	93133006	Letterer-Siwe disease of intra-abdominal lymph nodes (disorder)
	93134000	Letterer-Siwe disease of intrapelvic lymph nodes (disorder)
	93135004	Letterer-Siwe disease of intrathoracic lymph nodes (disorder)
	93136003	Letterer-Siwe disease of lymph nodes of axilla AND/OR upper limb (disorder)
	93137007	Letterer-Siwe disease of lymph nodes of head, face AND/OR neck (disorder)
	93138002	Letterer-Siwe disease of lymph nodes of inguinal region AND/OR lower limb (disorder)
	93139005	Letterer-Siwe disease of lymph nodes of multiple sites (disorder)
	93140007	Letterer-Siwe disease of spleen (disorder)
	93144003	Leukemic reticuloendotheliosis of intra-abdominal lymph nodes (disorder)
	93145002	Leukemic reticuloendotheliosis of intrapelvic lymph nodes (disorder)
	93146001	Leukemic reticuloendotheliosis of intrathoracic lymph nodes (disorder)
	93150008	Leukemic reticuloendotheliosis of lymph nodes of multiple sites (disorder)
	93182006	Malignant histiocytosis of intra-abdominal lymph nodes (disorder)
	93183001	Malignant histiocytosis of intrapelvic lymph nodes (disorder)
	93184007	Malignant histiocytosis of intrathoracic lymph nodes (disorder)

Lymphomas and Reticuloendothelial neoplasms	93188005	Malignant histiocytosis of lymph nodes of multiple sites (disorder)
	93189002	Malignant histiocytosis of spleen (disorder)
	93191005	Malignant lymphoma of intra-abdominal lymph nodes (disorder)
	93192003	Malignant lymphoma of intrapelvic lymph nodes (disorder)
	93193008	Malignant lymphoma of intrathoracic lymph nodes (disorder)
	93194002	Malignant lymphoma of lymph nodes of axilla AND/OR upper limb (disorder)
	93195001	Malignant lymphoma of lymph nodes of head, face AND/OR neck (disorder)
	93196000	Malignant lymphoma of lymph nodes of inguinal region AND/OR lower limb (disorder)
	93197009	Malignant lymphoma of lymph nodes of multiple sites (disorder)
	93198004	Malignant lymphoma of spleen (disorder)
	93493001	Hodgkin's disease, lymphocytic-histiocytic predominance of intra-abdominal lymph nodes (disorder)
	93494007	Hodgkin's disease, lymphocytic-histiocytic predominance of intrapelvic lymph nodes (disorder)
	93495008	Hodgkin's disease, lymphocytic-histiocytic predominance of intrathoracic lymph nodes (disorder)
	93500006	Hodgkin's disease, lymphocytic-histiocytic predominance of spleen (disorder)
	93509007	Hodgkin's disease, mixed cellularity of spleen (disorder)
	93518009	Hodgkin's disease, nodular sclerosis of spleen (disorder)
	93520007	Hodgkin's disease of intra-abdominal lymph nodes (disorder)
	93521006	Hodgkin's disease of intrapelvic lymph nodes (disorder)
	93522004	Hodgkin's disease of intrathoracic lymph nodes (disorder)
	93523009	Hodgkin's disease of lymph nodes of axilla AND/OR upper limb (disorder)
	93524003	Hodgkin's disease of lymph nodes of head, face AND/OR neck (disorder)
	93525002	Hodgkin's disease of lymph nodes of inguinal region AND/OR lower limb (disorder)
	93526001	Hodgkin's disease of lymph nodes of multiple sites (disorder)
	93527005	Hodgkin's disease of spleen (disorder)
	93530003	Hodgkin's granuloma of intrapelvic lymph nodes (disorder)

Lymphomas and Reticuloendothelial neoplasms	93531004	Hodgkin's granuloma of intrathoracic lymph nodes (disorder)
	93536009	Hodgkin's granuloma of spleen (disorder)
	93541001	Hodgkin's paragranuloma of lymph nodes of axilla AND/OR upper limb (disorder)
	93542008	Hodgkin's paragranuloma of lymph nodes of head, face AND/OR neck (disorder)
	93543003	Hodgkin's paragranuloma of lymph nodes of inguinal region AND/OR lower limb (disorder)
	93545005	Hodgkin's paragranuloma of spleen (disorder)
	93547002	Hodgkin's sarcoma of intra-abdominal lymph nodes (disorder)
	93548007	Hodgkin's sarcoma of intrapelvic lymph nodes (disorder)
	93549004	Hodgkin's sarcoma of intrathoracic lymph nodes (disorder)
	93554008	Hodgkin's sarcoma of spleen (disorder)
	94707004	Mycosis fungoides of intra-abdominal lymph nodes (disorder)
	94708009	Mycosis fungoides of intrapelvic lymph nodes (disorder)
	94709001	Mycosis fungoides of intrathoracic lymph nodes (disorder)
	94710006	Mycosis fungoides of lymph nodes of axilla AND/OR upper limb (disorder)
	94711005	Mycosis fungoides of lymph nodes of head, face AND/OR neck (disorder)
	94712003	Mycosis fungoides of lymph nodes of inguinal region AND/OR lower limb (disorder)
	94714002	Mycosis fungoides of spleen (disorder)
	95186006	Nodular lymphoma of intra-abdominal lymph nodes (disorder)
	95187002	Nodular lymphoma of intrapelvic lymph nodes (disorder)
	95188007	Nodular lymphoma of intrathoracic lymph nodes (disorder)
	95192000	Nodular lymphoma of lymph nodes of multiple sites (disorder)
	95193005	Nodular lymphoma of spleen (disorder)
	95263006	Sezary's disease of spleen (disorder)

Cancer Type	SNOMED Concep-tID	SNOMED concept Fully Specified Name
CNS and miscella-neous intracranial and intraspinal neoplasms	115240006	Glioma (morphologic abnormality)
	127024001	Neoplasm of pituitary gland (disorder)
	127579001	Meningeal neoplasm (morphologic abnormality)
	128665000	Pituitary carcinoma (morphologic abnormality)
	128839002	Papillary ependymoma (morphologic abnormality)
	128854008	Pilocytic astrocytoma (morphologic abnormality)
	12943006	Polar spongioblastoma (morphologic abnormality)
	1586004	Subependymal giant cell astrocytoma (morphologic ab-normality)
	1623000	Myxopapillary ependymoma (morphologic abnormality)
	18021007	Choroid plexus papilloma, no International Classification of Diseases for Oncology subtype (morphologic abnor-mality)
	188292007	Malignant tumor of choroid plexus (disorder)
	188296005	Malignant neoplasm of medulla oblongata (disorder)
	188297001	Malignant neoplasm of midbrain (disorder)
	189164002	Cerebral meningioma (disorder)
	189167009	Spinal meningioma (disorder)
	189179009	Craniopharyngioma (disorder)
	189919000	Primitive polar spongioblastoma (morphologic abnormal-ity)
	21589007	Ependymoma, anaplastic (morphologic abnormality)
	24604009	Medullomyoblastoma (morphologic abnormality)
	253071005	Mixed oligoastrocytoma (morphologic abnormality)
	26138003	Gliomatosis cerebri (morphologic abnormality)
	3102004	Oligodendroglioma, anaplastic (morphologic abnormality)
	32456001	Desmoplastic medulloblastoma (morphologic abnormal-ity)
	35262004	Gliosarcoma (morphologic abnormality)
	35701008	Hemangioblastic meningioma (morphologic abnormality)
	363483004	Malignant tumor of pineal gland (disorder)
	38431002	Psammomatous meningioma (morphologic abnormality)
	38713004	Astrocytoma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	39781001	Primitive neuroectodermal tumor (morphologic abnor-mality)
	40009002	Craniopharyngioma (morphologic abnormality)
	409681000000102	[M]Meningiomas (morphologic abnormality)
	443354006	Ependymoma - category (morphologic abnormality)
	443565000	Oligodendroglioma - category (morphologic abnormality)
	44529004	Giant cell glioblastoma (morphologic abnormality)
	4553004	Subependymal glioma (morphologic abnormality)
	48952003	Astroblastoma (morphologic abnormality)

CNS and miscellaneous intracranial and intraspinal neoplasms	511008	Fibrous meningioma (morphologic abnormality)
	55094006	Protoplasmic astrocytoma (morphologic abnormality)
	55353007	Astrocytoma, anaplastic (morphologic abnormality)
	57606003	Clear cell meningioma (morphologic abnormality)
	63634009	Glioblastoma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	64967004	Transitional meningioma (morphologic abnormality)
	68944005	Meningothelial meningioma (morphologic abnormality)
	71314006	Fibrillary astrocytoma (morphologic abnormality)
	73676002	Peripheral neuroectodermal tumor (morphologic abnormality)
	73918009	Angiomatous meningioma (morphologic abnormality)
	73982001	Gemistocytic astrocytoma (morphologic abnormality)
	74532006	Glioma, malignant (morphologic abnormality)
	78303004	Meningioma, malignant (morphologic abnormality)
	78838008	Pleomorphic xanthoastrocytoma (morphologic abnormality)
	80061003	Oligodendroblastoma (morphologic abnormality)
	83217000	Medulloblastoma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	88252006	Choroid plexus carcinoma (morphologic abnormality)
	92683008	Carcinoma in situ of pituitary gland (disorder)
	94808000	Neoplasm of uncertain behavior of craniopharyngeal duct (disorder)
	95002008	Neoplasm of uncertain behavior of pineal gland (disorder)
	95004009	Neoplasm of uncertain behavior of pituitary gland (disorder)

Cancer Type	**SNOMED ConceptID**	**SNOMED concept Fully Specified Name**
Neuroblastoma and other peripheral nervous cell tumours	10438002	Glomus tumor (morphologic abnormality)
	115222007	Paraganglioma AND/OR glomus tumor (morphologic abnormality)
	13875003	Glomangiosarcoma (morphologic abnormality)
	253075001	Sympathicoblastoma (morphologic abnormality)
	29370006	Malignant pheochromocytoma (morphologic abnormality)
	30699005	Carotid body tumor (morphologic abnormality)
	31794002	Sympathetic paraganglioma (morphologic abnormality)
	32037004	Glomus jugulare tumor (morphologic abnormality)
	32512003	Extra-adrenal paraganglioma, malignant (morphologic abnormality)
	34550005	Glomangiomyoma (morphologic abnormality)
	51747000	Extra-adrenal paraganglioma (morphologic abnormality)
	52205002	Parasympathetic paraganglioma (morphologic abnormality)

	53320004	Aortic body tumor (morphologic abnormality)
	64059005	Spongioneuroblastoma (morphologic abnormality)
	69515008	Ganglioneuroblastoma (morphologic abnormality)
	72787006	Gangliocytic paraganglioma (morphologic abnormality)
	76060004	Olfactory neuroblastoma (morphologic abnormality)
	803009	Paraganglioma (morphologic abnormality)
	85583005	Pheochromocytoma (morphologic abnormality)
	87364003	Neuroblastoma (morphologic abnormality)
	89880005	Ganglioglioma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	9903002	Paraganglioma, malignant (morphologic abnormality)

Cancer Type	SNOMED Concept ID	SNOMED concept Fully Specified Name
Renal Tumours	126880001	Neoplasm of kidney (disorder)
	128860008	Juxtaglomerular tumor (morphologic abnormality)
	188252005	Malignant tumor of renal calyx (disorder)
	188253000	Malignant tumor of pelviureteric junction (disorder)
	24007003	Clear cell sarcoma of kidney (morphologic abnormality)
	25081006	Nephroblastoma (morphologic abnormality)
	302849000	Nephroblastoma (disorder)
	31470003	Adenosarcoma (morphologic abnormality)
	363457009	Malignant tumor of renal pelvis (disorder)
	363518003	Malignant tumor of kidney (disorder)
	41607009	Renal cell carcinoma (morphologic abnormality)

Cancer Type	SNOMED Concept ID	SNOMED concept Fully Specified Name
Hepatic Tumours	109841003	Liver cell carcinoma (disorder)
	109843000	Hepatoblastoma (disorder)
	187769009	Primary carcinoma of liver (disorder)
	189243008	Carcinoma in situ of hepatic duct (disorder)
	189246000	Carcinoma in situ of sphincter of Oddi (disorder)
	25370001	Hepatocellular carcinoma (morphologic abnormality)
	271525004	Carcinoma in situ of liver and/or biliary system (disorder)
	45024009	Hepatoblastoma (morphologic abnormality)
	907381000006106	[RFC] Liver cancer (finding)
	92530007	Carcinoma in situ of ampulla of Vater (disorder)
	92545000	Carcinoma in situ of biliary tract (disorder)
	92570000	Carcinoma in situ of common bile duct (disorder)
	92574009	Carcinoma in situ of cystic duct (disorder)
	92599005	Carcinoma in situ of gallbladder (disorder)
	92618006	Carcinoma in situ of intrahepatic bile ducts (disorder)
	92644006	Carcinoma in situ of liver (disorder)

Cancer Type	SNOMED Concept ID	SNOMED concept Fully Specified Name
Gonadal Germ Cell Tumours	18854008	Struma ovarii, malignant (morphologic abnormality)
	189849004	Dermoid cyst with malignant transformation (morphologic abnormality)
	19467007	Teratoma, malignant, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	21912003	Malignant teratoma, intermediate (morphologic abnormality)
	24327009	Struma ovarii (morphologic abnormality)
	28047004	Embryonal carcinoma (morphologic abnormality)
	28307001	Germinoma (morphologic abnormality)
	28325008	Polyembryoma (morphologic abnormality)
	32071008	Strumal carcinoid (morphologic abnormality)
	32844007	Mixed germ cell tumor (morphologic abnormality)
	42717009	Teratoma, benign (morphologic abnormality)
	443677002	Seminoma - category (morphologic abnormality)
	55818009	Teratoma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	60718004	Malignant dysgerminoma (morphologic abnormality)
	67830002	Teratocarcinoma (morphologic abnormality)
	72907003	Seminoma, anaplastic (morphologic abnormality)
	74409009	Endodermal sinus tumor (morphologic abnormality)
	74751003	Gonadoblastoma (morphologic abnormality)
	83292005	Malignant teratoma, undifferentiated (morphologic abnormality)
	9294008	Spermatocytic seminoma (morphologic abnormality)

Cancer Type	SNOMED Concept ID	SNOMED concept Fully Specified Name
Bone, Soft Tissue and Extraosseous Tumours	10069009	Giant cell tumor of bone, malignant (morphologic abnormality)
	109385007	Kaposi's sarcoma (disorder)
	109386008	Kaposi's sarcoma of skin (disorder)
	109388009	Kaposi's sarcoma of palate (disorder)
	109392002	Kaposi's sarcoma of multiple organs (disorder)
	109844006	Angiosarcoma of liver (disorder)
	11073003	Mixed liposarcoma (morphologic abnormality)
	112682009	Fibroma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	112683004	Pleomorphic liposarcoma (morphologic abnormality)
	112685006	Carcinosarcoma, embryonal (morphologic abnormality)

Bone, Soft Tissue and Extraosseous Tumours	115224008	Soft tissue tumor AND/OR sarcoma (morphologic abnormality)
	115225009	Fibromatous neoplasm (morphologic abnormality)
	115237006	Osseous AND/OR chondromatous neoplasm (morphologic abnormality)
	115243008	Granular cell tumor AND/OR alveolar soft part sarcoma (morphologic abnormality)
	12169001	Granular cell tumor (morphologic abnormality)
	12302002	Small cell osteosarcoma (morphologic abnormality)
	12622007	Clear cell sarcoma (except of Kidney M-89643) (morphologic abnormality)
	12690005	Fibroblastic osteosarcoma (morphologic abnormality)
	128918008	Parosteal osteosarcoma (morphologic abnormality)
	13238004	Granular cell tumor, malignant (morphologic abnormality)
	13583002	Mast cell sarcoma (morphologic abnormality)
	14269005	Embryonal rhabdomyosarcoma (morphologic abnormality)
	14990007	Chondrosarcoma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	16090008	Myxoid leiomyosarcoma (morphologic abnormality)
	18588008	Synovial sarcoma, biphasic (morphologic abnormality)
	187821001	Angiosarcoma of spleen (disorder)
	187822008	Fibrosarcoma of spleen (disorder)
	188029000	Kaposi's sarcoma of soft tissue (disorder)
	19134004	Fascial fibrosarcoma (morphologic abnormality)
	20380000	Ameloblastic odontosarcoma (morphologic abnormality)
	20667008	Myosarcoma (morphologic abnormality)
	21708004	Osteosarcoma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	24007003	Clear cell sarcoma of kidney (morphologic abnormality)
	254601002	Sarcoma of liver (disorder)
	255067005	Sarcoma of bone and connective tissue (disorder)
	25889007	Benign fibrous histiocytoma (morphologic abnormality)
	26211003	Juxtacortical chondrosarcoma (morphologic abnormality)
	26496005	Atypical fibrous histiocytoma (morphologic abnormality)
	27092008	Ameloblastic fibrosarcoma (morphologic abnormality)
	27849002	Myxoid liposarcoma (morphologic abnormality)
	28655007	Liposarcoma, well differentiated (morphologic abnormality)
	302843004	Histiocytoma (morphologic abnormality)
	307576001	Osteosarcoma of bone (disorder)
	30924005	Rhabdomyosarcoma (morphologic abnormality)
	31470003	Adenosarcoma (morphologic abnormality)
	33681003	Osteosarcoma in Paget's disease of bone (morphologic abnormality)
	34360000	Fibrous histiocytoma, malignant (morphologic abnormality)
	37206003	Synovial sarcoma, spindle cell (morphologic abnormality)
	372151005	Sarcoma - category (morphologic abnormality)

Bone, Soft Tissue and Extraosseous Tumours	400007004	Fibrohistiocytic neoplasm - category (morphologic abnormality)
	400011005	Leiomyosarcoma - category (morphologic abnormality)
	42392001	Epithelioid leiomyosarcoma (morphologic abnormality)
	43296007	Round cell liposarcoma (morphologic abnormality)
	45187003	Abdominal fibromatosis (morphologic abnormality)
	47284001	Aggressive fibromatosis (morphologic abnormality)
	47623001	Soft tissue tumor, benign (morphologic abnormality)
	49430005	Liposarcoma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	49937004	Kaposi's sarcoma, morphology (morphologic abnormality)
	52040006	Infantile fibrosarcoma (morphologic abnormality)
	52399003	Fascial fibroma (morphologic abnormality)
	53305005	Periosteal fibroma (morphologic abnormality)
	53654007	Fibrosarcoma (morphologic abnormality)
	54443001	Fibrous mesothelioma, malignant (morphologic abnormality)
	55333008	Osteoblastoma (morphologic abnormality)
	56422000	Synovial sarcoma, epithelioid cell (morphologic abnormality)
	56565002	Mesenchymal chondrosarcoma (morphologic abnormality)
	59238007	Epithelioid sarcoma (morphologic abnormality)
	59583009	Embryonal sarcoma (morphologic abnormality)
	62383007	Mixed type rhabdomyosarcoma (morphologic abnormality)
	6250003	Fibromyxosarcoma (morphologic abnormality)
	63211008	Synovial sarcoma (morphologic abnormality)
	63264007	Carcinosarcoma (morphologic abnormality)
	63449009	Alveolar rhabdomyosarcoma (morphologic abnormality)
	65140001	Periosteal fibrosarcoma (morphologic abnormality)
	67280001	Dedifferentiated liposarcoma (morphologic abnormality)
	6842002	Desmoplastic fibroma (morphologic abnormality)
	70555003	Endometrial stromal sarcoma, high grade (morphologic abnormality)
	71666005	Osteoid osteoma (morphologic abnormality)
	73506006	Small cell sarcoma (morphologic abnormality)
	75622000	Myxoid chondrosarcoma (morphologic abnormality)
	76312009	Chondroblastic osteosarcoma (morphologic abnormality)
	76594008	Dermatofibrosarcoma (morphologic abnormality)
	76909002	Ewing's sarcoma (morphologic abnormality)
	77455004	Pleomorphic rhabdomyosarcoma (morphologic abnormality)
	78453009	Telangiectatic osteosarcoma (morphologic abnormality)
	83118000	Malignant rhabdoid tumor (morphologic abnormality)
	83612000	Osteoma, no International Classification of Diseases for Oncology subtype (morphologic abnormality)
	8664001	Fibromyxoma (morphologic abnormality)
	87992000	Giant cell sarcoma (except of Bone, M-92503) (morphologic abnormality)

	88195001	Alveolar soft part sarcoma (morphologic abnormality)
	9395006	Sarcomatosis (morphologic abnormality)
	9671003	Elastofibroma (morphologic abnormality)
	9801004	Spindle cell sarcoma (morphologic abnormality)

Supplementary Table 9-3 *National Cancer Registry Morphology Codes based on ICCC-3 (adapted from Public Health England, National Cancer Registry Analysis Services(1))*

Diagnostic group	Diagnostic sub group	Morphology codes	Behaviour codes
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	(a) Lymphoid leukaemias	9820, 9823, 9826, 9827, 9831-9836, 9940, 9948	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	(a) Lymphoid leukaemias	9811-9819, 9837	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	(a.1) Precursor cell leukaemias	9835, 9836	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	(a.1) Precursor cell leukaemias	9811-9819, 9837	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	(a.2) Mature B-cell leukaemias	9823, 9826, 9832, 9833, 9940	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	(b) Acute myeloid leukaemias	9840, 9861, 9865-9867, 9869-9874, 9877-9879, 9891, 9895-9898, 9910-9912, 9920, 9931	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	(c) Chronic myeloproliferative diseases	9863, 9875, 9876, 9960-9964	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	Ic (subset) Chronic myeloid leukaemia	9863, 9875, 9876	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	Ic (subset) Other chronic myeloproliferative diseases	9960-9964	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	(d) Myelodysplastic syndrome and other myeloproliferative diseases	9945, 9946, 9975, 9980, 9982-9987, 9989, 9991-9993	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	(d) (subset) Myelodysplastic syndrome	9980-9989, 9991-9993	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	(d) (subset) Juvenile myelomonocytic leukaemia & chronic myelomonocytic leukaemia	9945, 9946	3,6,9
I Leukaemias, myeloproliferative diseases, and myelodysplastic diseases	(e) Unspecified and other specified leukaemias	9800, 9801, 9805-9809, 9860, 9930, 9965-9968	3,6,9
II Lymphomas and reticuloendothelial neoplasms	(a) Hodgkin lymphomas	9650-9655, 9659, 9661-9665, 9667	3,6,9
II Lymphomas and reticuloendothelial neoplasms	(b)(c) Non-Hodgkin lymphomas (including Burkitt lymphoma)	9591, 9597, 9670, 9671, 9673, 9675, 9678-9680, 9684, 9687-9691, 9695, 9698-9702, 9705, 9708, 9709, 9712, 9714-9719, 9724-9729, 9731-9735, 9737, 9738, 9760-9762, 9764-9769, 9970, 9971	3,6,9
II Lymphomas and reticuloendothelial neoplasms	(b)(c) Non-Hodgkin lymphomas (including Burkitt lymphoma)	9811-9819, 9837	3,6,9
II Lymphomas and reticuloendothelial neoplasms	(b.1) Precursor cell lymphomas	9727-9729	3,6,9
II Lymphomas and reticuloendothelial neoplasms	(b.1) Precursor cell lymphomas	9811-9819, 9837	3,6,9

II Lymphomas and reticuloendothelial neoplasms	(b.2) Mature B-cell lymphomas (excluding Burkitt lymphoma)	9597, 9670, 9671, 9673, 9675, 9678-9680, 9684, 9688-9691, 9695, 9698, 9699, 9712, 9731-9735, 9737, 9738, 9761, 9762, 9764-9766, 9769, 9970, 9971	3,6,9
II Lymphomas and reticuloendothelial neoplasms	(b.3) Mature T-cell and NK-cell lymphomas	9700-9702, 9705, 9708, 9709, 9714-9719, 9724-9726, 9767, 9768	3,6,9
II Lymphomas and reticuloendothelial neoplasms	(b.4) Non-Hodgkin lymphomas, NOS	9591, 9760	3,6,9
II Lymphomas and reticuloendothelial neoplasms	(c) Burkitt lymphoma	9687	3,6,9
II Lymphomas and reticuloendothelial neoplasms	(d) Miscellaneous lymphoreticular neoplasms	9740-9742, 9749, 9750, 9755-9759	3,6,9
II Lymphomas and reticuloendothelial neoplasms	(e) Unspecified lymphomas	9590, 9596	3,6,10
III CNS and miscellaneous intracranial and intraspinal neoplasms	(a) Ependymomas and choroid plexus tumour	9383, 9390-9394, 9396	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(a.1) Ependymomas	9383, 9391-9394, 9396	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(a.2) (subset) Choroid plexus carcinoma	9390	3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(a.2) (subset) Choroid plexus papilloma	9390	0,1
III CNS and miscellaneous intracranial and intraspinal neoplasms	(b) Astrocytomas	9380	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(b) Astrocytomas	9384, 9400-9411, 9420-9425, 9440-9442, 9445	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(b.1) Pilocytic astrocytoma	9421	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(b.2) Other Astrocytomas	9380	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(b.2) Other Astrocytomas	9384, 9400-9411, 9420, 9422-9425, 9440-9442, 9445	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(c) Intracranial and intraspinal embryonal tumours	9470-9478, 9480, 9508	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(c) Intracranial and intraspinal embryonal tumours	9501-9504	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(c.1) Medulloblastomas	9470-9472, 9474-9477, 9480	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(c.2) CNS embryonal tumour NOS	9473, 9478	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(c.3) Medulloepithelioma	9501-9504	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(c.4) Atypical teratoid/rhabdoid tumour	9508	0,1,3,6,9

III CNS and miscellaneous intracranial and intraspinal neoplasms	(d) Other gliomas	9380	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(d) Other gliomas	9381, 9382, 9385,9430-9432, 9444, 9450, 9451, 9460	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(d.1) Oligodendrogliomas	9450, 9451, 9460	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(d.2) Mixed and unspecified gliomas	9380	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(d.2) Mixed and unspecified gliomas	9382, 9385	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(d.3) Neuroepithelial glial tumours of uncertain origin	9381, 9430-9432, 9444	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(e) Other specified intracranial and intraspinal neoplasms	8270-8281, 8300, 9350-9352, 9360-9362, 9395, 9412, 9413, 9492, 9493, 9505-9507, 9509, 9530-9539, 9582	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(e.1) Pituitary adenomas and carcinomas	8270-8281, 8300	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(e.2) Tumours of the sellar region (craniopharyngiomas)	9350-9352, 9582	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(e.3) Pineal parenchymal tumours	9360-9362, 9395	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(e.4) Neuronal and mixed neuronal-glial tumours	9412, 9413, 9492, 9493, 9505-9507, 9509	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(e.5) Meningiomas	9530-9539	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(f) Unspecified intracranial and intraspinal neoplasms	8000-8005	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(f) Unspecified intracranial and intraspinal neoplasms - (subset) Brain stem gliomas	9380-9382, 9384, 9385, 9400-9411, 9420-9425, 9430-9432, 9440-9442, 9444, 9450, 9451, 9460	0,1,3,6,9
III CNS and miscellaneous intracranial and intraspinal neoplasms	(f) Unspecified intracranial and intraspinal neoplasms - (subset) Optic glioma tumours	9380, 9384, 9400-9411, 9420-9425, 9440-9442	0,1,3,6,9
IV Neuroblastoma and other peripheral nervous cell tumours	(a) Neuroblastoma and ganglioneuroblastoma	9490, 9500	3,6,9
IV Neuroblastoma and other peripheral nervous cell tumours	(b) Other peripheral nervous cell tumours	8680-8683, 8690-8693, 8700, 9520-9523	3,6,9
IV Neuroblastoma and other peripheral nervous cell tumours	(b) Other peripheral nervous cell tumours	9501-9504	3,6,9
V Retinoblastoma	No sub group	9510-9514	3,6,9
VI Renal tumours	(a) Nephroblastoma and other nonepithelial renal tumours	8959, 8960, 8964-8967	3,6,9
VI Renal tumours	(a) Nephroblastoma and other nonepithelial renal tumours	8963, 9364	3,6,9
VI Renal tumours	(a.1) Nephroblastoma (Wilms tumour)	8959, 8960	3,6,9

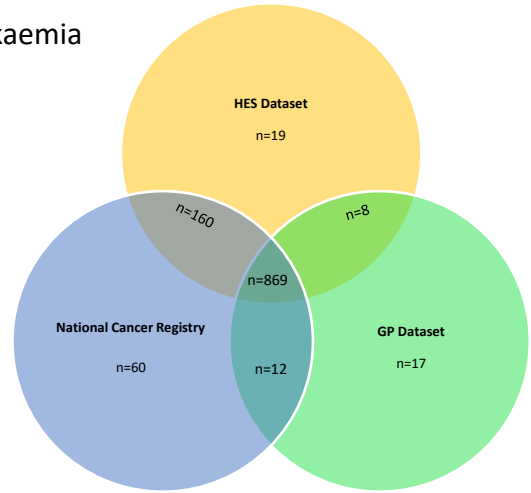
VI Renal tumours	(a.2) Rhabdoid renal tumour	8963	3,6,9
VI Renal tumours	(a.3) Kidney sarcomas	8964-8967	3,6,9
VI Renal tumours	(a.4) pPNET of kidney	9364	3,6,9
VI Renal tumours	(b) Renal carcinomas	8010-8041, 8050-8075, 8082, 8120-8122, 8130-8141, 8143, 8155, 8190-8201, 8210, 8211, 8221-8231, 8240, 8241, 8244-8246, 8260-8263, 8290, 8310, 8320, 8323, 8401, 8430, 8440, 8480-8490, 8504, 8510, 8550, 8560-8576	3,6,9
VI Renal tumours	(b) Renal carcinomas	8311, 8312, 8316-8319, 8361	3,6,9
VI Renal tumours	(c) Unspecified malignant renal tumours	8000-8005	3,6,9
VII Hepatic tumours	(a) Hepatoblastoma	8970	3,6,9
VII Hepatic tumours	(b) Hepatic carcinomas	8010-8041, 8050-8075, 8082, 8120-8122, 8140, 8141, 8143, 8155, 8190-8201, 8210, 8211, 8230, 8231, 8240, 8241, 8244-8246, 8260-8264, 8310, 8320, 8323, 8401, 8430, 8440, 8480-8490, 8504, 8510, 8550, 8560-8576	3,6,9
VII Hepatic tumours	(b) Hepatic carcinomas	8160-8180	3,6,9
VII Hepatic tumours	(c) Unspecified malignant hepatic tumours	8000-8005	3,6,9
VII Hepatic tumours	(a) Osteosarcomas	9180-9187, 9191-9195, 9200	3,6,9
VII Hepatic tumours	(b) Chondrosarcomas	9210, 9220, 9240	3,6,9
VII Hepatic tumours	(b) Chondrosarcomas	9221, 9230, 9241-9243	3,6,9
VII Hepatic tumours	(c) Ewing tumour and related sarcomas of bone	9260	3,6,9
VII Hepatic tumours	(c) Ewing tumour and related sarcomas of bone	9363-9365	3,6,9
VII Hepatic tumours	(d) Other specified malignant bone tumours	8810, 8811, 8814, 8823, 8830,	3,6,9
VII Hepatic tumours	(d) Other specified malignant bone tumours	8812, 9250, 9261, 9262, 9270-9275, 9280-9282, 9290, 9300-9302, 9310-9312, 9320-9322, 9330, 9340-9342, 9370-9372	3,6,9
VII Hepatic tumours	(e) Unspecified malignant bone tumours	8000-8005, 8800, 8801, 8803-8805	3,6,9
IX Soft tissue and other extraosseous sarcomas	(a) Rhabdomyosarcomas	8900-8905, 8910, 8912, 8920, 8991	3,6,9
IX Soft tissue and other extraosseous sarcomas	(a.1) Rhabdomyosarcomas - Embryonal	8910, 8991	3,6,9
IX Soft tissue and other extraosseous sarcomas	(a.2) Rhabdomyosarcomas - Alveolar	8920	3,6,9
IX Soft tissue and other extraosseous sarcomas	(a.3) Rhabdomyosarcomas - Other and unspecified	8900-8905, 8912	3,6,9
IX Soft tissue and other extraosseous sarcomas	(b) Fibrosarcomas, peripheral nerve sheath tumours, and other fibrous neoplasms	8810, 8811, 8813-8815, 8821, 8823, 8834-8835	3,6,9
IX Soft tissue and other extraosseous sarcomas	(b) Fibrosarcomas, peripheral nerve sheath tumours, and other fibrous neoplasms	8820, 8822, 8824-8827, 9150, 9160, 9491, 9540-9571, 9580	3,6,9
IX Soft tissue and other extraosseous sarcomas	(b.1) Fibroblastic and myofibroblastic tumours	8810, 8811, 8813-8815, 8821, 8823, 8834-8835	3,6,9

IX Soft tissue and other extraosseous sarcomas	(b.1) Fibroblastic and myofibroblastic tumours	8820, 8822, 8824-8827, 9150, 9160	3,6,9
IX Soft tissue and other extraosseous sarcomas	(b.2) Nerve sheath tumours	9540-9571	3,6,9
IX Soft tissue and other extraosseous sarcomas	(c)(d) Kaposis sarcoma and other specified soft tissue sarcomas	8587, 8710-8714, 8806, 8831-8833, 8836, 8840-8842, 8850-8858, 8860-8862, 8870, 8880, 8881, 8890-8898, 8921, 8990, 9040-9044, 9120-9125, 9130-9133, 9135, 9136, 9137, 9140-9142, 9161, 9170-9175, 9231, 9251, 9252, 9373, 9581	3,6,9
IX Soft tissue and other extraosseous sarcomas	(c)(d) Kaposis sarcoma and other specified soft tissue sarcomas	8830	3,6,9
IX Soft tissue and other extraosseous sarcomas	(c)(d) Kaposis sarcoma and other specified soft tissue sarcomas	8963	3,6,9
IX Soft tissue and other extraosseous sarcomas	(c)(d) Kaposis sarcoma and other specified soft tissue sarcomas	9180, 9210, 9220, 9240	3,6,9
IX Soft tissue and other extraosseous sarcomas	(c)(d) Kaposis sarcoma and other specified soft tissue sarcomas	9260	3,6,9
IX Soft tissue and other extraosseous sarcomas	(c)(d) Kaposis sarcoma and other specified soft tissue sarcomas	9364	3,6,9
IX Soft tissue and other extraosseous sarcomas	(c)(d) Kaposis sarcoma and other specified soft tissue sarcomas	9365	3,6,9
IX Soft tissue and other extraosseous sarcomas	(d.1) & (d.2) Extraosseous Ewing sarcoma family tumours	9260	3,6,9
IX Soft tissue and other extraosseous sarcomas	(d.1) & (d.2) Extraosseous Ewing sarcoma family tumours	9365	3,6,9
IX Soft tissue and other extraosseous sarcomas	(d.1) & (d.2) Extraosseous Ewing sarcoma family tumours	9364	3,6,9
IX Soft tissue and other extraosseous sarcomas	(d.3) Extrarenal rhabdoid tumour	8963	3,6,9
IX Soft tissue and other extraosseous sarcomas	(d.5) Fibrohistiocytic tumours	8830	3,6,9
IX Soft tissue and other extraosseous sarcomas	(d.5) Fibrohistiocytic tumours	8831-8833, 8836, 9251, 9252	3,6,9
IX Soft tissue and other extraosseous sarcomas	(d.7) Synovial sarcomas	9040-9044	3,6,9
IX Soft tissue and other extraosseous sarcomas	(e) Unspecified soft tissue sarcomas	8800-8805	3,6,9
IX Soft tissue and other extraosseous sarcomas	(subset) Hepatic sarcomas	8800-8805, 8991	3,6,9
IX Soft tissue and other extraosseous sarcomas	(subset) Ewing sarcoma family of tumours	9260	3,6,9
IX Soft tissue and other extraosseous sarcomas	(subset) Ewing sarcoma family of tumours	9363-9365	3,6,9
IX Soft tissue and other extraosseous sarcomas	(subset) Ewing sarcoma family of tumours	9260	3,6,9
IX Soft tissue and other extraosseous sarcomas	(subset) Ewing sarcoma family of tumours	9365	3,6,9
IX Soft tissue and other extraosseous sarcomas	(subset) Ewing sarcoma family of tumours	9364	3,6,9
X Germ cell tumours, trophoblastic tumours, and neoplasms of gonads	(b) Malignant extracranial and extragonadal germ cell tumours	9060-9065, 9070-9072, 9080-9086, 9100-9105	3,6,9

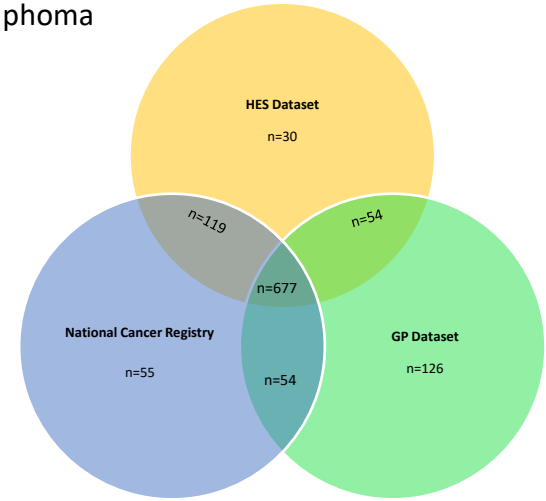
X Germ cell tumours, trophoblastic tumours, and neoplasms of gonads	(c) Malignant gonadal germ cell tumours	9060-9065, 9070-9073, 9080-9085, 9090-9093, 9100, 9101	3,6,9
X Germ cell tumours, trophoblastic tumours, and neoplasms of gonads	(d) Gonadal carcinomas	8010-8041, 8044, 8050-8075, 8082, 8120-8123, 8130-8141, 8143, 8190-8201, 8210, 8211, 8221-8241, 8244-8246, 8249, 8260-8263, 8290, 8310, 8313, 8320, 8323, 8380-8384, 8430, 8440, 8480-8490, 8504, 8510, 8550, 8560-8573, 9000, 9014, 9015	3,6,9
X Germ cell tumours, trophoblastic tumours, and neoplasms of gonads	(d) Gonadal carcinomas	8441-8444, 8450, 8451, 8460-8473	3,6,9
X Germ cell tumours, trophoblastic tumours, and neoplasms of gonads	(e) Other and unspecified malignant gonadal tumours	8590-8671	3,6,9
X Germ cell tumours, trophoblastic tumours, and neoplasms of gonads	(e) Other and unspecified malignant gonadal tumours	8000-8005	3,6,9

Supplementary Figure 9-1 *Venn diagrams depicting number of cancer diagnoses identified in each dataset for each cancer subtype.*

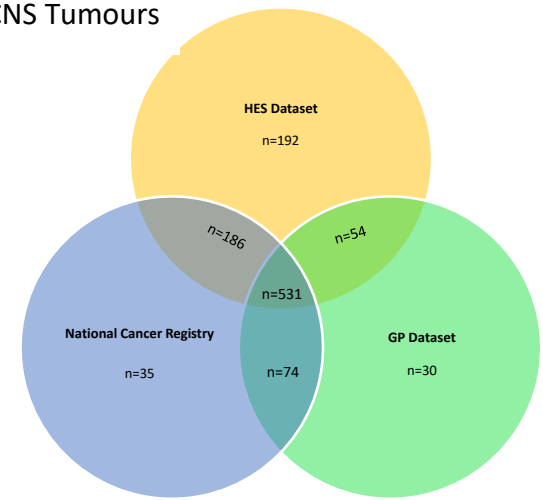
Leukaemia



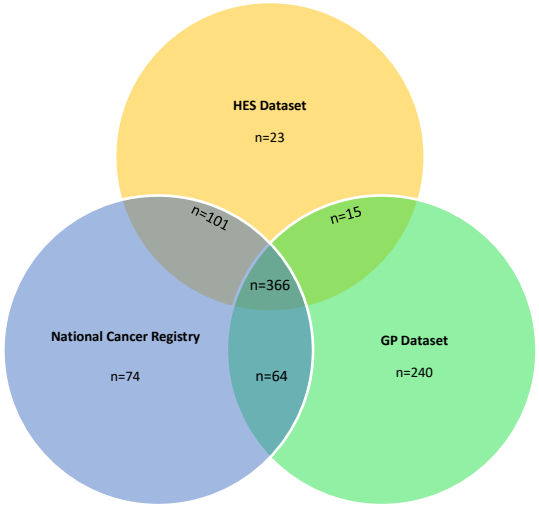
Lymphoma



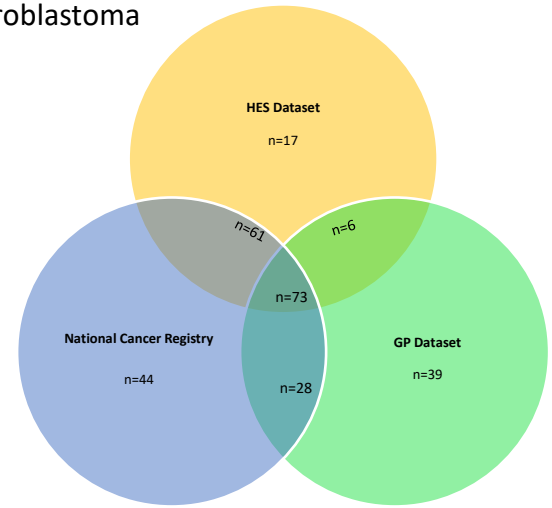
CNS Tumours



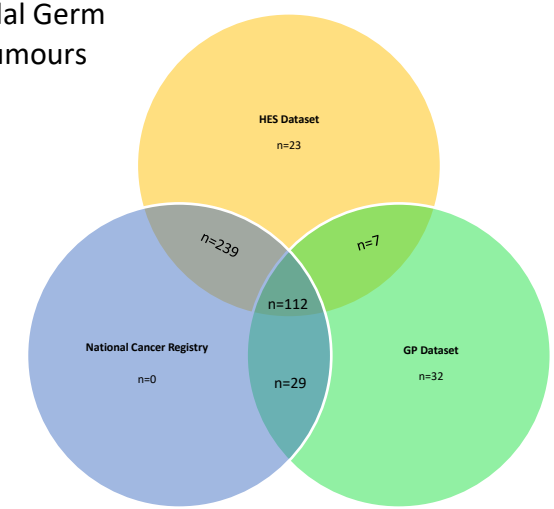
Sarcomas



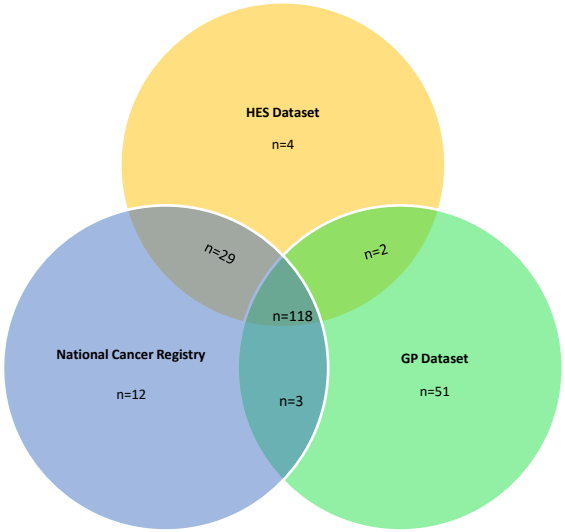
Neuroblastoma



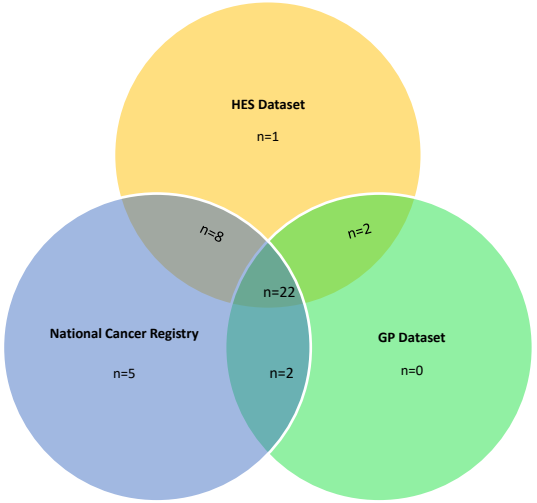
Gonadal Germ Cell Tumours



Renal Tumours



Hepatic Tumours



Supplementary Table 9-4 Measures of accuracy of SNOMED-CT/ICD-10 cancer codes using the National Cancer Registry as gold standard.

	Code Source	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
Leukaemias	SNOMED-CT	0.8	0.99	0.97	0.99
	ICD-10	0.93	0.99	0.92	0.99
Lymphomas	SNOMED-CT	0.83	0.99	0.94	0.97
	ICD-10	0.84	0.99	0.94	0.99
CNS Tumours	SNOMED-CT	0.72	0.99	0.77	0.99
	ICD-10	0.83	0.99	0.89	0.99
Sarcomas	SNOMED-CT	0.70	0.99	0.61	0.99
	ICD-10	0.74	0.99	0.89	0.99
Neuroblastomas	SNOMED-CT	0.51	0.99	0.66	0.99
	ICD-10	0.51	0.99	0.79	0.99
Renal Tumours	SNOMED-CT	0.77	0.99	0.75	0.99
	ICD-10	0.91	0.99	0.94	0.99
Hepatic tumours	SNOMED-CT	0.65	0.99	0.92	0.99

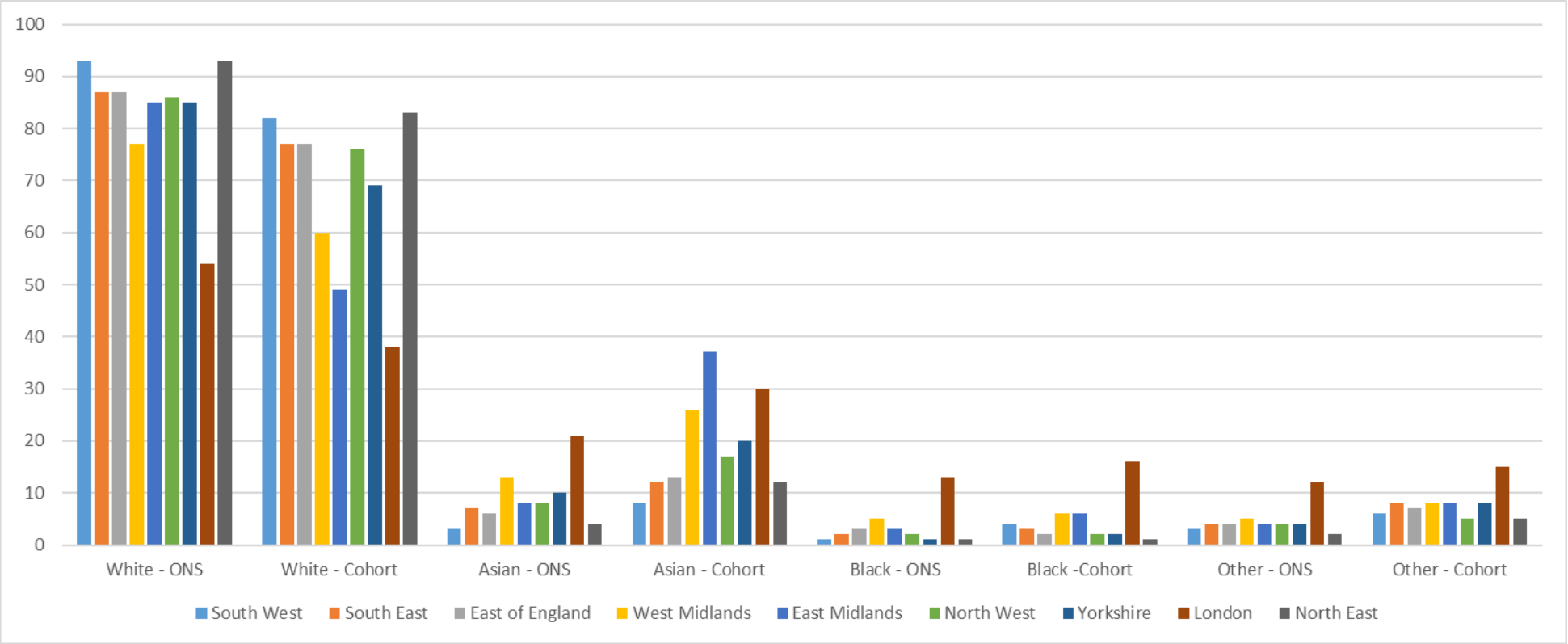
	ICD-10	0.81	0.99	0.91	0.99
Gonadal Germ Cell Tumours	SNOMED-CT	0.36	0.99	0.76	0.99
	ICD-10	0.93	0.99	0.92	0.99

Supplementary Table 9-5 Diagnostic codes of false positives identified in GP and HES datasets when National Cancer Registry is used as gold standard

	SNOMED-CT codes of false positives from GP dataset (n=564)		**ICD-10 codes of false positives from HES dataset (n=136)**
Leukaemias (n=17)	Acute lymphoid leukaemia Acute myeloid leukaemia B-cell acute lymphoblastic leukaemia Chronic lymphoid leukaemia Myelodysplastic syndrome	Leukaemias (n=19)	Acute lymphoid leukaemia Acute myeloblastic leukaemia Acute promyelocytic leukaemia Chronic lymphocytic leukaemia Lymphoid leukaemia Myeloid leukaemia
Lymphomas (n=30)	B-cell non-Hodgkin's Lymphoma Burkitt's Lymphoma Follicular non-Hodgkin's Lymphoma Hodgkin's Lymphoma Malignant Lymphoma Mycosis fungoides	Lymphomas (n=19)	B-cell lymphoma, unspecified Burkitt Lymphoma Diffuse large B-cell lymphoma Hodgkin lymphoma, unspecified Mature T/NK-cell lymphoma Non-Hodgkin lymphoma Other (classical) Hodgkin lymphoma
CNS Tumours (n=150)	Meningioma (angiomatous) Astrocytoma Pilocytic astrocytoma Cerebral meningioma Craniophryngioma Ependymoma	CNS Tumours (n=30)	Malignant neoplasm: Brain, unspecified Malignant neoplasm: Cerebellum Malignant neoplasm: Cerebral meninges Malignant neoplasm: Cerebrum Malignant neoplasm: Meninges, unspecified Malignant neoplasm: Occipital lobe

	Glioma (morphologic abnormality) Meningeal neoplasm Neoplasm of pituitary		Malignant neoplasm: Optic nerve Malignant neoplasm: Spinal cord
Sarcomas (n=245)	Rhabdomyosarcoma Dermatofibrosarcoma Fibrohistocytic neoplasm Granular cell tumour Kaposi's sarcoma Leiomyosarcoma Osteoblastoma (morphologic abnormality) Soft tissue tumour Synovial Sarcoma	Sarcomas (n=23)	Malignant neoplasm: Bone and articular Malignant neoplasm: Bones of skull Malignant neoplasm: Connective and soft tissue Malignant neoplasm: Long bones of lower limb Malignant neoplasm: Mandible Malignant neoplasm: Scapula and long bones of upper limb Malignant neoplasm: sort bones of upper limb
Neuroblastomas (n=39)	Neuroblastoma Paraganglioma	Neuroblastomas (n=12)	Malignant neoplasm: Peripheral nerves
Renal Tumours (n=51)	Adenoma/adenocarcinoma Nephroblastoma	Renal Tumours (n=4)	Malignant neoplasm of kidney
Hepatic Tumours (n=0)		Hepatic Tumours (n=1)	Malignant neoplasm: Liver cell carcinoma
Gonadal Germ Cell Tumours (n=32)	Teratoma Seminoma Germinoma Germ cell neoplasm N=32	Gonadal Germ Cell Tumours (n-23)	Malignant neoplasm of ovary Malignant neoplasm: Descended testis

Supplementary Figure 9-2 Proportion (%) of 4 ethnic categories for a) 2021 census population estimates b) QResearch CTYA cohort across 9 geographical regions in England (75).



Supplementary Table 9-6. *Definition of clinical features using SNOMED-CT clinical codes. Grouped into gastrointestinal, genitourinary, cardiorespiratory, neurological, systemic, haematological and musculoskeletal symptom groups.*

Clinical Feature	SNOMED Concept ID	SNOMED concept Fully Specified Name
Gastrointestinal Signs and Symptoms		
Abdominal Pain	102614006	Generalized abdominal pain (finding)
	162038003	Non-colicky abdominal pain (finding)
	162042000	Abdominal wall pain (finding)
	162046002	Central abdominal pain (finding)
	162049009	Left flank pain (finding)
	162050009	Right flank pain (finding)
	162051008	Right iliac fossa pain (finding)
	162052001	Left iliac fossa pain (finding)
	162053006	Suprapubic pain (finding)
	162147009	Complaining of pelvic pain (finding)
	1983921000006100	Manchester triage - Abdominal pain in child (finding)
	21522001	Abdominal pain (finding)
	247361008	Evening colic (finding)
	271857006	Loin pain (finding)
	271858001	Recurrent acute abdominal pain (finding)
	274671002	Pelvic and perineal pain (finding)
	301715003	Left upper quadrant pain (finding)
	301716002	Left lower quadrant pain (finding)
	301717006	Right upper quadrant pain (finding)
	301754002	Right lower quadrant pain (finding)
	304542004	Nonspecific abdominal pain (finding)
	308903002	On examination - epigastric pain on palpation (finding)
	35363006	Infantile colic (finding)
	43478001	Abdominal tenderness (finding)
	45979003	Abdominal wind pain (finding)
	51197009	Stomach cramps (finding)
	54586004	Lower abdominal pain (finding)
	79922009	Epigastric pain (finding)
	83132003	Upper abdominal pain (finding)
	853721000006104	Left loin pain (finding)
	88522004	Umbilical pain (finding)
	9209005	Acute abdomen (disorder)
	958291000006108	Colicky abdominal pain present (finding)
	9991008	Abdominal colic (finding)
Abdominal Mass	163278007	On examination - abdominal mass palpated (finding)
	163280001	On examination - abdominal mass less than 1 quadrant (finding)
	163281002	On examination - abdominal mass fills 1 quadrant (finding)

	163282009	On examination - abdominal mass fills half abdomen (finding)
	163283004	On examination - abdominal mass fills abdomen (finding)
	163285006	Right iliac fossa mass (finding)
	163290009	On examination - abdominal mass consistency (finding)
	163291008	On examination - abdominal mass - soft (finding)
	163292001	On examination - abdominal mass - hard (finding)
	163293006	On examination - abdominal mass-very hard (finding)
	163294000	On examination - pulsatile mass of abdomen (finding)
	163296003	On examination - abdominal mass shape (finding)
	163297007	On examination - abdominal mass - regular shape (finding)
	163298002	On examination - abdominal mass - irregular shape (finding)
	163302005	On examination - abdominal mass movement with respiration (finding)
	163303000	On examination - abdominal mass still with respiration (finding)
	163305007	On examination - abdominal mass -border defined (finding)
	163306008	On examination - abdominal mass - upper border defined (finding)
	163307004	On examination - abdominal mass - lower border defined (finding)
	163335000	On examination - tympany over abdominal mass (finding)
	163336004	On examination - dullness over abdominal mass (finding)
	163367002	On examination - pelvic mass palpated (finding)
	163368007	On examination - pelvic mass palpable-left iliac fossa (finding)
	163369004	On examination - pelvic mass palpable-right iliac fossa (finding)
	163370003	On examination - central pelvic mass (finding)
	1805241000006100	Right upper quadrant mass (finding)
	1805251000006100	Left upper quadrant mass (finding)
	1805271000006100	Right flank mass (finding)
	1805291000006100	Left flank mass (finding)
	271860004	Abdominal mass (finding)
	307135001	Left iliac fossa mass (finding)
	309082002	Epigastric mass (finding)
	312355005	On examination - left lower abdominal mass (finding)
	404200001	Central abdominal mass (finding)
	163254002	On examination - liver edge palpable (finding)
	163255001	On examination - liver moderately enlarged (disorder)
	163256000	On examination - liver grossly enlarged (disorder)
	163258004	On examination - liver character (finding)
	163259007	On examination - liver character normal (finding)

Organomegaly	163260002	On examination - liver soft (finding)
	163261003	On examination - liver firm (finding)
	163262005	On examination - liver very hard (finding)
	163263000	On examination - liver grossly irregular (finding)
	163264006	On examination - liver finely irregular (finding)
	163265007	On examination - liver smooth (finding)
	163266008	On examination - liver not painful (situation)
	163267004	On examination - liver tender (finding)
	163268009	On examination - liver pulsatile (finding)
	268943002	On examination - liver palpated (finding)
	275296001	On examination - hepatomegaly (disorder)
	407000	Congenital hepatomegaly (disorder)
	80515008	Large liver (disorder)
	16294009	Splenomegaly (disorder)
	164162005	On examination - splenomegaly (finding)
	164164006	On examination - spleen just palpable (finding)
	164165007	On examination - moderate splenomegaly (disorder)
	164166008	On examination - gross splenomegaly (disorder)
	164167004	On examination - painful splenomegaly (disorder)
	36760000	Hepatosplenomegaly (disorder)
Vomiting & Nausea	162057007	Nausea present (situation)
	167820005	Vomit examination (procedure)
	16932000	Nausea and vomiting (disorder)
	1777091000006100	Reason for referral: Vomiting/Nausea (finding)
	18773000	Cyclical vomiting syndrome (disorder)
	1959511000006100	24 hour vomitus output (finding)
	196746003	Persistent vomiting (disorder)
	249493007	Posseting (finding)
	249497008	Vomiting symptom (finding)
	267051003	Vomiting blood - fresh (disorder)
	269908006	Vomit: frank blood present (finding)
	271352005	Bilious vomit (finding)
	272044004	Complaining of vomiting (finding)
	275744000	Bilious vomit on examination (disorder)
	275786006	Blood in vomit on examination (finding)
	281102003	Blood in vomit - symptom (finding)
	300359004	Finding of vomiting (finding)
	300363006	Vomitus with coffee ground appearance (finding)
	40835002	Coffee ground vomiting (disorder)
	419219000	Drug-induced nausea and vomiting (disorder)
	422400008	Vomiting (disorder)
	422587007	Nausea (finding)
	71419002	Bilious vomiting (disorder)
	84480002	Retching (finding)
	8579004	Projectile vomiting (disorder)

	866121000000105	Frequency of vomiting (observable entity)
	8765009	Hematemesis (disorder)
	960221000006103	Nausea/vomiting (finding)
Abdominal Distension	162068007	Abdominal distension symptom (finding)
	248490000	Bloating symptom (finding)
	271835004	Abdominal distension, gaseous (finding)
Genito-urinary Symptoms/Signs		
Testicular abnormalities	163376009	On examination - scrotal swelling-transillumination (finding)
	163377000	On examination - scrotal swelling - no transillumination (disorder)
	163378005	On examination - swelling of testicle (finding)
	268946005	On examination - scrotal swelling (finding)
	390918003	On examination - right scrotal swelling (finding)
	390919006	On examination - left scrotal swelling (finding)
	438457000	Swelling of testicle (finding)
	773611000000105	Complaining of scrotal swelling (finding)
	87860000	Mass of testicle (finding)
Female reproductive system symptoms	386692008	Menorrhagia (finding)
	315224006	Heavy episode of vaginal bleeding (finding)
	275570002	Complaining of per vaginam bleeding (finding)
	301822002	Abnormal vaginal bleeding (finding)
Haematuria	167235001	Urine: red - blood (finding)
	167298001	Urine: trace non-hemolyzed blood (finding)
	167299009	Urine: trace hemolyzed blood (finding)
	167300001	Urine blood test = + (finding)
	167301002	Urine blood test = ++ (finding)
	167302009	Urine blood test = +++ (finding)
	1808141000006100	Urine blood test = moderate non-haemolysed blood (observable entity)
	197938001	Painless hematuria (disorder)
	197939009	Painful hematuria (disorder)
	197940006	Microscopic hematuria (disorder)
	197941005	Frank hematuria (disorder)
	197942003	Clot hematuria (disorder)
	281058009	Red blood cell- red blood cells in urine (finding)
	281860005	Recurrent hematuria (disorder)
	307821001	Urine blood test (procedure)
	34436003	Blood in urine (finding)
	962031000006109	Haematuria/abnormal colour (finding)
Cardio-respiratory symptoms/signs		

Wheeze	1834141000006100	Constant wheeze (finding)
	2761910000001007	Viral wheeze (disorder)
	704114003	On examination - inspiratory wheeze (finding)
	7555810000001011	Mild wheeze (finding)
	7555910000001004	Moderate wheeze (finding)
	7556010000001005	Severe wheeze (finding)
	7556110000001007	Very severe wheeze (finding)
Cardio-respira-tory pain	102588006	Chest wall pain (finding)
	102589003	Atypical chest pain (finding)
	161972006	Central chest pain (finding)
	161973001	Anterior chest wall pain (finding)
	161974007	Parasternal pain (finding)
	161977000	Costal margin chest pain (finding)
	2237002	Pleuritic pain (finding)
	274668005	Non-cardiac chest pain (finding)
	29857009	Chest pain (finding)
	279084009	Chest discomfort (finding)
	281245003	Musculoskeletal chest pain (finding)
	4568003	Retrosternal pain (finding)
	75483001	Breathing painful (finding)
	81953000	Chest pain on exertion (finding)
Cough	11833005	Dry cough (finding)
	161923004	Productive cough -clear sputum (finding)
	161924005	Productive cough -green sputum (finding)
	161925006	Productive cough-yellow sputum (finding)
	161927003	Night cough present (situation)
	161929000	Chesty cough (finding)
	161932002	Morning cough (finding)
	161933007	Evening cough (finding)
	161947006	Nocturnal cough and wheeze (finding)
	17986004	Barking cough (finding)
	272039006	Complaining of cough (finding)
	284523002	Persistent cough (finding)
	28743005	Productive cough (finding)
	315246003	Unexplained cough (finding)
	49727002	Cough (finding)
	68154008	Chronic cough (finding)
	66857006	Hemoptysis (finding)
	810511000000109	Episodic dry cough (finding)
Breathlessness	161939006	Breathless - moderate exertion (finding)
	161940008	Breathless - mild exertion (finding)
	161941007	Dyspnea at rest (finding)

	162890008	On examination - dyspnea (finding)
	162894004	On examination - expiratory wheeze (finding)
	1807841000006100	O/E - dyspnoea at rest (finding)
	1807851000006100	O/E - dyspnoea on exertion (finding)
	1861171000006100	Breathlessness causing difficulty eating (finding)
	1861181000006100	Breathlessness causing anxiety (finding)
	1984331000006100	Manchester triage - Shortness of breath in child (finding)
	230145002	Difficulty breathing (finding)
	23924001	Tight chest (finding)
	267036007	Dyspnea (finding)
	390870001	Short of breath dressing/undressing (finding)
	390871002	Breathless - strenuous exertion (finding)
	407588003	Unable to complete a sentence in one breath (finding)
	60845006	Dyspnea on exertion (finding)
	62744007	Orthopnea (finding)
	702581006	Difficulty eating due to breathlessness (finding)
	982351000006102	Dyspnoea (finding)
Symptoms/signs of the nervous system		
Papilloedema	163975003	On examination - optic disc margin blurred (finding)
	163976002	On examination - optic disc margin absent (finding)
	163978001	On examination - optic disc elevated (finding)
	423341008	Edema of optic disc (disorder)
	423488006	Papilledema - optic disc edema due to raised intracranial pressure (disorder)
	424497004	Optic disc edema due to hypotony (disorder)
	424669003	Optic disc edema associated with retinal disorder (disorder)
	103006007	Unilateral headache (situation)
	162297001	Headache site (finding)
	162299003	Generalized headache (finding)
	162301005	Bilateral headache (finding)
	162304002	Parietal headache (finding)
	162306000	Headache character (observable entity)
	162307009	Aching headache (finding)
	162308004	Throbbing headache (finding)
	162309007	Shooting headache (finding)
	162310002	Morning headache (finding)
	162311003	Heavy head (finding)
	193028008	Sick headache (disorder)
	193031009	Cluster headache syndrome (disorder)
	193039006	Complicated migraine (disorder)
	197785100000610	Unexplained recurrent headaches (finding)

Headache	230462002	Migraine with typical aura (disorder)
	230467008	Status migrainosus (disorder)
	230471006	Chronic tension-type headache (disorder)
	25064002	Headache (finding)
	267096005	Frontal headache (finding)
	272027003	Complaining of a headache (finding)
	330007	Occipital headache (finding)
	364760007	Finding of headache character (finding)
	37796009	Migraine (disorder)
	398057008	Tension-type headache (disorder)
	398126006	Muscular headache (disorder)
	41413006	Temporal headache (finding)
	429051000000106	[D]Pain in head NOS (situation)
	431237007	Chronic headache disorder (disorder)
	443094001	Paroxysmal hemicrania (disorder)
	443095000	Hemicrania continua (disorder)
	4473006	Migraine with aura (disorder)
	449814007	Trigeminal autonomic cephalalgia (disorder)
	49605003	Recurrent painful ophthalmoplegic neuropathy (disorder)
	4969004	Sinus headache (finding)
	56097005	Migraine without aura (disorder)
	59292006	Hemiplegic migraine (disorder)
	722981005	Infrequent episodic tension-type headache (disorder)
	722982003	Frequent episodic tension-type headache (disorder)
	725058003	Short-lasting unilateral neuralgiform headache attacks with conjunctival injection and tearing syndrome (disorder)
	75879005	Abdominal migraine (disorder)
	781381000000100	Migraine induced by oestrogen contraceptive (disorder)
	83351003	Basilar migraine (disorder)
	95654002	Chronic paroxysmal hemicrania (disorder)
	95655001	Ophthalmic migraine (disorder)
Focal motor weakness	162236007	Weakness present (finding)
	163600007	On examination - paresis (weakness) (finding)
	180571100000610 0	Weakness of upper arm (disorder)
	180705100000610 0	Weakness of lower arm (disorder)
	20022000	Hemiparesis (disorder)
	249944006	Monoparesis - arm (disorder)
	249945007	Monoparesis of lower limb (disorder)
	26544005	Muscle weakness (finding)
	278286009	Right hemiparesis (disorder)
	278287000	Left hemiparesis (disorder)
	855781000006104	Hemiparesis (finding)

	856411000006108	Right hemiparesis (disorder)
	857171000006108	Left hemiparesis (finding)
Torticollis	202767003	Intermittent torticollis (disorder)
	270476009	Wry neck/torticollis (disorder)
	70070008	Torticollis (disorder)
	74333002	Spasmodic torticollis (disorder)
Squint	111528000	Alternating esotropia with A pattern (disorder)
	111529008	Monocular exotropia with V-pattern strabismus (disorder)
	128602000	Heterotropia (disorder)
	164011009	On examination - strabismus - squint (finding)
	164012002	On examination - convergent squint (disorder)
	164013007	On examination - divergent squint (disorder)
	16596007	Esotropia (disorder)
	170196006	Child examination: squint (procedure)
	19373007	External ophthalmoplegia (disorder)
	194078004	Monocular esotropia with X pattern (disorder)
	194079007	Monocular esotropia with Y pattern (disorder)
	194083007	Alternating esotropia with X pattern (disorder)
	194084001	Alternating esotropia with Y pattern (disorder)
	194090002	Monocular exotropia with X pattern (disorder)
	194092005	Monocular exotropia with Y pattern (disorder)
	194096008	Alternating exotropia with X pattern (disorder)
	194097004	Alternating exotropia with Y pattern (disorder)
	194101008	Intermittent monocular esotropia (disorder)
	194102001	Intermittent alternating esotropia (disorder)
	194103006	Intermittent monocular exotropia (disorder)
	194104000	Intermittent alternating exotropia (disorder)
	194108002	Constant vertical heterotropia (disorder)
	194112008	Esotropia with accommodative compensation (disorder)
	194118007	Partial oculomotor nerve palsy (disorder)
	194119004	Total oculomotor nerve palsy (disorder)
	20636006	Vertical heterophoria (disorder)
	22066006	Strabismus (disorder)
	24566000	Alternating esotropia with V pattern (disorder)
	2473007	Intermittent vertical heterotropia (disorder)
	267089008	Has a squint (disorder)
	27590007	Intermittent esotropia (disorder)
	28633004	Microstrabismus (disorder)
	31254003	Alternating exotropia with A pattern (disorder)
	313157004	Pseudostrabismus (disorder)
	37214009	Alternating exotropia (disorder)
	39837002	Alternating esotropia (disorder)
	399054005	Exotropia (disorder)
	400942002	Paralytic strabismus (disorder)
	40608009	Hypertropia (disorder)
	45062007	Monocular exotropia with A pattern (disorder)

	5455000	Monocular esotropia (disorder)
	5725006	Monocular exotropia (disorder)
	575911000000100	Strabismus (squint) and other disorders of binocular eye movements (disorder)
	57805002	Alternating exotropia with V pattern (disorder)
	58409000	Monocular esotropia with A pattern (disorder)
	63988001	Intermittent exotropia (disorder)
	65215002	Monocular esotropia with V pattern (disorder)
	74025007	Intermittent tropia (disorder)
	811181000000104	Possible squint (situation)
	856261000006105	Possible squint present (finding)
	856391000006108	Squint present on examination (finding)
	907781000006105	[RFC] Sudden onset squints/double vision (finding)
	922671000006106	Childs exam. - squint (navigational concept)
Seizure	192981006	Atonic epileptic seizure (finding)
	192982004	Epileptic seizures - akinetic (finding)
	192991000	Clonic epileptic seizure (finding)
	192992007	Epileptic seizures - myoclonic (disorder)
	192993002	Tonic epileptic seizure (finding)
	199451000000106	Simple partial epileptic seizure (disorder)
	230434009	Seizures in response to acute event (disorder)
	230445007	Nocturnal epilepsy (disorder)
	246545002	Generalized onset epileptic seizure (finding)
	29753000	Focal onset epileptic seizure (finding)
	37356005	Myoclonic seizure (disorder)
	407675009	Focal onset impaired awareness epileptic seizure (finding)
	440443001	Reflex anoxic seizure (finding)
	445095002	Epileptic seizure witnessed by provider of history other than subject (finding)
	54200006	Tonic-clonic seizure (finding)
	79631006	Absence seizure (finding)
	91175000	Seizure (finding)
Abnormal visual acuity	163938008	On examination - visual acuity right eye =6/9 (finding)
	163939000	On examination - visual acuity right eye=6/12 (finding)
	163940003	On examination - visual acuity right eye=6/18 (finding)
	163941004	On examination - visual acuity right eye=6/24 (finding)
	163942006	On examination - visual acuity right eye=6/36 (finding)
	163943001	On examination - visual acuity right eye=6/60 (finding)
	163952005	On examination - visual acuity left eye =6/9 (finding)
	163953000	On examination - visual acuity left eye=6/12 (finding)
	163954006	On examination - visual acuity left eye=6/18 (finding)
	163955007	On examination - visual acuity left eye=6/24 (finding)
	163956008	On examination - visual acuity left eye=6/36 (finding)

	163957004	On examination - visual acuity left eye=6/60 (finding)
	163965001	On examination - near vision abnormal (finding)
	170728008	Poor visual acuity (finding)
	1757871000006100	O/E - visual acuity L-eye = 6/48 (finding)
	1757881000006100	O/E - visual acuity L-eye = 6/7.5 (finding)
	1757891000006100	O/E - visual acuity R-eye = 6/48 (finding)
	1757901000006100	O/E - visual acuity R-eye = 6/7.5 (finding)
	1780411000006100	O/E - visual acuity R-eye =6/20 (finding)
	1780421000006100	O/E - visual acuity R-eye =6/10 (finding)
	1780431000006100	O/E - visual acuity L-eye =6/20 (finding)
	1780441000006100	O/E - visual acuity L-eye =6/10 (finding)
	1800921000006100	O/E - visual acuity L-eye=6/7 (finding)
	1800931000006100	O/E - visual acuity R-eye=6/7 (finding)
	1953101000006100	Child exam - vision: problem identified (procedure)
	1953111000006100	Child exam - vision: problem suspected (procedure)
	202511000000100	On examination visual acuity left eye = 6/30 (finding)
	202531000000108	On examination visual acuity right eye = 6/30 (finding)
	225581002	Sight deteriorating (finding)
	240311000000103	Loss of vision (disorder)
	390804009	On examination - visual acuity right eye=3/60 (finding)
	390806006	On examination - visual acuity left eye=3/60 (finding)
	408295008	On examination - pinhole visual acuity left eye=6/9 (finding)
	408296009	On examination - pinhole visual acuity left eye=6/12 (finding)
	408297000	On examination - pinhole visual acuity left eye=6/18 (finding)
	408298005	On examination - pinhole visual acuity left eye =6/24 (finding)
	408299002	On examination - pinhole visual acuity left eye=6/36 (finding)
	408300005	On examination - pinhole visual acuity left eye=6/60 (finding)
	408303007	On examination - pinhole visual acuity right eye=6/9 (finding)
	408304001	On examination - pinhole visual acuity right eye=6/12 (finding)

	408305000	On examination - pinhole visual acuity right eye=6/18 (finding)
	408306004	On examination - pinhole visual acuity right eye=6/24 (finding)
	408307008	On examination - pinhole visual acuity right eye=6/36 (finding)
	408308003	On examination - pinhole visual acuity right eye=6/60 (finding)
	416793003	On examination - visual acuity right eye =6/15 (finding)
	416965009	On examination - visual acuity left eye =6/15 (finding)
	704109002	On examination - visual acuity left eye = 6/7.5 (finding)
	704110007	On examination - visual acuity left eye = 6/48 (finding)
	704111006	On examination - visual acuity right eye = 6/48 (finding)
	704112004	On examination - visual acuity right eye = 6/7.5 (finding)
	809751000000108	On examination visual acuity left eye = 6/20 (finding)
	810251000000105	On examination visual acuity right eye = 6/20 (finding)
	810331000000109	On examination visual acuity right eye = 6/10 (finding)
	810371000000106	On examination visual acuity left eye = 6/10 (finding)
	822201000000104	On examination visual acuity left eye = 6/7 (finding)
	822261000000100	On examination visual acuity right eye = 6/7 (finding)
	915041000006108	Visual impairment: unable to recognise people across street (disorder)
	915051000006105	Visual impairment: difficulty with close up work (disorder)
	915061000006107	Patient reports difficulty with peripheral vision (disorder)
	915071000006100	Visual impairment: difficulty using stairs (disorder)
	915081000006102	Visual impairment: avoids driving (disorder)
	915091000006104	Visual impairment: difficulty reading newspaper print (disorder)
	915221000006105	Patient concerned about eyesight (disorder)
	982201000006109	O/E - visual acuity R-eye=6/15 (finding)
	982211000006107	O/E - visual acuity L-eye=6/15 (finding)
Cranial nerve ab-normalities	164091005	On examination - cranial nerve 3 - palsy (disorder)
	164092003	On examination - cranial nerve 3-paralysis (finding)
	164094002	On examination - cranial nerve 4 - palsy (disorder)
	164095001	On examination - cranial nerve 4-paralysis (finding)
	164097009	On examination - cranial nerve 6 - palsy (disorder)
	164098004	On examination - cranial nerve 6-paralysis (finding)
	164103002	On examination - cranial nerve 5 - palsy (disorder)

	164104008	On examination - cranial nerve 5-paralysis (finding)
	164106005	On examination - cranial nerve 7-palsy-lower motor neuron (finding)
	164107001	On examination - cranial 7 -paralysis-lower motor neuron (finding)
	164109003	On examination - cranial nerve -palsy -upper motor neuron (disorder)
	164110008	On examination - cranial 7 -paralysis-upper motor neuron (disorder)
	164115003	On examination - cranials 9 + 10 palsy (finding)
	164116002	On examination - cranials 9+10 paralysis (finding)
	164118001	On examination - cranial nerve 11 - palsy (finding)
	164119009	On examination - cranial nerve 11 paralysis (finding)
	164121004	On examination - cranial nerve 12 - palsy (finding)
	164122006	On examination - cranial nerve 12 paralysis (disorder)
	1872111000006100	O/E cranial nerves abnormal (procedure)
	193102008	Multiple cranial nerve palsies in neoplastic disease (disorder)
	302884001	Multiple cranial nerve lesions and linked syndromes (disorder)
	60750009	Third cranial nerve disease (disorder)
	73013002	Cranial nerve disorder (disorder)
	78152008	Multiple cranial nerve palsy (disorder)
Precocious puberty	103021001	Premature adrenarche (finding)
	237815000	Precocious pubarche (disorder)
	277872000	Precocious puberty with adrenal hyperplasia (disorder)
	400179000	Precocious puberty (disorder)
Abnormal gait	109611000000108	On examination - abnormal body sway (finding)
	163685000	On examination - gait spastic (finding)
	163686004	On examination - gait ataxic (finding)
	163688003	On examination - waddling gait (finding)
	163689006	On examination - steppage gait (finding)
	163690002	On examination - hysterical gait (finding)
	163691003	On examination - limping gait (finding)
	163768005	On examination - arms ataxic (finding)
	163769002	On examination - legs ataxic (finding)
	163770001	On examination - Romberg test positive (finding)
	163772009	On examination - past pointing present (situation)
	163773004	On examination - dysdiadochokinesia present (situation)
	163774005	On examination - generally unsteady (finding)
	163775006	On examination - clumsy (finding)
	1818771000006100	O/E gait abnormal (finding)
	1916131000006100	CHSP-PS recall: gait abnormal (regime/therapy)
	1916141000006100	CHSP-PS recall: gait uncertain (regime/therapy)

	1919151000006100	CHSP-PS unsched. contact: gait abnormal (regime/therapy)
	1919161000006100	CHSP-PS unsched. contact: gait uncertain (regime/therapy)
	202567001	Walking difficulty due to pelvic region and thigh (finding)
	202569003	Walking difficulty due to lower leg (finding)
	202570002	Walking difficulty due to ankle and foot (finding)
	202572005	Walking difficulty due to multiple sites (finding)
	20262006	Ataxia (finding)
	22325002	Abnormal gait (finding)
	228158008	Walking disability (finding)
	249983008	Feels as though will fall (finding)
	250018006	Toe-walking gait (finding)
	25136009	Ataxic gait (finding)
	267083009	Incoordination symptom (finding)
	268964003	On examination - festination-Parkinson gait (finding)
	268968000	On examination - coordination (finding)
	275858008	On examination - ataxia (finding)
	275864001	On examination - intoeing (finding)
	281016006	Incoordination (finding)
	282195009	Difficulty walking up stairs (finding)
	282299006	Unable to balance (finding)
	299031000000106	Disorder of gait and/or balance present (situation)
	299061000000101	Disorder of gait and/or balance not present (situation)
	390789009	Head lag (finding)
	39384006	Muscular incoordination (finding)
	496121000000100	[D]Gait abnormality (situation)
	67141003	Antalgic gait (finding)
	719232003	Difficulty walking (finding)
	781091000000100	Worsening balance (finding)
	78691002	Staggering gait (finding)
	855481000006107	Child gait delayed (procedure)
	855501000006102	Child gait abnormal (procedure)
	89419008	Loss of equilibrium (finding)
	9447003	Spastic gait (finding)
	948391000000106	On examination - antalgic gait (finding)
	960101000006103	Gait/ambulation disturbance (finding)
Systemic Symptom/Signs		
Fever	135882008	Feverish cold (finding)
	161855003	Complaining of shivering (finding)
	163633002	On examination - skin temperature abnormal (finding)
	164285001	On examination - fever - general (finding)

	164288004	On examination - pyrexia of unknown origin (finding)
	164303007	On examination - temperature elevated (finding)
	164304001	On examination - hyperpyrexia - greater than 40.5 degrees Celsius (finding)
	164307008	On examination - fever - acute rise (finding)
	164308003	On examination - fever - gradual rise (finding)
	164309006	On examination - fever - continuous (finding)
	164311002	On examination - fever - remittent (finding)
	164312009	On examination - fever - intermittent (finding)
	164313004	On examination - staircase fever (finding)
	164314005	On examination - fever - irregular (finding)
	164315006	On examination - fever - fast fall-crisis (finding)
	164316007	On examination - fever-gradual fall-lysis (finding)
	186694006	Sweating fever (finding)
	248427009	Fever symptoms (finding)
	248457000	Rigor - symptom (finding)
	268984004	Examination of fever (procedure)
	271897009	On examination - fever (finding)
	274640006	Fever with chills (finding)
	309646008	On examination - temperature (finding)
	373904004	Feels hot/feverish (finding)
	386661006	Fever (finding)
	409702008	Hyperpyrexia (finding)
	495741000000101	[D]Persistent fever (situation)
	7520000	Pyrexia of unknown origin (finding)
	851081000006103	[D]Pyrexia (situation)
	981661000006101	Fever (finding)
	982701000006100	Pyrexia (finding)
Weight loss	161832001	Weight decreasing (finding)
	198511000000103	Complaining of weight loss (finding)
	267024001	Abnormal weight loss (finding)
	448765001	Unintentional weight loss (finding)
	960561000006106	Unexplained/progressive weight loss (finding)
Lethargy/tiredness	214264003	Lethargy (finding)
	248274002	Lack of energy (finding)
	267031002	Tiredness symptom (finding)
	267032009	Tired all the time (finding)
	271795006	Malaise and fatigue (finding)
	272060000	Fatigue - symptom (finding)
	272062008	Complaining of tired all the time (finding)
	367391008	Malaise (finding)
	60119000	Exhaustion (finding)
	84229001	Fatigue (finding)

Night Sweats	161857006	Sweating symptom (finding)
	42984000	Night sweats (finding)
	52613005	Excessive sweating (finding)
Haematological Symptoms/Signs		
Lymphadenopathy	127086001	Cervical lymphadenopathy (disorder)
	164148003	On examination - sub-mental lymphadenopathy (disorder)
	164149006	On examination - supraclavicular lymphadenopathy (disorder)
	164150006	On examination - axillary lymphadenopathy (disorder)
	164151005	On examination - inguinal lymphadenopathy (disorder)
	164152003	On examination - popliteal lymphadenopathy (disorder)
	164155001	On examination - lymph nodes firm (finding)
	164156000	On examination - lymph nodes hard (finding)
	164157009	On examination - lymph nodes tethered (finding)
	164158004	On examination - lymph nodes discrete (finding)
	164159007	On examination - lymph nodes fluctuant (finding)
	164160002	On examination - lymph nodes tender (finding)
	19471005	Lymphadenitis (disorder)
	274303007	On examination - lymph nodes (finding)
	274740001	Enlarged submandibular lymph gland (disorder)
	274741002	Generalized enlarged lymph nodes (disorder)
	274744005	Localized enlarged lymph nodes (disorder)
	30746006	Lymphadenopathy (disorder)
	394677008	On examination - post-auricular lymphadenopathy (disorder)
	445718006	Reactive lymphadenopathy (disorder)
	62683002	Mediastinal lymph node structure (body structure)
Bruising	10050004	Contusion of chest (disorder)
	105618004	Contusion with intact skin (disorder)
	110244000	Contusion of forehead (disorder)
	111721009	Contusion of multiple sites (disorder)
	111723007	Contusion of multiple sites of lower limb (disorder)
	11437003	Contusion of back (disorder)
	125668004	Contusion of face (disorder)
	161885008	Bruising symptom (finding)
	161887000	Spontaneous bruising (disorder)
	164331003	On examination - petechiae present (finding)
	164624008	On examination - bruising (disorder)
	17048005	Contusion of toenail (disorder)
	1954501000006100	Bruise/ contusion/ abrasion (secondary group) (administrative concept)
	211476000	Contusion, throat (disorder)
	211486004	Contusion of anterior abdominal wall (disorder)
	211492005	Contusion of coccyx (disorder)
	211500005	Contusion, shoulder or upper arm (disorder)
	211501009	Contusion of clavicular area (disorder)
	211502002	Contusion, shoulder and upper arm, multiple sites (disorder)

	211504001	Contusion, elbow or forearm (disorder)
	211506004	Contusion wrist or hand (disorder)
	211507008	Contusion, hand, excluding finger (disorder)
	211508003	Contusion hand, palm (disorder)
	211509006	Contusion hand, dorsum (disorder)
	211510001	Contusion wrist, volar (disorder)
	211511002	Contusion wrist, dorsum (disorder)
	211514005	Contusion, fingernail (includes subungual hematoma) (disorder)
	211515006	Contusion of multiple fingers (disorder)
	211519000	Contusion, hip and thigh (disorder)
	211521005	Contusion, knee and lower leg (disorder)
	211523008	Contusion, ankle and foot, excluding toe(s) (disorder)
	21643007	Contusion of groin (disorder)
	22562004	Contusion of scalp (disorder)
	22878006	Contusion of knee (disorder)
	231814009	Contusion of eyelid (disorder)
	262528003	Bruise of head (disorder)
	262637006	Contusion of gingivae (disorder)
	269214009	Contusion of face, scalp and neck, excluding eye(s) (disorder)
	270911002	Superficial bruising (disorder)
	271813007	Petechiae (disorder)
	283135004	Superficial bruising of head and neck (disorder)
	283137007	Contusion of pelvic region (disorder)
	284062002	Contusion of lower back (disorder)
	313262006	Contusion of sacral region (disorder)
	3649004	Contusion of multiple sites of trunk (disorder)
	367423000	Contusion of eye (disorder)
	36787001	Contusion of trunk (disorder)
	37907001	Contusion of abdominal wall (disorder)
	385494008	Hematoma (disorder)
	39812007	Contusion of forearm (disorder)
	40257000	Contusion of shoulder region (disorder)
	416179002	Contusion of eye AND ocular adnexa (disorder)
	417654008	Contusion of ocular adnexa and periocular tissues (disorder)
	424131007	Easy bruising (finding)
	42697001	Contusion of orbital tissues (disorder)
	44801007	Contusion of hip (disorder)
	45613006	Contusion of lower leg (disorder)
	48123002	Contusion of wrist (disorder)
	4821001	Contusion of cheek (disorder)
	50228009	Contusion of ear (disorder)
	52945003	Contusion of mandibular joint area (disorder)
	53759006	Contusion of axillary region (disorder)
	544681000000101	Contusion, eye NOS (disorder)
	54499003	Contusion of interscapular region (disorder)

	55042009	Contusion of ankle (disorder)
	58075000	Contusion of toe (disorder)
	58221004	Contusion of multiple sites of upper limb (disorder)
	60211009	Contusion of flank (disorder)
	60897004	Contusion of nose (disorder)
	609791000000105	Contusion, upper limb NOS (disorder)
	64458007	Contusion of lip (disorder)
	649491000000106	Black eye NOS (disorder)
	649551000000104	Contusion,trunk NOS (disorder)
	656761000000104	Other contusion neck (disorder)
	656771000000106	Contusion of face, scalp and neck NOS (disorder)
	669321000000105	Contusion, lower limb and other unspecified sites (disorder)
	669331000000107	Contusion, lower limb NOS (disorder)
	669341000000103	Contusion, site NOS (disorder)
	68142008	Contusion of upper limb (disorder)
	68893001	Contusion of buttock (disorder)
	69787006	Contusion of finger (disorder)
	74814004	Contusion of foot (disorder)
	76437000	Contusion of thumb (disorder)
	80907003	Contusion of scapular region (disorder)
	84416003	Contusion of thigh (disorder)
	8513005	Contusion of neck (disorder)
	85636003	Contusion of upper arm (disorder)
	90244007	Contusion of heel (disorder)
	91603007	Contusion of lower limb (disorder)
	91613004	Contusion of elbow (disorder)
	959941000006103	Bruising (finding)
	982381000006105	Purpura (finding)
Anaemia/pallor	162738007	On examination - color pale (finding)
	164139008	On examination - anemia (finding)
	164141009	On examination - equivocally anemic (disorder)
	164142002	On examination - clinically anemic (disorder)
	164143007	On examination - profoundly anemic (disorder)
	267029006	Pale - symptom (finding)
	371262000	Pale color saturation (qualifier value)
	961701000006107	Pallor/perspiration (finding)
Lumps and Swellings		
Lumps/Swellings	162410003	Complaining of a swelling (finding)
	162831005	On examination - parotid swelling (finding)

	162832003	On examination - submandibular swelling (finding)
	164620004	On examination - a lump (finding)
	164621000	On examination - a swelling (finding)
	248480008	Leg swelling symptom (finding)
	248521008	Lump on finger (finding)
	274745006	Localized swelling, mass and lump, lower limb (finding)
	274746007	Localized swelling, mass and lump, multiple sites (finding)
	274747003	Localized swelling, mass and lump, neck (finding)
	274750000	Localized swelling, mass and lump, upper limb (finding)
	275319005	Swollen legs (finding)
	278528006	Facial swelling (finding)
	281792000	Swollen calf (finding)
	299703001	Mass of neck (finding)
	300863000	Mass of axilla (finding)
	309576006	Lump on thigh (finding)
	315284009	Mass of parotid gland (finding)
	390761000000104	[D]Swelling, mass or lump of chest NOS (situation)
	432525007	Mass of head (finding)
	438581000000100	[D]Swelling, mass or lump in head or neck NOS (situation)
	440299000	Mass of thoracic structure (finding)
	441708007	Swelling of lower jaw region (finding)
	441975003	Mass of tongue (finding)
	441998006	Subungual swelling (finding)
	496471000000109	[D]Local superficial swelling, mass or lump (situation)
	496481000000106	[D]Swelling, local and superficial (situation)
	496491000000108	[D]Mass, localised and superficial (situation)
	496501000000102	[D]Lump, localised and superficial (situation)
	496561000000103	[D]Lump on back (situation)
	496571000000105	[D]Foot lump (situation)
	496581000000107	[D]Shoulder lump (situation)
	496591000000109	[D]Lump on hand (situation)
	496601000000103	[D]Lump on knee (situation)
	496611000000101	[D]Lump on leg (situation)
	496621000000107	[D]Lump on shin (situation)
	496651000000102	[D]Wrist lump (situation)

	496661000000104	[D]Toe lump (situation)
	497131000000101	[D]Swelling, mass or lump in head and neck (situation)
	49714100000010 5	[D]Swelling in head or neck (situation)
	49715100000010 8	[D]Mass in head or neck (situation)
	49716100000010 6	[D]Lump in head or neck (situation)
	49721100000010 1	[D]Lump on head (situation)
	49722100000010 7	[D]Lump on nose (situation)
	49723100000010 9	[D]Submental lump (situation)
	49724100000010 0	[D]Swelling of ear (situation)
	49826100000010 0	[D]Swelling, mass and lump of chest (situation)
	49828100000010 9	[D]Chest lump (situation)
	80068009	Swelling of limb (finding)
	853681000006106	Lump in axilla (situation)
Musculoskeletal symptoms/signs		
Limp	163691003	On examination - limping gait (finding)
	1984211000006100	Manchester triage - Limping child (finding)
	202567001	Walking difficulty due to pelvic region and thigh (finding)
	202569003	Walking difficulty due to lower leg (finding)
	202570002	Walking difficulty due to ankle and foot (finding)
	67141003	Antalgic gait (finding)
Limb pain	102482005	Growing pains (finding)
	102556003	Pain in upper limb (finding)
	10601006	Pain in lower limb (finding)
	1088071000000100	Pain in upper arm (finding)
	275318002	Aching leg syndrome (finding)
	287047008	Pain in left lower limb (finding)
	287048003	Pain in right lower limb (finding)
	288241003	Pain in limb - multiple (finding)
	300954003	Pain in calf (finding)
	444899003	Pain in forearm (finding)
	47933007	Foot pain (finding)
	53057004	Hand pain (finding)
	78514002	Thigh pain (finding)
	853861000006100	Painful left arm (finding)
	90834002	Pain in limb (finding)

Joint pain	202478007	Sternoclavicular joint pain (finding)
	202479004	Acromioclavicular joint pain (finding)
	202480001	Elbow joint pain (finding)
	202481002	Distal radioulnar joint pain (finding)
	202482009	Wrist joint pain (finding)
	202483004	Metacarpophalangeal joint pain (finding)
	202484005	Proximal interphalangeal joint of finger pain (finding)
	202485006	Distal interphalangeal joint of finger pain (finding)
	202487003	Sacroiliac joint pain (finding)
	202489000	Tibiofibular joint pain (finding)
	202490009	Ankle joint pain (finding)
	202491008	Subtalar joint pain (finding)
	202493006	Talonavicular joint pain (finding)
	202495004	First metatarsophalangeal joint pain (finding)
	202496003	Lesser metatarsophalangeal joint pain (finding)
	202497007	Interphalangeal joint of toe pain (finding)
	239733006	Anterior knee pain (finding)
	267949000	Shoulder joint pain (finding)
	267950000	Arthralgia of the upper arm (finding)
	267952008	Arthralgia of the pelvic region and thigh (finding)
	267953003	Arthralgia of the lower leg (finding)
	267954009	Arthralgia of the ankle and/or foot (finding)
	298857005	Shoulder joint painful on movement (finding)
	301813003	Irritable hip (disorder)
	30989003	Knee pain (finding)
	35678005	Multiple joint pain (finding)
	49218002	Hip pain (finding)
	56608008	Pain in wrist (finding)
	57676002	Joint pain (finding)
	629831000000105	Arthralgia of other specified site (finding)
	687011000000105	Arthralgia NOS (finding)
	853951000006106	Painful elbow (finding)
	921261000000101	Shoulder joint painful on external rotation (finding)

Supplementary Table 9-7 Minimum sample size required to detect an odds ratio of 2 for a single feature separately for childhood (0-14 years) and TYA (15-24 years). 95% CI= 95% Confidence Interval

Cancer Type	**Clinical Feature**	**Age Group**	**Minimum sample size Required to detect an odds ratio of 2 for my proposed study α=0.05 and 1-β=0.8**
Any	Fatigue	0-14 years	5295
		15-24 years	2660
Leukaemia	Lymphadenopathy	0-14 years	3538
		15-24 years	5295
Lymphoma	Lymphadenopathy	0-14 years	3538
		15-24 years	5295
CNS Tumours	Headache	0-14 years	2133
		15-24 years	1781
Sarcoma	Musculoskeletal symptoms	0-14 years	1079
		15-24 years	1079

Supplementary Table 9-8 Minimum sample size required to detect odds ratio of 2 for two interacting dependent variables x_1 and x_2 (where x_1 is the clinical feature detailed in Supplementary Table 6.3-3)

Cancer Type	Age Group	Sample Size required to detect an odds ratio of 2 for a=0.05 and 1-B=0.8	
		Low correlation x_1 and x_2	High correlation x_1 and x_2
Any	0-14 years	5552	7636
	15-24 years	1788	3832
Leukaemia	0-14 years	3709	5100
	15-24 years	5552	7636
Lymphoma	0-14 years	3709	5100
	15-24 years	5552	7636
CNS Tumours	0-14 years	2235	3071
	15-24 years	1867	2564
Sarcomas	0-14 years	1130	1550
	15-24 years	1130	1550

Supplementary Table 9-9 *Frequency of selected clinical features in cancer cases and controls per cancer subtype. Association between individual clinical feature and a subsequent cancer diagnosis in univariable conditional logical regression analyses. OR= Odds Ratio. CI=Confidence interval.*

Clinical Feature	**Leukaemia (n, %) N=1,127**	**OR (95%CI) in univariable analysis**	**Lymphoma (n, %) N=919**	**OR (95%CI) in univariable analysis**	**CNS Tumours (n, %) N=933**	**OR (95%CI) in univariable analysis**	**Sarcoma (n, %) N=637**	**OR (95%CI) in univariable analysis**
Organomegaly	11 (1.0)	55 (17.5-172.7)	<5	-				
Lymphadenopathy	53 (4.7)	9.4 (6.7-13.3)	201 (21.9)	97.2 (68.4-138.3)			12 (1.9)	4.8 (2.5-9.2)
Fever	127 (11.2)	7.33 (5.84-9.19)	26 (2.8)	5.0 (3.2-7.9)	33 (3.5)	2.1 (1.6-3.4)	8 (1.3)	1.4 (0.6-3.0)
URTI	247 (21.9)	2.3 (1.9-2.7)	141 (15.4)	3.1 (2.5-3.8)				
Limb pain	59 (5.2)	7.63 (5.60-10.40)	32 (3.5)	2.8 (1.9-4.1)			95 (15.1)	15.8 (11.9-21.0)
Joint pain	41 (3.6)	9.25 (6.32-13.54)	17 (1.9)	2.3 (1.4-3.8)			59 (9.3)	16.1 (11.3-22.9)
Limp	8 (0.7)	10.61 (4.43 - 25.38)	<5	-	6 (0.7)	15 (5.2-43.2)	5 (0.8)	49.6 (9.6-255.8)
Anaemic/pallor	32 (2.8)	62.70 (30.80 – 127.61)	8 (0.9)	31.6 (10.3-96.6)				
Tiredness	79 (7.0)	15.96 (11.76 – 21.68)	20 (2.2)	3.1 (1.8-5.1)	24 (2.6)	5.3 (3.4-8.4)	10 (1.6)	2.3 (1.2-4.4)
Abdominal pain	59 (5.2)	4.70 (3.50-6.31)	58 (6.3)	4.4 (3.2-5.9)			28 (4.4)	3.4 (2.3-5.4)
Anorexia	7 (0.6)	9.33 (3.81 – 22.89)	<5	-	<5	-	<5	-
Rash	78 (6.9)	2.06 (1.61-2.64)	64 (7.0)	4.6 (3.4-6.1)			11 (1.7)	0.8 (0.4-1.5)
Abnormal skin lesion	<5	1.6 (0.5-5.1)	11 (1.2)	8.0 (3.9-16.2)			8 (1.2)	9.4 (3.4-26.0)
Chest pain	11 (1.0)	3.75 (1.95 to 7.18)	31 (3.4)	8.0 (5.2-12.2)			9 (1.4)	3.4 (1.7-7.1)
Cough	106 (9.4)	2.5 (2.0-3.1)	70 (7.6)	4.5 (3.4-6.0)			20 (3.1)	1.3 (0.8-2.1)
Breathlessness	11 (1.0)	1.5 (0.81-2.83)	16 (1.7)	4.6 (2.6-8.0)			5 (0.8)	2.2 (0.9-5.6)
Headache	22 (2.0)	3.16 (2.00-5.02)	11 (1.2)	1.1 (0.6-1.9)	174 (18.7)	32.2 (25.0-41.6)	11 (1.7)	1.8 (0.9-3.4)
Bruising	60 (5.3)	29.17 (19.54 – 43.56)	5 (0.5)	3.3 (1.2-8.9)			7 (1.1)	5.4 (2.4-12.5)
Vomiting	17 (2.0)	2.8 (1.41-5.74)	10 (1.1)	7.1 (3.4-14.8)	61 (6.5)	28.5 (19.1-42.6)	<5	-
Abdominal mass	17 (2.0)	-	<5	-			<5	-

Head/neck lump	6 (0.5)	7.15 (2.74 – 18.69)	110 (12.0)	357.3 (157.1-812.8)			5 (0.8)	13.3 (4.2-42.1)
Lump on body	12 (1.1)	4.55 (2.41 to 8.58)	67 (7.3)	20.9 (14.7-29.8)			49 (9.3)	34.1 (22.0-52.9)
Weight loss	6 (0.5)	9.6 (3.58 to 25.56)	14 (1.5)	27.3 (12.1-61.5)	5 (0.5)	11.2 (3.6 – 34.6)	5 (0.8)	14.9 (4.5-49.3)
Pruritus			14 (1.5)	27.2 (12.0-61.2)				
Papilloedema					16 (1.7)	-		
Hemiparesis					8 (0.9)	51.5 (13.7-194.4)	<5	-
Torticollis	<5	12.0 (2.9-50.2)			6 (0.6)	10.1 (3.7-27.5)	<5	-
Squint					13 (1.4)	7.8 (4.0-15.1)	<5	-
Abnormal eye movements					6 (0.6)	25.1 (7.0-90.5)	<5	-
Visual acuity abnormalities					20 (2.1)	35.9 (16.1-79.8)	<5	-
Diplopia					10 (1.1)	94.2 (20.6-431.1)	<5	-
Exopthalmos					<5	-	<5	-
Behavioural Changes					<5	1.9 (0.6-6.4)		
Irritability					<5	4.1 (0.4-37.2)		
Developmental delay					6 (0.6)	7.1 (2.6-19.6)		
Abnormal cranial nerve signs (any)					<5	-	<5	-
First seizure					43 (4.6)	164.6 (65.2-415.8)		
Precocious puberty					<5	19.9 (5.0-79.5)		
Haematuria	8 (0.9)	3.7 (1.7-7.9)					<5	1.4 (0.4-4.6)
Constipation							14 (2.2)	4.6 (2.6-8.4)
>1 presentation	224 (19.8)	5.6 (4.7 – 6.7)	284 (30.9)	20.4 (16.9-24.7)	172 (18.4)	6.1 (5.0-7.4)	124 (19.6)	8.8 (6.9-11.1)

Supplementary Table 9-10 *Frequency of selected clinical features in cancer cases and controls per cancer subtype in children. Association between individual clinical feature and a subsequent cancer diagnosis in univariable conditional logical regression analyses. OR= Odds Ratio. CI=Confidence interval.*

Clinical Feature	Leukaemia (n, %) N=819	OR (95%CI) in univariable analysis	Lymphoma (n, %) N=266	OR (95%CI) in univariable analysis	CNS Tumours (n, %) N=625	OR (95%CI) in univariable analysis	Sarcoma (n, %) N=294	OR (95%CI) in univariable analysis	Abdominal tumours (n, %) N=295	OR (95%CI) in univariable analysis
Organomegaly	7 (0.8)	70 (14.5-336.9)	<5	-						
Lymphadenopathy	23 (3.7)	6.36 (3.92 - 10.33)	48 (18.1)	52.3 (29.5-92.5)			6 (2.0)	3.3 (1.3-8.0)		
Fever	115 (14.0)	6.92 (5.46 – 8.77)	10 (3.8)	3.0 (1.5-6.0)	31 (5.0)	2.5 (1.7-3.7)	7 (2.4)	1.3 (0.5-3.0)	28 (9.5)	3.6 (2.3-5.8)
URTI	204 (24.9)	2.12 (1.78-2.53)	50 (18.8)	2.5 (1.7-3.5)						
Limb pain	45 (5.5)	9.86 (6.82-14.24)	5 (1.9)	1.8 (0.7-4.5) p=0.2			44 (15.0)	18.8 (12.1-29.4)	7 (2.4)	4.9 (2.1-11.4)
Joint pain	34 (4.1)	13.87 (8.89-21.66)	7	3.8 (1.7-8.9)			23 (7.2)	15.6 (8.9-27.3)	7 (2.4)	9.8 (3.9 - 24.7)
Limp	8 (1.0)	12.31 (5.01 - 30.22)	<5	-	18 (2.9)	7.9 (4.5-13.7)	5 (1.7)	99.1 (11.6-848.7)	<5	20 (5.0-80.0)
Anaemic/pallor	27 (3.3)	88.14 (36.37 – 213.62)	<5	-						
Tiredness	64 (7.8)	21.80 (15.13 - 31.39)	8 (3.0)	10.9 (4.5-26.0)	6 (1.2)	6.9 (2.7-17.8)	<5	-	13 (4.4)	10.6 (5.2-21.8)
Abdominal pain	43 (5.3)	5.16 (3.64 - 7.31)	21 (7.9)	5.9 (3.5-9.7)			12 (4.1)	3.2 (1.7-6.1)	37 (12.5)	20.8 (12.8-33.7)
Anorexia	<5	5.71 (1.88 – 17.36)	<5	-	<5	-	<5	-		
Rash	66 (8.1)	1.91 (1.46-2.51)	22 (8.3)	3.5 (2.2-5.7)			8 (2.7)	1.0 (0.5-2.1)		
Abnormal skin lesion	<5	2.0 (0.58 – 6.42)	<5	8.3 (2.1-32.0)			<5	15.7 (4.2-58.5)		
Chest pain	<5	4.3 (1.5-12.8)	<5	-			<5	2.9 (0.6-12.6)		

Cough	91 (11.1)	2.40 (1.90-3.00)	24 (9.0)	3.3 (2.0-5.2)			12 (4.1)	1.1 (0.6-2.1)		
Breathlessness	10 (1.2)	1.58 (0.82-3.05) p=0.2	<5	2.7 (0.8-9.2)			<5	2.7 (0.8-9.0)		
Headache	8 (1.0)	2.50 (1.18-5.26) p=0.02	<5	0.7 (0.09-4.8) p=0.7	101 (16.2)	40.7 (28.6-57.7)	<5	2.4 (0.8-6.7)	<5	9.1 (2.3-36.9)
Bruising	22 (3.6)	21.3 (11.5-39.4)	<5	2.7 (0.3-25.4)			<5	3.4 (0.4-29.9)	6 (2.3)	8.7 (3.3-22.8)
Vomiting	12 (1.9)	2.67 (1.21-5.88)	5 (1.9)	10.4 (3.4-31.2)	54 (8.6)	35.6 (22.5-56.4)	<5	1.3 (0.3-5.4) p=0.3	<5	3.2 (0.9-10.8)
Abdominal mass	<5	-	<5	-			<5	-	31 (10.5)	582.5 (79.4-4271.6)
Head/neck lump	6 (0.7)	10.87 (3.88-30.42)	25 (9.4)	500 (67.8-3690.0)			<5	51.1 (5.2-500.4)		
Lump on body	10 (1.2)	6.82 (3.31 to 14.08)	11 (4.1)	16.2 (7.2-36.2)			21 (7.1)	42.4 (19.3-93.2)	7 (2.3)	22.2 (7.4-66.3)
Weight loss	<5	14.72 (3.92 to 55.32)	<5	-	<5	22.2 (4.9-100.4)	<5	-	<5	-
Pruritus			<5	8.7 (1.6-48.5)						
Papilloedema					9 (1.4)	-				
Hemiparesis					7 (1.1)	67.7 (14.0-325.5)	<5	-		
Torticollis	<5	13.3 (2.2-79.8)			6 (1.0)	23.4 (7.1-76.7)	<5	-	<5	-
Squint					10 (1.6)	5.9 (2.9-12.3)	<5	-	<5	5.8 (0.9-18.0)
Abnormal eye movements					5 (0.8)	29.8 (7.0-126.7)	<5	-	<5	-
Visual acuity abnormalities					8 (1.3)	18.3 (6.5-51.6)	<5	-	<5	-
Diplopia					5 (0.8)	100 (11.7-855.9)	<5	-	<5	-

Exophthalmos					<5	-	<5	-	<5	-
Behavioural Changes					<5	1.4 (0.3-5.9)			<5	-
Irritability					<5	7.9 (0.6 -86.2) p=0.1			<5	-
Developmental delay					6 (1.0)	7.5 (2.7-21.4)			<5	-
Abnormal cranial nerve signs (any)					<5	-	<5	-	<5	-
First seizure					20 (3.2)	129.8 (38.5-436.8)			<5	-
Precocious puberty					<5	30.0 (5.0-179.5)			<5	-
Haematuria	<5	2.5 (0.8-8.3)					<5	5.1 (1.4-18.5)	15 (5.5)	32.9 (14.3-75.2)
Constipation							10 (3.4)	4.2 (2.1-8.6)	20 (6.7)	5.8 (3.5-9.8)
Testicular swelling										
>1 presentation			65 (24.4)	10.3 (7.3-14.5)	118 (24.4)		54 (18.3)	6.6 (4.6-9.4)	176 (59.6)	5.3 (4.1-6.8)

Supplementary Table 9-11 *Frequency of selected clinical features in cancer cases and controls per cancer subtype in TYA. Association between individual clinical feature and a subsequent cancer diagnosis in univariable conditional logical regression analyses. OR= Odds Ratio. CI=Confidence interval.*

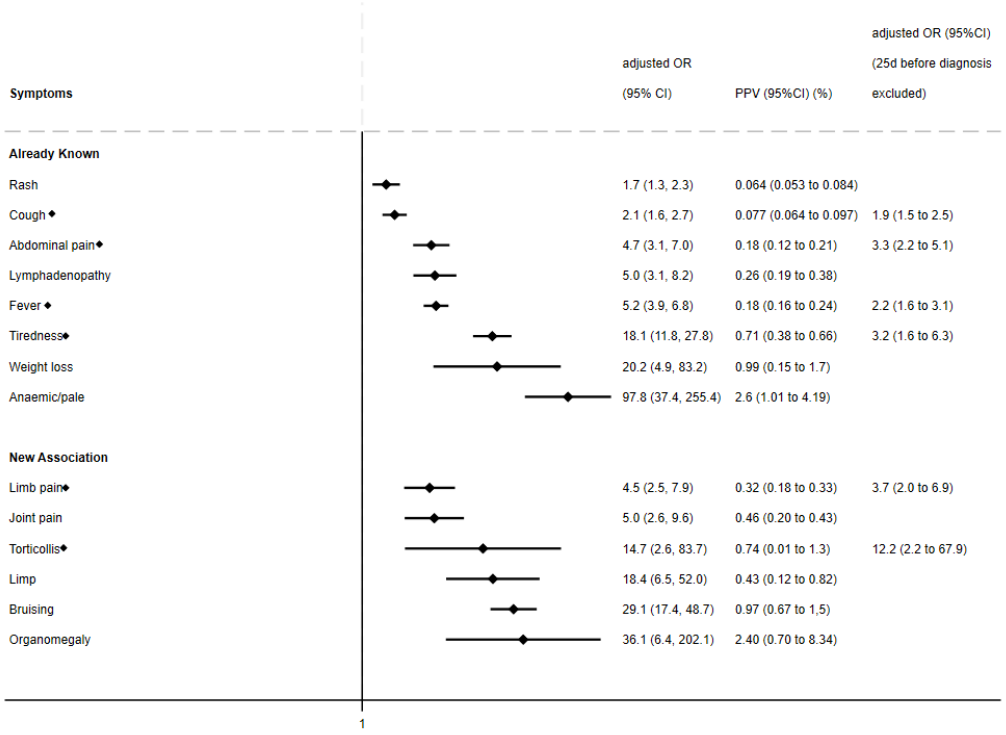
Clinical Feature	**Leukaemia (n, %) N=308**	**OR (95%CI) in univariable analysis**	**Lymphoma (n, %) N=652**	**OR (95%CI) in univariable analysis**	**CNS Tumours (n, %) N=308**	**OR (95%CI) in univariable analysis**	**Sarcoma (n, %) N=343**	**OR (95%CI) in univariable analysis**	**Gonadal germ cell N=366**	**OR (95%CI) in univariable analysis**
Organomegaly	<5	40 (7.32 - 218.8)	<5	-						
Lymphadenopathy	8 (3.1)	10.18 (4.17 – 24.85)	153 (23.5)	131.5 (83.2-207.8)			6 (1.8)	8.3 (3.2-21.5)		
Fever	12 (3.9)	14.28 (6.52-31.23)	16 (2.5)	8.3 (4.5-15.4)	<5	1.2 (0.3-5.3)	<5	0.8 (0.09-5.5)	7 (1.9)	3.2 (1.4-7.3) p=0.008
URTI	43 (14.0)	3.46 (2.43-4.94)	91 (14.0)	3.5 (2.7-4.4)						
Limb pain	14 (4.5)	4.37 (2.41-7.92)	27 (4.2)	3.1 (2.1-4.8)			52 (15.2)	14.0 (9.6-20.3)		
Joint pain	7 (2.3)	3.35 (1.47-7.65)	10	1.8 (0.9-3.4)			36 (10.5)	16.4 (10.4-25.8)		
Limp	<5	-	<5	-	<5	-	<5	-		
Anaemic/pallor	5 (1.7)	24.56 (6.59 - 91.48)	7	34.5 (10.1-117.9)						
Tiredness	15 (4.7)	7.25 (3.92 – 13.38)	12 (1.8)	2.1 (1.1-3.8)	6 (2.0)	2.6 (1.1-6.3)	<5	1.4 (0.4-4.7)	5 (1.5)	1.6 (0.6-3.9)
Abdominal pain	16 (5.2)	3.78 (2.18 – 6.58)	37 (5.7)	3.8 (2.6-5.5)			16 (4.7)	3.7 (2.1-6.4)	29 (7.9)	5.1 (3.3-7.7)
Anorexia	<5	-	<5	-	<5	-	<5	-	<5	-
Rash	15 (3.9)	3.34 (1.78-6.28)	42 (6.4)	5.3 (3.7-7.6)			<5	0.7 (0.2-2.2)		
Abnormal skin lesion	<5	2.0 (0.2-16.8) p=0.3	8 (1.3)	7.9 (3.4-18.1)			<5	6.0 (1.9-19.4)		

Chest pain	8 (2.6)	5.78 (2.58-12.95)	29 (4.5)	8.6 (5.5-13.4)			7 (2.0)	3.7 (1.6-8.4)	<5	1.8 (0.7-5.3) p=0.2
Cough	15 (4.9)	3.45 (1.95 to 6.10)	46 (7.1)	5.6 (4.0-7.8)			8 (2.3)	1.7 (0.8-3.5)		
Breathlessness	<5	1.10 (0.15 to 8.25)	13. (2.0)	5.4 (2.9-10.2)			<5	1.8 (0.4-7.7)		
Headache	14 (4.6)	3.75 (2.07 – 6.82)	9 (1.4)	1.0 (0.5-2.0)	73 (23.7)	24.5 (17.0-35.4)	7 (2.0)	1.6 (0.7-3.4)		
Bruising	9 (3.6)	52.8 (14.1-197.3)	<5	3.5 (1.2-10.4)			6 (1.8)	7.7 (3.0-19.8)		
Vomiting	5 (2.0)	4.4 (1.7-11.6) p=0.003	5 (0.8)	5.4 (2.0-14.7)	7 (2.3)	11.7 (4.6-29.6)	<5	-	9 (2.5)	5.0 (2.4-10.5)
Abdominal mass	<5	-	<5	-			<5	-	10 (2.7)	-
Head/neck lump	<5	-	85 (13.0)	328.8 (133.4-810.6)			<5	6.6 (1.3-32.6)		
Other lump	<5	1.7 (0.4-7.2)	56 (8.6)	22.3 (15.0-33.1)			38 (11.1)	30.8 (18.1-52.2)		
Weight loss	<5	5.72 (1.19 to 27.50)	12 (1.8)	23.3 (10.0-54.0)	<5	4.0 (0.5-33.9)	<5	14.9 (4.5-49.3)	<5	9.6 (1.5-62.9) p=0.02
Pruritus			12 (1.8)	39.9 (15.0 – 106.3)						
Papilloedema					7 (2.3)	-	<5	-		
Hemiparesis					<5	-	<5	-		
Torticollis	<5	6 (1.65 to 21.80)			<5	-	-			
Squint					<5	-				
Abnormal eye movements					<5	-	<5	-		
Visual acuity abnormalities					12 (3.9)	98.6 (21.7-443.8)	<5	-		
Diplopia					5 (1.6)	88.5 (10.3-762.9)	<5	-		
Exopthalmos					<5	-	<5	-		

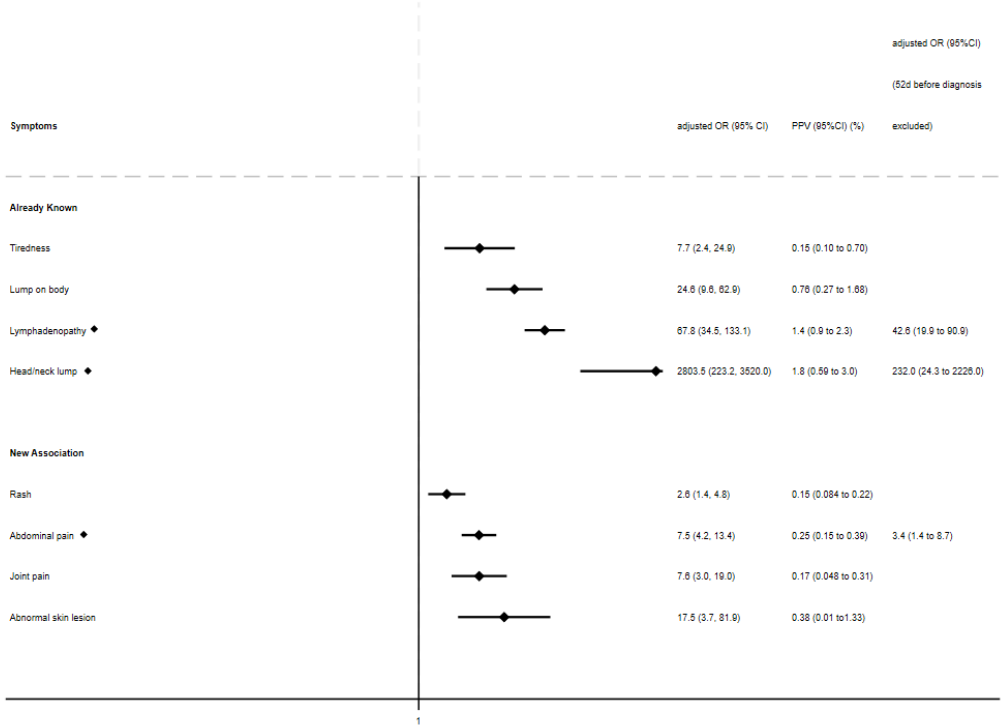
Behavioural Changes					<5	-				
Irritability					<5	-				
Developmental delay					<5	-				
Abnormal cranial nerve signs (any)					<5	-	<5	-		
First seizure					23 (7.5)	216.8 (51.1-920.4)				
Precocious puberty					<5	-				
Haematuria	5 (2.0)	1.32 (0.31 to 5.61)					<5	-	<5	0.8 (0.1-6.2) p=0.7
Constipation							<5	5.8 (1.9-18.2	<5	2.1 (0.5-8.9) p=0.3
Testicular swelling							<5	-	93 (25.4)	198.7 (46.8 to 841.0)
>1 presentation	597 (53.0)	5.85 (5.11 to 6.71)	219 (33.5)	27.8 (22.0-35.2)	46 (17.6)		70 (20.4)	11.2 (8.1-15.5)	76 (20.8)	8.7 (6.4--11.7)

Supplementary Figure 9-3 *Clinical features encountered in primary care associated with a) leukaemias, b) lymphomas, c) CNS tumours, d) sarcomas in children (0-14 years) in multivariable conditional logistic regression models. Early symptoms denoted with *. OR= Odds ratio CI= Confidence interval PPV=Positive Predictive Value.*

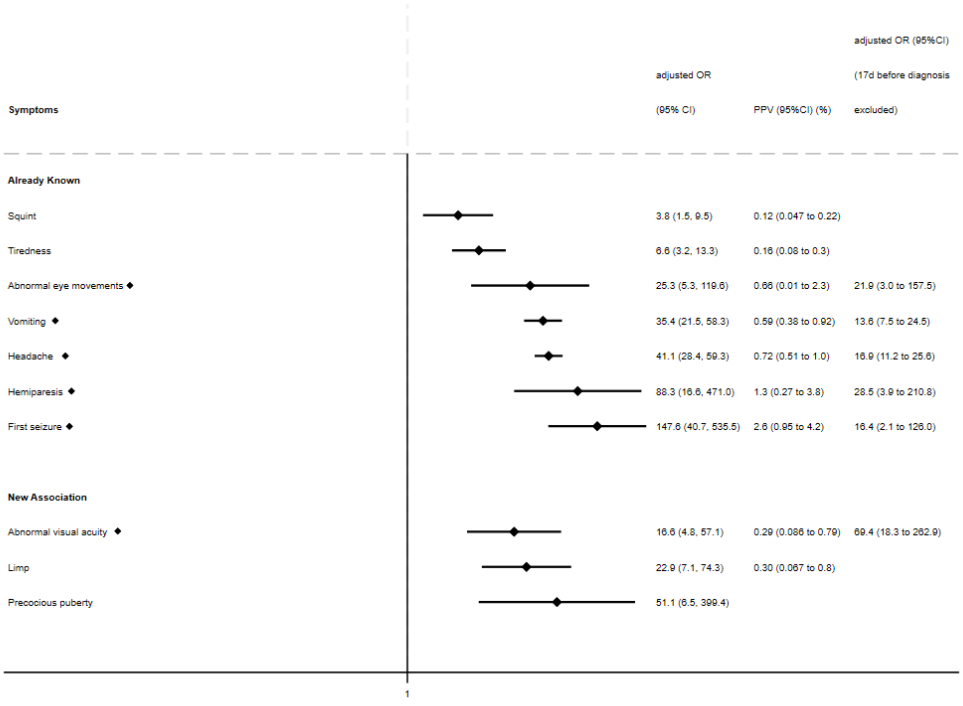
a)



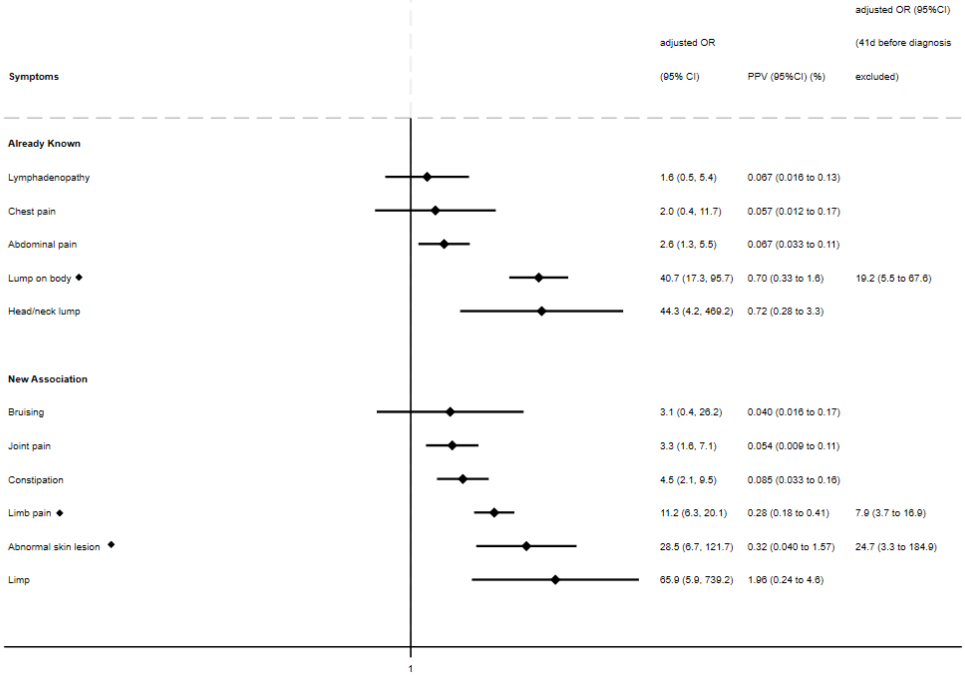
b)



c)

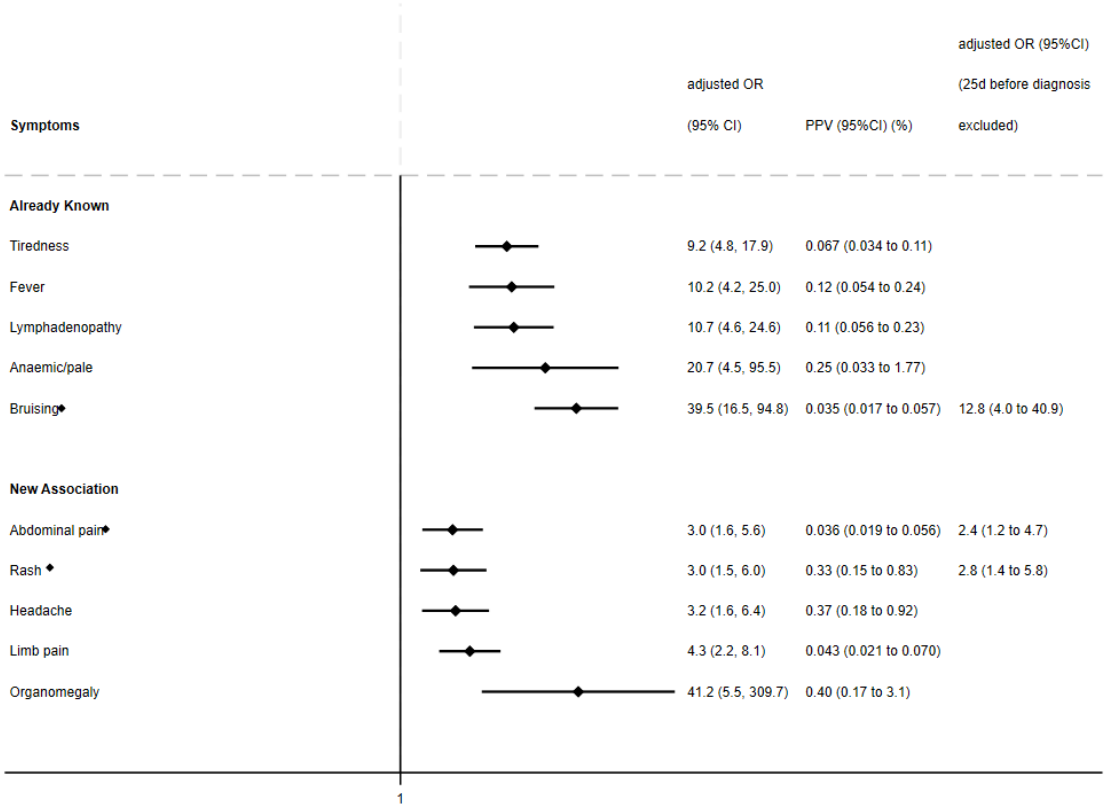

Symptoms
adjusted OR (95% CI)
PPV (95%CI) (%)
adjusted OR (95%CI) (17d before diagnosis excluded)
Already Known
Squint 3.8 (1.5, 9.5) 0.12 (0.047 to 0.22)
Tiredness 6.6 (3.2, 13.3) 0.16 (0.08 to 0.3)
Abnormal eye movements ◆ 25.3 (5.3, 119.6) 0.66 (0.01 to 2.3) 21.9 (3.0 to 157.5)
Vomiting ◆ 35.4 (21.5, 58.3) 0.59 (0.38 to 0.92) 13.6 (7.5 to 24.5)
Headache ◆ 41.1 (28.4, 59.3) 0.72 (0.51 to 1.0) 16.9 (11.2 to 25.6)
Hemiparesis ◆ 88.3 (16.6, 471.0) 1.3 (0.27 to 3.8) 28.5 (3.9 to 210.8)
First seizure ◆ 147.6 (40.7, 535.5) 2.6 (0.95 to 4.2) 16.4 (2.1 to 126.0)
New Association
Abnormal visual acuity ◆ 16.6 (4.8, 57.1) 0.29 (0.086 to 0.79) 69.4 (18.3 to 262.9)
Limp 22.9 (7.1, 74.3) 0.30 (0.067 to 0.8)
Precocious puberty 51.1 (6.5, 399.4)
1

d)

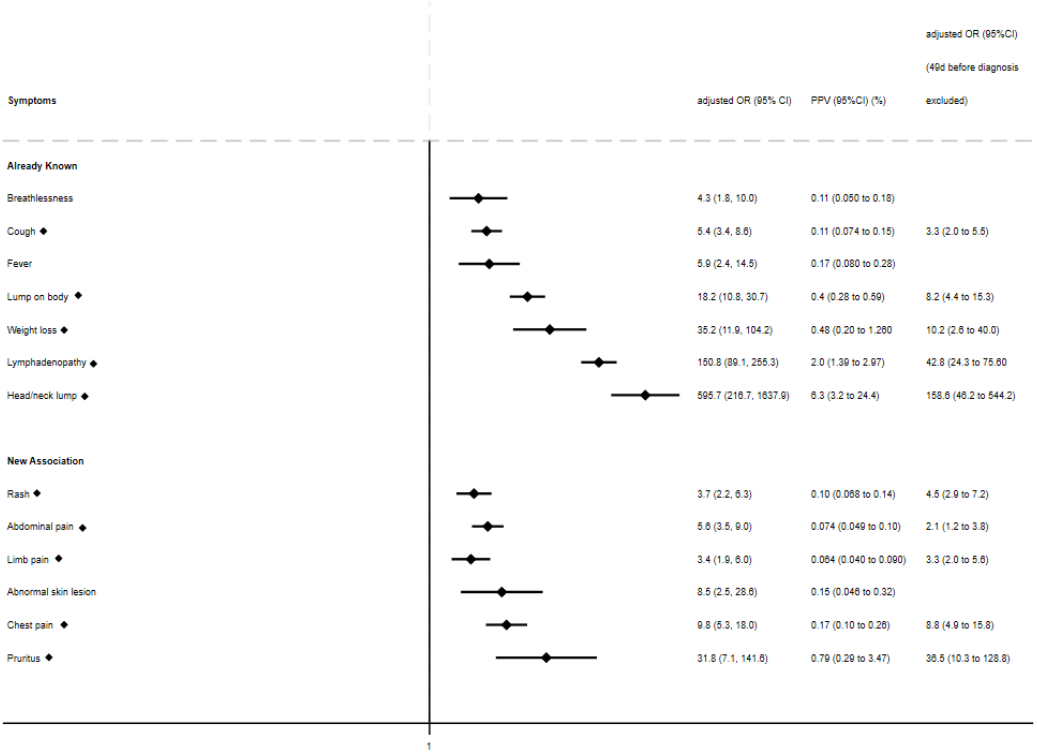

Symptoms
adjusted OR (95% CI)
PPV (95%CI) (%)
adjusted OR (95%CI) (41d before diagnosis excluded)
Already Known
Lymphadenopathy 1.6 (0.5, 5.4) 0.067 (0.016 to 0.13)
Chest pain 2.0 (0.4, 11.7) 0.057 (0.012 to 0.17)
Abdominal pain 2.6 (1.3, 5.5) 0.067 (0.033 to 0.11)
Lump on body ◆ 40.7 (17.3, 95.7) 0.70 (0.33 to 1.6) 19.2 (5.5 to 67.6)
Head/neck lump 44.3 (4.2, 469.2) 0.72 (0.28 to 3.3)
New Association
Bruising 3.1 (0.4, 26.2) 0.040 (0.016 to 0.17)
Joint pain 3.3 (1.6, 7.1) 0.054 (0.009 to 0.11)
Constipation 4.5 (2.1, 9.5) 0.085 (0.033 to 0.16)
Limb pain ◆ 11.2 (6.3, 20.1) 0.28 (0.18 to 0.41) 7.9 (3.7 to 16.9)
Abnormal skin lesion ◆ 28.5 (6.7, 121.7) 0.32 (0.040 to 1.57) 24.7 (3.3 to 184.9)
Limp 65.9 (5.9, 739.2) 1.95 (0.24 to 4.6)
1

Supplementary Figure 9-4*. Clinical features encountered in primary care associated with a) leukaemia, b) lymphoma, c) CNS tumours, d) sarcomas in TYA (15-24 years) in multivariable conditional logistic regression models. Early symptoms denoted with *. OR= Odds ratio CI= Confidence interval PPV=Positive Predictive Value.*

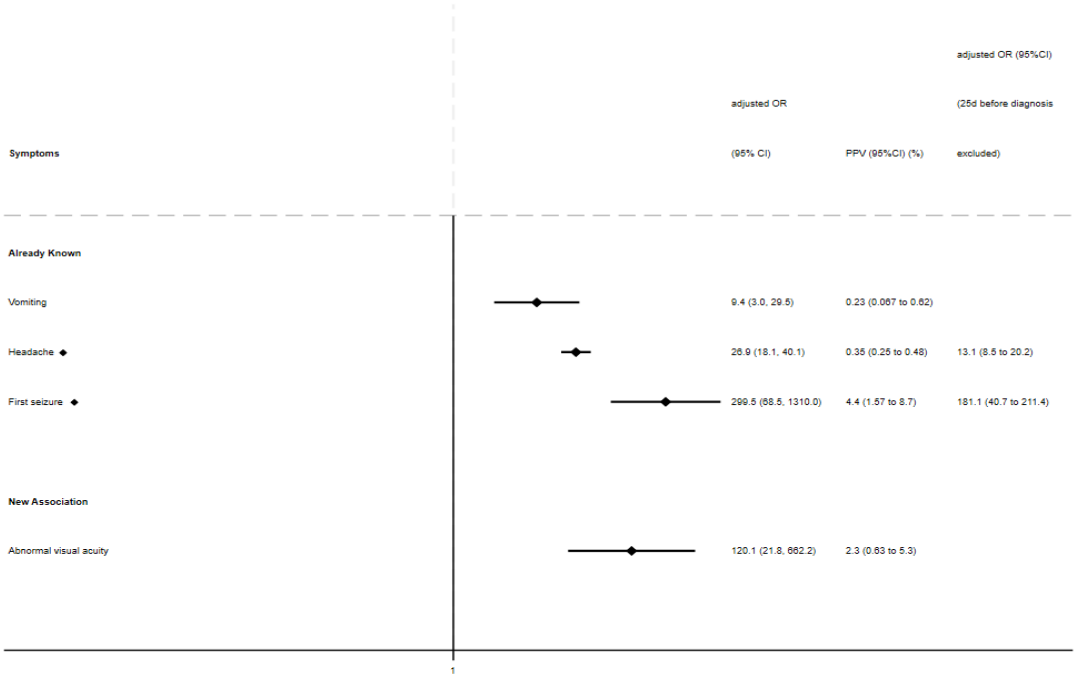
a)



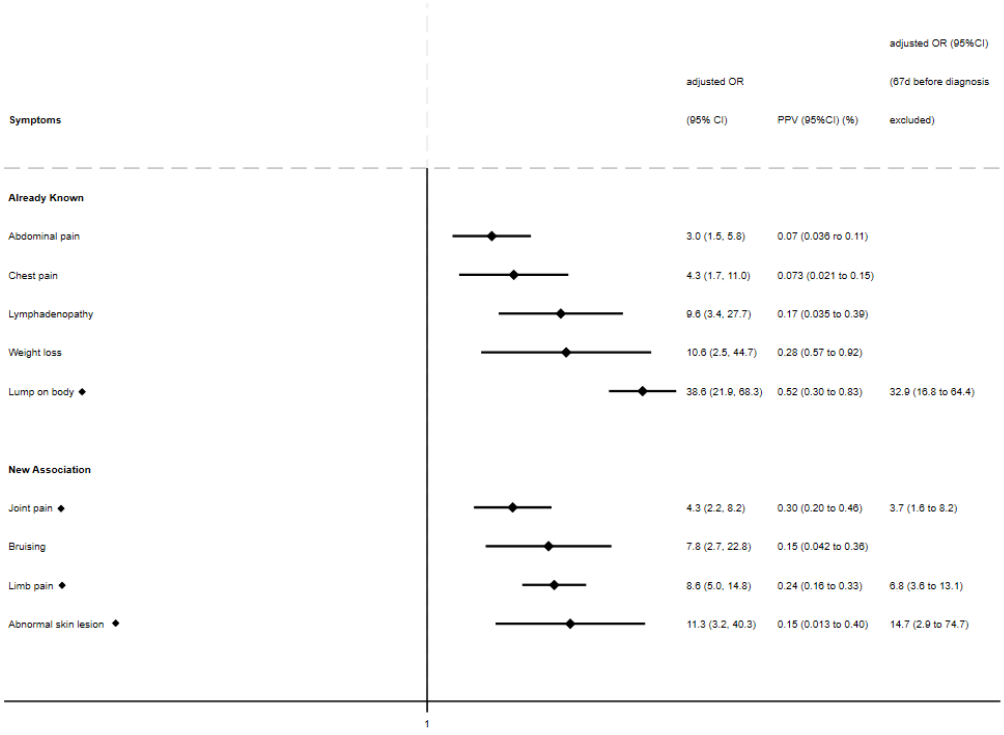
b)


Symptoms
adjusted OR (95% CI)
PPV (95%CI) (%)
adjusted OR (95%CI) (49d before diagnosis excluded)
Already Known
Breathlessness
4.3 (1.8, 10.0)
0.11 (0.050 to 0.18)
Cough
5.4 (3.4, 8.6)
0.11 (0.074 to 0.15)
3.3 (2.0 to 5.5)
Fever
5.9 (2.4, 14.5)
0.17 (0.080 to 0.28)
Lump on body
18.2 (10.8, 30.7)
0.4 (0.28 to 0.59)
8.2 (4.4 to 15.3)
Weight loss
35.2 (11.9, 104.2)
0.48 (0.20 to 1.260
10.2 (2.6 to 40.0)
Lymphadenopathy
150.8 (89.1, 255.3)
2.0 (1.39 to 2.97)
42.8 (24.3 to 75.60
Head/neck lump
595.7 (216.7, 1637.9)
6.3 (3.2 to 24.4)
158.6 (46.2 to 544.2)
New Association
Rash
3.7 (2.2, 6.3)
0.10 (0.068 to 0.14)
4.5 (2.9 to 7.2)
Abdominal pain
5.6 (3.5, 9.0)
0.074 (0.049 to 0.10)
2.1 (1.2 to 3.8)
Limb pain
3.4 (1.9, 6.0)
0.064 (0.040 to 0.090)
3.3 (2.0 to 5.6)
Abnormal skin lesion
8.5 (2.5, 28.6)
0.15 (0.046 to 0.32)
Chest pain
9.8 (5.3, 18.0)
0.17 (0.10 to 0.26)
8.8 (4.9 to 15.8)
Pruritus
31.8 (7.1, 141.6)
0.79 (0.29 to 3.47)
36.5 (10.3 to 128.8)
1

c)


Symptoms
adjusted OR (95% CI)
PPV (95%CI) (%)
adjusted OR (95%CI) (25d before diagnosis excluded)
Already Known
Vomiting
9.4 (3.0, 29.5)
0.23 (0.067 to 0.62)
Headache
26.9 (18.1, 40.1)
0.35 (0.25 to 0.48)
13.1 (8.5 to 20.2)
First seizure
299.5 (68.5, 1310.0)
4.4 (1.57 to 8.7)
181.1 (40.7 to 211.4)
New Association
Abnormal visual acuity
120.1 (21.8, 662.2)
2.3 (0.63 to 5.3)
1

d)


Symptoms
adjusted OR (95% CI)
PPV (95%CI) (%)
adjusted OR (95%CI) (67d before diagnosis excluded)
Already Known
Abdominal pain
3.0 (1.5, 5.8)
0.07 (0.036 ro 0.11)
Chest pain
4.3 (1.7, 11.0)
0.073 (0.021 to 0.15)
Lymphadenopathy
9.6 (3.4, 27.7)
0.17 (0.035 to 0.39)
Weight loss
10.6 (2.5, 44.7)
0.28 (0.57 to 0.92)
Lump on body
38.6 (21.9, 68.3)
0.52 (0.30 to 0.83)
32.9 (16.8 to 64.4)
New Association
Joint pain
4.3 (2.2, 8.2)
0.30 (0.20 to 0.46)
3.7 (1.6 to 8.2)
Bruising
7.8 (2.7, 22.8)
0.15 (0.042 to 0.36)
Limb pain
8.6 (5.0, 14.8)
0.24 (0.16 to 0.33)
6.8 (3.6 to 13.1)
Abnormal skin lesion
11.3 (3.2, 40.3)
0.15 (0.013 to 0.40)
14.7 (2.9 to 74.7)
1

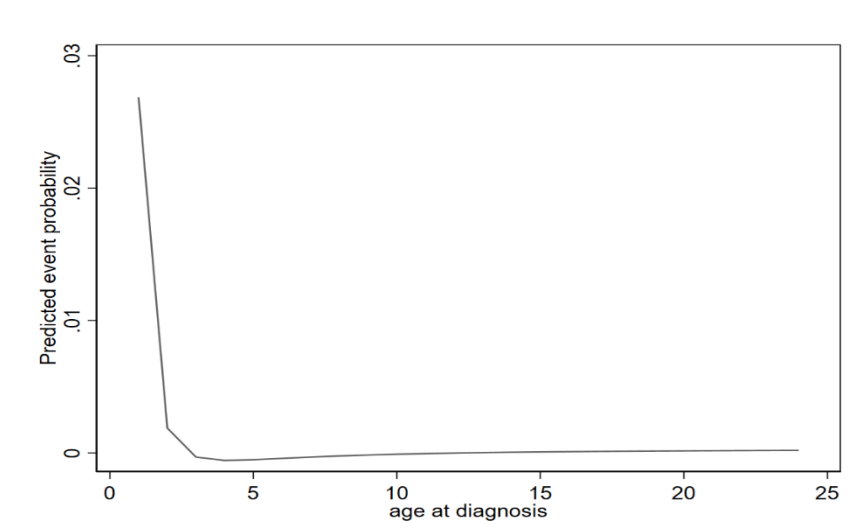
Supplementary Table 9-12 *Proportional hazard test using Schoenfeld residuals demonstrating that proportionally assumption is violated (p<0.05)*

	Global test	**p-value**
Blood cancers	32.55	0.005
Solid Cancers	21.8	0.04

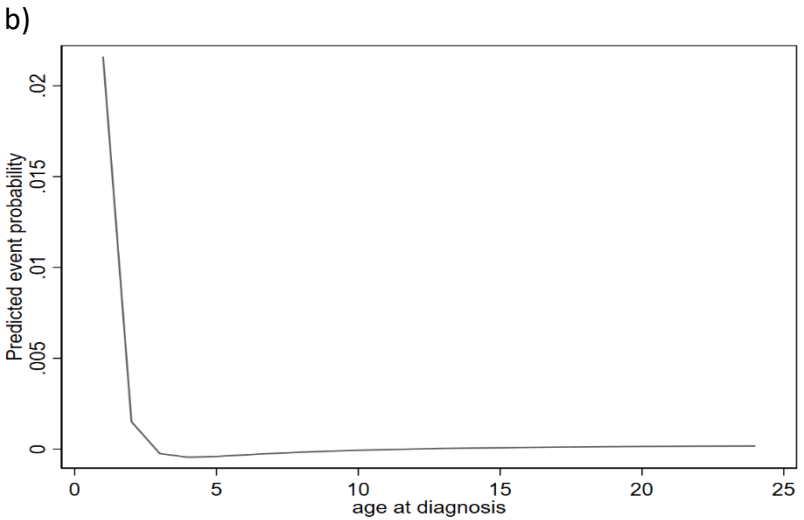
Supplementary Table 9-13 *Sample size requirements for clinical prediction model development using Riley et al., 2020(128)*

Cohort	**R^2_{cs}**	**Parameters**	**Diagnosis Rate**	**Required Sample size**	**Events per predictor parameter**
Blood cancers (0-24 years)	0.00027	30	0.000089 (89 per million)	999850	3
Solid Tumours (0-24 years)	0.0003	30	0.0001 (100 per million)	870818	3
Leukaemias (0-24 years)	0.0001	30	0.000033 (33 per million)	2699850	3
Lymphomas (0-24 years)	0.0001	30	0.000031 (31 per million)	2699850	3
CNS Tumours (0-24 years)	0.00008	30	0.000026 (26 per million)	3374850	3
Sarcomas (0-24 years)	0.00005	30	0.000015 (15 per million)	5399850	3
Abdominal tumours (0-24 years)	0.00005	30	0.000017 (17 per million)	5399850	3

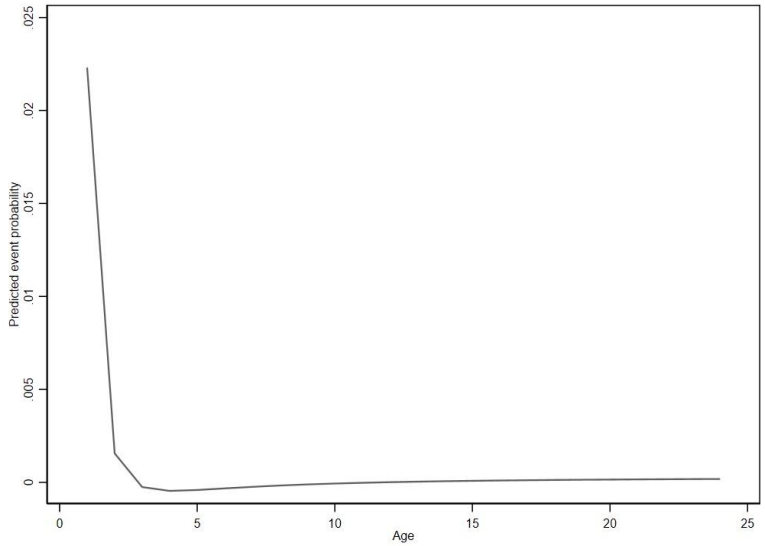
Supplementary Figure 9-4 *Fractional polynomial terms for age for a) the logistic model for solid cancers and b) logistic model for haematological cancers c) logistic model for leukaemias, d) logistic model for CNS tumours*



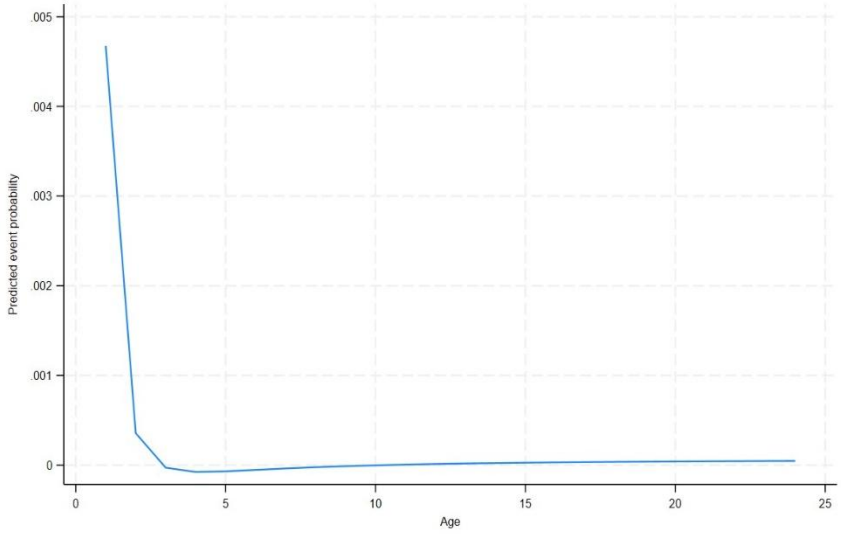
b)



c)



d)



Supplementary Table 9-14 *Meta-regression analysis for age-at-diagnosis, ethnicity, Townsend deprivation score, described as Townsend Quintile Coefficient in this table. 95%CI = 95% confidence interval. AUC= area-under-the-cure*

	Age Coefficient (95%CI)	p-value	Variance accounted	Ethnicity Coefficient (95%CI)	p-value	Variance accounted	Townsend Quintile Coefficient (95%CI)	p-value	Variance accounted (%)
Solid Cancers									
AUC	0.00015 (-0.0047 to 0.0049)	0.95	0%	0.00014 (-0.0045 to 0.0048)	0.95	0%	0.00015 (-0.004 to 0.0049)	0.83	0%
Calibration slope	-0.002 (-0.012 to 0.0077)	0.66	0%	0.0022 (-0.011 to 0.0074)	0.65	0%	-0.0022 (-0.12 to 0.0078)	0.63	0%
Calibration-in-the-large	-0.002 (-0.012 to 0.0077)	0.66	0%	0.0022 (-0.011 to 0.0074)	0.65	0%	-0.0022 (-0.12 to 0.0078)	0.63	0%
Haematological Cancers									
AUC	-0.0018 (-0.0071 to 0.0034)	0.49	0%	-0.0018 (-0.0069 to 0.0032)	0.49	0%	-0.0018 (-0.0069 to 0.0032)	0.49	0%
Calibration slope	0.0037 (-0.0061 to 0.0068)	0.91	0%	0.0029 (-0.0059 to 0.0065)	0.93	0%	0.00037 (0.9961 to 0.0068)	0.91	0%
Calibration-in-the-large	0.0037 (-0.0061 to 0.0068)	0.91	0%	0.0029 (-0.0059 to 0.0065)	0.93	0%	0.00037 (0.9961 to 0.0068)	0.91	0%
Haematological Cancers- symptom only									
AUC	-0.0020 (-0.0078 to 0.0039)	0.51	0%	-0.0019 (-0.0076 to 0.0038)	0.52	0%	-0.0020 (-0.0078 to 0.0039)	0.5	0%
Calibration slope	-0.0019 (-0.0098 to 0.0060)	0.6	0%	-0.0019 (-0.0096 to 0.0058)	0.6	0%	-0.0019 (-0.010 to 0.006)	0.6	0%

Calibration-in-the-large	-0.0019 (-0.0098 to 0.0060)	0.6	0%	-0.0019 (-0.0096 to 0.0058)	0.6	0%	-0.0019 (-0.010 to 0.006)	0.6	0%
Leukaemias (symptom + test model)									
AUC	-0.004 (-0.0089 to 0.00086)	0.1	13.7%	-0.004 (-0.0089 to 0.00086)	0.11	13.1%	-0.004 (-0.0089 to 0.00086)	0.1	13.7%
Calibration slope	-0.0061 (-0.017 to 0.005)	0.3	5.1%	-0.0059 (-0.17 to 0.004)	0.3	5.3%	-0.0061 (-0.17 to 0.005)	0.3	5.1%
Calibration-in-the-large	-0.0061 (-0.017 to 0.005)	0.3	5.1%	-0.0059 (-0.17 to 0.004)	0.3	5.3%	-0.0061 (-0.17 to 0.005)	0.3	5.1%
Leukaemias (symptom model)									
AUC	-0.0063 (-0.011 to -0.0011)	0.02	39.1%	-0.0062 (-0.011 to -0.0012)	0.02	39.8%	-0.0063 (-0.011 to -0.0011)	0.02	39.1%
Calibration slope	-0.0045 (-0.015 to 0.0059)	0.4	0%	-0.0044 (-0.015 to 0.0056)	0.4	1.0%	-0.0045 (-0.015 to 0.0059)	0.4	0%
Calibration-in-the-large	-0.0045 (-0.015 to 0.0059)	0.4	0%	-0.0044 (-0.015 to 0.0056)	0.4	1.0%	-0.0045 (-0.015 to 0.0059)	0.4	0%
Lymphomas (symptom + test model)									
AUC	-0.005 (-0.012 to 0.0012)	0.1	17.5%	-0.005 (-0.011 to 0.0011)	0.1	18.0%	-0.0052 (-0.012 to 0.0012)	0.1	17.5%
Calibration slope	-0.004 (-0.013 to 0.005)	0.4	0 %	-0.004 (-0.013 to 0.005)	0.4	0 %	-0.004 (-0.013 to 0.005)	0.4	0 %
Calibration-in-the-large	-0.004 (-0.013 to 0.005)	0.4	0 %	-0.004 (-0.013 to 0.005)	0.4	0 %	-0.004 (-0.013 to 0.005)	0.4	0 %
Lymphomas (symptom model)									

AUC	-0.0056 (-0.013 to 0.0012)	0.1	17.5%	-0.0056 (-0.012 to 0.0011)	0.1	18.1%	-0.0056 (-0.013 to 0.0012)	0.1	17.5%
Calibration slope	-0.0018 (-0.0099 to 0.0063)	0.7	0%	-0.0018 (-0.0096 to 0.0061)	0.7	0%	-0.0018 (-0.0099 to 0.0063)	0.7	0%
Calibration-in-the-large	-0.0018 (-0.0099 to 0.0063)	0.7	0%	-0.0018 (-0.0096 to 0.0061)	0.7	0%	-0.0018 (-0.0099 to 0.0063)	0.7	0%
CNS tumours									
AUC	0.00015 (-0.0032 to 0.0058)	0.82	0%	0.00013 (-0.0029 to 0.0061)	0.80	0%	0.00015 (-0.0032 to 0.0058)	0.85	0%
Calibration slope	-0.005 (-0.018 to 0.0082)	0.46	0%	0.004 (-0.017 to 0.0080)	0.48	0%	-0.005 (-0.18 to 0.0083)	0.44	0%
Calibration-in-the-large	-0.005 (-0.017 to 0.0082)	0.47	0%	0.0046 (-0.017 to 0.0080)	0.47	0%	-0.0048 (-0.18 to 0.0082)	0.47	0%

Supplementary Table 9-15 *Search Strategies for Literature Review in Chapter 5*

Search Strategy for All Cancers		
Population (1)	(cancer* OR malignan* OR tum?or* OR oncolog*)	10,592,672
Outcome (2)	(diagnos* OR diagnos* adj1 time OR diagnos* adj interval* OR emergenc* adj diagnos*)	15,992
Combined	1 AND 2	4,062

Search Strategy for CTYA Cancers		
Population (1)	(p?ediatric* OR child* OR adolescen* OR teenage* OR young adj1 adult*") AND (cancer* OR malignan* OR tum?or*)	197,669
Outcome (2)	(diagnos* OR diagnos* adj1 time OR diagnos* adj interval* OR emergenc* adj diagnos*)	15,992
Combined	1 AND 2	279

Supplementary Table 9-16 *Histological classification of high risk cancer biology. ICD-O-3=International Classification of Diseases for Oncology 3.*

Cancer subtype	**High-risk Histological Codes as per ICD-O-3**
Leukaemias	9811, 9812, 9813, 9814, 9815, 9816, 9817, 9818, 9835, 9836, 9861, 9871, 9872, 9873, 9874, 9877, 9879, 9896, 9897, 9898
Lymphomas	9596, 9679, 9680, 9684, 9685, 9686, 9687, 9688, 9702, 9714, 9716, 9719, 9727, 9728, 9729, 9737, 9823, 9837
CNS Tumours	9505, 9508, 9440, 9441, 9442, 9445, 9470, 9471, 9472, 9473, 9474, 9424, 9393, 9380, 9385
Sarcomas	8800, 8804, 8810, 8830, 8900, 8910, 8920, 9040, 9041, 9043, 9180, 9181, 9184, 9185, 9186, 9187, 9192, 9260, 9364, 9365
Abdominal tumours	8954, 8963, 8964, 8970, 8170, 9500 (neuroblastoma/hepatoblastoma restricted to >18 months for high risk)
Gonadal germ cells	9070, 9071, 9100, 9101, 9102

Supplementary Table 9-17 *Baseline characteristics of study cohort for COVID-19 pandemic for time period 1st January – 1st May 2017-2021. All % are column percentages. Age rounded up to the nearest integer. IQR – interquartile range*

	2017-2019 (Col %)	**2020 (Col %)**	**2021 (Col %)**
Total	613	232	270
Age (median ± IQR)	15 (6,21)	14 (6,20)	12 [5, 19]
Age Group			
0-14 years	302 (49.3)	97 (41.8)	129 (47.8)
15-24 years	311 (50.7)	135 (58.2)	141 (52.2)
Sex			
Female	272 (44.3)	93 (40.1)	129 (47.8)
Male	341 (55.7)	139 (59.9)	141 (57.2)
Ethnicity			
White	205 (33.4)	84 (36.2)	92 (34.1)
Asian	73 (11.9)	22 (9.5)	24 (8.9)
Black	25 (4.1)	13 (5.6)	6 (2.2)
Other	33 (5.4)	16 (6.9)	14 (5.2)
Not Recorded	277 (45.2)	97 (41.8)	97 (41.8)
Townsend Deprivation Score			
1 (lowest deprivation)	125 (20.4)	46 (20.1)	53 (19.8)
2	121 (19.7)	41 (17.9)	68 (25.4)
3	120 (19.6)	48 (21.0)	59 (22.0)
4	127 (20.7)	54 (23.6)	45 (16.8)
5 (highest deprivation)	112 (18.3)	40 (17.5)	43 (16.0)
Not Recorded	8 (1.3)	<5	<5
Tumour Type			
Leukaemias	148 (24.1)	67 (29.0)	65 (25.1)
Lymphomas	139 (22.7)	47 (20.3)	59 (21.9)
CNS Tumours	174 (28.4)	68 (29.3)	92 (34.1)
Sarcoma	127 (20.7)	48 (20.7)	50 (18.5)
Renal Tumours	24 (4.1)	<5	<5

Supplementary Table 9-18 *Baseline characteristics of study cohort for COVID-19 pandemic for time period 2nd May- 31st December 2017-2021. All % are column percentages. Age rounded up to the nearest integer. IQR – interquartile range*

	2017-2019 (Col %)	**2020 (Col %)**	**2021**
Total	1342	513	536
Age (median ± IQR)	15 (7,21)	12 [4,18]	11 [4, 19]
Age Group			
0-14 years	622 (46.4)	264 (51.5)	266 (49.5)
15-24 years	720 (53.6)	249 (48.5)	270 (50.4)
Sex			
Female	614 (45.8)	231 (45.0)	237 (44.6)
Male	728 (54.3)	282 (55.0)	299 (55.4)
Ethnicity			
White	473 (35.2)	183 (35.7)	214 (39.9)
Asian	136 (10.1)	50 (9.8)	47 (8.8)
Black	46 (3.4)	14 (2.7)	21 (3.9)
Other	68 (5.1)	37 (7.2)	26 (4.9)
Not Recorded	620 (46.2)	229 (44.6)	228 (42.5)
Townsend Deprivation Score			
1 (lowest deprivation)	241 (18.0)	103 (20.1)	119 (22.2)
2	259 (19.3)	114 (22.2)	100 (18.7)
3	295 (22.0)	105 (20.5)	93 (17.4)
4	279 (20.8)	86 (16.8)	110 (20.5)
5 (highest deprivation)	252 (18.8)	99 (19.3)	106 (19.8)
Not Recorded	16 (1.2)	6 (1.2)	8 (1.5)
Tumour Type			
Leukaemias	301 (22.4)	138 (26.9)	148 (27.6)
Lymphomas	301 (22.4)	100 (19.5)	117 (21.8)
CNS Tumours	424 (31.6)	154 (30.0)	148 (27.6)
Sarcomas	260 (19.4)	94 (18.3)	111 (20.7)
Renal Tumours	56 (4.2)	27 (5.3)	12 (2.2)

Supplementary Table 9-19 *List of pre-defined clinical features associated with a diagnosis of CTYA cancer (adapted from NICE Guideline [NG12] – Suspected Cancer: Recognition and Referral)*

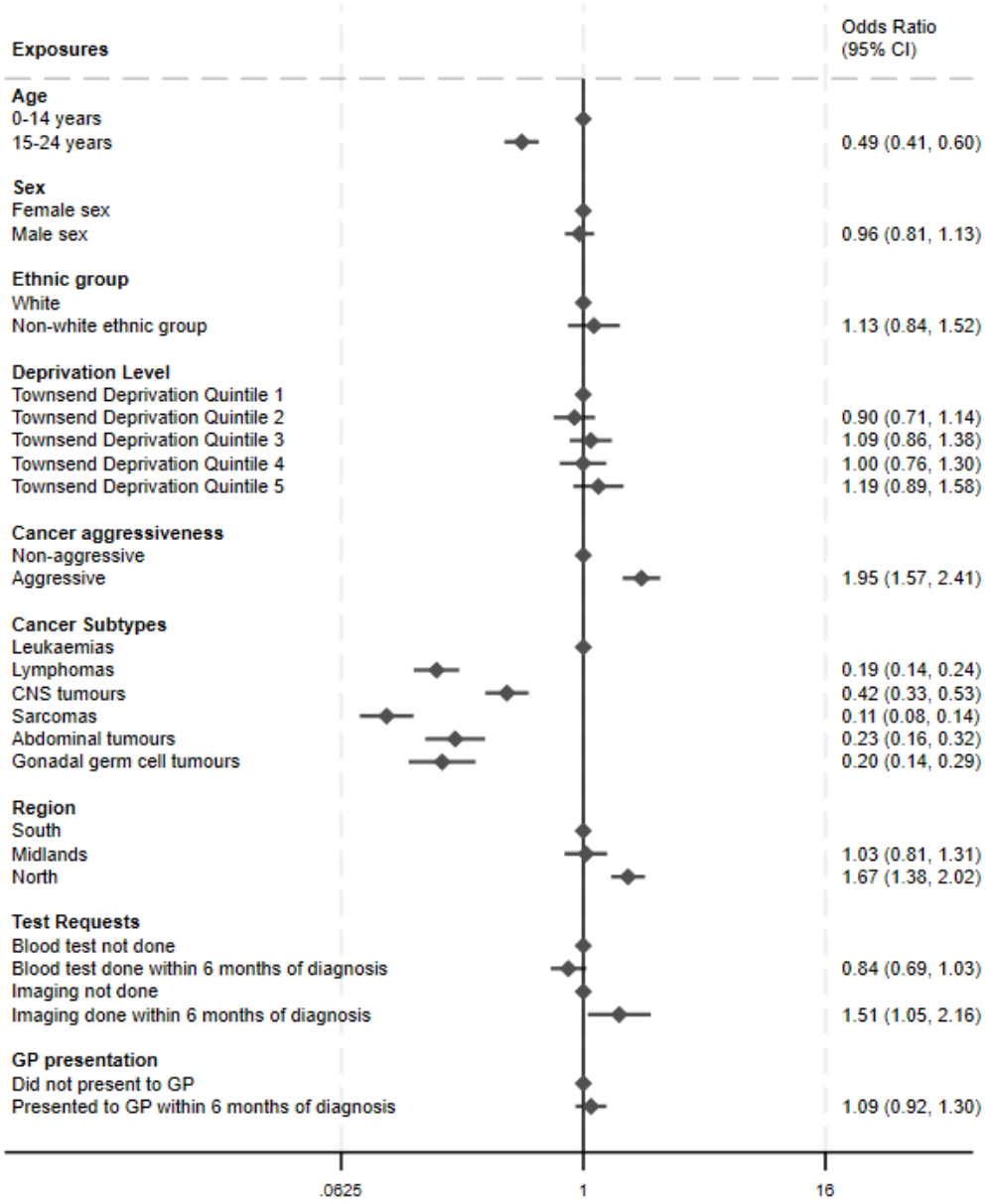
Cancer	Clinical Features
CNS Tumours	Raised intracranial pressure (vomiting, headache)
	Abnormal neurological or cerebellar function (including changes in gait, coordination, vision)
Lymphoma	Abdominal mass (including hepatomegaly, splenomegaly)
	B-symptoms (fever/sweats/weight loss)
	Breathlessness
	Pruritus
	Unexplained lymphadenopathy
Leukaemia	Abdominal mass (including hepatomegaly/splenomegaly)
	Petechiae (abnormal bruising/bleeding)
	B-symptoms (fever/sweats/weight loss)
	Fatigue
	Generalised lymphadenopathy
	Bone pain
	Pallor
Sarcoma	Bone swelling or pain
	Unexplained lump
Renal Tumours	Abdominal pain/mass

Supplementary Table 9-20 *The association between exposures-of-interest and 25th centile, median and 75th centile diagnostic intervals in adjusted quantile regression models following multiple imputation. Adjusted for age, sex, cancer subtype and biology. When variable of interest was an adjusting factor, that factor was excluded from regression models. 95%CI = 95% confidence interval.*

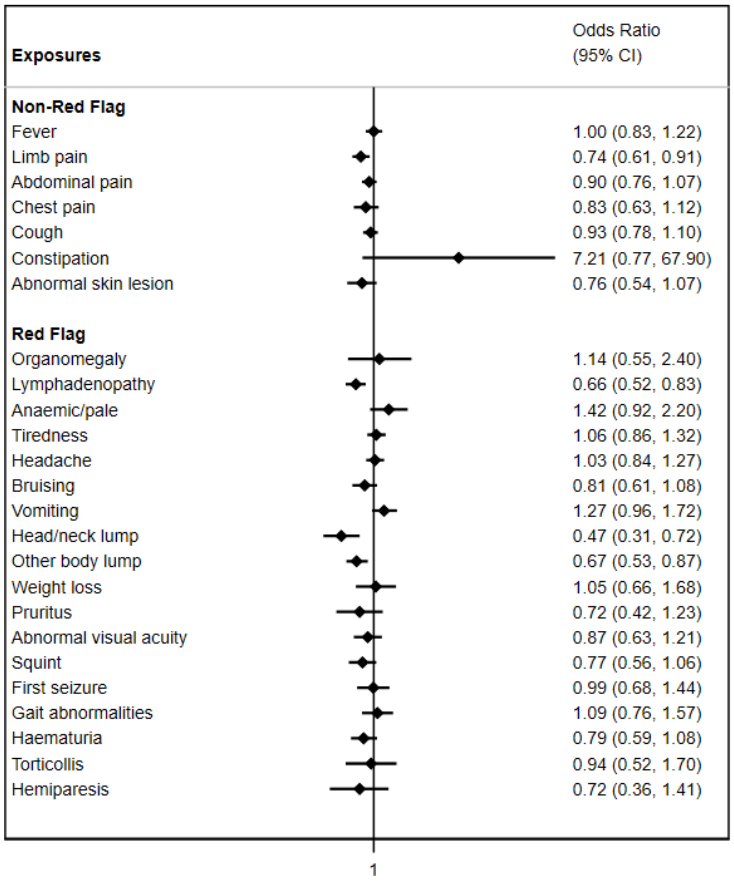
Percentile	Adjusted (days, 95%CI)	p-value
Age category		
25th	+2 [-1, +3]	0.06
50th	+5 [-1,+9]	0.06
75th	+7 [-2,+16]	0.1
Sex		
25th	-1 [-2, +1]	0.3
50th	-1 [-5, +3]	0.6
75th	-4 [-12,+4]	0.3
Ethnicity		
25th	0 [-2, +2]	0.7
50th	-1 [-5, +3]	0.6
75th	-1 [-13,+11]	0.8
Deprivation		
Townsend 1 vs Townsend 2		
25th	0 [-1, +2]	0.8
50th	-1 [-6, +5]	0.8
75th	-2 [-12, +9]	0.7
Townsend 1 vs Townsend 3		
25th	0 [-1 +2]	0.8
50th	0 [-5, +6]	0.9
75th	+2 [-8, +13]	0.7
Townsend 1 vs Townsend 4		
25th	+2 [-1, +4]	0.1
50th	+4 [-2, +11]	0.2
75th	+8 [-4, +21]	0.2
Townsend 1 vs Townsend 5		
25th	+1 [-2 +3]	0.5
50th	+2 [-5, +8]	0.6
75th	+7 [-8 +20]	0.4
Cancer Subtype		
Leukaemia vs lymphoma		
25th	+10 [+7, +13]	<0.001
50th	+22 [+15, +28]	<0.001
75th	+39 [+28, +50]	<0.001
Leukaemia vs CNS tumours		
25th	+3 [+1, +5]	0.002
50th	+8 {+3, +13]	0.004
75th	+15 [+3, +27]	0.01

Leukaemia vs sarcomas		
25th	+14 [+8, +20]	<0.001
50th	+32 [+4, +23]	<0.001
75th	+56 [+37, +76]	<0.001
Leukaemia vs abdominal tumours		
25th	+5 [+1, +9]	0.008
50th	+14 [+4, +23]	0.005
75th	+23 [+4, +42]	0.02
Leukaemia vs gonadal tumours		
25th	+1 [-3 +5]	0.6
50th	+2 [-10, +14]	0.7
75th	+2 [-22, +27]	0.8
Histological aggressiveness		
25th	-2 [-4, -1]	0.02
50th	-7 (-12, -2)	0.004
75th	-13 [-23, -3]	0.01
Red flags		
25th	0 [-1, +2]	0.8
50th	+1 [-4, +6]	0.7
75th	+4 [-6, +15]	0.4
Blood test request		
25th	+3 [+1, +6]	0.007
50th	+8 [+3, +14]	0.004
75th	+12 [+3, +22]	0.01
Imaging test request		
25th	+3 [-2, +8]	0.3
50th	+8 [-6, +21]	0.3
75th	+12 [-20, +32]	0.5
Region		
South vs North		
25th	0 [-2, +2]	0.8
50th	-1 [-5, +4]	0.9
75th	-3 [-9, +4]	0.4

Supplementary Table 9-21 *Forest plot demonstrating adjusted odds of emergency presentation compared to non-emergency presentation in CTYA cancers for each factor/exposure in multiple imputation models. 95% CI = 95% confidence interval.*



Supplementary Table 9-22 *Forest plot demonstrating adjusted odds of emergency presentation compared to non-emergency presentation in CTYA cancers for symptoms only in multiple imputation models. 95% CI = 95% confidence interval.*

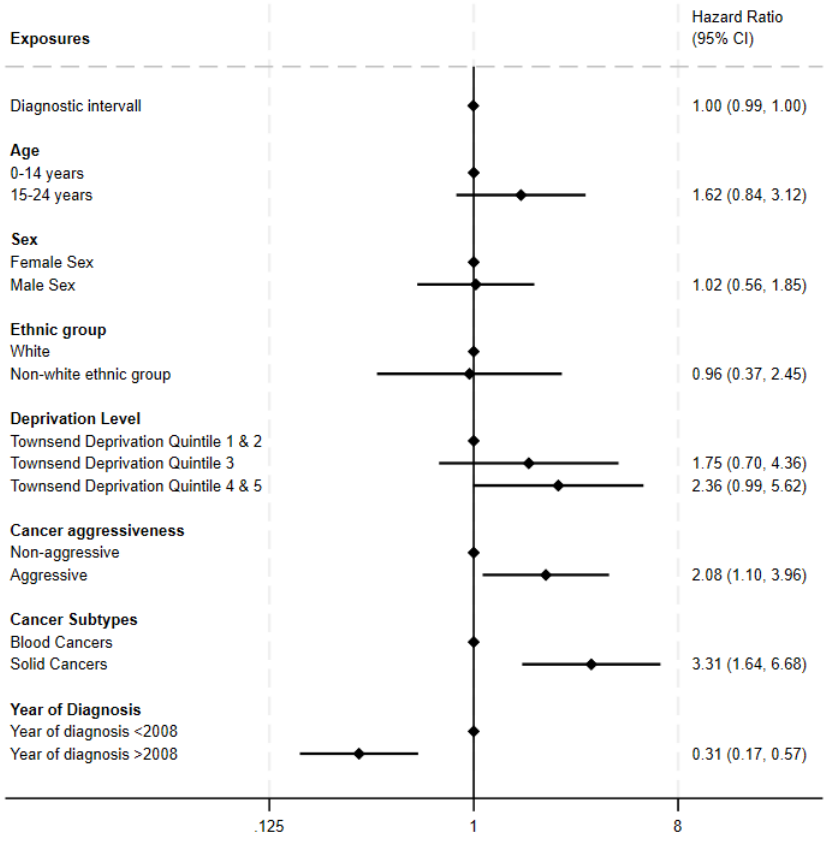


Supplementary Table 9-23 *Crude incidence rates per million population by sex in generated cohort restricted to exit date of 31st January 2017 and nationally reported world age-standardised incidence rates per million population by sex in England.*

Cancer type	**Cohort incidence rate per million population (0-24 years)**	**World age-standardised incidence rate per million population (0-24 years)**
Leukaemias, myeloproliferative disease and myelodysplastic diseases	Female: 37.0 (95%CI 34.0-42.0)	Female: 36.1
	Male:40.0 (95% 35.7-43.8)	Male 44.1
Lymphomas & reticuloendothelial neoplasms	Female: 26.0 (95%CI 23.0-29.4)	Female: 24.2
	Male: 35.3 (95% 32.3-39.3)	Male 33.5
CNS Tumours & miscellaneous intracranial and intraspinal tumours	Female: 30.6 (95%CI 27.3-34.2)	Female: 33.6
	Male: 37.6 (95% 34.1-41.3)	Male 36.9
Sarcomas (bone, soft tissue and other extraosseous tumours)	Female:20.0 (95%CI 17.4-23.1)	Female: 16.5
	Male:23.6 (95% 20.9-26.6)	Male 20.9
Neuroblastoma & other peripheral nervous system tumours	Female: 5.0 (95%3.7-6.6)	Female: 6.5
	Male: 6.5 (95%5.2-8.3)	Male 6.9
Renal tumours	Female: 7.4 (95%5.9-9.3)	Female: 7.0
	Male: 4.3 (95%3.3-5.7)	Male 6.0
Hepatic tumours	Female: 1.1 (95% 0.6-2.0)	Female: 1.6
	Male: 1.2 (95% 0.7-2.1)	Male 1.9
Gonadal Germ Cell Tumours	Female:3.6 (95%CI 2.6-5.0)	Female: 3.9
	Male: 22.7 (95%20.1-25.7)	Male: 22.8

Supplementary Figure 9-5 *Forest plots demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and diagnostic intervals following Cox proportional regression* and multiple imputation *using Approach 1.*

a)



b)

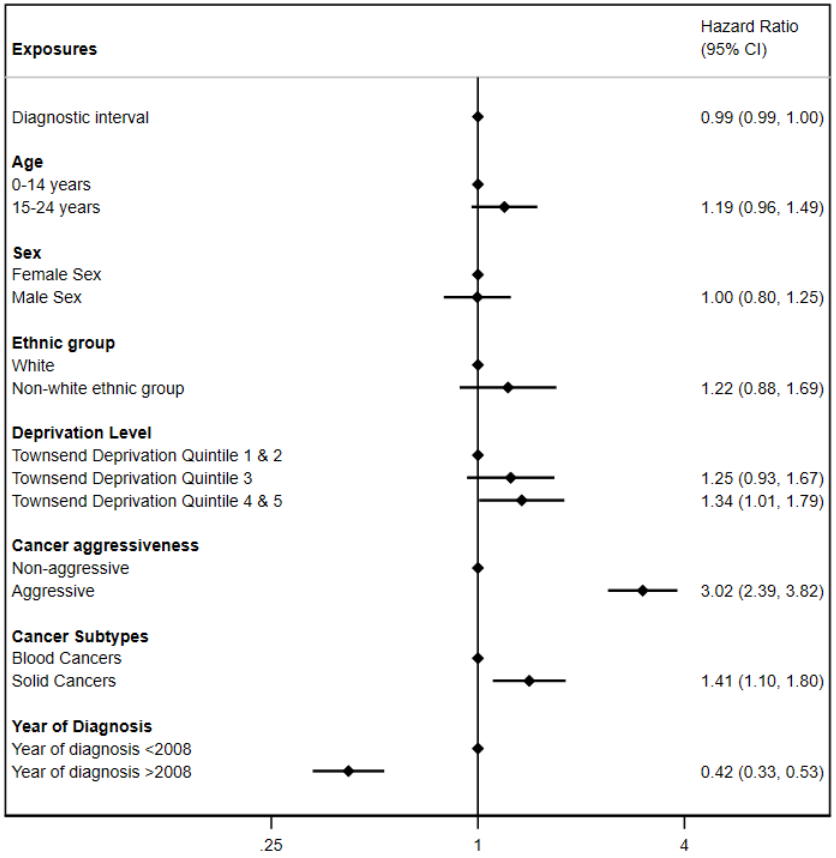
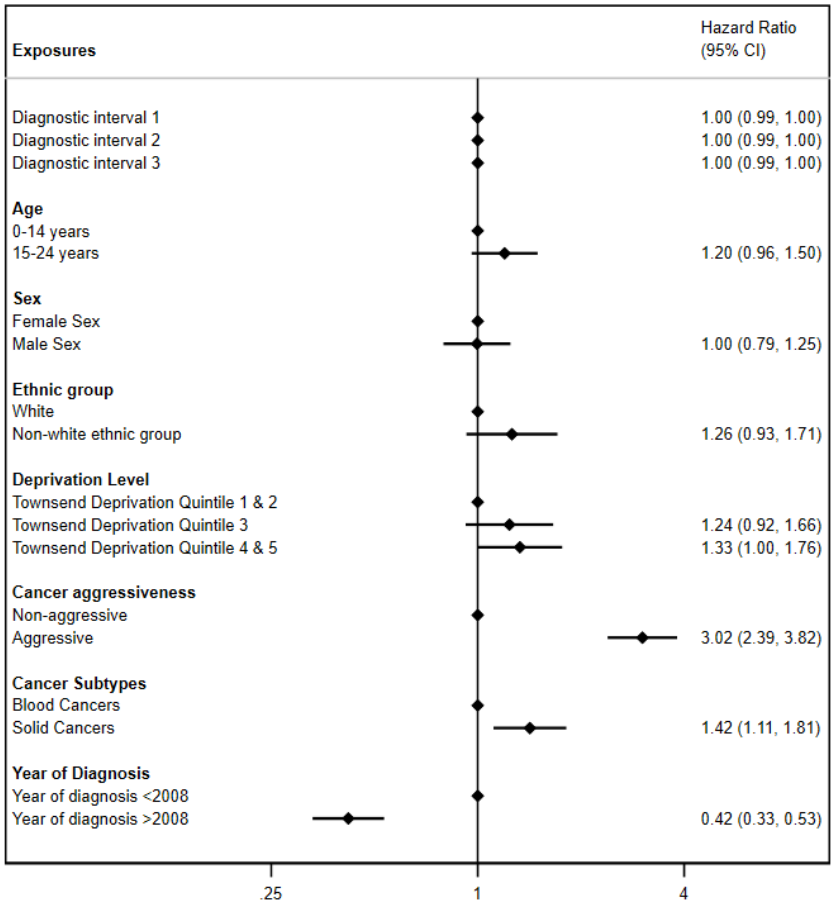
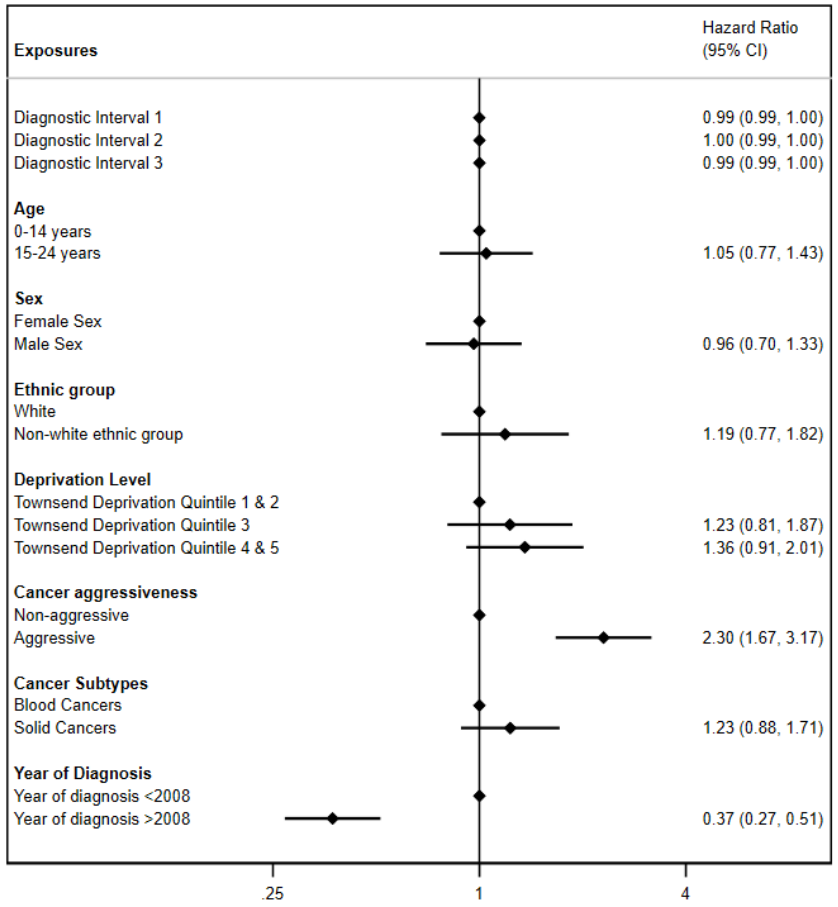


Figure 9-6 Forest plots demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and diagnostic intervals following Cox proportional regression and multiple imputation using Approach 2 (i.e., restricted cubic spline approach).

a)

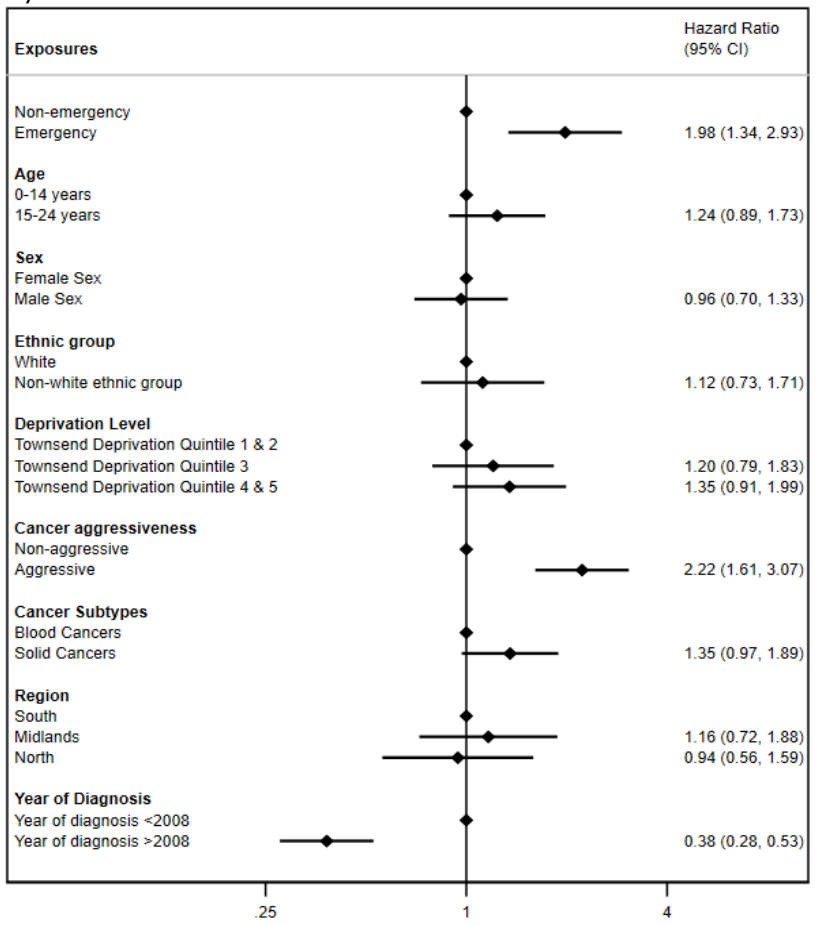


b)

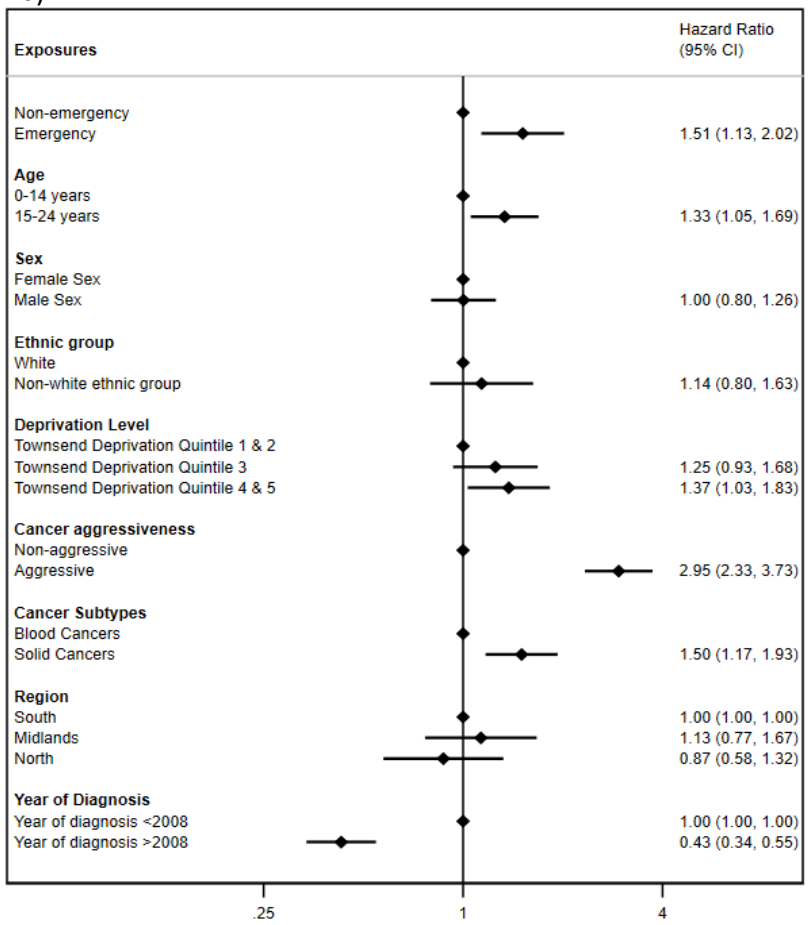


Supplementary Figure 9-7 Forest plot demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and emergency presentations following Cox proportional regression.

a)



b)



Supplementary Table 9-24 *SNOMED-CT and ICD-10 Codes for each mental health outcome*

Disorder	SNOMED Concept ID	Snomed Concept FullySpecifiedName
Depression	15639000	Moderate major depression, single episode (disorder)
	162722001	On examination - depressed (disorder)
	182388100000 06100	Depression confirmed (situation)
	191604000	Single major depressive episode, severe, with psychosis (disorder)
	191659001	Atypical depressive disorder (disorder)
	191676002	Reactive depressive psychosis (disorder)
	192046006	Brief depressive adjustment reaction (disorder)
	192049004	Prolonged depressive adjustment reaction (disorder)
	19527009	Single episode of major depression in full remission (disorder)
	197211100000 06100	Organic depressive disorder (disorder)
	197598100000 06100	Mild depressive episode, without somatic syndrome (disorder)
	197599100000 06100	Mild depressive episode, with somatic syndrome (disorder)
	197602100000 06100	Moderate depressive episode, without somatic syndrome (disorder)
	197605100000 06100	Moderate depressive episode, with somatic syndrome (disorder)
	231499006	Endogenous depression first episode (disorder)
	231500002	Masked depression (disorder)
	231542000	Depressive conduct disorder (disorder)
	247803002	Seasonal affective disorder (disorder)
	251000119105	Severe major depression, single episode (disorder)
	272022009	Complaining of feeling depressed (finding)
	272024005	Complaining of feeling unhappy (finding)
	300706003	Endogenous depression (disorder)
	310495003	Mild depression (disorder)
	310496002	Moderate depression (disorder)
	310497006	Severe depression (disorder)
	35489007	Depressive disorder (disorder)
	361761000000106	On full dose long term treatment for depression (finding)
	366979004	Depressed mood (finding)
	36923009	Major depression, single episode (disorder)
	394924000	Symptoms of depression (finding)
	397701000000102	[X]Severe depressive episode without psychotic symptoms (disorder)
	430421000000104	[X]Mild depressive episode (disorder)
	465441000000108	[X]Moderate depressive episode (disorder)
	473126001	Suspected depressive disorder (situation)
	58703003	Postpartum depression (disorder)

	609311000000100	Depressive disorder NEC (disorder)
	70747007	Major depression single episode, in partial remission (disorder)
	73867007	Severe major depression with psychotic features (disorder)
	75084000	Severe major depression without psychotic features (disorder)
	755321000000106	Single major depressive episode, severe, with psychosis, psychosis in remission (disorder)
	78667006	Dysthymia (disorder)
	790961000000101	Antenatal depression (disorder)
	79298009	Mild major depression, single episode (disorder)
	832007	Moderate major depression (disorder)
	83458005	Agitated depression (disorder)
	87414006	Reactive depression (situational) (disorder)
	87512008	Mild major depression (disorder)
	909681000006106	[RFC] Depression (finding)
	959741000006101	Sadness/hopelessness/decreased self-esteem (finding)
	981611000006104	Abnormal depressed feelings (finding)
Self-Harm	182129100006100	Recent suicide attempt (situation)
	200026100006100	MHIP behavioural history - previous suicidal behaviour (procedure)
	200027100006100	MHIP behavioural history - severity of suicidal behaviour (procedure)
	200028100006100	MHIP behavioural history - intent of suicidal behaviour (procedure)
	200043100006100	MHIP comprehensive assessment - suicidal ideation (procedure)
	200076100006100	MHIP comprehensive assessment - suicidal thoughts - strong (procedure)
	200077100006100	MHIP comprehensive assessment - suicidal thoughts - serious (procedure)
	200078100006100	MHIP comprehensive assessment - suicidal thoughts - frequent (procedure)
	200079100006100	MHIP comprehensive assessment - suicidal thoughts - intense (procedure)
	200080100006100	MHIP comprehensive assessment - suicidal thoughts - persistent (procedure)
	200081100006100	MHIP comprehensive assessment - motivated to act on suicidal thoughts (procedure)
	200101100006100	MHIP comprehensive assessment - current suicide plans (procedure)
	200102100006100	MHIP - specific/serious/feasible suicide plans (procedure)
	200106100006100	MHIP - suicide plans prepared (procedure)
	200107100006100	MHIP - resources to act on suicide plans (procedure)

	2001101000006100	MHIP - confidence to complete suicide plans (procedure)
	2001341000006100	MHIP comprehensive assessment - medium risk of suicide (procedure)
	2001351000006100	MHIP comprehensive assessment - high risk of suicide (procedure)
	2001461000006100	MHIP - intention to act on suicide plans (procedure)
	219108005	Suicide and selfinflicted poisoning by analgesics, antipyretics and anti-rheumatics (navigational concept)
	219122005	Self poisoning by corrosive or caustic substance (disorder)
	219123000	Self poisoning caused by arsenic and/or arsenic compound (disorder)
	219125007	Suicide and selfinflicted poisoning by gases in domestic use (disorder)
	219126008	Suicide and selfinflicted poisoning by gas distributed by pipeline (disorder)
	219127004	Suicide and selfinflicted poisoning by liquefied petroleum gas distributed in mobile containers (disorder)
	219131005	Self poisoning caused by motor vehicle exhaust gas (disorder)
	219137009	Suicide or attempted suicide by hanging (navigational concept)
	219141008	Suicide and selfinflicted injury by drowning (navigational concept)
	219142001	Suicide or selfinflicted injury by firearms and explosives (navigational concept)
	219150005	Suicide or selfinflicted injury by cutting and stabbing (navigational concept)
	219151009	Suicide or selfinflicted injury by cutting (navigational concept)
	219152002	Self inflicted lacerations to wrist (disorder)
	219153007	Suicide or selfinflicted injury by stabbing (navigational concept)
	219155000	Suicide or self injury by jumping from a height (navigational concept)
	219168008	Suicide or self injury by electric current (navigational concept)
	242836008	Self poisoning by non-drug solid or liquid agents (disorder)
	242839001	Self poisoning by gas (disorder)
	248062006	Self-injurious behavior (finding)
	267073005	Suicidal (finding)
	269725004	Suicide and selfinflicted injury (navigational concept)
	269726003	Suicide and selfinflicted injury by suffocation by plastic bag (navigational concept)
	269808005	Attempted suicide - car exhaust (event)
	271982007	Intentional self poisoning (disorder)
	274912004	Self poisoning by agricultural chemical (disorder)
	276853009	Self inflicted injury (disorder)
	287181000	Attempted suicide - hanging (event)
	287182007	Attempted suicide - suffocation (event)
	287183002	Attempted suicide - drowning (event)
	287184008	Attempted suicide - firearms (event)
	287185009	Attempted suicide - cut/stab (event)
	287186005	Attempted suicide - jumping from a high place (event)

	288311002	Attempted suicide - domestic gas (event)
	296036006	Barbiturate overdose (disorder)
	296039004	Intentional amylobarbitone overdose (disorder)
	304594002	Suicidal intent (finding)
	403582001	Self-inflicted caustic burn (disorder)
	418420002	Intentionally harming self (event)
	431307001	Intentional poisoning caused by drug (disorder)
	44301001	Suicide (event)
	48981002	Parasuicide (event)
	59274003	Intentional drug overdose (disorder)
	6471006	Suicidal thoughts (finding)
	681111000000102	Suicide and self inflicted poisoning by drug or medicine NOS (navigational concept)
	684281000000100	Suicide and self inflicted injury NOS (navigational concept)
	82313006	Suicide attempt (event)
Anxiety Disorders	1038231000000100	howRu rating score - Feeling low or worried (observable entity)
	13438001	Overanxious disorder of childhood (disorder)
	162723006	On examination - anxious (finding)
	181811100000106100	Feeling anxious (finding)
	191708009	Chronic anxiety (finding)
	191722009	Agoraphobia with panic attacks (disorder)
	192108001	Disturbance of anxiety and fearfulness in childhood and adolescence (disorder)
	192110004	Childhood and adolescent fearfulness disturbance (disorder)
	192611004	Childhood phobic anxiety disorder (disorder)
	197480006	Anxiety disorder (disorder)
	197649100000106100	Mixed anxiety and depressive reaction (disorder)
	198288003	Anxiety state (finding)
	207363009	Anxiety neurosis (finding)
	21897009	Generalized anxiety disorder (disorder)
	225624000	Panic attack (finding)
	25501002	Social phobia (disorder)
	371631005	Panic disorder (disorder)
	386810004	Phobic disorder (disorder)
	48694002	Anxiety (finding)
	54307006	Zoophobia (finding)
	54587008	Simple phobia (disorder)
	56576003	Panic disorder without agoraphobia (disorder)
	64165008	Avoidant disorder of childhood (disorder)
	70691001	Agoraphobia (disorder)
	79015004	Worried (finding)
	853241000006103	Phobic anxiety (disorder)
	13746004	Bipolar disorder (disorder)

Psychotic disorders	162004	Severe manic bipolar I disorder without psychotic features (disorder)
	191618007	Bipolar affective disorder, current episode manic (disorder)
	191620005	Bipolar affective disorder, currently manic, mild (disorder)
	191621009	Bipolar affective disorder, currently manic, moderate (disorder)
	191623007	Bipolar affective disorder, currently manic, severe, with psychosis (disorder)
	191625000	Bipolar affective disorder, currently manic, in full remission (disorder)
	191627008	Bipolar affective disorder, current episode depression (disorder)
	191629006	Bipolar affective disorder, currently depressed, mild (disorder)
	191630001	Bipolar affective disorder, currently depressed, moderate (disorder)
	191634005	Bipolar affective disorder, currently depressed, in full remission (disorder)
	191636007	Mixed bipolar affective disorder (disorder)
	191638008	Mixed bipolar affective disorder, mild (disorder)
	191639000	Mixed bipolar affective disorder, moderate (disorder)
	191641004	Mixed bipolar affective disorder, severe, with psychosis (disorder)
	191643001	Mixed bipolar affective disorder, in full remission (disorder)
	192362008	Bipolar affective disorder, current episode mixed (disorder)
	1972101000006100	Organic bipolar affective disorder (disorder)
	1972121000006100	Organic mixed affective disorder (disorder)
	1975931000006100	Bipolar affect disorder cur epi mild or moderate depressn, without somatic syndrome (disorder)
	1975941000006100	Bipolar affect disorder cur epi mild or moderate depressn, with somatic syndrome (disorder)
	231494001	Mania (disorder)
	231496004	Hypomania (disorder)
	268619003	Manic disorder, single episode (disorder)
	270901009	Schizoaffective disorder, mixed type (disorder)
	271428004	Schizoaffective disorder, manic type (disorder)
	31446002	Bipolar I disorder, most recent episode hypomanic (disorder)
	371596008	Bipolar I disorder (disorder)
	405273008	Manic mood (finding)
	49512000	Depressed bipolar I disorder in partial remission (disorder)
	58329000	Organic mood disorder of manic type (disorder)
	61403008	Severe depressed bipolar I disorder without psychotic features (disorder)
	63249007	Manic bipolar I disorder in partial remission (disorder)
	68890003	Schizoaffective disorder (disorder)
	760721000000109	Mixed bipolar affective disorder, in partial remission (disorder)
	764591000000108	Mixed bipolar affective disorder, severe (disorder)
	765176007	Psychosis and severe depression co-occurrent and due to bipolar affective disorder (disorder)

	83225003	Bipolar II disorder (disorder)
	84760002	Schizoaffective disorder, depressive type (disorder)
	85248005	Bipolar disorder in remission (disorder)
	111482003	Subchronic schizophrenia with acute exacerbations (disorder)
	111483008	Catatonic schizophrenia in remission (disorder)
	111484002	Undifferentiated schizophrenia (disorder)
	12939007	Chronic disorganized schizophrenia (disorder)
	14291003	Subchronic disorganized schizophrenia with acute exacerbations (disorder)
	16990005	Subchronic schizophrenia (disorder)
	191525009	Non-organic psychosis (disorder)
	191526005	Schizophrenic disorders (disorder)
	191527001	Simple schizophrenia (disorder)
	191542003	Catatonic schizophrenia (disorder)
	191559008	Latent schizophrenia (disorder)
	191561004	Subchronic latent schizophrenia (disorder)
	191562006	Chronic latent schizophrenia (disorder)
	191565008	Latent schizophrenia in remission (disorder)
	191567000	Schizoaffective schizophrenia (disorder)
	191569002	Subchronic schizoaffective schizophrenia (disorder)
	191570001	Chronic schizoaffective schizophrenia (disorder)
	191574005	Schizoaffective schizophrenia in remission (disorder)
	191577003	Cenesthopathic schizophrenia (disorder)
	191680007	Psychogenic paranoid psychosis (disorder)
	191687005	Psychosis with origin in childhood (disorder)
	191696005	Atypical childhood psychoses (disorder)
	191697001	Borderline psychosis of childhood (disorder)
	19743510000 06100	Paranoid schizophrenia, continuous (disorder)
	19743910000 06100	Paranoid schizophrenia, episodic with progressive deficit (disorder)
	19744310000 06100	Paranoid schizophrenia, episodic with stable deficit (disorder)
	19746010000 06100	Hebephrenic schizophrenia, continuous (disorder)
	19746110000 06100	Hebephrenic schizophrenia, episodic with progressive deficit (disorder)
	19746210000 06100	Hebephrenic schizophrenia, episodic with stable deficit (disorder)
	19747510000 06100	Catatonic schizophrenia, continuous (disorder)
	19747810000 06100	Catatonic schizophrenia, episodic with progressive deficit (disorder)
	19748010000 06100	Catatonic schizophrenia, episodic with stable deficit (disorder)
	19749510000 06100	Undifferentiated schizophrenia, continuous (disorder)
	19749710000 06100	Undifferentiated schizophrenia, episodic with progressive deficit (disorder)

	1975001000006100	Undifferentiated schizophrenia, episodic with stable deficit (disorder)
	1975191000006100	Post-schizophrenic depression, continuous (disorder)
	1975351000006100	Residual schizophrenia, continuous (disorder)
	1975381000006100	Residual schizophrenia, episodic with progressive deficit (disorder)
	1975401000006100	Residual schizophrenia, episodic with stable deficit (disorder)
	1975491000006100	Simple schizophrenia, continuous (disorder)
	1975501000006100	Simple schizophrenia, episodic with progressive deficit (disorder)
	1975521000006100	Simple schizophrenia, episodic with stable deficit (disorder)
	1975621000006100	Acute polymorphic psychot disord without symp of schizoph, without associated acute stress (disorder)
	1975641000006100	Simple schizophrenia, course uncertain, period of observation too short (disorder)
	1975671000006100	Acute polymorphic psychot disord without symp of schizoph, with associated acute stress (disorder)
	1975681000006100	Acute polymorphic psychot disord with symp of schizophren, without associated acute stress (disorder)
	1975691000006100	Acute polymorphic psychot disord with symp of schizophren, with associated acute stress (disorder)
	1975711000006100	Acute schizophrenia-like psychotic disorder, without associated acute stress (disorder)
	1975731000006100	Acute schizophrenia-like psychotic disorder, with associated acute stress (disorder)
	1975751000006100	Other schizophrenia, continuous (disorder)
	1975761000006100	Other acute predominantly delusional psychotic disorders, without associated acute stress (disorder)
	1975771000006100	Other schizophrenia, episodic with progressive deficit (disorder)
	1975781000006100	Other acute predominantly delusional psychotic disorders, with associated acute stress (disorder)
	1975801000006100	Other schizophrenia, episodic with stable deficit (disorder)
	1975821000006100	Other acute and transient psychotic disorders, without associated acute stress (disorder)
	1975831000006100	Other schizophrenia, episodic remittent (disorder)
	1975841000006100	Other schizophrenia, incomplete remission (disorder)
	1975851000006100	Other acute and transient psychotic disorders, with associated acute stress (disorder)
	1975871000006100	Acute and transient psychotic disorder, unspecified, without associated acute stress (disorder)
	1975881000006100	Other schizophrenia, complete remission (disorder)

	1975901000006100	Acute and transient psychotic disorder, unspecified, with associated acute stress (disorder)
	1975911000006100	Other schizophrenia, course uncertain, period of observation too short (disorder)
	1976001000006100	Schizophrenia, unspecified, continuous (disorder)
	1976061000006100	Schizophrenia, unspecified, episodic with progressive deficit (disorder)
	1976101000006100	Schizophrenia, unspecified, episodic with stable deficit (disorder)
	1976121000006100	Schizophrenia, unspecified, episodic remittent (disorder)
	1976141000006100	Schizophrenia, unspecified, incomplete remission (disorder)
	1976151000006100	Schizophrenia, unspecified, complete remission (disorder)
	1976171000006100	Schizophrenia, unspecified, course uncertain, period of observation too short (disorder)
	1976881000006100	Paranoid schizophrenia, other (disorder)
	1976891000006100	Hebephrenic schizophrenia, other (disorder)
	1976901000006100	Catatonic schizophrenia, other (disorder)
	1976911000006100	Undifferentiated schizophrenia, other (disorder)
	1976931000006100	Residual schizophrenia, other (disorder)
	1976941000006100	Simple schizophrenia, other (disorder)
	1976951000006100	Other schizophrenia, other (disorder)
	1976961000006100	Schizophrenia, unspecified, other (disorder)
	231437006	Reactive psychoses (disorder)
	231487004	Persistent delusional disorder (disorder)
	231489001	Acute transient psychotic disorder (disorder)
	247804008	Schizophrenic prodrome (disorder)
	26025008	Residual schizophrenia (disorder)
	268617001	Acute schizophrenic episode (disorder)
	270901009	Schizoaffective disorder, mixed type (disorder)
	271428004	Schizoaffective disorder, manic type (disorder)

Drug Group	Drug Medicine ID	EMIS ID (Prep Code ID)	Ingredient	Code Description
Anxiolytics	12141711000001100	11275941000033100	Oxazepam	Oxazepam 10mg/5ml oral suspension
	13893211000001100	2762441000033110	Diazepam	Diazepam 10mg/5ml oral suspension
	13893311000001100	3345241000033110	Diazepam	Diazepam 2.5mg/5ml oral suspension
	142611000001108	1383241000033110	Diazepam	Stesolid 10mg rectal tube (Accord-UK Ltd)
	142611000001108	1383241000033110	Diazepam	Stesolid 10mg rectal tube (Accord-UK Ltd)
	153811000001102	2621341000033110	Chlordiazepoxide Hydrochloride	Tropium 10mg tablets (Dr Reddy's Laboratories (UK) Ltd)
	19677811000001100	3993641000033110	Diazepam	Diazepam 2mg/5ml oral suspension
	204011000001103	2621241000033110	Chlordiazepoxide Hydrochloride	Tropium 5mg tablets (Dr Reddy's Laboratories (UK) Ltd)
	218811000001104	827841000033112	Chlordiazepoxide Hydrochloride	Librium 10mg capsules (Viatris UK Healthcare Ltd)
	26811000001102	4579410000331 12	Diazepam	Diazepam 10mg RecTubes (Wockhardt UK Ltd)
	26811000001102	4579410000331 12	Diazepam	Diazepam 10mg RecTubes (Wockhardt UK Ltd)
	283311000001103	458041000033110	Diazepam	Diazepam 2.5mg RecTubes (Wockhardt UK Ltd)
	283311000001103	458041000033110	Diazepam	Diazepam 2.5mg RecTubes (Wockhardt UK Ltd)
	307711000001103	1455941000033110	Chlordiazepoxide Hydrochloride	Tropium 5mg capsules (Dr Reddy's Laboratories (UK) Ltd)
	34697811000001100	12503341000033100	Lorazepam	Lorazepam 1mg/ml oral solution sugar free

	36040511000001100	844141000033111	Lorazepam,Lorazepam	Lorazepam 4mg/1ml solution for injection ampoules
	36097511000001100	3033041000033110	Diazepam,Soya Oil,Egg Protein	Diazepam 10mg/2ml emulsion for injection ampoules
	36097511000001100	3033041000033110	Diazepam,Soya Oil,Egg Protein	Diazepam 10mg/2ml emulsion for injection ampoules
	36097611000001100	3033141000033110	Diazepam,Diazepam	Diazepam 10mg/2ml solution for injection ampoules
	36097611000001100	3033141000033110	Diazepam,Diazepam	Diazepam 10mg/2ml solution for injection ampoules
	365311000001106	174141000033116	Buspirone Hydrochloride	Buspar 5mg tablets (IXL Pharma Ltd)
	37236111000001100	130583410000033100	Buspirone Hydrochloride	Buspirone 7.5mg tablets
	37600511000001100	131594410000033100	Lorazepam	Lorazepam 1mg orodispersible tablets
	37600611000001100	131595410000033100	Lorazepam	Lorazepam 2.5mg orodispersible tablets
	38243110000011100	150294100003311 0	Diazepam	Valclair 10mg suppositories (Durbin Plc)
	38894311000001100	574864100003311 0	Diazepam	Diazepam 2mg/5ml oral solution sugar free
	39258110000011100	88941000033110	Benzyl Alcohol,Lorazepam,Benzyl Alcohol,Lorazepam	Ativan 4mg/1ml solution for injection ampoules (Pfizer Ltd)
	39709411000001100	458641000033116	Diazepam	Diazepam 5mg/2.5ml rectal solution tube
	39709411000001100	458641000033116	Diazepam	Diazepam 5mg/2.5ml rectal solution tube
	39711110000011100	446441000033119	Diazepam,Soya Oil,Egg Protein	Diazemuls 10mg/2ml emulsion for injection ampoules (Accord-UK Ltd)

	3971111000001100	446441000033119	Diazepam,Soya Oil,Egg Protein	Diazemuls 10mg/2ml emulsion for injection ampoules (Accord-UK Ltd)
	3973431100000100	172141000033114	Buspirone Hydrochloride	Buspirone 10mg tablets
	3973441100000100	172041000033110	Buspirone Hydrochloride	Buspirone 5mg tablets
	4054861100000100	140159410000331100	Buspirone Hydrochloride	Buspirone 30mg tablets
	4054871100000100	140158410000331100	Buspirone Hydrochloride	Buspirone 15mg tablets
	4155681100000100	141906410000331100	Lorazepam	Lorazepam 250microgram tablets
	4212211100000100	41141000033110	Alprazolam	Alprazolam 250microgram tablets
	4212521100000100	41241000033115	Alprazolam	Alprazolam 500microgram tablets
	4212651100000100	238241000033116	Chlordiazepoxide Hydrochloride	Chlordiazepoxide 10mg capsules
	4212661100000100	250641000033119	Chlordiazepoxide Hydrochloride	Chlordiazepoxide 10mg tablets
	4212681100000100	238141000033111	Chlordiazepoxide	Chlordiazepoxide 5mg capsules
	4212691100000100	250841000033118	Chlordiazepoxide Hydrochloride	Chlordiazepoxide 5mg tablets
	4212761100000100	460041000033115	Diazepam	Diazepam 10mg suppositories
	4212771100000100	466741000033115	Diazepam	Diazepam 10mg tablets
	4212781100000100	458541000033117	Diazepam	Diazepam 10mg/2.5ml rectal solution tube
	4212781100000100	458541000033117	Diazepam	Diazepam 10mg/2.5ml rectal solution tube
	4212791100000100	458341000033112	Diazepam	Diazepam 2.5mg/1.25ml rectal solution tube
	4212791100000100	458341000033112	Diazepam	Diazepam 2.5mg/1.25ml rectal solution tube
	4212801100000100	3979041000033110	Diazepam	Diazepam 25mg/5ml oral solution
	4212811100000100	466841000033113	Diazepam	Diazepam 2mg tablets

	42128211000001100	2762341000033110	Diazepam	Diazepam 2mg/5ml oral solution
	42128311000001100	4669410000331 7	Diazepam	Diazepam 5mg tablets
	42128411000001100	1728141000033110	Diazepam	Diazepam 5mg/5ml oral solution
	42131511000001100	8511410000331 1	Lorazepam	Lorazepam 1mg tablets
	42131611000001100	8512410000331 6	Lorazepam	Lorazepam 2.5mg tablets
	42131711000001100	1305814100003 3100	Lorazepam	Lorazepam 500microgram tablets
	42132011000001100	9019410000331 1	Meprobamate	Meprobamate 400mg tablets
	42132611000001100	1025741000033110	Oxazepam	Oxazepam 10mg tablets
	42132711000001100	1025841000033110	Oxazepam	Oxazepam 15mg tablets
	42495311000001100	1426574100003 3100	Diazepam	Diazepam 2mg/5ml oral suspension sugar free (Special Order)
	4658511000001100	1543541000033110	Alprazolam	Xanax 250microgram tablets (Viatris UK Healthcare Ltd)
	4659411000001100	1543641000033110	Alprazolam	Xanax 500microgram tablets (Viatris UK Healthcare Ltd)
	4902211000001100	1454441000033110	Clorazepate Dipotassium	Tranxene 7.5mg capsules (Boehringer Ingelheim Ltd)
	4902711000001100	1454341000033110	Clorazepate Dipotassium	Tranxene 15mg capsules (Boehringer Ingelheim Ltd)
	5113110000011 08	2621141000033110	Chlordiazepoxide Hydrochloride	Tropium 10mg capsules (Dr Reddy's Laboratories (UK) Ltd)
	6038110000011 08	4582410000331 9	Diazepam	Diazepam 5mg RecTubes (Wockhardt UK Ltd)
	6038110000011 08	4582410000331 9	Diazepam	Diazepam 5mg RecTubes (Wockhardt UK Ltd)

	737211000001100	172241000033119	Buspirone Hydrochloride	Buspar 10mg tablets (IXL Pharma Ltd)
	82011000001104	1383341000033110	Diazepam	Stesolid 5mg rectal tube (Accord-UK Ltd)
	82011000001104	1383341000033110	Diazepam	Stesolid 5mg rectal tube (Accord-UK Ltd)
	830441100000110000	126792410000331000	Buspirone Hydrochloride	Buspirone 5mg/5ml oral solution
	835821100000110000	443414100003311000	Chlordiazepoxide Hydrochloride	Chlordiazepoxide 5mg/5ml oral suspension
	849211000001108	827941000033116	Chlordiazepoxide Hydrochloride	Librium 5mg capsules (Viatris UK Healthcare Ltd)
	862431100000110000	589244100003311000	Lorazepam	Lorazepam 1mg/5ml oral solution
	862451100000110000	392384100003311000	Lorazepam	Lorazepam 1mg/5ml oral suspension
	865921100000110000	466334100003311000	Lorazepam	Lorazepam 250micrograms/5ml oral suspension
	865961100000110000	589234100003311000	Lorazepam	Lorazepam 500micrograms/5ml oral solution
	865981100000110000	330804100003311000	Lorazepam	Lorazepam 500micrograms/5ml oral suspension
	879271100000110000	599164100003311000	Diazepam	Diazepam 10mg/5ml oral solution
	879291100000110000	599174100003311000	Diazepam	Diazepam 2.5mg/5ml oral solution
	910681100000110000	142814100003311000	Diazepam	Tensium 10mg tablets (DDSA Pharmaceuticals Ltd)
	910721100000110000	142824100003311000	Diazepam	Tensium 2mg tablets (DDSA Pharmaceuticals Ltd)
4.2.1 Antipsychotic Drugs	106646110000001100	395064100003311000	Aripiprazole	Abilify 10mg orodispersible tab-

				lets (Otsuka Pharmaceuticals (U.K.) Ltd)
	10664911000001100	3950741000033110	Aripiprazole	Abilify 15mg orodispersible tablets (Otsuka Pharmaceuticals (U.K.) Ltd)
	10667211000001100	3950841000033110	Aripiprazole	Abilify 1mg/ml oral solution (Otsuka Pharmaceuticals (U.K.) Ltd)
	10675311000001100	3950341000033110	Aripiprazole	Aripiprazole 10mg orodispersible tablets sugar free
	10675411000001100	3950441000033110	Aripiprazole	Aripiprazole 15mg orodispersible tablets sugar free
	10878511000001100	3996341000033110	Olanzapine	Zyprexa 20mg tablets (Neon Healthcare Ltd)
	11134311000001100	4012341000033110	Risperidone	Risperdal Quicklet 4mg orodispersible tablets (Janssen-Cilag Ltd)
	11134611000001100	4012241000033110	Risperidone	Risperdal Quicklet 3mg orodispersible tablets (Janssen-Cilag Ltd)
	11140811000001100	4012041000033110	Risperidone	Risperidone 3mg orodispersible tablets sugar free
	11140911000001100	4012141000033110	Risperidone	Risperidone 4mg orodispersible tablets sugar free
	113611000001108	1276041000033110	Haloperidol	Serenace 10mg tablets (Teva UK Ltd)
	11749311000001100	4200941000033110	Paliperidone	Invega 3mg modified-release tablets (Janssen-Cilag Ltd)
	11749811000001100	4201041000033110	Paliperidone	Invega 6mg modified-release tablets (Janssen-Cilag Ltd)

	11751011000001100	42011410000033110	Paliperidone	Invega 9mg modified-release tablets (Janssen-Cilag Ltd)
	11763211000001100	42006410000033110	Paliperidone	Paliperidone 3mg modified-release tablets
	11763311000001100	42007410000033110	Paliperidone	Paliperidone 6mg modified-release tablets
	11763411000001100	42008410000033110	Paliperidone	Paliperidone 9mg modified-release tablets
	117911000001106	661541000033117	Haloperidol	Haldol 2mg/ml oral solution (Essential Pharma Ltd)
	12504911000001100	59694410000033110	Haloperidol	Haloperidol 1.5mg/5ml oral solution
	12505011000001100	59695410000033110	Haloperidol	Haloperidol 1.5mg/5ml oral suspension
	12505511000001100	59696410000033110	Haloperidol	Haloperidol 250micrograms/5ml oral solution
	12505611000001100	59697410000033110	Haloperidol	Haloperidol 250micrograms/5ml oral suspension
	12505911000001100	44245410000033110	Haloperidol	Haloperidol 5mg/5ml oral solution
	12506011000001100	59935410000033110	Haloperidol	Haloperidol 5mg/5ml oral suspension
	12646911000001100	64456410000033110	Levomepromazine Maleate	Levomepromazine 6mg/5ml oral suspension
	12797811000001100	10253841000033100	Melperone Hydrochloride	Melperone hydrochloride 100mg tablets
	13011611000001100	58897410000033110	Quetiapine Fumarate	Quetiapine 12.5mg/5ml oral solution
	13011711000001100	58898410000033110	Quetiapine Fumarate	Quetiapine 12.5mg/5ml oral suspension
	13012211000001100	59748410000033110	Quetiapine Fumarate	Quetiapine 25mg/5ml oral solution

	13012311000001100	5974941000033110	Quetiapine Fumarate	Quetiapine 25mg/5ml oral suspension
	13059411000001100	11798841000033100	Pimozide	Pimozide 2mg/5ml oral suspension
	13451111000001100	4456841000033110	Aripiprazole	Abilify 9.75mg/1.3ml solution for injection vials (Otsuka Pharmaceuticals (U.K.) Ltd)
	13456411000001100	4456741000033110	Aripiprazole	Aripiprazole 9.75mg/1.3ml solution for injection vials
	14015811000001100	5997341000033110	Olanzapine	Olanzapine 2.5mg/5ml oral solution
	14015911000001100	3892541000033110	Olanzapine	Olanzapine 2.5mg/5ml oral suspension
	14018611000001100	5891141000033110	Quetiapine Fumarate	Quetiapine 100mg/5ml oral solution
	14018711000001100	4422541000033110	Quetiapine Fumarate	Quetiapine 100mg/5ml oral suspension
	14025311000001100	4592241000033110	Quetiapine Fumarate	Seroquel XL 200mg tablets (Luye Pharma Ltd)
	14025711000001100	4592441000033110	Quetiapine Fumarate	Seroquel XL 400mg tablets (Luye Pharma Ltd)
	14026011000001100	4592341000033110	Quetiapine Fumarate	Seroquel XL 300mg tablets (Luye Pharma Ltd)
	14026511000001100	4592141000033110	Quetiapine Fumarate	Seroquel XL 50mg tablets (Luye Pharma Ltd)
	14042811000001100	4591741000033110	Quetiapine Fumarate	Quetiapine 50mg modified-release tablets
	145511000001103	2658941000033110	Sulpiride	Sulpor 200mg/5ml oral solution (Rosemont Pharmaceuticals Ltd)

	152111000001103	663541000033116	Haloperidol	Haldol 5mg tablets (Janssen-Cilag Ltd)
	153111000001109	117764100003311 0	Risperidone	Risperdal 6mg tablets (Janssen-Cilag Ltd)
	153611000001101	287041000033119	Zuclopenthixol Dihydrochloride	Clopixol 2mg tablets (Lundbeck Ltd)
	154520110000001 100	597504100003311 0	Quetiapine Fumarate	Quetiapine 50mg/5ml oral solution
	154521110000001 100	597514100003311 0	Quetiapine Fumarate	Quetiapine 50mg/5ml oral suspension
	154989110000001 100	507824100003311 0	Clozapine	Denzapine 50mg tablets (Britannia Pharmaceuticals Ltd)
	155057110000001 100	507814100003311 0	Clozapine	Denzapine 200mg tablets (Britannia Pharmaceuticals Ltd)
	155060110000001 100	507774100003311 0	Clozapine	Denzapine 50mg/ml oral suspension (Britannia Pharmaceuticals Ltd)
	155062110000001 100	507794100003311 0	Clozapine	Clozapine 200mg tablets
	155063110000001 100	507764100003311 0	Clozapine	Clozapine 50mg/ml oral suspension sugar free
	160211000001102	896041000033111	Thioridazine	Melleril 25mg tablets (Novartis Pharmaceuticals UK Ltd)
	176898110000001 100	590944100003311 0	Quetiapine Fumarate	Seroquel XL 150mg tablets (Luye Pharma Ltd)
	178286110000001 100	590934100003311 0	Quetiapine Fumarate	Quetiapine 150mg modified-release tablets
	181723110000001 100	645574100003311 0	Ziprasidone	Ziprasidone 40mg capsules
	189566110000001 100	118767410000331 00	Chlorpromazine Hydrochloride	Chlorpromazine 25mg/5ml oral solution sugar free
	195259110000001 100	645584100003311 0	Ziprasidone	Ziprasidone 80mg capsules

	19587711000001100	6456941000033110	Olanzapine	Zalasta 2.5mg tablets (Krka UK Ltd)
	19587911000001100	6457141000033110	Olanzapine	Zalasta 5mg tablets (Krka UK Ltd)
	19588111000001100	6457241000033110	Olanzapine	Zalasta 7.5mg tablets (Krka UK Ltd)
	19588311000001100	6456741000033110	Olanzapine	Zalasta 10mg tablets (Krka UK Ltd)
	19588511000001100	6456841000033110	Olanzapine	Zalasta 15mg tablets (Krka UK Ltd)
	19588711000001100	6457041000033110	Olanzapine	Zalasta 20mg tablets (Krka UK Ltd)
	19608211000001100	6465641000033110	Olanzapine	Zalasta 5mg orodispersible tablets (Krka UK Ltd)
	19609111000001100	6466041000033110	Olanzapine	Olanzapine 10mg orodispersible tablets sugar free
	19609311000001100	6466141000033110	Olanzapine	Olanzapine 15mg orodispersible tablets sugar free
	19609411000001100	6466241000033110	Olanzapine	Olanzapine 20mg orodispersible tablets sugar free
	19609511000001100	6465941000033110	Olanzapine	Olanzapine 5mg orodispersible tablets sugar free
	19609711000001100	6465341000033110	Olanzapine	Zalasta 10mg orodispersible tablets (Krka UK Ltd)
	19610111000001100	6465441000033110	Olanzapine	Zalasta 15mg orodispersible tablets (Krka UK Ltd)
	19610411000001100	6465541000033110	Olanzapine	Zalasta 20mg orodispersible tablets (Krka UK Ltd)
	19626711000001100	7681641000033110	Olanzapine	Olanzapine 10mg orodispersible tablets
	19626811000001100	7681741000033110	Olanzapine	Olanzapine 15mg orodispersible tablets
	19626911000001100	7681841000033110	Olanzapine	Olanzapine 20mg orodispersible tablets
	19627011000001100	7681941000033110	Olanzapine	Olanzapine 5mg orodispersible tablets

	19725311000001100	6469841000033110	Clotiapine	Clotiapine 40mg tablets
	19809011000001100	2653841000033110	Levomepromazine Maleate	Levinan 6mg tablets (Kyowa Kirin Ltd)
	19890011000001100	6515041000033110	Olanzapine	Arkolamyl 10mg orodispersible tablets (Viatris UK Healthcare Ltd)
	19890311000001100	6515141000033110	Olanzapine	Arkolamyl 15mg orodispersible tablets (Viatris UK Healthcare Ltd)
	19890611000001100	6514941000033110	Olanzapine	Arkolamyl 5mg orodispersible tablets (Viatris UK Healthcare Ltd)
	19890911000001100	6515241000033110	Olanzapine	Arkolamyl 20mg orodispersible tablets (Viatris UK Healthcare Ltd)
	20176811000001100	8886041000033110	Chlorpromazine Hydrochloride	Chlorpromazine 10mg capsules
	20356011000001100	7687841000033110	Quetiapine Fumarate	Sondate XL 50mg tablets (Teva UK Ltd)
	20356211000001100	7687941000033110	Quetiapine Fumarate	Sondate XL 200mg tablets (Teva UK Ltd)
	20356411000001100	7688041000033110	Quetiapine Fumarate	Sondate XL 300mg tablets (Teva UK Ltd)
	20356611000001100	7688141000033110	Quetiapine Fumarate	Sondate XL 400mg tablets (Teva UK Ltd)
	21248811000001100	8112941000033110	Quetiapine Fumarate	Tenprolide XL 50mg tablets (Actavis UK Ltd)
	21249311000001100	8113041000033110	Quetiapine Fumarate	Tenprolide XL 200mg tablets (Actavis UK Ltd)
	21249511000001100	8113141000033110	Quetiapine Fumarate	Tenprolide XL 300mg tablets (Actavis UK Ltd)
	213711000001101	1385341000033110	Trifluoperazine	Stelazine 10mg Spansules (Mercury Pharma Group Ltd)

	21714411000001100	82462410000033110	Quetiapine Fumarate	Sondate XL 150mg tablets (Teva UK Ltd)
	22221211000001100	82983410000033110	Quetiapine Fumarate	Seotiapim XL 200mg tablets (Sandoz Ltd)
	22221411000001100	82985410000033110	Quetiapine Fumarate	Seotiapim XL 400mg tablets (Sandoz Ltd)
	22221611000001100	82984410000033110	Quetiapine Fumarate	Seotiapim XL 300mg tablets (Sandoz Ltd)
	22409511000001100	83491410000033110	Quetiapine Fumarate	Tenprolide XL 400mg tablets (Actavis UK Ltd)
	22635511000001100	11730041000033100	Quetiapine Fumarate	Quetiapine 200mg/5ml oral suspension
	228411000001106	12762410000033110	Haloperidol	Serenace 5mg tablets (Teva UK Ltd)
	24311000001100	13667410000033110	Amisulpride	Solian 50 tablets (Sanofi)
	246411000001108	13965410000033110	Sulpiride	Sulpitil 200mg tablets (Pfizer Ltd)
	254011000001107	22167410000033110	Amisulpride	Solian 400 tablets (Sanofi)
	25803211000001100	96044410000033110	Quetiapine Fumarate	Ebesque XL 50mg tablets (Ethypharm UK Ltd)
	25804311000001100	96055410000033110	Quetiapine Fumarate	Ebesque XL 200mg tablets (Ethypharm UK Ltd)
	25804911000001100	96059410000033110	Quetiapine Fumarate	Ebesque XL 300mg tablets (Ethypharm UK Ltd)
	25806011000001100	96063410000033110	Quetiapine Fumarate	Ebesque XL 400mg tablets (Ethypharm UK Ltd)
	26654411000001100	98092410000033110	Lurasidone Hydrochloride	Latuda 18.5mg tablets (CNX Therapeutics Ltd)
	26654711000001100	98093410000033110	Lurasidone Hydrochloride	Latuda 37mg tablets (CNX Therapeutics Ltd)
	26655011000001100	98094410000033110	Lurasidone Hydrochloride	Latuda 74mg tablets (CNX Therapeutics Ltd)

	26655311000001100	9808941000033110	Lurasidone Hydrochloride	Lurasidone 18.5mg tablets
	26655411000001100	9809041000033110	Lurasidone Hydrochloride	Lurasidone 37mg tablets
	26655511000001100	9809141000033110	Lurasidone Hydrochloride	Lurasidone 74mg tablets
	271211000001102	6634410000331 17	Haloperidol	Haldol 10mg tablets (Janssen-Cilag Ltd)
	27842211000001100	10615141000033100	Quetiapine Fumarate	Atrolak XL 50mg tablets (Accord-UK Ltd)
	27842411000001100	10615241000033100	Quetiapine Fumarate	Atrolak XL 200mg tablets (Accord-UK Ltd)
	27842611000001100	10615341000033100	Quetiapine Fumarate	Atrolak XL 300mg tablets (Accord-UK Ltd)
	27842811000001100	10615441000033100	Quetiapine Fumarate	Atrolak XL 400mg tablets (Accord-UK Ltd)
	27946311000001100	13719141000033100	Levomepromazine Maleate	Levomepromazine 100mg tablets
	28400211000001100	10041141000033100	Quetiapine Fumarate	Biquelle XL 50mg tablets (Aspire Pharma Ltd)
	28400411000001100	10041241000033100	Quetiapine Fumarate	Biquelle XL 150mg tablets (Aspire Pharma Ltd)
	28400611000001100	10041441000033100	Quetiapine Fumarate	Biquelle XL 300mg tablets (Aspire Pharma Ltd)
	28400811000001100	10041541000033100	Quetiapine Fumarate	Biquelle XL 400mg tablets (Aspire Pharma Ltd)
	28401011000001100	10041341000033100	Quetiapine Fumarate	Biquelle XL 200mg tablets (Aspire Pharma Ltd)
	28411511000001100	10041741000033100	Quetiapine Fumarate	Zaluron XL 50mg tablets (Fontus Health Ltd)
	28411711000001100	10041841000033100	Quetiapine Fumarate	Zaluron XL 150mg tablets (Fontus Health Ltd)
	28411911000001100	10041941000033100	Quetiapine Fumarate	Zaluron XL 200mg tablets

				(Fontus Health Ltd)
	28412111000001100	10042041000033100	Quetiapine Fumarate	Zaluron XL 300mg tablets (Fontus Health Ltd)
	28412311000001100	10042141000033100	Quetiapine Fumarate	Zaluron XL 400mg tablets (Fontus Health Ltd)
	28419110000011100	28352410000331 10	Amisulpride	Solian 100 tablets (Sanofi)
	28805110000011100	81344100003311 4	Chlorpromazine Hydrochloride	Largactil Forte 100mg/5ml oral suspension (Sanofi)
	28994711000001100	10254741000033100	Quetiapine Fumarate	Mintreleq XL 50mg tablets (Aristo Pharma Ltd)
	28994911000001100	10254841000033100	Quetiapine Fumarate	Mintreleq XL 150mg tablets (Aristo Pharma Ltd)
	28995111000001100	10254941000033100	Quetiapine Fumarate	Mintreleq XL 200mg tablets (Aristo Pharma Ltd)
	28995611000001100	10255041000033100	Quetiapine Fumarate	Mintreleq XL 300mg tablets (Aristo Pharma Ltd)
	28995811000001100	10255141000033100	Quetiapine Fumarate	Mintreleq XL 400mg tablets (Aristo Pharma Ltd)
	30799611000001100	27581410000331 10	Haloperidol	Haloperidol 200mi-crograms/ml oral solution sugar free
	31951100000110 3	13666410000331 10	Amisulpride	Solian 200 tablets (Sanofi)
	32241100000110 2	29555410000331 10	Trifluoperazine	Terrazine 1mg tablets (Dr Red-dy's Laboratories (UK) Ltd)
	32457110000011100	12716410000331 10	Haloperi-dol,Haloperidol	Serenace 5mg/1ml solution for injection am-poules (IVAX Pharmaceuticals UK Ltd)

	32469110000011 00	29298410000331 10	Haloperidol	Haldol 5mg/1ml solution for injec-tion ampoules (Janssen-Cilag Ltd)
	32503211000001 100	11578941000033 100	Quetiapine Fumarate	Psyquet XL 50mg tablets (Sandoz Ltd)
	32503411000001 100	11579241000033 100	Quetiapine Fumarate	Psyquet XL 150mg tablets (Sandoz Ltd)
	32503611000001 100	11579341000033 100	Quetiapine Fumarate	Psyquet XL 200mg tablets (Sandoz Ltd)
	32503811000001 100	11579441000033 100	Quetiapine Fumarate	Psyquet XL 300mg tablets (Sandoz Ltd)
	32504011000001 100	11579541000033 100	Quetiapine Fumarate	Psyquet XL 400mg tablets (Sandoz Ltd)
	32531100000110 3	13863410000331 10	Trifluoperazine	Stelazine 1mg tablets (Mercury Pharma Group Ltd)
	32687711000001 100	11583741000033 100	Loxapine	Loxapine 9.1mg/dose inha-lation powder
	32687811000001 100	11583841000033 100	Loxapine	Adasuve 9.1mg/dose inha-lation powder (Galen Ltd)
	33391011000001 100	11753641000033 100	Quetiapine Fumarate	Quetiapine 20mg/ml oral suspension sugar free
	33568811000001 100	11782141000033 100	Quetiapine Fumarate	Brancico XL 300mg tablets (Zentiva Pharma UK Ltd)
	33569011000001 100	11782041000033 100	Quetiapine Fumarate	Brancico XL 200mg tablets (Zentiva Pharma UK Ltd)
	33569711000001 100	11781941000033 100	Quetiapine Fumarate	Brancico XL 150mg tablets (Zentiva Pharma UK Ltd)
	33569911000001 100	11782341000033 100	Quetiapine Fumarate	Brancico XL 50mg tablets (Zentiva Pharma UK Ltd)
	33570111000001 100	11782241000033 100	Quetiapine Fumarate	Brancico XL 400mg tablets

				(Zentiva Pharma UK Ltd)
	33690711000001100	11821541000033100	Levomepromazine Maleate	Levomepromazine 62.5mg/5ml oral suspension
	33731100000110 9	17139410000331 10	Olanzapine	Zyprexa 2.5mg tablets (Neon Healthcare Ltd)
	34001100000110 9	96554100003311 9	Pericyazine	Neulactil 10mg tablets (Sanofi)
	34091100000110 8	12776410000331 10	Quetiapine Fumarate	Seroquel 100mg tablets (Luye Pharma Ltd)
	34331100000110 1	11738410000331 10	Risperidone	Risperdal 1mg/ml oral solution (Janssen-Cilag Ltd)
	34349110000011 00	27597410000331 10	Olanzapine,Olanzapine	Olanzapine 15mg oral lyophilisates sugar free
	34936811000001100	12428041000033100	Quetiapine Fumarate	Atrolak XL 150mg tablets (Accord-UK Ltd)
	35753411000001100	13155441000033100	Haloperidol	Halkid 200micrograms/ml oral solution (Thame Laboratories Ltd)
	35771811000001100	12665741000033100	Cariprazine Hydrochloride	Reagila 1.5mg capsules (Recordati Pharmaceuticals Ltd)
	35772111000001100	12665841000033100	Cariprazine Hydrochloride	Reagila 3mg capsules (Recordati Pharmaceuticals Ltd)
	35772411000001100	12665941000033100	Cariprazine Hydrochloride	Reagila 4.5mg capsules (Recordati Pharmaceuticals Ltd)
	35772711000001100	12666041000033100	Cariprazine Hydrochloride	Reagila 6mg capsules (Recordati Pharmaceuticals Ltd)
	35853811000001100	12665341000033100	Cariprazine Hydrochloride	Cariprazine 1.5mg capsules
	35853911000001100	12665441000033100	Cariprazine Hydrochloride	Cariprazine 3mg capsules
	35854011000001100	12665541000033100	Cariprazine Hydrochloride	Cariprazine 4.5mg capsules
	35854111000001100	12665641000033100	Cariprazine Hydrochloride	Cariprazine 6mg capsules
	35901111000001100	26448410000331 10	Amisulpride	Amisulpride 100mg/ml oral

				solution sugar free
	35931111000001100	16984410000033110	Sulpiride	Sulpiride 200mg/5ml oral solution sugar free
	35931311000001100	11296410000033110	Promazine Hydrochloride	Promazine 25mg/5ml oral solution
	36027911000001100	10635410000033110	Pericyazine	Pericyazine 10mg/5ml oral solution
	36053411000001100	65964100003311 1	Haloperidol,Haloperidol	Haloperidol 20mg/2ml solution for injection ampoules
	36053511000001100	29297410000033110	Haloperidol,Haloperidol	Haloperidol 5mg/1ml solution for injection ampoules
	36137911000001100	24664100003311 8	Chlorpromazine Hydrochloride	Chlorpromazine 100mg/5ml oral suspension sugar free
	36138211000001100	38589410000033110	Chlorpromazine Hydrochloride,Chlorpromazine Hydrochloride	Chlorpromazine 25mg/1ml solution for injection ampoules
	36138311000001100	24004100003311 0	Chlorpromazine Hydrochloride	Chlorpromazine 25mg/5ml oral solution
	36138411000001100	24134100003311 6	Chlorpromazine Hydrochloride	Chlorpromazine 50mg/2ml solution for injection ampoules
	36143111000001100	14669410000033110	Trifluoperazine	Trifluoperazine 10mg modified-release capsules
	36143211000001100	14670410000033110	Trifluoperazine	Trifluoperazine 15mg modified-release capsules
	36143311000001100	14676410000033110	Trifluoperazine	Trifluoperazine 1mg/5ml oral solution sugar free
	36143511000001100	14671410000033110	Trifluoperazine	Trifluoperazine 2mg modified-release capsules
	36151811000001100	15634410000033110	Zuclopenthixol Acetate	Zuclopenthixol acetate 100mg/2ml solution for injection ampoules

	36152011000001100	1563541000033110	Zuclopenthixol Acetate	Zuclopenthixol acetate 50mg/1ml solution for injection ampoules
	36796110000011100	4317410000331 8	Flupentixol	Depixol 3mg tablets (Lundbeck Ltd)
	36796110000011100	4317410000331 8	Flupentixol	Depixol 3mg tablets (Lundbeck Ltd)
	36803110000011100	5709410000331 7	Perphenazine	Fentazin 2mg tablets (Advanz Pharma)
	36809110000011100	5710410000331 0	Perphenazine	Fentazin 4mg tablets (Advanz Pharma)
	36861110000011100	9350410000331 5	Fluphenazine Hydrochloride	Moditen 1mg tablets (Sanofi)
	36868411000001100	1299034100003 3100	Aripiprazole	Arpoya 5mg tablets (Torrent Pharma (UK) Ltd)
	36868711000001100	1299044100003 3100	Aripiprazole	Arpoya 10mg tablets (Torrent Pharma (UK) Ltd)
	36869111000001100	1299054100003 3100	Aripiprazole	Arpoya 15mg tablets (Torrent Pharma (UK) Ltd)
	36869311000001100	1299064100003 3100	Aripiprazole	Arpoya 30mg tablets (Torrent Pharma (UK) Ltd)
	36880110000011100	1016541000033110	Pimozide	Orap 4mg tablets (Eumedica Pharma Ltd)
	36883110000011100	2846441000033110	Benperidol	Benquil 250microgram tablets (Concord Pharmaceuticals Ltd)
	36905011000001100	1298874100003 3100	Olanzapine	Olanzapine 2.5mg orodispersible tablets sugar free
	36905111000001100	1298884100003 3100	Olanzapine	Olanzapine 7.5mg orodispersible tablets sugar free
	37249011000001100	1335324100003 3100	Aripiprazole	Aripiprazole 1mg/ml oral solution sugar free
	37687211000001100	1315894100003 3100	Clozapine	Zaponex 12.5mg orodispersible tablets (Leyden Delta B.V.)

	37687511000001100	13159141000033100	Clozapine	Zaponex 25mg orodispersible tablets (Leyden Delta B.V.)
	37687811000001100	13159241000033100	Clozapine	Zaponex 50mg orodispersible tablets (Leyden Delta B.V.)
	37688211000001100	13158841000033100	Clozapine	Zaponex 100mg orodispersible tablets (Leyden Delta B.V.)
	37688811000001100	13159041000033100	Clozapine	Zaponex 200mg orodispersible tablets (Leyden Delta B.V.)
	37697511000001100	13158341000033100	Clozapine	Clozapine 100mg orodispersible tablets sugar free
	37697711000001100	13158441000033100	Clozapine	Clozapine 12.5mg orodispersible tablets sugar free
	37697811000001100	13158541000033100	Clozapine	Clozapine 200mg orodispersible tablets sugar free
	37697911000001100	13158641000033100	Clozapine	Clozapine 25mg orodispersible tablets sugar free
	37698011000001100	13158741000033100	Clozapine	Clozapine 50mg orodispersible tablets sugar free
	379911000001101	813741000033119	Chlorpromazine Hydrochloride	Largactil 25mg/5ml syrup (Sanofi)
	38257110000011100	935141000033116	Fluphenazine Hydrochloride	Moditen 2.5mg tablets (Sanofi)
	38260110000011100	935241000033111	Fluphenazine Hydrochloride	Moditen 5mg tablets (Sanofi-Synthelabo Ltd)
	38303110000011100	284941000033113	Clozapine	Clozaril 25mg tablets (Viatris UK Healthcare Ltd)
	38320110000011100	285041000033113	Clozapine	Clozaril 100mg tablets (Viatris UK Healthcare Ltd)
	38696711000001100	13581641000033100	Risperidone	Risperidone 250microgram tablets
	38728110000011100	842041000033117	Loxapine Succinate	Loxapac 10mg capsules (Wyeth Pharmaceuticals)

	387311000001101	1276141000033110	Haloperidol	Serenace 20mg tablets (Teva UK Ltd)
	387361100000110 0	8422410000331 13	Loxapine Succinate	Loxapac 50mg capsules (Wyeth Pharmaceuticals)
	387507110000001 100	1464141000033110	Trifluoperazine	Trifluoperazine 5mg/5ml oral solution sugar free
	387535110000001 100	2640141000033110	Promazine Hydrochloride	Promazine 50mg/5ml oral solution
	388953110000001 100	3063441000033110	Levomepromazine Hydrochloride,Levomepromazine Hydrochloride,Levomepromazine Hydrochloride	Levomepromazine 25mg/1ml solution for injection ampoules
	389318110000001 100	1371904100003 3100	Levomepromazine Maleate	Levomepromazine 50mg tablets
	390050110000001 100	1373494100003 3100	Quetiapine Fumarate	Biquelle XL 600mg tablets (Aspire Pharma Ltd)
	390274110000001 100	1373484100003 3100	Quetiapine Fumarate	Quetiapine 600mg modified-release tablets
	390481100000110 0	1765841000033110	Zotepine	Zoleptil 25 tablets (Movianto UK Ltd)
	390561100000110 0	1765941000033110	Zotepine	Zoleptil 50 tablets (Movianto UK Ltd)
	391257110000001 100	1374554100003 3100	Haloperidol	Haloperidol 5mg/5ml oral solution sugar free (Special Order)
	391488110000001 100	1374844100003 3100	Trifluoperazine	Trifluoperazine 1mg/5ml oral suspension (Special Order)
	392751100000110 0	1271541000033110	Haloperidol,Haloperidol	Serenace 20mg/2ml solution for injection ampoules (IVAX Pharmaceuticals UK Ltd)
	393381100000110 0	2644941000033110	Amisulpride	Solian 100mg/ml oral solution (Sanofi)
	396924110000001 100	3091241000033110	Flupentixol	Flupentixol 3mg tablets

	39699511000001100	840741000033115	Loxapine Succinate	Loxapine 10mg capsules
	39699611000001100	840841000033113	Loxapine Succinate	Loxapine 25mg capsules
	39699711000001100	840941000033117	Loxapine Succinate	Loxapine 50mg capsules
	39703111000001100	184464100003311 0	Quetiapine Fumarate	Quetiapine 150mg tablets
	39706011000001100	147064100003311 0	Trifluoperazine	Trifluoperazine 1mg tablets
	39706211000001100	147074100003311 0	Trifluoperazine	Trifluoperazine 5mg tablets
	39707711000001100	156404100003311 0	Zuclopenthixol Dihydrochloride	Zuclopenthixol 10mg tablets
	39707911000001100	156414100003311 0	Zuclopenthixol Dihydrochloride	Zuclopenthixol 2mg tablets
	39708111000001100	156424100003311 0	Zuclopenthixol Dihydrochloride	Zuclopenthixol 25mg tablets
	40230110000011 00	842141000033118	Loxapine Succinate	Loxapac 25mg capsules (Wyeth Pharmaceuticals)
	40392110000011 00	286904100003311 0	Risperidone	Risperdal Quicklet 1mg orodispersible tablets (Janssen-Cilag Ltd)
	40397110000011 00	286914100003311 0	Risperidone	Risperdal Quicklet 2mg orodispersible tablets (Janssen-Cilag Ltd)
	40540110000011 00	286884100003311 0	Risperidone	Risperidone 1mg orodispersible tablets sugar free
	40541110000011 00	286894100003311 0	Risperidone	Risperidone 2mg orodispersible tablets sugar free
	40884711000001100	140634410000331 00	Aripiprazole	Aripiprazole 10mg tablets (Imported)
	4111000001106	184474100003311 0	Quetiapine Fumarate	Seroquel 150mg tablets (AstraZeneca UK Ltd)
	41245110000011 00	176574100003311 0	Zotepine	Zoleptil 100 tablets (Movianto UK Ltd)
	41299110000011 00	273641000033114	Zuclopenthixol Acetate	Clopixol Acuphase 50mg/1ml solution for injection ampoules (Lundbeck Ltd)
	41304110000011 00	274841000033112	Zuclopenthixol Acetate	Clopixol Acuphase

				100mg/2ml solution for injection ampoules (Lundbeck Ltd)
	41442811000001100	14174441000033100	Levomepromazine Hydrochloride	Levorol 5mg/ml oral solution (Galvany Pharma Ltd)
	41482411000001100	14174341000033100	Levomepromazine Hydrochloride	Levomepromazine hydrochloride 25mg/5ml oral solution sugar free
	41641100000110 1	47984100003311 2	Sulpiride	Dolmatil 400mg tablets (Sanofi)
	41829110000011 00	98164100003311 7	Levomepromazine Hydrochloride,Levomepromazine Hydrochloride,Levomepromazine Hydrochloride	Nozinan 25mg/1ml solution for injection ampoules (Neuraxpharm UK Ltd)
	42267611000001100	28351410000331 10	Amisulpride	Amisulpride 100mg tablets
	42267711000001100	58941000033117	Amisulpride	Amisulpride 200mg tablets
	42267811000001100	22166410000331 10	Amisulpride	Amisulpride 400mg tablets
	42267911000001100	59041000033114	Amisulpride	Amisulpride 50mg tablets
	42268111000001100	31375410000331 10	Aripiprazole	Aripiprazole 10mg tablets
	42268211000001100	31376410000331 10	Aripiprazole	Aripiprazole 15mg tablets
	42268311000001100	39505410000331 10	Aripiprazole	Aripiprazole 1mg/ml oral solution
	42268411000001100	31377410000331 10	Aripiprazole	Aripiprazole 30mg tablets
	42268511000001100	32252410000331 10	Aripiprazole	Aripiprazole 5mg tablets
	42268611000001100	14514100003311 9	Benperidol	Benperidol 250microgram tablets
	42268711000001100	12667641000033100	Chlorpromazine	Chlorpromazine 100mg suppositories
	42268811000001100	24914100003311 8	Chlorpromazine Hydrochloride	Chlorpromazine 100mg tablets
	42268911000001100	24904100003311 7	Chlorpromazine Hydrochloride	Chlorpromazine 10mg tablets

	42269011000001100	8885941000033110	Chlorpromazine	Chlorpromazine 25mg suppositories
	42269111000001100	2492410000331113	Chlorpromazine Hydrochloride	Chlorpromazine 25mg tablets
	42269211000001100	2493410000331115	Chlorpromazine Hydrochloride	Chlorpromazine 50mg tablets
	42270011000001100	2875410000331112	Clozapine	Clozapine 100mg tablets
	42270111000001100	2876410000331113	Clozapine	Clozapine 25mg tablets
	42270211000001100	5078041000033110	Clozapine	Clozapine 50mg tablets
	42272111000001100	5954410000331112	Fluphenazine Hydrochloride	Fluphenazine 1mg tablets
	42272211000001100	5956410000331114	Fluphenazine Hydrochloride	Fluphenazine 2.5mg tablets
	42272311000001100	5957410000331117	Fluphenazine Hydrochloride	Fluphenazine 5mg tablets
	42272611000001100	6623410000331115	Haloperidol	Haloperidol 1.5mg tablets
	42272711000001100	6632410000331118	Haloperidol	Haloperidol 10mg tablets
	42272811000001100	6613410000331112	Haloperidol	Haloperidol 10mg/5ml oral solution sugar free
	42272911000001100	6624410000331114	Haloperidol	Haloperidol 20mg tablets
	42273011000001100	6583410000331110	Haloperidol	Haloperidol 500microgram capsules
	42273111000001100	6633410000331111	Haloperidol	Haloperidol 500microgram tablets
	42273211000001100	6625410000331110	Haloperidol	Haloperidol 5mg tablets
	42273311000001100	6603410000331116	Haloperidol	Haloperidol 5mg/5ml oral solution sugar free
	42273511000001100	3063541000033110	Levomepromazine Maleate	Levomepromazine 25mg tablets
	42275411000001100	2077841000033110	Olanzapine,Olanzapine	Olanzapine 10mg oral lyophilisates sugar free
	42275511000001100	1003641000033110	Olanzapine	Olanzapine 10mg tablets
	42275611000001100	2759641000033110	Olanzapine	Olanzapine 15mg tablets
	42275711000001100	1661141000033110	Olanzapine	Olanzapine 2.5mg tablets

	42275811000001100	38665410000033110	Olanzap-ine,Olanzapine	Olanzapine 20mg oral lyophilisates sugar free
	42275911000001100	39962410000033110	Olanzapine	Olanzapine 20mg tablets
	42276011000001100	20777410000033110	Olanzap-ine,Olanzapine	Olanzapine 5mg oral lyophilisates sugar free
	42276111000001100	10037410000033110	Olanzapine	Olanzapine 5mg tablets
	42276211000001100	10038410000033110	Olanzapine	Olanzapine 7.5mg tablets
	42276311000001100	10227410000033110	Oxypertine	Oxypertine 10mg capsules
	42276411000001100	10269410000033110	Oxypertine	Oxypertine 40mg tablets
	42276811000001100	10667410000033110	Pericyazine	Pericyazine 10mg tablets
	42276911000001100	10670410000033110	Pericyazine	Pericyazine 2.5mg tablets
	42277011000001100	10651410000033110	Perphenazine	Perphenazine 2mg tablets
	42277111000001100	10652410000033110	Perphenazine	Perphenazine 4mg tablets
	42277411000001100	10909410000033110	Pimozide	Pimozide 2mg tablets
	42277511000001100	10910410000033110	Pimozide	Pimozide 4mg tablets
	42277611000001100	11343410000033110	Promazine Hy-drochloride	Promazine 25mg tablets
	42277711000001100	11344410000033110	Promazine Hy-drochloride	Promazine 50mg tablets
	42278011000001100	11500410000033110	Quetiapine Fumarate	Quetiapine 100mg tablets
	42278111000001100	45918410000033110	Quetiapine Fumarate	Quetiapine 200mg modified-release tablets
	42278211000001100	11501410000033110	Quetiapine Fumarate	Quetiapine 200mg tablets
	42278311000001100	11502410000033110	Quetiapine Fumarate	Quetiapine 25mg tablets
	42278411000001100	45919410000033110	Quetiapine Fumarate	Quetiapine 300mg modified-release tablets
	42278511000001100	29758410000033110	Quetiapine Fumarate	Quetiapine 300mg tablets
	42278611000001100	45920410000033110	Quetiapine Fumarate	Quetiapine 400mg modified-release tablets
	42278711000001100	11766410000033110	Risperidone	Risperidone 1mg tablets

	42278811000001100	1173941000033110	Risperidone	Risperidone 1mg/ml oral solution sugar free
	42278911000001100	1176741000033110	Risperidone	Risperidone 2mg tablets
	42279011000001100	1176841000033110	Risperidone	Risperidone 3mg tablets
	42279111000001100	1176941000033110	Risperidone	Risperidone 4mg tablets
	42279211000001100	2188241000033110	Risperidone	Risperidone 500microgram tablets
	42279311000001100	1177741000033110	Risperidone	Risperidone 6mg tablets
	42279411000001100	1396441000033110	Sulpiride	Sulpiride 200mg tablets
	42279611000001100	1397641000033110	Sulpiride	Sulpiride 400mg tablets
	42279711000001100	1433741000033110	Thioridazine	Thioridazine 100mg tablets
	42279811000001100	1432341000033110	Thioridazine	Thioridazine 100mg/5ml oral suspension
	42279911000001100	1433641000033110	Thioridazine	Thioridazine 10mg tablets
	42280011000001100	1433841000033110	Thioridazine	Thioridazine 25mg tablets
	42280111000001100	1432541000033110	Thioridazine	Thioridazine 25mg/5ml oral solution
	42280211000001100	1432441000033110	Thioridazine	Thioridazine 25mg/5ml oral suspension
	42280311000001100	1433941000033110	Thioridazine	Thioridazine 50mg tablets
	42281411000001100	1766041000033110	Zotepine	Zotepine 100mg tablets
	42281511000001100	1766141000033110	Zotepine	Zotepine 25mg tablets
	42281611000001100	1766241000033110	Zotepine	Zotepine 50mg tablets
	42405911000001100	14255041000033100	Olanzapine	Xyquila 10mg powder for solution for injection vials (Eramol (UK) Ltd)
	42547611000001100	14277941000033100	Levomepromazine Maleate	Levorol 6.25mg tablets (Galvany Pharma Ltd)
	42565811000001100	14277841000033100	Levomepromazine Maleate	Levomepromazine 6.25mg tablets

	42684011000001100	14397841000033100	Quetiapine Fumarate	Quetiapine 400mg tablets
	42684111000001100	14397741000033100	Quetiapine Fumarate	Quetiapine 50mg tablets
	429611000001104	11765410000033110	Risperidone	Risperdal 4mg tablets (Janssen-Cilag Ltd)
	42986711000001100	14506241000033100	Olanzapine	Ceyxa 10mg powder for solution for injection vials (Nordic Pharma Ltd)
	43185011000001100	14515941000033100	Aripiprazole	Elozar 10mg orodispersible tablets (Novumgen Ltd)
	43185211000001100	14516041000033100	Aripiprazole	Elozar 15mg orodispersible tablets (Novumgen Ltd)
	432111000001100	81584100003311 3	Chlorpromazine Hydrochloride	Largactil 100mg tablets (Sanofi)
	43761100000110 6	27599410000331 10	Olanzapine,Olanzapine	Zyprexa 15mg Velotabs (Neon Healthcare Ltd)
	44050811000001100	14852541000033100	Aripiprazole	Aripiprazole 1mg tablets
	44050911000001100	14852641000033100	Aripiprazole	Aripiprazole 2.5mg tablets
	44411000001109	48094100003311 6	Sulpiride	Dolmatil 200mg tablets (Sanofi)
	45449110000011 00	10164410000331 10	Pimozide	Orap 2mg tablets (Janssen-Cilag Ltd)
	45476110000011 00	75754100003311 1	Oxypertine	Integrin 10mg capsules (Sterling-Winthrop)
	45811000001107	13864410000331 10	Trifluoperazine	Stelazine 5mg tablets (Mercury Pharma Group Ltd)
	463911000001104	47654100003311 9	Haloperidol	Dozic 5mg/5ml oral solution (Rosemont Pharmaceuticals Ltd)
	467011000001109	15652410000331 10	Olanzapine	Zyprexa 5mg tablets (Neon Healthcare Ltd)
	47296110000011 00	12771410000331 10	Sertindole	Serdolect 4mg tablets (Imported (Denmark))

	47317110000011 00	12768410000331 10	Sertindole	Serdolect 12mg tablets (Imported (Denmark))
	47320110000011 00	12769410000331 10	Sertindole	Serdolect 16mg tablets (Imported (Denmark))
	47323110000011 00	12770410000331 10	Sertindole	Serdolect 20mg tablets (Imported (Denmark))
	47482110000011 00	12772410000331 10	Sertindole	Sertindole 12mg tablets
	47483110000011 00	12773410000331 10	Sertindole	Sertindole 16mg tablets
	47484110000011 00	12774410000331 10	Sertindole	Sertindole 20mg tablets
	47485110000011 00	12775410000331 10	Sertindole	Sertindole 4mg tablets
	47641100000110 8	12777410000331 10	Quetiapine Fumarate	Seroquel 200mg tablets (Luye Pharma Ltd)
	48661100000110 1	15651410000331 10	Olanzapine	Zyprexa 10mg tablets (Neon Healthcare Ltd)
	49891110000011 00	29759410000331 10	Quetiapine Fumarate	Seroquel 300mg tablets (Luye Pharma Ltd)
	50261100000110 7	12697410000331 10	Haloperidol	Serenace 2mg/ml liquid (Teva UK Ltd)
	50751100000110 6	59354100003311 2	Flupentixol	Fluanxol 500mi-crogram tablets (Lundbeck Ltd)
	50751100000110 6	59354100003311 2	Flupentixol	Fluanxol 500mi-crogram tablets (Lundbeck Ltd)
	511000001103	81594100003311 7	Chlorpromazine Hydrochloride	Largactil 25mg tablets (Sanofi)
	51111100000110 6	27598410000331 10	Olanzapine	Zyprexa 15mg tablets (Neon Healthcare Ltd)
	51891100000110 0	11764410000331 10	Risperidone	Risperdal 3mg tablets (Janssen-Cilag Ltd)
	52271100000110 6	12683410000331 10	Haloperidol	Serenace 500mi-crogram capsules (Teva UK Ltd)
	52661100000110 1	21881410000331 10	Risperidone	Risperdal 500mi-crogram tablets (Janssen-Cilag Ltd)
	55511100000110 2	89214100003311 4	Thioridazine	Melleril 100mg/5ml oral

				suspension (Novartis Pharmaceuticals UK Ltd)
	567111000001103	2955641000033110	Trifluoperazine	Terrazine 5mg tablets (Dr Reddy's Laboratories (UK) Ltd)
	569411100000110 0	3030941000033110	Olanzapine	Zyprexa 10mg powder for solution for injection vials (Nordic Pharma Ltd)
	581831100000110 0	3030841000033110	Olanzapine	Olanzapine 10mg powder for solution for injection vials
	58711000001108	8922410000331 19	Thioridazine	Melleril 25mg/5ml oral suspension (Novartis Pharmaceuticals UK Ltd)
	589911000001105	1176241000033110	Risperidone	Risperdal 1mg tablets (Janssen-Cilag Ltd)
	596211000001109	1277841000033110	Quetiapine Fumarate	Seroquel 25mg tablets (Luye Pharma Ltd)
	618211000001102	2871410000331 5	Zuclopenthixol Dihydrochloride	Clopixol 25mg tablets (Lundbeck Ltd)
	636711000001103	1385941000033110	Trifluoperazine	Stelazine 1mg/5ml syrup (Mercury Pharma Group Ltd)
	701911000001108	1512741000033110	Levomepromazine Maleate	Nozinan 25mg tablets (Neuraxpharm UK Ltd)
	710411000001106	1385441000033110	Trifluoperazine	Stelazine 15mg Spansules (Mercury Pharma Group Ltd)
	711411000001102	8958410000331 4	Thioridazine	Melleril 10mg tablets (Novartis Pharmaceuticals UK Ltd)
	714111000001105	1385541000033110	Trifluoperazine	Stelazine 2mg Spansules (Mercury Pharma Group Ltd)
	746611000001104	9656410000331 8	Pericyazine	Neulactil 2.5mg tablets (Sanofi)

	7637111000001100	3137841000033110	Aripiprazole	Abilify 10mg tablets (Otsuka Pharmaceuticals (U.K.) Ltd)
	7637411000001100	3137941000033110	Aripiprazole	Abilify 15mg tablets (Otsuka Pharmaceuticals (U.K.) Ltd)
	7637811000001100	3138041000033110	Aripiprazole	Abilify 30mg tablets (Otsuka Pharmaceuticals (U.K.) Ltd)
	7740511000001100	3141141000033110	Trifluoperazine	Stelazine Forte 1mg/ml oral solution (Mercury Pharma Group Ltd)
	7769110000011109	1565341000033110	Olanzapine	Zyprexa 7.5mg tablets (Neon Healthcare Ltd)
	7885110000011105	8937410000331115	Thioridazine	Melleril 25mg/5ml syrup (Novartis Pharmaceuticals UK Ltd)
	7972911000001100	3158941000033110	Clozapine	Denzapine 25mg tablets (Britannia Pharmaceuticals Ltd)
	7974011000001100	3159041000033110	Clozapine	Denzapine 100mg tablets (Britannia Pharmaceuticals Ltd)
	8275110000011102	2869410000331115	Zuclopenthixol Dihydrochloride	Clopixol 10mg tablets (Lundbeck Ltd)
	8277211000001100	4940141000033110	Amisulpride	Amisulpride 25mg/5ml oral solution
	8277311000001100	5890941000033110	Amisulpride	Amisulpride 25mg/5ml oral suspension
	8359411000001100	1719341000033110	Chlorpromazine Hydrochloride	Chlorpromazine 100mg/5ml oral solution
	8359711000001100	3849441000033110	Chlorpromazine Hydrochloride	Chlorpromazine 50mg/5ml oral solution
	8430110000011101	5934410000331111	Flupentixol	Fluanxol 1mg tablets (Lundbeck Ltd)

	843011000001101	593441000033111	Flupentixol	Fluanxol 1mg tablets (Lundbeck Ltd)
	852791100000110000	505464100003311000	Haloperidol	Haloperidol 1mg/5ml oral solution
	852801100000110000	589074100003311000	Haloperidol	Haloperidol 1mg/5ml oral suspension
	852831100000110000	104933410000331000	Haloperidol	Haloperidol 500micrograms/5ml oral solution
	852841100000110000	104934410000331000	Haloperidol	Haloperidol 500micrograms/5ml oral suspension
	858281100000110000	581434100003311000	Levomepromazine Maleate	Levomepromazine 2.5mg/5ml oral suspension
	858911000001109	117634100003311000	Risperidone	Risperdal 2mg tablets (Janssen-Cilag Ltd)
	869511000001104	811041000033116	Chlorpromazine Hydrochloride	Largactil 50mg/2ml solution for injection ampoules (Sanofi)
	883891100000110000	322634100003311000	Clozapine	Zaponex 25mg tablets (Leyden Delta B.V.)
	883911000001106	896141000033110	Thioridazine	Melleril 50mg tablets (Novartis Pharmaceuticals UK Ltd)
	883951100000110000	322644100003311000	Clozapine	Zaponex 100mg tablets (Leyden Delta B.V.)
	884511000001101	895941000033118	Thioridazine	Melleril 100mg tablets (Novartis Pharmaceuticals UK Ltd)
	897111000001104	207804100003311000	Olanzapine,Olanzapine	Zyprexa 10mg Velotabs (Neon Healthcare Ltd)
	899811000001103	127594100003311000	Haloperidol	Serenace 1.5mg tablets (Teva UK Ltd)
	903791100000110000	322534100003311000	Aripiprazole	Abilify 5mg tablets (Otsuka Pharmaceuticals (U.K.) Ltd)

	9044211000001100	3246141000033110	Risperidone	Risperdal Quicklet 500microgram orodispersible tablets (Janssen-Cilag Ltd)
	9045911000001100	3246041000033110	Risperidone	Risperidone 500microgram orodispersible tablets sugar free
	9106110000011102	8160410000331 10	Chlorpromazine Hydrochloride	Largactil 50mg tablets (Sanofi)
	9113110000011102	8157410000331 15	Chlorpromazine Hydrochloride	Largactil 10mg tablets (Sanofi)
	91711000001108	2077941000033110	Olanzapine,Olanzapine	Zyprexa 5mg Velotabs (Neon Healthcare Ltd)
	9178110000011104	9650410000331 12	Pericyazine	Neulactil Forte syrup (Sanofi)
	9483411000001100	69041000033115	Benperidol	Anquil 250microgram tablets (Neon Healthcare Ltd)
	9536311000001100	3063641000033110	Levomepromazine Maleate	Levomepromazine 6mg tablets
	9891011000001100	3866641000033110	Olanzapine,Olanzapine	Zyprexa 20mg Velotabs (Neon Healthcare Ltd)
	1066461100000 1100	3950641000033110	Aripiprazole	Abilify 10mg orodispersible tablets (Otsuka Pharmaceuticals (U.K.) Ltd)
	1066491100000 1100	3950741000033110	Aripiprazole	Abilify 15mg orodispersible tablets (Otsuka Pharmaceuticals (U.K.) Ltd)
	1066721100000 1100	3950841000033110	Aripiprazole	Abilify 1mg/ml oral solution (Otsuka Pharmaceuticals (U.K.) Ltd)
	1067531100000 1100	3950341000033110	Aripiprazole	Aripiprazole 10mg orodispersible tablets sugar free
	1067541100000 1100	3950441000033110	Aripiprazole	Aripiprazole 15mg orodispersible tablets sugar free
	1087851100000 1100	3996341000033110	Olanzapine	Zyprexa 20mg tablets (Neon Healthcare Ltd)

	11134311000001100	4012341000033110	Risperidone	Risperdal Quicklet 4mg orodispersible tablets (Janssen-Cilag Ltd)
	11134611000001100	4012241000033110	Risperidone	Risperdal Quicklet 3mg orodispersible tablets (Janssen-Cilag Ltd)
	11140811000001100	4012041000033110	Risperidone	Risperidone 3mg orodispersible tablets sugar free
	11140911000001100	4012141000033110	Risperidone	Risperidone 4mg orodispersible tablets sugar free
	113611000001108	1276041000033110	Haloperidol	Serenace 10mg tablets (Teva UK Ltd)
	11749311000001100	4200941000033110	Paliperidone	Invega 3mg modified-release tablets (Janssen-Cilag Ltd)
	11749811000001100	4201041000033110	Paliperidone	Invega 6mg modified-release tablets (Janssen-Cilag Ltd)
	11751011000001100	4201141000033110	Paliperidone	Invega 9mg modified-release tablets (Janssen-Cilag Ltd)
	11763211000001100	4200641000033110	Paliperidone	Paliperidone 3mg modified-release tablets
	11763311000001100	4200741000033110	Paliperidone	Paliperidone 6mg modified-release tablets
	11763411000001100	4200841000033110	Paliperidone	Paliperidone 9mg modified-release tablets
	117911000001106	6615410000331 17	Haloperidol	Haldol 2mg/ml oral solution (Essential Pharma Ltd)
	12504911000001100	5969441000033110	Haloperidol	Haloperidol 1.5mg/5ml oral solution
	12505011000001100	5969541000033110	Haloperidol	Haloperidol 1.5mg/5ml oral suspension

	12505511000001100	5969641000033110	Haloperidol	Haloperidol 250micrograms/5ml oral solution
	12505611000001100	5969741000033110	Haloperidol	Haloperidol 250micrograms/5ml oral suspension
	12505911000001100	4424541000033110	Haloperidol	Haloperidol 5mg/5ml oral solution
	12506011000001100	5993541000033110	Haloperidol	Haloperidol 5mg/5ml oral suspension
	12646911000001100	6445641000033110	Levomepromazine Maleate	Levomepromazine 6mg/5ml oral suspension
	12797811000001100	1025384100003310 0	Melperone Hydrochloride	Melperone hydrochloride 100mg tablets
	13011611000001100	5889741000033110	Quetiapine Fumarate	Quetiapine 12.5mg/5ml oral solution
	13011711000001100	5889841000033110	Quetiapine Fumarate	Quetiapine 12.5mg/5ml oral suspension
	13012211000001100	5974841000033110	Quetiapine Fumarate	Quetiapine 25mg/5ml oral solution
	13012311000001100	5974941000033110	Quetiapine Fumarate	Quetiapine 25mg/5ml oral suspension
	13059411000001100	1179884100003310 0	Pimozide	Pimozide 2mg/5ml oral suspension
	13451111000001100	4456841000033110	Aripiprazole	Abilify 9.75mg/1.3ml solution for injection vials (Otsuka Pharmaceuticals (U.K.) Ltd)
	13456411000001100	4456741000033110	Aripiprazole	Aripiprazole 9.75mg/1.3ml solution for injection vials
	14015811000001100	5997341000033110	Olanzapine	Olanzapine 2.5mg/5ml oral solution
	14015911000001100	3892541000033110	Olanzapine	Olanzapine 2.5mg/5ml oral suspension

	14018611000001100	5891141000033110	Quetiapine Fumarate	Quetiapine 100mg/5ml oral solution
	14018711000001100	4422541000033110	Quetiapine Fumarate	Quetiapine 100mg/5ml oral suspension
	14025311000001100	4592241000033110	Quetiapine Fumarate	Seroquel XL 200mg tablets (Luye Pharma Ltd)
	14025711000001100	4592441000033110	Quetiapine Fumarate	Seroquel XL 400mg tablets (Luye Pharma Ltd)
	14026011000001100	4592341000033110	Quetiapine Fumarate	Seroquel XL 300mg tablets (Luye Pharma Ltd)
	14026511000001100	4592141000033110	Quetiapine Fumarate	Seroquel XL 50mg tablets (Luye Pharma Ltd)
	14042811000001100	4591741000033110	Quetiapine Fumarate	Quetiapine 50mg modified-release tablets
	145511000001103	2658941000033110	Sulpiride	Sulpor 200mg/5ml oral solution (Rosemont Pharmaceuticals Ltd)
	152111000001103	6635410000331116	Haloperidol	Haldol 5mg tablets (Janssen-Cilag Ltd)
	153111000001109	1177641000033110	Risperidone	Risperdal 6mg tablets (Janssen-Cilag Ltd)
	153611000001101	2870410000331119	Zuclopenthixol Dihydrochloride	Clopixol 2mg tablets (Lundbeck Ltd)
	15452011000001100	5975041000033110	Quetiapine Fumarate	Quetiapine 50mg/5ml oral solution
	15452111000001100	5975141000033110	Quetiapine Fumarate	Quetiapine 50mg/5ml oral suspension
	15498911000001100	5078241000033110	Clozapine	Denzapine 50mg tablets (Britannia Pharmaceuticals Ltd)
	15505711000001100	5078141000033110	Clozapine	Denzapine 200mg tablets (Britannia Pharmaceuticals Ltd)

	15506011000001100	5077741000033110	Clozapine	Denzapine 50mg/ml oral suspension (Britannia Pharmaceuticals Ltd)
	15506211000001100	5077941000033110	Clozapine	Clozapine 200mg tablets
	15506311000001100	5077641000033110	Clozapine	Clozapine 50mg/ml oral suspension sugar free
	160211000001102	8960410000331 1	Thioridazine	Melleril 25mg tablets (Novartis Pharmaceuticals UK Ltd)
	17689811000001100	5909441000033110	Quetiapine Fumarate	Seroquel XL 150mg tablets (Luye Pharma Ltd)
	17828611000001100	5909341000033110	Quetiapine Fumarate	Quetiapine 150mg modified-release tablets
	18172311000001100	6455741000033110	Ziprasidone	Ziprasidone 40mg capsules
	18956611000001100	1187674100003 3100	Chlorpromazine Hydrochloride	Chlorpromazine 25mg/5ml oral solution sugar free
	19525911000001100	6455841000033110	Ziprasidone	Ziprasidone 80mg capsules
	19587711000001100	6456941000033110	Olanzapine	Zalasta 2.5mg tablets (Krka UK Ltd)
	19587911000001100	6457141000033110	Olanzapine	Zalasta 5mg tablets (Krka UK Ltd)
	19588111000001100	6457241000033110	Olanzapine	Zalasta 7.5mg tablets (Krka UK Ltd)
	19588311000001100	6456741000033110	Olanzapine	Zalasta 10mg tablets (Krka UK Ltd)
	19588511000001100	6456841000033110	Olanzapine	Zalasta 15mg tablets (Krka UK Ltd)
	19588711000001100	6457041000033110	Olanzapine	Zalasta 20mg tablets (Krka UK Ltd)
	19608211000001100	6465641000033110	Olanzapine	Zalasta 5mg orodispersible tablets (Krka UK Ltd)
	19609111000001100	6466041000033110	Olanzapine	Olanzapine 10mg orodispersible tablets sugar free
	19609311000001100	6466141000033110	Olanzapine	Olanzapine 15mg orodispersible tablets sugar free

	19609411000001100	6466241000033110	Olanzapine	Olanzapine 20mg orodispersible tablets sugar free
	19609511000001100	6465941000033110	Olanzapine	Olanzapine 5mg orodispersible tablets sugar free
	19609711000001100	6465341000033110	Olanzapine	Zalasta 10mg orodispersible tablets (Krka UK Ltd)
	19610111000001100	6465441000033110	Olanzapine	Zalasta 15mg orodispersible tablets (Krka UK Ltd)
	19610411000001100	6465541000033110	Olanzapine	Zalasta 20mg orodispersible tablets (Krka UK Ltd)
	19626711000001100	7681641000033110	Olanzapine	Olanzapine 10mg orodispersible tablets
	19626811000001100	7681741000033110	Olanzapine	Olanzapine 15mg orodispersible tablets
	19626911000001100	7681841000033110	Olanzapine	Olanzapine 20mg orodispersible tablets
	19627011000001100	7681941000033110	Olanzapine	Olanzapine 5mg orodispersible tablets
	19725311000001100	6469841000033110	Clotiapine	Clotiapine 40mg tablets
	19809011000001100	2653841000033110	Levomeproma-zine Maleate	Levinan 6mg tab-lets (Kyowa Kirin Ltd)
	19890011000001100	6515041000033110	Olanzapine	Arkolamyl 10mg orodispersible tablets (Viatris UK Healthcare Ltd)
	19890311000001100	6515141000033110	Olanzapine	Arkolamyl 15mg orodispersible tablets (Viatris UK Healthcare Ltd)
	19890611000001100	6514941000033110	Olanzapine	Arkolamyl 5mg orodispersible tablets (Viatris UK Healthcare Ltd)
	19890911000001100	6515241000033110	Olanzapine	Arkolamyl 20mg orodispersible tablets (Viatris

				UK Healthcare Ltd)
	20176811000001100	88860410000033110	Chlorpromazine Hydrochloride	Chlorpromazine 10mg capsules
	20356011000001100	76878410000033110	Quetiapine Fumarate	Sondate XL 50mg tablets (Teva UK Ltd)
	20356211000001100	76879410000033110	Quetiapine Fumarate	Sondate XL 200mg tablets (Teva UK Ltd)
	20356411000001100	76880410000033110	Quetiapine Fumarate	Sondate XL 300mg tablets (Teva UK Ltd)
	20356611000001100	76881410000033110	Quetiapine Fumarate	Sondate XL 400mg tablets (Teva UK Ltd)
	21248811000001100	81129410000033110	Quetiapine Fumarate	Tenprolide XL 50mg tablets (Actavis UK Ltd)
	21249311000001100	81130410000033110	Quetiapine Fumarate	Tenprolide XL 200mg tablets (Actavis UK Ltd)
	21249511000001100	81131410000033110	Quetiapine Fumarate	Tenprolide XL 300mg tablets (Actavis UK Ltd)
	213711000001101	13853410000033110	Trifluoperazine	Stelazine 10mg Spansules (Mercury Pharma Group Ltd)
	21714411000001100	82462410000033110	Quetiapine Fumarate	Sondate XL 150mg tablets (Teva UK Ltd)
	22221211000001100	82983410000033110	Quetiapine Fumarate	Seotiapim XL 200mg tablets (Sandoz Ltd)
	22221411000001100	82985410000033110	Quetiapine Fumarate	Seotiapim XL 400mg tablets (Sandoz Ltd)
	22221611000001100	82984410000033110	Quetiapine Fumarate	Seotiapim XL 300mg tablets (Sandoz Ltd)
	22409511000001100	83491410000033110	Quetiapine Fumarate	Tenprolide XL 400mg tablets (Actavis UK Ltd)
	22635511000001100	11730041000033100	Quetiapine Fumarate	Quetiapine 200mg/5ml oral suspension
	228411000001106	12762410000033110	Haloperidol	Serenace 5mg tablets (Teva UK Ltd)
	24311000001100	13667410000033110	Amisulpride	Solian 50 tablets (Sanofi)

	246411000001108	1396541000033110	Sulpiride	Sulpitil 200mg tablets (Pfizer Ltd)
	254011000001107	2216741000033110	Amisulpride	Solian 400 tablets (Sanofi)
	258032110000001100	9604441000033110	Quetiapine Fumarate	Ebesque XL 50mg tablets (Ethypharm UK Ltd)
	258043110000001100	9605541000033110	Quetiapine Fumarate	Ebesque XL 200mg tablets (Ethypharm UK Ltd)
	258049110000001100	9605941000033110	Quetiapine Fumarate	Ebesque XL 300mg tablets (Ethypharm UK Ltd)
	258060110000001100	9606341000033110	Quetiapine Fumarate	Ebesque XL 400mg tablets (Ethypharm UK Ltd)
	266544110000001100	9809241000033110	Lurasidone Hydrochloride	Latuda 18.5mg tablets (CNX Therapeutics Ltd)
	266547110000001100	9809341000033110	Lurasidone Hydrochloride	Latuda 37mg tablets (CNX Therapeutics Ltd)
	266550110000001100	9809441000033110	Lurasidone Hydrochloride	Latuda 74mg tablets (CNX Therapeutics Ltd)
	266553110000001100	9808941000033110	Lurasidone Hydrochloride	Lurasidone 18.5mg tablets
	266554110000001100	9809041000033110	Lurasidone Hydrochloride	Lurasidone 37mg tablets
	266555110000001100	9809141000033110	Lurasidone Hydrochloride	Lurasidone 74mg tablets
	271211000001102	6634410000331 17	Haloperidol	Haldol 10mg tablets (Janssen-Cilag Ltd)
	278422110000001100	1061514100003 3100	Quetiapine Fumarate	Atrolak XL 50mg tablets (Accord-UK Ltd)
	278424110000001100	1061524100003 3100	Quetiapine Fumarate	Atrolak XL 200mg tablets (Accord-UK Ltd)
	278426110000001100	1061534100003 3100	Quetiapine Fumarate	Atrolak XL 300mg tablets (Accord-UK Ltd)
	278428110000001100	1061544100003 3100	Quetiapine Fumarate	Atrolak XL 400mg tablets (Accord-UK Ltd)
	279463110000001100	1371914100003 3100	Levomepromazine Maleate	Levomepromazine 100mg tablets

	28400211000001100	10041141000033100	Quetiapine Fumarate	Biquelle XL 50mg tablets (Aspire Pharma Ltd)
	28400411000001100	10041241000033100	Quetiapine Fumarate	Biquelle XL 150mg tablets (Aspire Pharma Ltd)
	28400611000001100	10041441000033100	Quetiapine Fumarate	Biquelle XL 300mg tablets (Aspire Pharma Ltd)
	28400811000001100	10041541000033100	Quetiapine Fumarate	Biquelle XL 400mg tablets (Aspire Pharma Ltd)
	28401011000001100	10041341000033100	Quetiapine Fumarate	Biquelle XL 200mg tablets (Aspire Pharma Ltd)
	28411511000001100	10041741000033100	Quetiapine Fumarate	Zaluron XL 50mg tablets (Fontus Health Ltd)
	28411711000001100	10041841000033100	Quetiapine Fumarate	Zaluron XL 150mg tablets (Fontus Health Ltd)
	28411911000001100	10041941000033100	Quetiapine Fumarate	Zaluron XL 200mg tablets (Fontus Health Ltd)
	28412111000001100	10042041000033100	Quetiapine Fumarate	Zaluron XL 300mg tablets (Fontus Health Ltd)
	28412311000001100	10042141000033100	Quetiapine Fumarate	Zaluron XL 400mg tablets (Fontus Health Ltd)
	28419110000011100	28352410000331 10	Amisulpride	Solian 100 tablets (Sanofi)
	28805110000011100	81344100003311 4	Chlorpromazine Hydrochloride	Largactil Forte 100mg/5ml oral suspension (Sanofi)
	28994711000001100	10254741000033100	Quetiapine Fumarate	Mintreleq XL 50mg tablets (Aristo Pharma Ltd)
	28994911000001100	10254841000033100	Quetiapine Fumarate	Mintreleq XL 150mg tablets (Aristo Pharma Ltd)

	28995111000001100	10254941000033100	Quetiapine Fumarate	Mintreleq XL 200mg tablets (Aristo Pharma Ltd)
	28995611000001100	10255041000033100	Quetiapine Fumarate	Mintreleq XL 300mg tablets (Aristo Pharma Ltd)
	28995811000001100	10255141000033100	Quetiapine Fumarate	Mintreleq XL 400mg tablets (Aristo Pharma Ltd)
	30799611000001100	27581410000331 10	Haloperidol	Haloperidol 200mi-crograms/ml oral solution sugar free
	31951100000110 3	13666410000331 10	Amisulpride	Solian 200 tablets (Sanofi)
	32241100000110 2	29555410000331 10	Trifluoperazine	Terrazine 1mg tablets (Dr Red-dy's Laboratories (UK) Ltd)
	32457110000011 00	12716410000331 10	Haloperi-dol,Haloperidol	Serenace 5mg/1ml solution for injection am-poules (IVAX Pharmaceuticals UK Ltd)
	32469110000011 00	29298410000331 10	Haloperidol	Haldol 5mg/1ml solution for injec-tion ampoules (Janssen-Cilag Ltd)
	32503211000001100	11578941000033100	Quetiapine Fumarate	Psyquet XL 50mg tablets (Sandoz Ltd)
	32503411000001100	11579241000033100	Quetiapine Fumarate	Psyquet XL 150mg tablets (Sandoz Ltd)
	32503611000001100	11579341000033100	Quetiapine Fumarate	Psyquet XL 200mg tablets (Sandoz Ltd)
	32503811000001100	11579441000033100	Quetiapine Fumarate	Psyquet XL 300mg tablets (Sandoz Ltd)
	32504011000001100	11579541000033100	Quetiapine Fumarate	Psyquet XL 400mg tablets (Sandoz Ltd)
	32531100000110 3	13863410000331 10	Trifluoperazine	Stelazine 1mg tablets (Mercury Pharma Group Ltd)

	32687711000001100	11583741000033100	Loxapine	Loxapine 9.1mg/dose inhalation powder
	32687811000001100	11583841000033100	Loxapine	Adasuve 9.1mg/dose inhalation powder (Galen Ltd)
	33391011000001100	11753641000033100	Quetiapine Fumarate	Quetiapine 20mg/ml oral suspension sugar free
	33568811000001100	11782141000033100	Quetiapine Fumarate	Brancico XL 300mg tablets (Zentiva Pharma UK Ltd)
	33569011000001100	11782041000033100	Quetiapine Fumarate	Brancico XL 200mg tablets (Zentiva Pharma UK Ltd)
	33569711000001100	11781941000033100	Quetiapine Fumarate	Brancico XL 150mg tablets (Zentiva Pharma UK Ltd)
	33569911000001100	11782341000033100	Quetiapine Fumarate	Brancico XL 50mg tablets (Zentiva Pharma UK Ltd)
	33570111000001100	11782241000033100	Quetiapine Fumarate	Brancico XL 400mg tablets (Zentiva Pharma UK Ltd)
	33690711000001100	11821541000033100	Levomepromazine Maleate	Levomepromazine 62.5mg/5ml oral suspension
	337311000001109	171394100003311 0	Olanzapine	Zyprexa 2.5mg tablets (Neon Healthcare Ltd)
	340011000001109	965541000033119	Pericyazine	Neulactil 10mg tablets (Sanofi)
	340911000001108	127764100003311 0	Quetiapine Fumarate	Seroquel 100mg tablets (Luye Pharma Ltd)
	343311000001101	117384100003311 0	Risperidone	Risperdal 1mg/ml oral solution (Janssen-Cilag Ltd)
	343491100000110 0	275974100003311 0	Olanzapine,Olanzapine	Olanzapine 15mg oral lyophilisates sugar free
	34936811000001100	12428041000033100	Quetiapine Fumarate	Atrolak XL 150mg tablets (Accord-UK Ltd)
	35753411000001100	13155441000033100	Haloperidol	Halkid 200micrograms/ml oral

				solution (Thame Laboratories Ltd)
	35771811000001100	12665741000033100	Cariprazine Hydrochloride	Reagila 1.5mg capsules (Recordati Pharmaceuticals Ltd)
	35772111000001100	12665841000033100	Cariprazine Hydrochloride	Reagila 3mg capsules (Recordati Pharmaceuticals Ltd)
	35772411000001100	12665941000033100	Cariprazine Hydrochloride	Reagila 4.5mg capsules (Recordati Pharmaceuticals Ltd)
	35772711000001100	12666041000033100	Cariprazine Hydrochloride	Reagila 6mg capsules (Recordati Pharmaceuticals Ltd)
	35853811000001100	12665341000033100	Cariprazine Hydrochloride	Cariprazine 1.5mg capsules
	35853911000001100	12665441000033100	Cariprazine Hydrochloride	Cariprazine 3mg capsules
	35854011000001100	12665541000033100	Cariprazine Hydrochloride	Cariprazine 4.5mg capsules
	35854111000001100	12665641000033100	Cariprazine Hydrochloride	Cariprazine 6mg capsules
	35901111000001100	26448410000331 10	Amisulpride	Amisulpride 100mg/ml oral solution sugar free
	35931111000001100	16984410000331 10	Sulpiride	Sulpiride 200mg/5ml oral solution sugar free
	35931311000001100	11296410000331 10	Promazine Hydrochloride	Promazine 25mg/5ml oral solution
	36027911000001100	10635410000331 10	Pericyazine	Pericyazine 10mg/5ml oral solution
	36053411000001100	65964100003311 1	Haloperidol,Haloperidol	Haloperidol 20mg/2ml solution for injection ampoules
	36053511000001100	29297410000331 10	Haloperidol,Haloperidol	Haloperidol 5mg/1ml solution for injection ampoules
	36137911000001100	24664100003311 8	Chlorpromazine Hydrochloride	Chlorpromazine 100mg/5ml oral suspension sugar free

	36138211000001100	3858941000033110	Chlorpromazine Hydrochloride,Chlorpromazine Hydrochloride	Chlorpromazine 25mg/1ml solution for injection ampoules
	36138311000001100	2400410000331 10	Chlorpromazine Hydrochloride	Chlorpromazine 25mg/5ml oral solution
	36138411000001100	2413410000331 16	Chlorpromazine Hydrochloride	Chlorpromazine 50mg/2ml solution for injection ampoules
	36143111000001100	1466941000033110	Trifluoperazine	Trifluoperazine 10mg modified-release capsules
	36143211000001100	1467041000033110	Trifluoperazine	Trifluoperazine 15mg modified-release capsules
	36143311000001100	1467641000033110	Trifluoperazine	Trifluoperazine 1mg/5ml oral solution sugar free
	36143511000001100	1467141000033110	Trifluoperazine	Trifluoperazine 2mg modified-release capsules
	36151811000001100	1563441000033110	Zuclopenthixol Acetate	Zuclopenthixol acetate 100mg/2ml solution for injection ampoules
	36152011000001100	1563541000033110	Zuclopenthixol Acetate	Zuclopenthixol acetate 50mg/1ml solution for injection ampoules
	36796110000011000	4317410000331 18	Flupentixol	Depixol 3mg tablets (Lundbeck Ltd)
	36796110000011000	4317410000331 18	Flupentixol	Depixol 3mg tablets (Lundbeck Ltd)
	36803110000011000	5709410000331 17	Perphenazine	Fentazin 2mg tablets (Advanz Pharma)
	36809110000011000	5710410000331 10	Perphenazine	Fentazin 4mg tablets (Advanz Pharma)
	36861110000011000	9350410000331 15	Fluphenazine Hydrochloride	Moditen 1mg tablets (Sanofi)
	36868411000001100	1299034100003 3100	Aripiprazole	Arpoya 5mg tablets (Torrent Pharma (UK) Ltd)

	36868711000001100	12990441000033100	Aripiprazole	Arpoya 10mg tablets (Torrent Pharma (UK) Ltd)
	36869111000001100	12990541000033100	Aripiprazole	Arpoya 15mg tablets (Torrent Pharma (UK) Ltd)
	36869311000001100	12990641000033100	Aripiprazole	Arpoya 30mg tablets (Torrent Pharma (UK) Ltd)
	36880110000011100	10165410000033110	Pimozide	Orap 4mg tablets (Eumedica Pharma Ltd)
	36883110000011100	28464410000033110	Benperidol	Benquil 250mi-crogram tablets (Concord Phar-maceuticals Ltd)
	36905011000001100	12988741000033100	Olanzapine	Olanzapine 2.5mg orodis-persible tablets sugar free
	36905111000001100	12988841000033100	Olanzapine	Olanzapine 7.5mg orodis-persible tablets sugar free
	37249011000001100	13353241000033100	Aripiprazole	Aripiprazole 1mg/ml oral solu-tion sugar free
	37687211000001100	13158941000033100	Clozapine	Zaponex 12.5mg orodispersible tablets (Leyden Delta B.V.)
	37687511000001100	13159141000033100	Clozapine	Zaponex 25mg orodispersible tablets (Leyden Delta B.V.)
	37687811000001100	13159241000033100	Clozapine	Zaponex 50mg orodispersible tablets (Leyden Delta B.V.)
	37688211000001100	13158841000033100	Clozapine	Zaponex 100mg orodispersible tablets (Leyden Delta B.V.)
	37688811000001100	13159041000033100	Clozapine	Zaponex 200mg orodispersible tablets (Leyden Delta B.V.)
	37697511000001100	13158341000033100	Clozapine	Clozapine 100mg orodispersible tablets sugar free
	37697711000001100	13158441000033100	Clozapine	Clozapine 12.5mg orodispersible tablets sugar free

	37697811000001100	13158541000033100	Clozapine	Clozapine 200mg orodispersible tablets sugar free
	37697911000001100	13158641000033100	Clozapine	Clozapine 25mg orodispersible tablets sugar free
	37698011000001100	13158741000033100	Clozapine	Clozapine 50mg orodispersible tablets sugar free
	379911000001101	813741000033119	Chlorpromazine Hydrochloride	Largactil 25mg/5ml syrup (Sanofi)
	382571100000110 00	935141000033116	Fluphenazine Hydrochloride	Moditen 2.5mg tablets (Sanofi)
	382601100000110 00	935241000033111	Fluphenazine Hydrochloride	Moditen 5mg tablets (Sanofi-Synthelabo Ltd)
	383031100000110 00	284941000033113	Clozapine	Clozaril 25mg tablets (Viatris UK Healthcare Ltd)
	383201100000110 00	285041000033113	Clozapine	Clozaril 100mg tablets (Viatris UK Healthcare Ltd)
	38696711000001100	13581641000033100	Risperidone	Risperidone 250microgram tablets
	387281100000110 00	842041000033117	Loxapine Succinate	Loxapac 10mg capsules (Wyeth Pharmaceuticals)
	387311000001101	1276141000033110	Haloperidol	Serenace 20mg tablets (Teva UK Ltd)
	387361100000110 00	842241000033113	Loxapine Succinate	Loxapac 50mg capsules (Wyeth Pharmaceuticals)
	38750711000001100	1464141000033110	Trifluoperazine	Trifluoperazine 5mg/5ml oral solution sugar free
	38753511000001100	2640141000033110	Promazine Hydrochloride	Promazine 50mg/5ml oral solution
	38895311000001100	3063441000033110	Levomepromazine Hydrochloride,Levomepromazine Hydrochloride,Levomepromazine Hydrochloride	Levomepromazine 25mg/1ml solution for injection ampoules
	38931811000001100	13719041000033100	Levomepromazine Maleate	Levomepromazine 50mg tablets

	39005011000001100	13734941000033100	Quetiapine Fumarate	Biquelle XL 600mg tablets (Aspire Pharma Ltd)
	39027411000001100	13734841000033100	Quetiapine Fumarate	Quetiapine 600mg modified-release tablets
	39048110000011100	17658410000331 10	Zotepine	Zoleptil 25 tablets (Movianto UK Ltd)
	39056110000011100	17659410000331 10	Zotepine	Zoleptil 50 tablets (Movianto UK Ltd)
	39125711000001100	13745541000033100	Haloperidol	Haloperidol 5mg/5ml oral solution sugar free (Special Order)
	39148811000001100	13748441000033100	Trifluoperazine	Trifluoperazine 1mg/5ml oral suspension (Special Order)
	39275110000011100	12715410000331 10	Haloperidol,Haloperidol	Serenace 20mg/2ml solution for injection ampoules (IVAX Pharmaceuticals UK Ltd)
	39338110000011100	26449410000331 10	Amisulpride	Solian 100mg/ml oral solution (Sanofi)
	39692411000001100	30912410000331 10	Flupentixol	Flupentixol 3mg tablets
	39699511000001100	84074100003311 5	Loxapine Succinate	Loxapine 10mg capsules
	39699611000001100	84084100003311 3	Loxapine Succinate	Loxapine 25mg capsules
	39699711000001100	84094100003311 7	Loxapine Succinate	Loxapine 50mg capsules
	39703111000001100	18446410000331 10	Quetiapine Fumarate	Quetiapine 150mg tablets
	39706011000001100	14706410000331 10	Trifluoperazine	Trifluoperazine 1mg tablets
	39706211000001100	14707410000331 10	Trifluoperazine	Trifluoperazine 5mg tablets
	39707711000001100	15640410000331 10	Zuclopenthixol Dihydrochloride	Zuclopenthixol 10mg tablets
	39707911000001100	15641410000331 10	Zuclopenthixol Dihydrochloride	Zuclopenthixol 2mg tablets
	39708111000001100	15642410000331 10	Zuclopenthixol Dihydrochloride	Zuclopenthixol 25mg tablets
	40230110000011100	84214100003311 8	Loxapine Succinate	Loxapac 25mg capsules (Wyeth Pharmaceuticals)

	4039211000001100	2869041000033110	Risperidone	Risperdal Quicklet 1mg orodispersible tablets (Janssen-Cilag Ltd)
	4039711000001100	2869141000033110	Risperidone	Risperdal Quicklet 2mg orodispersible tablets (Janssen-Cilag Ltd)
	4054011000001100	2868841000033110	Risperidone	Risperidone 1mg orodispersible tablets sugar free
	4054111000001100	2868941000033110	Risperidone	Risperidone 2mg orodispersible tablets sugar free
	4088471100000100	1406344100003100	Aripiprazole	Aripiprazole 10mg tablets (Imported)
	4111000001106	1844741000033110	Quetiapine Fumarate	Seroquel 150mg tablets (AstraZeneca UK Ltd)
	4124511000001100	1765741000033110	Zotepine	Zoleptil 100 tablets (Movianto UK Ltd)
	4129911000001100	2736410000331 4	Zuclopenthixol Acetate	Clopixol Acuphase 50mg/1ml solution for injection ampoules (Lundbeck Ltd)
	4130411000001100	2748410000331 2	Zuclopenthixol Acetate	Clopixol Acuphase 100mg/2ml solution for injection ampoules (Lundbeck Ltd)
	4144281100000100	1417444100003100	Levomepromazine Hydrochloride	Levorol 5mg/ml oral solution (Galvany Pharma Ltd)
	4148241100000100	1417434100003100	Levomepromazine Hydrochloride	Levomepromazine hydrochloride 25mg/5ml oral solution sugar free
	4164110000011 01	4798410000331 2	Sulpiride	Dolmatil 400mg tablets (Sanofi)
	4182911000001100	9816410000331 7	Levomepromazine Hydrochloride,Levomepromazine Hy-	Nozinan 25mg/1ml solution for injection ampoules (Neuraxpharm UK Ltd)

			drochloride,Levomepromazine Hydrochloride	
	42267611000001100	2835141000033110	Amisulpride	Amisulpride 100mg tablets
	42267711000001100	58941000033117	Amisulpride	Amisulpride 200mg tablets
	42267811000001100	2216641000033110	Amisulpride	Amisulpride 400mg tablets
	42267911000001100	59041000033114	Amisulpride	Amisulpride 50mg tablets
	42268111000001100	3137541000033110	Aripiprazole	Aripiprazole 10mg tablets
	42268211000001100	3137641000033110	Aripiprazole	Aripiprazole 15mg tablets
	42268311000001100	3950541000033110	Aripiprazole	Aripiprazole 1mg/ml oral solution
	42268411000001100	3137741000033110	Aripiprazole	Aripiprazole 30mg tablets
	42268511000001100	3225241000033110	Aripiprazole	Aripiprazole 5mg tablets
	42268611000001100	145141000033119	Benperidol	Benperidol 250microgram tablets
	42268711000001100	1266764100003100	Chlorpromazine	Chlorpromazine 100mg suppositories
	42268811000001100	2491410000331 18	Chlorpromazine Hydrochloride	Chlorpromazine 100mg tablets
	42268911000001100	2490410000331 17	Chlorpromazine Hydrochloride	Chlorpromazine 10mg tablets
	42269011000001100	8885941000033110	Chlorpromazine	Chlorpromazine 25mg suppositories
	42269111000001100	2492410000331 13	Chlorpromazine Hydrochloride	Chlorpromazine 25mg tablets
	42269211000001100	2493410000331 15	Chlorpromazine Hydrochloride	Chlorpromazine 50mg tablets
	42270011000001100	2875410000331 12	Clozapine	Clozapine 100mg tablets
	42270111000001100	2876410000331 13	Clozapine	Clozapine 25mg tablets
	42270211000001100	5078041000033110	Clozapine	Clozapine 50mg tablets
	42272111000001100	5954410000331 12	Fluphenazine Hydrochloride	Fluphenazine 1mg tablets
	42272211000001100	5956410000331 14	Fluphenazine Hydrochloride	Fluphenazine 2.5mg tablets
	42272311000001100	5957410000331 17	Fluphenazine Hydrochloride	Fluphenazine 5mg tablets
	42272611000001100	6623410000331 15	Haloperidol	Haloperidol 1.5mg tablets

	42272711000001100	663241000033118	Haloperidol	Haloperidol 10mg tablets
	42272811000001100	661341000033112	Haloperidol	Haloperidol 10mg/5ml oral solution sugar free
	42272911000001100	662441000033114	Haloperidol	Haloperidol 20mg tablets
	42273011000001100	658341000033110	Haloperidol	Haloperidol 500microgram capsules
	42273111000001100	663341000033111	Haloperidol	Haloperidol 500microgram tablets
	42273211000001100	662541000033110	Haloperidol	Haloperidol 5mg tablets
	42273311000001100	660341000033116	Haloperidol	Haloperidol 5mg/5ml oral solution sugar free
	42273511000001100	306354100003311 0	Levomepromazine Maleate	Levomepromazine 25mg tablets
	42275411000001100	207784100003311 0	Olanzapine,Olanzapine	Olanzapine 10mg oral lyophilisates sugar free
	42275511000001100	100364100003311 0	Olanzapine	Olanzapine 10mg tablets
	42275611000001100	275964100003311 0	Olanzapine	Olanzapine 15mg tablets
	42275711000001100	166114100003311 0	Olanzapine	Olanzapine 2.5mg tablets
	42275811000001100	386654100003311 0	Olanzapine,Olanzapine	Olanzapine 20mg oral lyophilisates sugar free
	42275911000001100	399624100003311 0	Olanzapine	Olanzapine 20mg tablets
	42276011000001100	207774100003311 0	Olanzapine,Olanzapine	Olanzapine 5mg oral lyophilisates sugar free
	42276111000001100	100374100003311 0	Olanzapine	Olanzapine 5mg tablets
	42276211000001100	100384100003311 0	Olanzapine	Olanzapine 7.5mg tablets
	42276311000001100	102274100003311 0	Oxypertine	Oxypertine 10mg capsules
	42276411000001100	102694100003311 0	Oxypertine	Oxypertine 40mg tablets
	42276811000001100	106674100003311 0	Pericyazine	Pericyazine 10mg tablets
	42276911000001100	106704100003311 0	Pericyazine	Pericyazine 2.5mg tablets
	42277011000001100	106514100003311 0	Perphenazine	Perphenazine 2mg tablets

	42277111000001100	10652410000033110	Perphenazine	Perphenazine 4mg tablets
	42277411000001100	10909410000033110	Pimozide	Pimozide 2mg tablets
	42277511000001100	10910410000033110	Pimozide	Pimozide 4mg tablets
	42277611000001100	11343410000033110	Promazine Hydrochloride	Promazine 25mg tablets
	42277711000001100	11344410000033110	Promazine Hydrochloride	Promazine 50mg tablets
	42278011000001100	11500410000033110	Quetiapine Fumarate	Quetiapine 100mg tablets
	42278111000001100	45918410000033110	Quetiapine Fumarate	Quetiapine 200mg modified-release tablets
	42278211000001100	11501410000033110	Quetiapine Fumarate	Quetiapine 200mg tablets
	42278311000001100	11502410000033110	Quetiapine Fumarate	Quetiapine 25mg tablets
	42278411000001100	45919410000033110	Quetiapine Fumarate	Quetiapine 300mg modified-release tablets
	42278511000001100	29758410000033110	Quetiapine Fumarate	Quetiapine 300mg tablets
	42278611000001100	45920410000033110	Quetiapine Fumarate	Quetiapine 400mg modified-release tablets
	42278711000001100	11766410000033110	Risperidone	Risperidone 1mg tablets
	42278811000001100	11739410000033110	Risperidone	Risperidone 1mg/ml oral solution sugar free
	42278911000001100	11767410000033110	Risperidone	Risperidone 2mg tablets
	42279011000001100	11768410000033110	Risperidone	Risperidone 3mg tablets
	42279111000001100	11769410000033110	Risperidone	Risperidone 4mg tablets
	42279211000001100	21882410000033110	Risperidone	Risperidone 500microgram tablets
	42279311000001100	11777410000033110	Risperidone	Risperidone 6mg tablets
	42279411000001100	13964410000033110	Sulpiride	Sulpiride 200mg tablets
	42279611000001100	13976410000033110	Sulpiride	Sulpiride 400mg tablets
	42279711000001100	14337410000033110	Thioridazine	Thioridazine 100mg tablets
	42279811000001100	14323410000033110	Thioridazine	Thioridazine 100mg/5ml oral suspension

	42279911000001100	1433641000033110	Thioridazine	Thioridazine 10mg tablets
	42280011000001100	1433841000033110	Thioridazine	Thioridazine 25mg tablets
	42280111000001100	1432541000033110	Thioridazine	Thioridazine 25mg/5ml oral solution
	42280211000001100	1432441000033110	Thioridazine	Thioridazine 25mg/5ml oral suspension
	42280311000001100	1433941000033110	Thioridazine	Thioridazine 50mg tablets
	42281411000001100	1766041000033110	Zotepine	Zotepine 100mg tablets
	42281511000001100	1766141000033110	Zotepine	Zotepine 25mg tablets
	42281611000001100	1766241000033110	Zotepine	Zotepine 50mg tablets
	42405911000001100	1425504100003310 0	Olanzapine	Xyquila 10mg powder for solution for injection vials (Eramol (UK) Ltd)
	42547611000001100	1427794100003310 0	Levomepromazine Maleate	Levorol 6.25mg tablets (Galvany Pharma Ltd)
	42565811000001100	1427784100003310 0	Levomepromazine Maleate	Levomepromazine 6.25mg tablets
	42684011000001100	1439784100003310 0	Quetiapine Fumarate	Quetiapine 400mg tablets
	42684111000001100	1439774100003310 0	Quetiapine Fumarate	Quetiapine 50mg tablets
	429611000001104	1176541000033110	Risperidone	Risperdal 4mg tablets (Janssen-Cilag Ltd)
	42986711000001100	1450624100003310 0	Olanzapine	Ceyxa 10mg powder for solution for injection vials (Nordic Pharma Ltd)
	43185011000001100	1451594100003310 0	Aripiprazole	Elozar 10mg orodispersible tablets (Novumgen Ltd)
	43185211000001100	1451604100003310 0	Aripiprazole	Elozar 15mg orodispersible tablets (Novumgen Ltd)
	432111000001100	8158410000331 13	Chlorpromazine Hydrochloride	Largactil 100mg tablets (Sanofi)

	437611000001106	2759941000033110	Olanzap-ine,Olanzapine	Zyprexa 15mg Velotabs (Neon Healthcare Ltd)
	440508110000001100	1485254100003310 0	Aripiprazole	Aripiprazole 1mg tablets
	440509110000001100	1485264100003310 0	Aripiprazole	Aripiprazole 2.5mg tablets
	44411000001109	4809410000331 16	Sulpiride	Dolmatil 200mg tablets (Sanofi)
	454491100000110 0	1016441000033110	Pimozide	Orap 2mg tablets (Janssen-Cilag Ltd)
	454761100000110 0	7575410000331 11	Oxypertine	Integrin 10mg capsules (Ster-ling-Winthrop)
	45811000001107	1386441000033110	Trifluoperazine	Stelazine 5mg tablets (Mercury Pharma Group Ltd)
	463911000001104	4765410000331 19	Haloperidol	Dozic 5mg/5ml oral solution (Rosemont Phar-maceuticals Ltd)
	467011000001109	1565241000033110	Olanzapine	Zyprexa 5mg tab-lets (Neon Healthcare Ltd)
	472961100000110 0	1277141000033110	Sertindole	Serdolect 4mg tablets (Imported (Denmark))
	473171100000110 0	1276841000033110	Sertindole	Serdolect 12mg tablets (Imported (Denmark))
	473201100000110 0	1276941000033110	Sertindole	Serdolect 16mg tablets (Imported (Denmark))
	473231100000110 0	1277041000033110	Sertindole	Serdolect 20mg tablets (Imported (Denmark))
	474821100000110 0	1277241000033110	Sertindole	Sertindole 12mg tablets
	474831100000110 0	1277341000033110	Sertindole	Sertindole 16mg tablets
	474841100000110 0	1277441000033110	Sertindole	Sertindole 20mg tablets
	474851100000110 0	1277541000033110	Sertindole	Sertindole 4mg tablets
	476411000001108	1277741000033110	Quetiapine Fumarate	Seroquel 200mg tablets (Luye Pharma Ltd)
	486611000001101	1565141000033110	Olanzapine	Zyprexa 10mg tablets (Neon Healthcare Ltd)

	4989111000001100	2975941000033110	Quetiapine Fumarate	Seroquel 300mg tablets (Luye Pharma Ltd)
	5026110000011007	1269741000033110	Haloperidol	Serenace 2mg/ml liquid (Teva UK Ltd)
	5075110000011006	5935410000331112	Flupentixol	Fluanxol 500microgram tablets (Lundbeck Ltd)
	5075110000011006	5935410000331112	Flupentixol	Fluanxol 500microgram tablets (Lundbeck Ltd)
	511000001103	8159410000331107	Chlorpromazine Hydrochloride	Largactil 25mg tablets (Sanofi)
	5111110000011006	2759841000033110	Olanzapine	Zyprexa 15mg tablets (Neon Healthcare Ltd)
	5189110000011000	1176441000033110	Risperidone	Risperdal 3mg tablets (Janssen-Cilag Ltd)
	5227110000011006	1268341000033110	Haloperidol	Serenace 500microgram capsules (Teva UK Ltd)
	5266110000011001	2188141000033110	Risperidone	Risperdal 500microgram tablets (Janssen-Cilag Ltd)
	5551110000011002	8921410000331104	Thioridazine	Melleril 100mg/5ml oral suspension (Novartis Pharmaceuticals UK Ltd)
	5671110000011003	2955641000033110	Trifluoperazine	Terrazine 5mg tablets (Dr Reddy's Laboratories (UK) Ltd)
	5694111000001100	3030941000033110	Olanzapine	Zyprexa 10mg powder for solution for injection vials (Nordic Pharma Ltd)
	5818311000001100	3030841000033110	Olanzapine	Olanzapine 10mg powder for solution for injection vials
	5871100000110 8	8922410000331119	Thioridazine	Melleril 25mg/5ml oral suspension (Novartis Pharmaceuticals UK Ltd)
	5899110000011005	1176241000033110	Risperidone	Risperdal 1mg tablets (Janssen-Cilag Ltd)

	596211000001109	1277841000033110	Quetiapine Fumarate	Seroquel 25mg tablets (Luye Pharma Ltd)
	618211000001102	2871410000331 15	Zuclopenthixol Dihydrochloride	Clopixol 25mg tablets (Lundbeck Ltd)
	636711000001103	1385941000033110	Trifluoperazine	Stelazine 1mg/5ml syrup (Mercury Pharma Group Ltd)
	701911000001108	1512741000033110	Levomepromazine Maleate	Nozinan 25mg tablets (Neuraxpharm UK Ltd)
	710411000001106	1385441000033110	Trifluoperazine	Stelazine 15mg Spansules (Mercury Pharma Group Ltd)
	711411000001102	8958410000331 4	Thioridazine	Melleril 10mg tablets (Novartis Pharmaceuticals UK Ltd)
	714111000001105	1385541000033110	Trifluoperazine	Stelazine 2mg Spansules (Mercury Pharma Group Ltd)
	746611000001104	9656410000331 8	Pericyazine	Neulactil 2.5mg tablets (Sanofi)
	763711100000110 00	3137841000033110	Aripiprazole	Abilify 10mg tablets (Otsuka Pharmaceuticals (U.K.) Ltd)
	763741100000110 00	3137941000033110	Aripiprazole	Abilify 15mg tablets (Otsuka Pharmaceuticals (U.K.) Ltd)
	763781100000110 00	3138041000033110	Aripiprazole	Abilify 30mg tablets (Otsuka Pharmaceuticals (U.K.) Ltd)
	774051100000110 00	3141141000033110	Trifluoperazine	Stelazine Forte 1mg/ml oral solution (Mercury Pharma Group Ltd)
	776911000001109	1565341000033110	Olanzapine	Zyprexa 7.5mg tablets (Neon Healthcare Ltd)
	788511000001105	8937410000331 15	Thioridazine	Melleril 25mg/5ml syrup (Novartis Pharmaceuticals UK Ltd)

	79729110000011 00	31589410000331 10	Clozapine	Denzapine 25mg tablets (Britannia Pharmaceuticals Ltd)
	79740110000011 00	31590410000331 10	Clozapine	Denzapine 100mg tablets (Britannia Pharmaceuticals Ltd)
	82751100000110 2	28694100003311 5	Zuclopenthixol Dihydrochloride	Clopixol 10mg tablets (Lundbeck Ltd)
	82772110000011 00	49401410000331 10	Amisulpride	Amisulpride 25mg/5ml oral solution
	82773110000011 00	58909410000331 10	Amisulpride	Amisulpride 25mg/5ml oral suspension
	83594110000011 00	17193410000331 10	Chlorpromazine Hydrochloride	Chlorpromazine 100mg/5ml oral solution
	83597110000011 00	38494410000331 10	Chlorpromazine Hydrochloride	Chlorpromazine 50mg/5ml oral solution
	84301100000110 1	59344100003311 1	Flupentixol	Fluanxol 1mg tablets (Lundbeck Ltd)
	84301100000110 1	59344100003311 1	Flupentixol	Fluanxol 1mg tablets (Lundbeck Ltd)
	85279110000011 00	50546410000331 10	Haloperidol	Haloperidol 1mg/5ml oral solution
	85280110000011 00	58907410000331 10	Haloperidol	Haloperidol 1mg/5ml oral suspension
	85283110000011 00	10493341000033 100	Haloperidol	Haloperidol 500micrograms/5ml oral solution
	85284110000011 00	10493441000033 100	Haloperidol	Haloperidol 500micrograms/5ml oral suspension
	85828110000011 00	58143410000331 10	Levomepromazine Maleate	Levomepromazine 2.5mg/5ml oral suspension
	85891100000110 9	11763410000331 10	Risperidone	Risperdal 2mg tablets (Janssen-Cilag Ltd)
	86951100000110 4	81104100003311 6	Chlorpromazine Hydrochloride	Largactil 50mg/2ml solution for injection

				ampoules (Sanofi)
	8838911000001100	3226341000033110	Clozapine	Zaponex 25mg tablets (Leyden Delta B.V.)
	8839110000011106	8961410000331 10	Thioridazine	Melleril 50mg tablets (Novartis Pharmaceuticals UK Ltd)
	8839511000001100	3226441000033110	Clozapine	Zaponex 100mg tablets (Leyden Delta B.V.)
	8845110000011101	8959410000331 18	Thioridazine	Melleril 100mg tablets (Novartis Pharmaceuticals UK Ltd)
	8971110000011104	2078041000033110	Olanzap-ine,Olanzapine	Zyprexa 10mg Velotabs (Neon Healthcare Ltd)
	8998110000011103	1275941000033110	Haloperidol	Serenace 1.5mg tablets (Teva UK Ltd)
	9037911000001100	3225341000033110	Aripiprazole	Abilify 5mg tab-lets (Otsuka Phar-maceuticals (U.K.) Ltd)
	9044211000001100	3246141000033110	Risperidone	Risperdal Quick-let 500microgram orodispersible tablets (Janssen-Cilag Ltd)
	9045911000001100	3246041000033110	Risperidone	Risperidone 500microgram orodispersible tablets sugar free
	9106110000011102	8160410000331 0	Chlorpromazine Hydrochloride	Largactil 50mg tablets (Sanofi)
	9113110000011102	8157410000331 5	Chlorpromazine Hydrochloride	Largactil 10mg tablets (Sanofi)
	91711000001108	2077941000033110	Olanzap-ine,Olanzapine	Zyprexa 5mg Ve-lotabs (Neon Healthcare Ltd)
	9178110000011104	9650410000331 2	Pericyazine	Neulactil Forte syrup (Sanofi)
	9483411000001100	69041000033115	Benperidol	Anquil 250mi-crogram tablets (Neon Healthcare Ltd)
	9536311000001100	3063641000033110	Levomeproma-zine Maleate	Levomeproma-zine 6mg tablets
	9891011000001100	3866641000033110	Olanzap-ine,Olanzapine	Zyprexa 20mg Velotabs (Neon Healthcare Ltd)

	119311000001100	925341000033116	Sesame Oil,Fluphenazine Decanoate	Modecate 50mg/2ml solution for injection ampoules (Sanofi)
	153511000001100	924341000033117	Fluphenazine Decanoate	Modecate 50mg/2ml solution for injection pre-filled syringes (Sanofi-Synthelabo Ltd)
	170376110000011100	557574100003311 0	Olanzapine Pamoate Monohydrate	Olanzapine embonate 210mg powder and solvent for suspension for injection vials
	170378110000011100	556504100003311 0	Olanzapine Pamoate Monohydrate	Zypadhera 210mg powder and solvent for suspension for injection vials (Neon Healthcare Ltd)
	170382110000011100	557584100003311 0	Olanzapine Pamoate Monohydrate	Olanzapine embonate 300mg powder and solvent for suspension for injection vials
	170384110000011100	556514100003311 0	Olanzapine Pamoate Monohydrate	Zypadhera 300mg powder and solvent for suspension for injection vials (Neon Healthcare Ltd)
	170387110000011100	557594100003311 0	Olanzapine Pamoate Monohydrate	Olanzapine embonate 405mg powder and solvent for suspension for injection vials
	170390110000011100	556524100003311 0	Olanzapine Pamoate Monohydrate	Zypadhera 405mg powder and solvent for suspension for injection vials (Neon Healthcare Ltd)
	178611000001106	276041000033113	Zuclopenthixol Decanoate	Clopixol 200mg/1ml solution for injection

				ampoules (Lundbeck Ltd)
	19208911000001100	6359841000033110	Paliperidone	Xeplion 50mg/0.5ml prolonged-release suspension for injection pre-filled syringes (Janssen-Cilag Ltd)
	19210011000001100	6359941000033110	Paliperidone	Xeplion 75mg/0.75ml prolonged-release suspension for injection pre-filled syringes (Janssen-Cilag Ltd)
	19210611000001100	6360041000033110	Paliperidone	Xeplion 100mg/1ml prolonged-release suspension for injection pre-filled syringes (Janssen-Cilag Ltd)
	19210911000001100	6360141000033110	Paliperidone	Xeplion 150mg/1.5ml prolonged-release suspension for injection pre-filled syringes (Janssen-Cilag Ltd)
	19223511000001100	6359641000033110	Paliperidone	Paliperidone 100mg/1ml prolonged-release suspension for injection pre-filled syringes
	19223611000001100	6359741000033110	Paliperidone	Paliperidone 150mg/1.5ml prolonged-release suspension for injection pre-filled syringes
	19223711000001100	6359441000033110	Paliperidone	Paliperidone 50mg/0.5ml prolonged-release suspension for injection pre-filled syringes

	19223811000001100	6359541000033110	Paliperidone	Paliperidone 75mg/0.75ml prolonged-release suspension for injection pre-filled syringes
	23636811000001100	8960441000033110	Aripiprazole	Abilify Maintena 400mg powder and solvent for prolonged-release suspension for injection vials (Otsuka Pharmaceuticals (U.K.) Ltd)
	23661211000001100	8960241000033110	Aripiprazole	Aripiprazole 400mg powder and solvent for prolonged-release suspension for injection vials
	31007110000011100	4204410000331 9	Flupentixol Decanoate	Depixol Conc 50mg/0.5ml solution for injection ampoules (Lundbeck Ltd)
	31013110000011100	4230410000331 5	Flupentixol Decanoate	Depixol Conc 100mg/1ml solution for injection ampoules (Lundbeck Ltd)
	31021110000011100	4201410000331 0	Flupentixol Decanoate	Depixol 40mg/2ml solution for injection ampoules (Lundbeck Ltd)
	31029110000011100	4229410000331 3	Flupentixol Decanoate	Depixol 20mg/1ml solution for injection ampoules (Lundbeck Ltd)
	31040110000011100	4199410000331 3	Flupentixol Decanoate	Depixol Low Volume 200mg/1ml solution for injection ampoules (Lundbeck Ltd)
	31078011000001100	1094644100003 3100	Aripiprazole	Abilify Maintena 400mg powder and solvent for prolonged-release suspension for injection pre-filled syringes

				(Otsuka Pharma-ceuticals (U.K.) Ltd)
	31087211000001100	10946341000033100	Aripiprazole	Aripiprazole 400mg powder and solvent for prolonged-release suspension for injection pre-filled syringes
	32866911000001100	11694741000033100	Paliperidone	Trevicta 175mg/0.875ml prolonged-release suspension for injection pre-filled syringes (Janssen-Cilag Ltd)
	32867811000001100	11694841000033100	Paliperidone	Trevicta 263mg/1.315ml prolonged-release suspension for injection pre-filled syringes (Janssen-Cilag Ltd)
	32868111000001100	11694941000033100	Paliperidone	Trevicta 350mg/1.75ml prolonged-release suspension for injection pre-filled syringes (Janssen-Cilag Ltd)
	32868411000001100	11695041000033100	Paliperidone	Trevicta 525mg/2.625ml prolonged-release suspension for injection pre-filled syringes (Janssen-Cilag Ltd)
	32878911000001100	11694341000033100	Paliperidone	Paliperidone 175mg/0.875ml prolonged-release suspension for injection pre-filled syringes
	32879011000001100	11694441000033100	Paliperidone	Paliperidone 263mg/1.315ml prolonged-release suspension

				for injection pre-filled syringes
	32879111000001100	11694541000033100	Paliperidone	Paliperidone 350mg/1.75ml prolonged-release suspension for injection pre-filled syringes
	32879211000001100	11694641000033100	Paliperidone	Paliperidone 525mg/2.625ml prolonged-release suspension for injection pre-filled syringes
	32911000001106	276141000033112	Zuclopenthixol Decanoate	Clopixol Conc 500mg/1ml solution for injection ampoules (Lundbeck Ltd)
	36016711000001100	3103841000033110	Pipotiazine Palmitate	Pipotiazine 100mg/2ml solution for injection ampoules
	36016911000001100	3104041000033110	Pipotiazine Palmitate	Pipotiazine 50mg/1ml solution for injection ampoules
	36053611000001100	659741000033119	Haloperidol Decanoate	Haloperidol decanoate 100mg/1ml solution for injection ampoules
	36053711000001100	659841000033112	Haloperidol Decanoate	Haloperidol decanoate 50mg/1ml solution for injection ampoules
	36058011000001100	3090541000033110	Flupentixol Decanoate	Flupentixol 50mg/0.5ml solution for injection ampoules
	36058111000001100	587741000033113	Fluphenazine Decanoate	Fluphenazine decanoate 100mg/1ml solution for injection ampoules
	36058211000001100	586641000033119	Fluphenazine Decanoate	Fluphenazine decanoate 12.5mg/0.5ml solution for injection ampoules
	36058311000001100	586741000033111	Fluphenazine Decanoate	Fluphenazine decanoate

				250mg/10ml solution for injection vials
	36058411000001100	58464100003311 1	Fluphenazine Decanoate	Fluphenazine decanoate 25mg/1ml solution for injection ampoules
	36058511000001100	58204100003311 2	Fluphenazine Decanoate	Fluphenazine decanoate 25mg/1ml solution for injection pre-filled syringes
	36058611000001100	58694100003311 4	Fluphenazine Decanoate	Fluphenazine decanoate 50mg/0.5ml solution for injection ampoules
	36058711000001100	58684100003311 8	Fluphenazine Decanoate	Fluphenazine decanoate 50mg/2ml solution for injection ampoules
	36058811000001100	58214100003311 1	Fluphenazine Decanoate	Fluphenazine decanoate 50mg/2ml solution for injection pre-filled syringes
	36068611000001100	30906410000331 10	Flupentixol Decanoate	Flupentixol 100mg/1ml solution for injection ampoules
	36068711000001100	30907410000331 10	Flupentixol Decanoate	Flupentixol 200mg/1ml solution for injection ampoules
	36068811000001100	30908410000331 10	Flupentixol Decanoate	Flupentixol 20mg/1ml solution for injection ampoules
	36068911000001100	30909410000331 10	Flupentixol Decanoate	Flupentixol 40mg/2ml solution for injection ampoules
	36152111000001100	15636410000331 10	Zuclopenthixol Decanoate	Zuclopenthixol decanoate 200mg/1ml solution for injection ampoules
	36152211000001100	15637410000331 10	Zuclopenthixol Decanoate	Zuclopenthixol decanoate

				500mg/1ml solution for injection ampoules
	3747311000001100	659541000033110	Haloperidol,Sesame Oil	Haldol Decanoate 50mg/1ml solution for injection ampoules (Essential Pharma Ltd)
	3748011000001100	659441000033114	Haloperidol,Sesame Oil	Haldol Decanoate 100mg/1ml solution for injection ampoules (Essential Pharma Ltd)
	3822011000001100	108674100003311 0	Pipotiazine Palmitate,Sesame Oil	Piportil Depot 50mg/1ml solution for injection ampoules (Sanofi)
	3913221100000100	137466410000331 00	Fluphenazine Decanoate	Modecate 25mg/1ml solution for injection ampoules (Imported (Germany))
	3913241100000100	137465410000331 00	Fluphenazine Decanoate	Modecate 25mg/1ml solution for injection ampoules (Imported (Denmark))
	3913281100000100	137467410000331 00	Fluphenazine Decanoate	Modecate Concentrate 100mg/1ml solution for injection ampoules (Imported (Germany))
	4061721100000100	140241410000331 00	Paliperidone	Byannli 1000mg/5ml prolonged-release suspension for injection pre-filled syringes (Janssen-Cilag Ltd)
	4061751100000100	140242410000331 00	Paliperidone	Byannli 700mg/3.5ml prolonged-release suspension for injection pre-filled syringes (Janssen-Cilag Ltd)

	40617811000001100	14023941000033100	Paliperidone	Paliperidone 1g/5ml prolonged-release suspension for injection pre-filled syringes
	40617911000001100	14024041000033100	Paliperidone	Paliperidone 700mg/3.5ml prolonged-release suspension for injection pre-filled syringes
	40625611000001100	14025841000033100	Risperidone	Okedi 75mg powder and solvent for prolonged-release suspension for injection pre-filled syringes (ROVI Biotech Ltd)
	40625911000001100	14025941000033100	Risperidone	Okedi 100mg powder and solvent for prolonged-release suspension for injection pre-filled syringes (ROVI Biotech Ltd)
	40627211000001100	14025741000033100	Risperidone	Risperidone 100mg powder and solvent for prolonged-release suspension for injection pre-filled syringes
	40627311000001100	14025641000033100	Risperidone	Risperidone 75mg powder and solvent for prolonged-release suspension for injection pre-filled syringes
	41667110000011100	27799410000033110	Risperidone	Risperdal Consta 37.5mg powder and solvent for prolonged-release suspension for injection vials (Janssen-Cilag Ltd)
	41670110000011100	27800410000033110	Risperidone	Risperdal Consta 50mg powder and solvent for

				prolonged-release suspension for injection vials (Janssen-Cilag Ltd)
	41769110000011100	27802410000331 10	Risperidone	Risperidone 37.5mg powder and solvent for prolonged-release suspension for injection vials
	41770110000011100	27803410000331 10	Risperidone	Risperidone 50mg powder and solvent for prolonged-release suspension for injection vials
	42799711000001100	14416841000033100	Aripiprazole	Abilify Maintena 960mg/3.2ml prolonged-release suspension for injection pre-filled syringes (Otsuka Pharmaceuticals (U.K.) Ltd)
	42826411000001100	14416741000033100	Aripiprazole	Aripiprazole 960mg/3.2ml prolonged-release suspension for injection pre-filled syringes
	42878811000001100	14429541000033100	Aripiprazole	Abilify Maintena 720mg/2.4ml prolonged-release suspension for injection pre-filled syringes (Otsuka Pharmaceuticals (U.K.) Ltd)
	42880111000001100	14429441000033100	Aripiprazole	Aripiprazole 720mg/2.4ml prolonged-release suspension for injection pre-filled syringes
	45207110000011100	27798410000331 10	Risperidone	Risperdal Consta 25mg powder and solvent for prolonged-release suspension for injection vials

				(Janssen-Cilag Ltd)
	4559111000001100	2780141000033110	Risperidone	Risperidone 25mg powder and solvent for prolonged-release suspension for injection vials
	5561110000011008	9348410000331111	Fluphenazine Decanoate,Sesame Oil	Modecate 25mg/1ml solution for injection ampoules (Sanofi)
	5652110000011011	9259410000331117	Sesame Oil,Fluphenazine Decanoate	Modecate Concentrate 100mg/1ml solution for injection ampoules (Sanofi)
	66211000001107	9242410000331110	Fluphenazine Decanoate	Modecate 25mg/1ml solution for injection pre-filled syringes (Sanofi-Synthelabo Ltd)
	7903110000011000	9258410000331113	Fluphenazine Decanoate,Sesame Oil	Modecate 12.5mg/0.5ml solution for injection ampoules (Sanofi)
	8378110000011002	9278410000331117	Fluphenazine Decanoate,Sesame Oil	Modecate Concentrate 50mg/0.5ml solution for injection ampoules (Sanofi)
	8823211000001100	1087341000033110	Sesame Oil,Pipotiazine Palmitate	Piportil Depot 100mg/2ml solution for injection ampoules (Sanofi)
	9284110000011007	9277410000331110	Fluphenazine Decanoate	Modecate 250mg/10ml solution for injection vials (Sanofi-Synthelabo Ltd)
	9490711000001100	8540041000033110	Flupentixol Decanoate	Psytixol 50mg/0.5ml solution for injection ampoules (Viatris UK Healthcare Ltd)

4.2.3 Lithium Salts	19496311000001100	6443341000033110	Asenapine	Sycrest 10mg sublingual tablets (Organon Pharma (UK) Ltd)
	19496611000001100	6443241000033110	Asenapine	Sycrest 5mg sublingual tablets (Organon Pharma (UK) Ltd)
	19504311000001100	6443141000033110	Asenapine	Asenapine 10mg sublingual tablets sugar free
	19504411000001100	6443041000033110	Asenapine	Asenapine 5mg sublingual tablets sugar free
	250211000001105	2155410000331 7	Lithium Carbonate	Camcolit 400 modified-release tablets (Essential Pharma Ltd)
	30760011000001100	1067124100003 3100	Lithium Carbonate	Lithium carbonate 250mg tablets (Essential Pharma Ltd)
	36040211000001100	8376410000331 4	Lithium Carbonate	Lithium carbonate 450mg modified-release tablets
	36040311000001100	8329410000331 7	Lithium Citrate	Lithium citrate 520mg/5ml oral solution sugar free
	37573711000001100	1314954100003 3100	Valproic Acid (as semisodium valproate)	Belvo 250mg gastro-resistant tablets (Consilient Health Ltd)
	37573711000001100	1314954100003 3100	Valproic Acid (as semisodium valproate)	Belvo 250mg gastro-resistant tablets (Consilient Health Ltd)
	37574111000001100	1314964100003 3100	Valproic Acid (as semisodium valproate)	Belvo 500mg gastro-resistant tablets (Consilient Health Ltd)
	37574111000001100	1314964100003 3100	Valproic Acid (as semisodium valproate)	Belvo 500mg gastro-resistant tablets (Consilient Health Ltd)
	37982211000001100	1334954100003 3100	Valproic Acid (as semisodium valproate)	Syonell 500mg gastro-resistant tablets (Lupin Healthcare (UK) Ltd)

	37982211000001100	13349541000033100	Valproic Acid (as semisodium valproate)	Syonell 500mg gastro-resistant tablets (Lupin Healthcare (UK) Ltd)
	37982411000001100	13348941000033100	Valproic Acid (as semisodium valproate)	Syonell 250mg gastro-resistant tablets (Lupin Healthcare (UK) Ltd)
	37982411000001100	13348941000033100	Valproic Acid (as semisodium valproate)	Syonell 250mg gastro-resistant tablets (Lupin Healthcare (UK) Ltd)
	38323110000011100	11238410000331 10	Lithium Citrate	Priadel 520mg/5ml liquid (Essential Pharma Ltd)
	38464110000011100	22714410000331 10	Valproic Acid (as semisodium valproate)	Depakote 250mg gastro-resistant tablets (Sanofi)
	38464110000011100	22714410000331 10	Valproic Acid (as semisodium valproate)	Depakote 250mg gastro-resistant tablets (Sanofi)
	38471110000011100	22715410000331 10	Valproic Acid (as semisodium valproate)	Depakote 500mg gastro-resistant tablets (Sanofi)
	38471110000011100	22715410000331 10	Valproic Acid (as semisodium valproate)	Depakote 500mg gastro-resistant tablets (Sanofi)
	39110011000001100	83744100003311 2	Lithium Car-bonate	Lithium car-bonate 200mg modified-release tablets
	39110111000001100	83594100003311 0	Lithium Car-bonate	Lithium car-bonate 400mg modified-release tablets
	39209110000011100	32584100003311 9	Valproic Acid	Convulex 150mg gastro-resistant capsules (G.L. Pharma UK Ltd)
	39209110000011100	32584100003311 9	Valproic Acid	Convulex 150mg gastro-resistant capsules (G.L. Pharma UK Ltd)
	39213110000011100	32594100003311 0	Valproic Acid	Convulex 300mg gastro-resistant capsules (G.L. Pharma UK Ltd)
	39213110000011100	32594100003311 0	Valproic Acid	Convulex 300mg gastro-resistant

				capsules (G.L. Pharma UK Ltd)
	3921711000001100	3260410000331 7	Valproic Acid	Convulex 500mg gastro-resistant capsules (G.L. Pharma UK Ltd)
	3921711000001100	3260410000331 7	Valproic Acid	Convulex 500mg gastro-resistant capsules (G.L. Pharma UK Ltd)
	4213801100000 1100	2711241000033 110	Valproic Acid (as semisodium valproate)	Valproic acid 250mg gastro-resistant tablets
	4213831100000 1100	2711341000033 110	Valproic Acid (as semisodium valproate)	Valproic acid 500mg gastro-resistant tablets
	4229571100000 1100	8358410000331 9	Lithium Carbonate	Lithium carbonate 250mg tablets
	4229581100000 1100	2761041000033 110	Lithium Citrate	Lithium citrate 1.018g/5ml oral solution
	4229591100000 1100	2760941000033 110	Lithium Citrate	Lithium citrate 509mg/5ml oral solution
	4543511000001100	8357410000331 2	Lithium Citrate	Litarex 564mg modified-release tablets (Actavis UK Ltd)
	4559411000001100	8377410000331 7	Lithium Citrate	Lithium citrate 564mg modified-release tablets
	4888611000001100	2761241000033 110	Lithium Citrate	Li-Liquid 1.018g/5ml oral solution (Rosemont Pharmaceuticals Ltd)
	5650110000011 06	1132941000033 110	Lithium Carbonate	Priadel 200mg modified-release tablets (Essential Pharma Ltd)
	7043110000011 05	8356410000331 5	Lithium Carbonate	Liskonum 450mg modified-release tablets (Teofarma S.r.l.)
	7663110000011 00	1133041000033 110	Lithium Carbonate	Priadel 400mg modified-release tablets (Essential Pharma Ltd)
	8788711000001100	2761141000033 110	Lithium Citrate	Li-Liquid 509mg/5ml oral

				solution (Rosemont Pharmaceuticals Ltd)
	879511000001105	213841000033116	Lithium Carbonate	Camcolit 250 tablets (Essential Pharma Ltd)
	101711000001102	1389441000033110	Trimipramine	Surmontil 50mg capsules (Sanofi)
	110443110000001100	401044100003311 0	Doxepin Hydrochloride,Doxepin Hydrochloride	Sinepin 25mg capsules (Marlborough Pharmaceuticals Ltd)
	110446110000001100	401054100003311 0	Doxepin Hydrochloride	Sinepin 50mg capsules (Marlborough Pharmaceuticals Ltd)
	119411000001107	702410000331 14	Clomipramine Hydrochloride	Anafranil SR 75mg tablets (Novartis Pharmaceuticals UK Ltd)
	125111000001109	938941000033114	Fluphenazine Hydrochloride,Nortriptyline	Motival 10mg/500microgram tablets (Sanofi)
	128127110000001100	599664100003311 0	Lofepramine	Lofepramine 70mg/5ml oral solution
	128128110000001100	599674100003311 0	Lofepramine	Lofepramine 70mg/5ml oral suspension
	138924110000001100	549014100003311 0	Amitriptyline Hydrochloride	Amitriptyline 10mg/5ml oral solution
	14911000001107	472141000033110	Dosulepin Hydrochloride	Dothapax 25 capsules (Ashbourne Pharmaceuticals Ltd)
	150006110000001100	110728410000331 00	Nortriptyline	Nortriptyline 10mg/5ml oral solution
	150007110000001100	122578410000331 00	Nortriptyline	Nortriptyline 10mg/5ml oral suspension
	164234110000001100	545414100003311 0	Imipramine Hydrochloride	Imipramine 25mg/5ml oral solution sugar free
	1811000001101	938741000033111	Trazodone Hydrochloride	Molipaxin 150mg tablets (Zentiva Pharma UK Ltd)

	208811000001101	1111541000033110	Dosulepin Hydrochloride	Prothiaden 25mg capsules (Teofarma S.r.l.)
	214311000001103	1449441000033110	Imipramine Hydrochloride	Tofranil 25mg/5ml syrup (Novartis Pharmaceuticals UK Ltd)
	259211000001103	1294541000033110	Doxepin	Sinequan 10mg capsules (Pfizer Ltd)
	276511000001108	1294841000033110	Doxepin	Sinequan 75mg capsules (Pfizer Ltd)
	296867110000001100	1033244100003300	Amitriptyline Hydrochloride	Amitriptyline 10mg/5ml oral solution sugar free
	329511000001107	1135941000033110	Dosulepin Hydrochloride	Prepadine 75mg tablets (Teva UK Ltd)
	340341110000001100	1229234100003300	Nortriptyline	Nortriptyline 50mg tablets
	351133110000001100	1247654100003300	Trazodone Hydrochloride	Trazodone 100mg/5ml oral solution sugar free
	359012110000001100	45841000033119	Amitriptyline Hydrochloride	Amitriptyline 25mg modified-release capsules
	359014110000001100	55441000033119	Amitriptyline Hydrochloride	Amitriptyline 25mg/5ml oral solution sugar free
	359015110000001100	45941000033110	Amitriptyline Hydrochloride	Amitriptyline 50mg modified-release capsules
	359016110000001100	55541000033118	Amitriptyline Hydrochloride	Amitriptyline 50mg/5ml oral solution sugar free
	360404110000001100	8492410000331 15	Lofepramine	Lofepramine 70mg/5ml oral suspension sugar free
	360465110000001100	5328741000033110	Imipramine Hydrochloride	Imipramine 25mg/5ml oral solution
	360891110000001100	2857410000331 11	Clomipramine Hydrochloride	Clomipramine 75mg modified-release tablets
	368401100000110 00	8535410000331 18	Maprotiline Hydrochloride	Ludiomil 25mg tablets (Novartis

				Pharmaceuticals UK Ltd)
	369211000001100	923041000033117	Trazodone Hydrochloride	Molipaxin 50mg capsules (Zentiva Pharma UK Ltd)
	378689110000011100	133008410000331100	Nortriptyline Hydrochloride	Nortriptyline 10mg/5ml oral solution sugar free
	378690110000011100	133009410000331100	Nortriptyline Hydrochloride	Nortriptyline 25mg/5ml oral solution sugar free
	379861100000111100	87941000033111	Amoxapine	Asendis 50mg tablets (Mercury Pharma Group Ltd)
	380881100000111100	85374100003311 4	Maprotiline Hydrochloride	Ludiomil 75mg tablets (Novartis Pharmaceuticals UK Ltd)
	382858110000011100	97924100003311 0	Nortriptyline	Nortriptyline 10mg capsules
	382859110000011100	97934100003311 7	Nortriptyline	Nortriptyline 25mg capsules
	386491100000111100	85344100003311 9	Maprotiline Hydrochloride	Ludiomil 10mg tablets (Novartis Pharmaceuticals UK Ltd)
	386521100000111100	85364100003311 7	Maprotiline Hydrochloride	Ludiomil 50mg tablets (Novartis Pharmaceuticals UK Ltd)
	389631110000011100	137460410000331100	Mianserin Hydrochloride	Mianserin 10mg tablets (Imported)
	389638110000011100	137461410000331100	Mianserin Hydrochloride	Mianserin 30mg tablets (Imported)
	390451100000111100	92954100003311 3	Trazodone Hydrochloride	Molipaxin 50mg/5ml oral liquid (Sanofi)
	390941100000111100	146194100003311 10	Trazodone Hydrochloride	Trazodone 50mg/5ml oral solution sugar free
	393451100000111100	87641000033116	Amoxapine	Asendis 100mg tablets (Mercury Pharma Group Ltd)

	39692911000001100	3179941000033110	Fluphenazine Hydrochloride,Nortriptyline	Nortriptyline 10mg / Fluphenazine 500microgram tablets
	39699211000001100	8501410000331 16	Lofepramine	Lofepramine 70mg tablets
	39701411000001100	9891410000331 11	Nortriptyline	Nortriptyline 10mg tablets
	39701511000001100	9892410000331 16	Nortriptyline	Nortriptyline 25mg tablets
	39731711000001100	3180541000033110	Amitriptyline Hydrochloride,Perphenazine	Amitriptyline 10mg / Perphenazine 2mg tablets
	39731811000001100	3180441000033110	Amitriptyline Hydrochloride,Perphenazine	Amitriptyline 25mg / Perphenazine 2mg tablets
	39771100000110 1	1111641000033110	Dosulepin Hydrochloride	Prepadine 25mg capsules (Teva UK Ltd)
	40061100000110 2	1294641000033110	Doxepin	Sinequan 25mg capsules (Pfizer Ltd)
	40713110000011 00	2078841000033110	Lofepramine	Lomont 70mg/5ml oral suspension (Rosemont Pharmaceuticals Ltd)
	42206311000001100	61041000033116	Amoxapine	Amoxapine 100mg tablets
	42268011000001100	60941000033114	Amoxapine	Amoxapine 50mg tablets
	42269711000001100	2620410000331 15	Clomipramine Hydrochloride	Clomipramine 10mg capsules
	42269811000001100	2621410000331 16	Clomipramine Hydrochloride	Clomipramine 25mg capsules
	42269911000001100	2622410000331 11	Clomipramine Hydrochloride	Clomipramine 50mg capsules
	42270611000001100	3088941000033110	Dosulepin Hydrochloride	Dosulepin 25mg capsules
	42270711000001100	3089041000033110	Dosulepin Hydrochloride	Dosulepin 75mg tablets
	42270811000001100	4730410000331 19	Doxepin	Doxepin 10mg capsules
	42270911000001100	4731410000331 15	Doxepin	Doxepin 25mg capsules
	42271011000001100	4732410000331 10	Doxepin	Doxepin 50mg capsules
	42271111000001100	4733410000331 17	Doxepin	Doxepin 75mg capsules
	42273711000001100	8704410000331 14	Maprotiline Hydrochloride	Maprotiline 10mg tablets
	42273811000001100	8707410000331 19	Maprotiline Hydrochloride	Maprotiline 25mg tablets

	42273911000001100	87084100003311 2	Maprotiline Hydrochloride	Maprotiline 50mg tablets
	42274011000001100	87094100003311 6	Maprotiline Hydrochloride	Maprotiline 75mg tablets
	42274111000001100	91754100003311 0	Mianserin Hydrochloride	Mianserin 10mg tablets
	42274211000001100	91764100003311 1	Mianserin Hydrochloride	Mianserin 20mg tablets
	42274311000001100	91774100003311 9	Mianserin Hydrochloride	Mianserin 30mg tablets
	42280411000001100	14552410000331 10	Trazodone Hydrochloride	Trazodone 100mg capsules
	42280511000001100	13444341000033 100	Trazodone Hydrochloride	Trazodone 100mg tablets
	42280611000001100	14736410000331 10	Trazodone Hydrochloride	Trazodone 150mg tablets
	42280711000001100	14546410000331 10	Trazodone Hydrochloride	Trazodone 50mg capsules
	42280811000001100	13444241000033 100	Trazodone Hydrochloride	Trazodone 50mg tablets
	42280911000001100	14713410000331 10	Trimipramine	Trimipramine 10mg tablets
	42281011000001100	14714410000331 10	Trimipramine	Trimipramine 25mg tablets
	42281111000001100	14547410000331 10	Trimipramine	Trimipramine 50mg capsules
	42291111000001100	60441000033116	Amitriptyline Hydrochloride	Amitriptyline 10mg tablets
	42291211000001100	60641000033119	Amitriptyline Hydrochloride	Amitriptyline 25mg tablets
	42291311000001100	60541000033115	Amitriptyline Hydrochloride	Amitriptyline 50mg tablets
	42295311000001100	75554100003311 9	Imipramine Hydrochloride	Imipramine 10mg tablets
	42295411000001100	75424100003311 5	Imipramine Hydrochloride	Imipramine 25mg tablets
	45336110000011 00	81984100003311 0	Amitriptyline Hydrochloride	Lentizol 50mg modified-release capsules (Pfizer Ltd)
	45343110000011 00	81974100003311 7	Amitriptyline Hydrochloride	Lentizol 25mg modified-release capsules (Pfizer Ltd)
	45349110000011 00	68641000033117	Clomipramine Hydrochloride	Anafranil 25mg/5ml syrup (Novartis Pharmaceuticals UK Ltd)
	45495110000011 00	92984100003311 0	Trazodone Hydrochloride	Molipaxin CR 150mg tablets (Aventis Pharma)

	4557211000001100	1462241000033110	Trazodone Hydrochloride	Trazodone 150mg modified-release tablets
	4557811000001100	2847410000331 0	Clomipramine Hydrochloride	Clomipramine 25mg/5ml oral solution
	4794110000011 01	1473441000033110	Amitriptyline Hydrochloride,Perphenazine	Triptafen-M tablets (Mercury Pharma Group Ltd)
	4918110000011 08	9388410000331 8	Fluphenazine Hydrochloride,Nortripyline	Motipress tablets (Sanofi-Synthelabo Ltd)
	4982110000011 03	1396841000033110	Trimipramine	Surmontil 25mg tablets (Sanofi)
	5146110000011 01	1473341000033110	Amitriptyline Hydrochloride,Perphenazine	Triptafen tablets (Advanz Pharma)
	53511000001107	1450841000033110	Imipramine	Tofranil 25mg tablets (Novartis Pharmaceuticals UK Ltd)
	5511000001101	62141000033113	Clomipramine Hydrochloride	Anafranil 50mg capsules (Novartis Pharmaceuticals UK Ltd)
	6217110000011 07	1294741000033110	Doxepin	Sinequan 50mg capsules (Pfizer Ltd)
	6297110000011 06	38841000033112	Nortriptyline	Allegron 25mg tablets (King Pharmaceuticals Ltd)
	6471110000011 02	62041000033114	Clomipramine Hydrochloride	Anafranil 25mg capsules (Novartis Pharmaceuticals UK Ltd)
	6940110000011 02	2215541000033110	Lofepramine	Feprapax 70mg tablets (Ashbourne Pharmaceuticals Ltd)
	7012110000011 04	61941000033115	Clomipramine Hydrochloride	Anafranil 10mg capsules (Novartis Pharmaceuticals UK Ltd)
	7331110000011 03	1135541000033110	Dosulepin Hydrochloride	Prothiaden 75mg tablets (Teofarma S.r.l.)
	75511000001108	9232410000331 13	Trazodone Hydrochloride	Molipaxin 100mg capsules (Zentiva Pharma UK Ltd)

	82778110000011 00	58886410000331 10	Amitriptyline Hydrochloride	Amitriptyline 10mg/5ml oral suspension
	83776110000011 00	44315410000331 10	Clomipramine Hydrochloride	Clomipramine 50mg/5ml oral suspension
	84540110000011 00	59922410000331 10	Dosulepin Hydrochloride	Dosulepin 75mg/5ml oral suspension
	84541110000011 00	30892410000331 10	Dosulepin Hydrochloride	Dosulepin 75mg/5ml oral solution
	84542110000011 00	58910410000331 10	Dosulepin Hydrochloride	Dosulepin 25mg/5ml oral suspension
	84543110000011 00	30891410000331 10	Dosulepin Hydrochloride	Dosulepin 25mg/5ml oral solution
	84848110000011 00	49540410000331 10	Doxepin	Doxepin 50mg/5ml oral solution
	89381100000110 0	38741000033119	Nortriptyline	Allegron 10mg tablets (King Pharmaceuticals Ltd)
	90801100000110 8	13967410000331 10	Trimipramine	Surmontil 10mg tablets (Sanofi)
	92701100000110 4	47964100003311 1	Dosulepin Hydrochloride	Dothapax 75 tablets (Ashbourne Pharmaceuticals Ltd)
	95011000001108	62624100003311 5	Lofepramine	Gamanil 70mg tablets (Merck Serono Ltd)
4.3.2 Monoamine-Oxidase Inhibitors	32511000001104	95524100003311 0	Phenelzine	Nardil 15mg tablets (Neon Healthcare Ltd)
	38961911000001 100	13746441000033 100	Moclobemide	Moclobemide 300mg tablets (Imported)
	38976911000001 100	13719241000033 100	Phenelzine	Phenelzine 15mg capsules
	39000111000001 100	13747441000033 100	Phenelzine	Phenelzine 15mg capsules (Imported)
	39708811000001 100	14694410000331 10	Soya Oil,Tranylcypromine	Tranylcypromine 10mg tablets
	42273411000001 100	78714100003311 0	Isocarboxazid	Isocarboxazid 10mg tablets
	42274911000001 100	93724100003311 8	Moclobemide	Moclobemide 150mg tablets
	42275011000001 100	93784100003311 9	Moclobemide	Moclobemide 300mg tablets

	42277211000001100	1079341000033110	Phenelzine	Phenelzine 15mg tablets
	48311000001107	8689410000331 1	Moclobemide	Manerix 300mg tablets (Viatris UK Healthcare Ltd)
	538111000001106	8688410000331 5	Moclobemide	Manerix 150mg tablets (Viatris UK Healthcare Ltd)
4.3.3 Compound Antidepressant Preparations	10065611000001100	2770541000033110	Fluoxetine Hydrochloride	Prozit 20mg/5ml oral solution (Pinewood Healthcare)
	10364011000001100	3928941000033110	Escitalopram	Cipralex 10mg/ml oral drops (Lundbeck Ltd)
	10364811000001100	3928841000033110	Escitalopram	Escitalopram 10mg/ml oral drops sugar free
	11270611000001100	4079141000033110	Paroxetine Hydrochloride	Seroxat 10mg tablets (GlaxoSmithKline UK Ltd)
	11581211000001100	4498641000033110	Fluoxetine Hydrochloride	Prozep 20mg/5ml oral solution (Rosemont Pharmaceuticals Ltd)
	11590211000001100	5749441000033110	Fluoxetine Hydrochloride	Fluoxetine 20mg/5ml oral solution sugar free
	1211000001102	1900741000033110	Citalopram hydrochloride	Cipramil 40mg/ml drops (Lundbeck Ltd)
	13045911000001100	1290094100003 3100	Sertraline Hydrochloride	Sertraline 25mg/5ml oral suspension
	142311000001103	1111941000033110	Fluoxetine Hydrochloride	Prozac 60mg capsules (Eli Lilly and Company Ltd)
	15453311000001100	5128641000033110	Escitalopram	Cipralex 20mg/ml oral drops (Lundbeck Ltd)
	15499311000001100	5128541000033110	Escitalopram	Escitalopram 20mg/ml oral drops sugar free
	16811000001109	1686141000033110	Paroxetine Hydrochloride	Seroxat 20mg/10ml liquid (GlaxoSmithKline UK Ltd)

	24397511000001100	9120741000033110	Fluoxetine Hydrochloride	Olena 20mg dispersible tablets (Advanz Pharma)
	24407811000001100	9120641000033110	Fluoxetine Hydrochloride	Fluoxetine 20mg dispersible tablets sugar free
	32960811000001100	11707441000033100	Fluoxetine Hydrochloride	Fluoxetine 30mg capsules
	33885110000011100	2749641000033110	Escitalopram	Cipralex 10mg tablets (Lundbeck Ltd)
	36565911000001100	1662441000033110	Paroxetine Hydrochloride	Paroxetine 10mg/5ml oral suspension sugar free
	38962111000001100	13740541000033100	Fluoxetine Hydrochloride	Fluoxetine 10mg tablets (Imported)
	38962311000001100	13740641000033100	Fluoxetine Hydrochloride	Fluoxetine 30mg capsules (Imported)
	38962811000001100	13740741000033100	Fluoxetine Hydrochloride	Fluoxetine 40mg capsules (Imported)
	38996511000001100	13740341000033100	Fluoxetine Hydrochloride	Fluoxetine 40mg capsules (Special Order)
	39134011000001100	13743641000033100	Sertraline Hydrochloride	Sertraline 25mg tablets
	39701911000001100	4079041000033110	Paroxetine Hydrochloride	Paroxetine 10mg tablets
	39704011000001100	1278741000033110	Sertraline Hydrochloride	Sertraline 100mg tablets
	39704111000001100	1278841000033110	Sertraline Hydrochloride	Sertraline 50mg tablets
	40746711000001100	14061741000033100	Sertraline Hydrochloride	Contulen 50mg tablets (HFA Healthcare Ltd)
	40746911000001100	14061841000033100	Sertraline Hydrochloride	Contulen 100mg tablets (HFA Healthcare Ltd)
	41166911000001100	14116141000033100	Sertraline Hydrochloride	Sertraline 150mg tablets
	41167011000001100	14116241000033100	Sertraline Hydrochloride	Sertraline 200mg tablets
	41528711000001100	14183841000033100	Fluoxetine Hydrochloride	Fluoxetine 20mg/5ml oral solution (Special Order)
	41535111000001100	14183941000033100	Fluoxetine Hydrochloride	Fluoxetine 20mg/5ml oral suspension

	41549011000001100	14287541000033100	Sertraline Hydrochloride	Sertraline 100mg/5ml concentrate for oral solution sugar free (Thame Laboratories Ltd)
	41557011000001100	14190441000033100	Sertraline Hydrochloride	Sertraline 100mg/5ml concentrate for oral solution sugar free
	42269311000001100	1583341000033110	Citalopram Hydrobromide	Citalopram 10mg tablets
	42269411000001100	2592410000331 1	Citalopram Hydrobromide	Citalopram 20mg tablets
	42269511000001100	1583141000033110	Citalopram Hydrobromide	Citalopram 40mg tablets
	42269611000001100	1900641000033110	Citalopram hydrochloride	Citalopram 40mg/ml oral drops sugar free
	42271211000001100	2749541000033110	Escitalopram	Escitalopram 10mg tablets
	42271311000001100	2951341000033110	Escitalopram	Escitalopram 20mg tablets
	42271411000001100	3030341000033110	Escitalopram	Escitalopram 5mg tablets
	42271511000001100	11603841000033100	Fluoxetine Hydrochloride	Fluoxetine 10mg capsules
	42271611000001100	6496341000033110	Fluoxetine Hydrochloride	Fluoxetine 10mg tablets
	42271711000001100	5778410000331 4	Fluoxetine Hydrochloride	Fluoxetine 20mg capsules
	42271811000001100	2770641000033110	Fluoxetine Hydrochloride	Fluoxetine 20mg/5ml oral solution
	42271911000001100	11707541000033100	Fluoxetine Hydrochloride	Fluoxetine 40mg capsules
	42272011000001100	5790410000331 1	Fluoxetine Hydrochloride	Fluoxetine 60mg capsules
	42272411000001100	5928410000331 6	Fluvoxamine Maleate	Fluvoxamine 100mg tablets
	42272511000001100	5955410000331 3	Fluvoxamine Maleate	Fluvoxamine 50mg tablets
	42276511000001100	1042141000033110	Paroxetine Hydrochloride	Paroxetine 20mg tablets
	42276611000001100	1042241000033110	Paroxetine Hydrochloride	Paroxetine 30mg tablets
	42276711000001100	11755441000033100	Paroxetine Hydrochloride	Paroxetine 40mg tablets
	43159711000001100	14507641000033100	Sertraline Hydrochloride	Sertraline 50mg/5ml oral suspension sugar free

	456311000001103	1923141000033110	Fluoxetine Hydrochloride	Felicium 20mg capsules (Opus Pharmaceuticals Ltd)
	477511000001109	1278341000033110	Paroxetine Hydrochloride	Seroxat 20mg tablets (GlaxoSmithKline UK Ltd)
	494251100000110 0	2951441000033110	Escitalopram	Cipralex 20mg tablets (Lundbeck Ltd)
	542511000001103	5633410000331 15	Fluvoxamine Maleate	Faverin 50mg tablets (Viatris UK Healthcare Ltd)
	573011000001101	1112441000033110	Fluoxetine Hydrochloride	Prozac 20mg capsules (Eli Lilly and Company Ltd)
	595611000001101	5619410000331 15	Fluvoxamine Maleate	Faverin 100mg tablets (Viatris UK Healthcare Ltd)
	625110000011 09	1583241000033110	Citalopram Hydrobromide	Cipramil 10mg tablets (Lundbeck Ltd)
	683911000001104	2710441000033110	Fluoxetine Hydrochloride	Ranflutin 20mg capsules (Sun Pharma UK Ltd)
	687611000001102	8541410000331 13	Sertraline Hydrochloride	Lustral 100mg tablets (Viatris UK Healthcare Ltd)
	689411000001109	2602041000033110	Fluoxetine Hydrochloride	Oxactin 20mg capsules (Discovery Pharmaceuticals)
	695611000001105	1583041000033110	Citalopram Hydrobromide	Cipramil 40mg tablets (Lundbeck Ltd)
	695811000001109	1129541000033110	Fluoxetine Hydrochloride	Prozac 20mg/5ml liquid (Eli Lilly and Company Ltd)
	733181100000110 0	3030541000033110	Escitalopram	Cipralex 5mg tablets (Lundbeck Ltd)
	738111000001107	1278441000033110	Paroxetine Hydrochloride	Seroxat 30mg tablets (GlaxoSmithKline UK Ltd)
	847211000001102	2591410000331 16	Citalopram Hydrobromide	Cipramil 20mg tablets (Lundbeck Ltd)

	87217110000011100	59753410000331 10	Sertraline Hydrochloride	Sertraline 100mg/5ml oral suspension
	87219110000011100	41483410000331 10	Sertraline Hydrochloride	Sertraline 50mg/5ml oral suspension
	93641100000110 8	85424100003311 8	Sertraline Hydrochloride	Lustral 50mg tablets (Viatris UK Healthcare Ltd)
4.3.4 Other Antidepressant Drugs	30248111000001100	10617341000033100	Vortioxetine Hydrobromide	Brintellix 10mg tablets (Lundbeck Ltd)
	30738611000001100	10617041000033100	Vortioxetine Hydrobromide	Vortioxetine 10mg tablets
	30738711000001100	10617141000033100	Vortioxetine Hydrobromide	Vortioxetine 20mg tablets
	30738811000001100	10616941000033100	Vortioxetine Hydrobromide	Vortioxetine 5mg tablets
	30990411000001100	10701841000033100	Duloxetine	Duciltia 30mg gastro-resistant capsules (Pharmathen S.A.)
	30990611000001100	10701941000033100	Duloxetine	Duciltia 60mg gastro-resistant capsules (Pharmathen S.A.)
	30991111000001100	10730141000033100	Venlafaxine Hydrochloride	Efexor XL 225mg capsules (Viatris UK Healthcare Ltd)
	31016011000001100	10730041000033100	Venlafaxine Hydrochloride	Venlafaxine 225mg modified-release capsules
	32833611000001100	11693641000033100	Venlafaxine Hydrochloride	Vensir XL 225mg capsules (Morningside Healthcare Ltd)
	34180611000001100	12194641000033100	Venlafaxine Hydrochloride	Vencarm XL 225mg capsules (Aspire Pharma Ltd)
	34180811000001100	12194541000033100	Venlafaxine Hydrochloride	Vencarm XL 150mg capsules (Aspire Pharma Ltd)
	34181011000001100	12194841000033100	Venlafaxine Hydrochloride	Vencarm XL 75mg capsules (Aspire Pharma Ltd)
	34181211000001100	12194741000033100	Venlafaxine Hydrochloride	Vencarm XL 37.5mg capsules (Aspire Pharma Ltd)

	34740711000001100	12390341000033100	Venlafaxine Hydrochloride	Venlasov XL 150mg capsules (Sovereign Medical Ltd)
	34740911000001100	12389941000033100	Venlafaxine Hydrochloride	Venlasov XL 75mg capsules (Sovereign Medical Ltd)
	35034711000001100	12475941000033100	Duloxetine	Depalta 30mg gastro-resistant capsules (GlucoRx Ltd)
	35035211000001100	12476041000033100	Duloxetine	Depalta 60mg gastro-resistant capsules (GlucoRx Ltd)
	35638211000001100	12605541000033100	Venlafaxine Hydrochloride	Venladex XL 225mg tablets (Dexcel-Pharma Ltd)
	35797111000001100	12641541000033100	Venlafaxine Hydrochloride	Venzip XL 150mg capsules (Milpharm Ltd)
	35797611000001100	12641641000033100	Venlafaxine Hydrochloride	Venzip XL 75mg capsules (Milpharm Ltd)
	35839111000001100	14265941000033100	Duloxetine	Duloxetine 30mg/5ml oral suspension
	36606311000001100	12901041000033100	Venlafaxine Hydrochloride	Majoven XL 37.5mg capsules (Bristol Laboratories Ltd)
	36606511000001100	12901141000033100	Venlafaxine Hydrochloride	Majoven XL 75mg capsules (Bristol Laboratories Ltd)
	36606711000001100	12901241000033100	Venlafaxine Hydrochloride	Majoven XL 150mg capsules (Bristol Laboratories Ltd)
	36656911000001100	12904541000033100	Venlafaxine Hydrochloride	ViePax XL 225mg tablets (Dexcel-Pharma Ltd)
	36796110000011100	431741000033118	Flupentixol	Depixol 3mg tablets (Lundbeck Ltd)
	36796110000011100	431741000033118	Flupentixol	Depixol 3mg tablets (Lundbeck Ltd)
	36845311000001100	12986241000033100	Venlafaxine Hydrochloride	Apclaven XL 37.5mg capsules (Torrent Pharma (UK) Ltd)

	36845711000001100	12986341000033100	Venlafaxine Hydrochloride	Apclaven XL 75mg capsules (Torrent Pharma (UK) Ltd)
	36846211000001100	12986441000033100	Venlafaxine Hydrochloride	Apclaven XL 150mg capsules (Torrent Pharma (UK) Ltd)
	36871511000001100	12989441000033100	Duloxetine	Dutor 30mg gastro-resistant capsules (Torrent Pharma (UK) Ltd)
	36871911000001100	12989541000033100	Duloxetine	Dutor 60mg gastro-resistant capsules (Torrent Pharma (UK) Ltd)
	37088110000011100	10089410000033110	L-Tryptophan	Optimax 500mg tablets (Esteve Pharmaceuticals Ltd)
	37997111000001100	13318441000033100	Venlafaxine Hydrochloride	Venlalic XL 300mg tablets (Ethypharm UK Ltd)
	38003311000001100	13318341000033100	Venlafaxine Hydrochloride	Venlafaxine 300mg modified-release tablets
	38499811000001100	13495041000033100	Esketamine Hydrochloride	Spravato 28mg/0.2ml nasal spray unit dose (Janssen-Cilag Ltd)
	38515511000001100	13494941000033100	Esketamine Hydrochloride	Esketamine 28mg/0.2ml nasal spray unit dose
	38750311000001100	15093410000033110	Venlafaxine Hydrochloride	Venlafaxine 75mg modified-release capsules
	38750411000001100	15092410000033110	Venlafaxine Hydrochloride	Venlafaxine 150mg modified-release capsules
	38960711000001100	13746241000033100	Mirtazapine	Mirtazapine 15mg orodispersible tablets (Imported)
	38960911000001100	13746341000033100	Mirtazapine	Mirtazapine 45mg orodispersible tablets (Imported)
	39373110000011100	49004100003311 8	Nefazodone Hydrochloride	Dutonin tablets treatment initiation pack (Bristol-

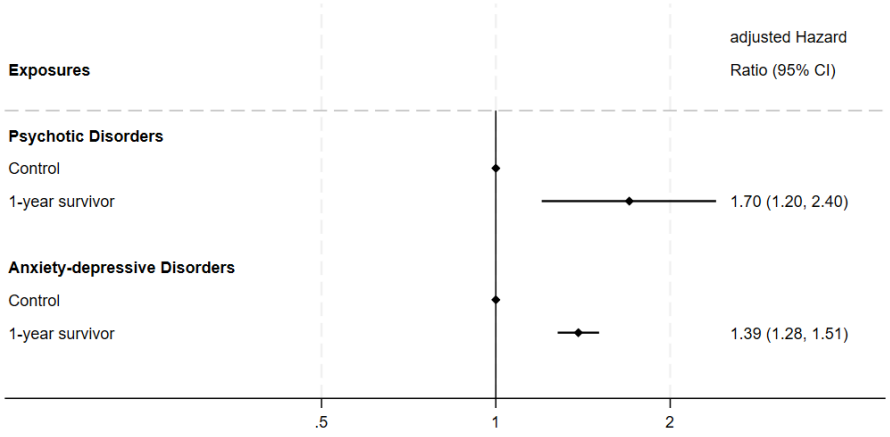
				Myers Squibb Pharmaceuticals Ltd)
	39692311000001100	3091041000033110	Flupentixol	Flupentixol 1mg tablets
	39692511000001100	3091141000033110	Flupentixol	Flupentixol 500microgram tablets
	39703311000001100	1165641000033110	Reboxetine	Reboxetine 4mg tablets
	39707311000001100	1512441000033110	Venlafaxine Hydrochloride	Venlafaxine 50mg tablets
	39707411000001100	1512341000033110	Venlafaxine Hydrochloride	Venlafaxine 75mg tablets
	39707611000001100	1512241000033110	Venlafaxine Hydrochloride	Venlafaxine 37.5mg tablets
	39964711000001100	1389874100003310 0	Duloxetine	Duloxetine 120mg gastro-resistant capsules
	39964811000001100	1389864100003310 0	Duloxetine	Duloxetine 90mg gastro-resistant capsules
	40111011000001100	1391624100003310 0	Venlafaxine Hydrochloride	Venlafaxine 37.5mg/5ml oral solution sugar free
	40111111000001100	1391634100003310 0	Venlafaxine Hydrochloride	Venlafaxine 75mg/5ml oral solution sugar free
	413311000001104	4985410000331 13	Venlafaxine Hydrochloride	Efexor XL 150mg capsules (Viatris UK Healthcare Ltd)
	41966611000001100	8248041000033110	Tryptophan	Optimax 500mg capsules (Esteve Pharmaceuticals Ltd)
	41989911000001100	1474141000033110	Tryptophan	Tryptophan 500mg tablets
	42163111000001100	1422804100003310 0	Reboxetine	Reboxetine 4mg tablets (Imported)
	42274411000001100	3029641000033110	Mirtazapine	Mirtazapine 15mg orodispersible tablets
	42274511000001100	3256741000033110	Mirtazapine	Mirtazapine 15mg tablets
	42274611000001100	2974841000033110	Mirtazapine	Mirtazapine 30mg orodispersible tablets
	42274711000001100	9196410000331 15	Mirtazapine	Mirtazapine 30mg tablets

	42274811000001100	3256841000033110	Mirtazapine	Mirtazapine 45mg tablets
	42275111000001100	9661410000331116	Nefazodone Hydrochloride	Nefazodone 100mg tablets
	42275211000001100	9662410000331111	Nefazodone Hydrochloride	Nefazodone 200mg tablets
	42294111000001100	3226641000033110	Duloxetine	Duloxetine 30mg gastro-resistant capsules
	42294211000001100	3226741000033110	Duloxetine	Duloxetine 60mg gastro-resistant capsules
	48597110000011100	3078541000033110	Mirtazapine	Zispin SolTab 30mg orodispersible tablets (Organon Pharma (UK) Ltd)
	50751100000110 6	5935410000331 2	Flupentixol	Fluanxol 500microgram tablets (Lundbeck Ltd)
	50751100000110 6	5935410000331 2	Flupentixol	Fluanxol 500microgram tablets (Lundbeck Ltd)
	52361100000110 7	1558041000033110	Mirtazapine	Zispin 30mg tablets (Organon Laboratories Ltd)
	55958110000011100	3029941000033110	Mirtazapine	Zispin SolTab 45mg orodispersible tablets (Organon Pharma (UK) Ltd)
	56088110000011100	3029841000033110	Mirtazapine	Zispin SolTab 15mg orodispersible tablets (Organon Pharma (UK) Ltd)
	56279110000011100	3029741000033110	Mirtazapine	Mirtazapine 45mg orodispersible tablets
	57671100000110 6	4967410000331 8	Reboxetine	Edronax 4mg tablets (Pfizer Ltd)
	62851100000110 7	4992410000331 5	Venlafaxine Hydrochloride	Efexor 50mg tablets (Wyeth Pharmaceuticals)
	65181100000110 0	4984410000331 2	Venlafaxine Hydrochloride	Efexor XL 75mg capsules (Viatris UK Healthcare Ltd)
	66621100000110 4	4898410000331 5	Nefazodone Hydrochloride	Dutonin 200mg tablets (Bristol-Myers Squibb Pharmaceuticals Ltd)

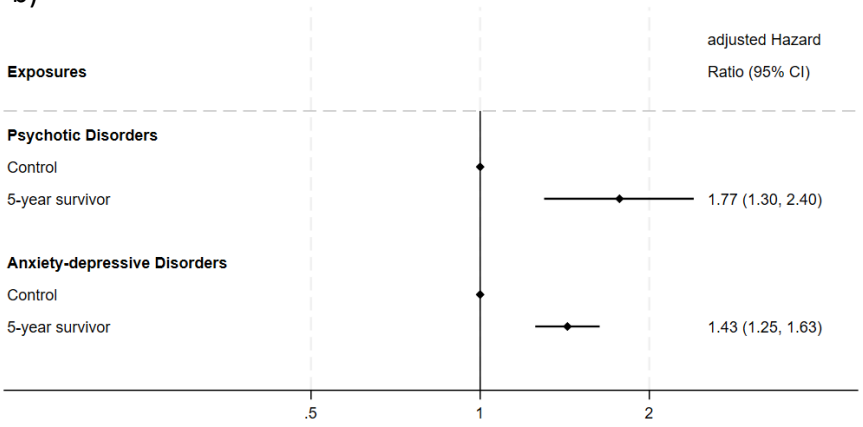
	843011000001101	593441000033111	Flupentixol	Fluanxol 1mg tablets (Lundbeck Ltd)
	843011000001101	593441000033111	Flupentixol	Fluanxol 1mg tablets (Lundbeck Ltd)
	864911000001105	489741000033113	Nefazodone Hydrochloride	Dutonin 100mg tablets (Bristol-Myers Squibb Pharmaceuticals Ltd)
	872271100000110000	449854100003311000	Venlafaxine Hydrochloride	Venlafaxine 37.5mg/5ml oral suspension
	903841100000110000	322684100003311000	Duloxetine	Cymbalta 30mg gastro-resistant capsules (Eli Lilly and Company Ltd)
	903931100000110000	322694100003311000	Duloxetine	Cymbalta 60mg gastro-resistant capsules (Eli Lilly and Company Ltd)
	910351100000110000	305874100003311000	Mirtazapine	Mirtazapine 15mg/ml oral solution sugar free

Supplementary Figure 9-8 *Forest plot showing risk (measured using adjusted hazard ratio) of psychotic and anxiety-depressive disorders in a) 1 year and b) 5-year CTYA cancer survivors following multiple imputation. 95% CI = 95% confidence interval*

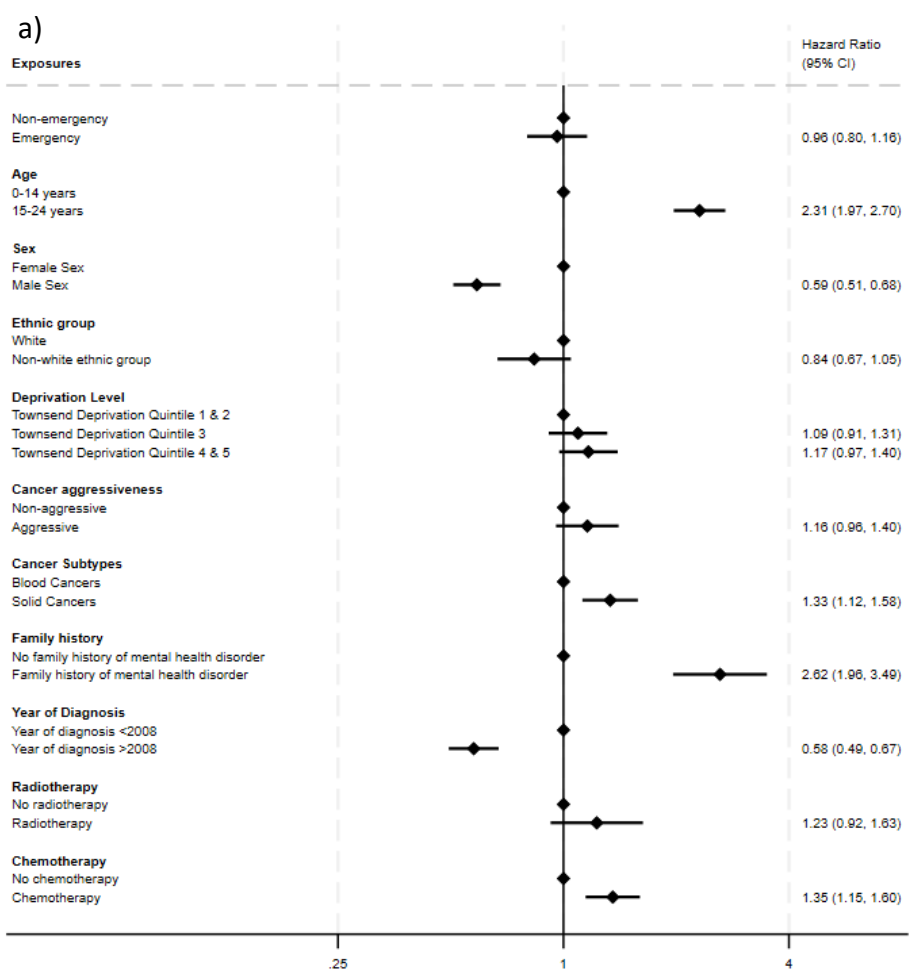
a)

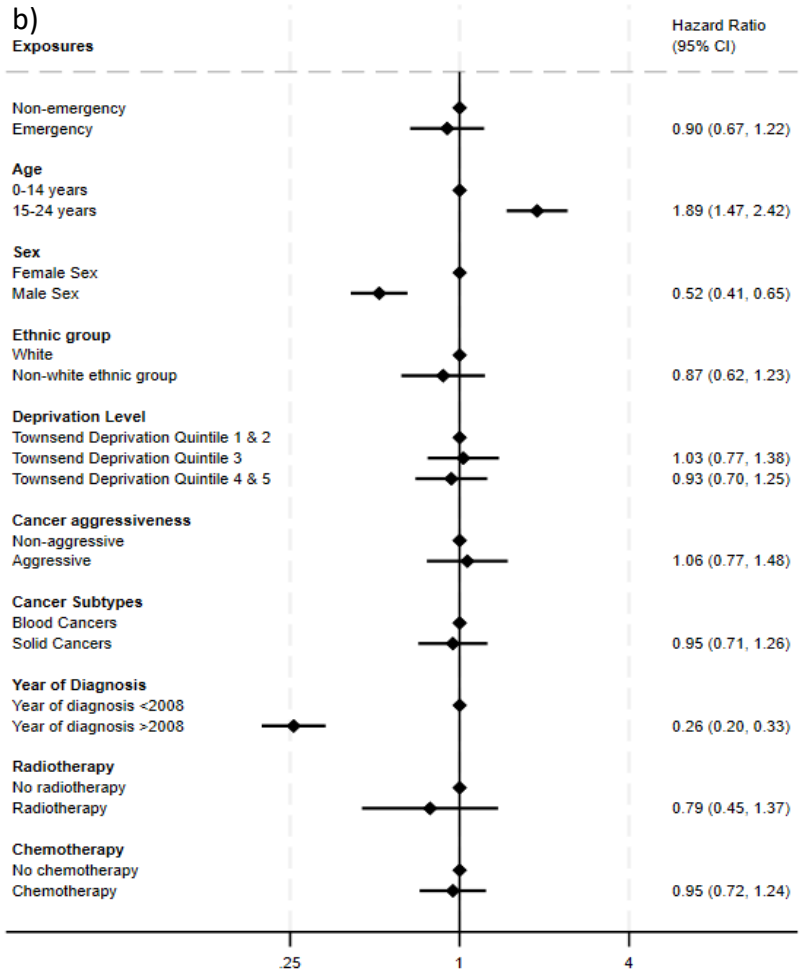


b)

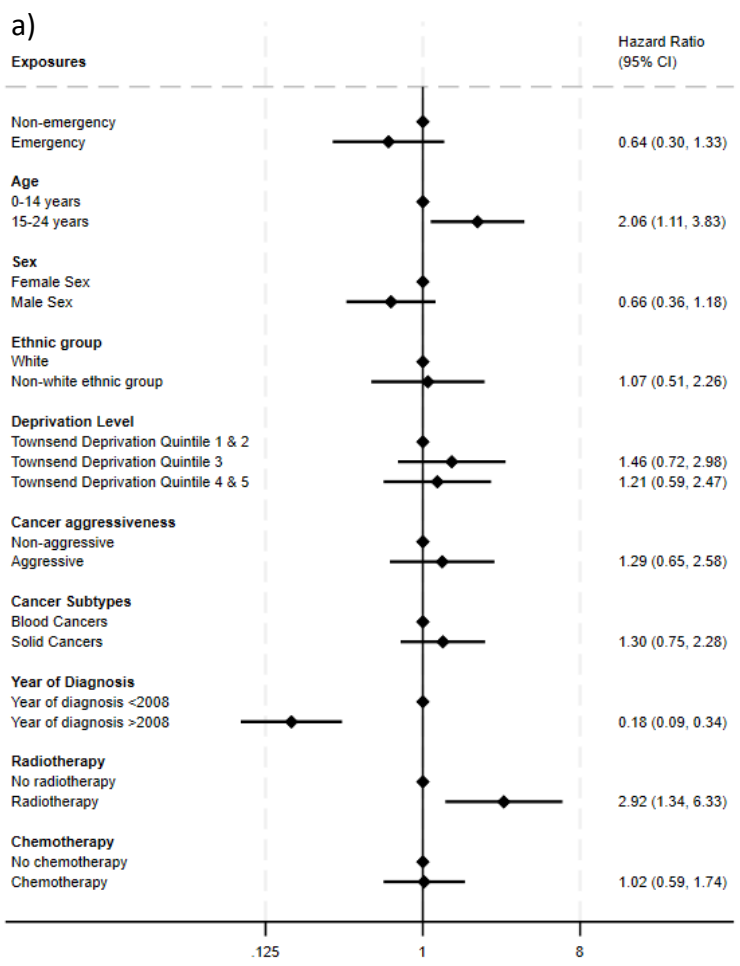


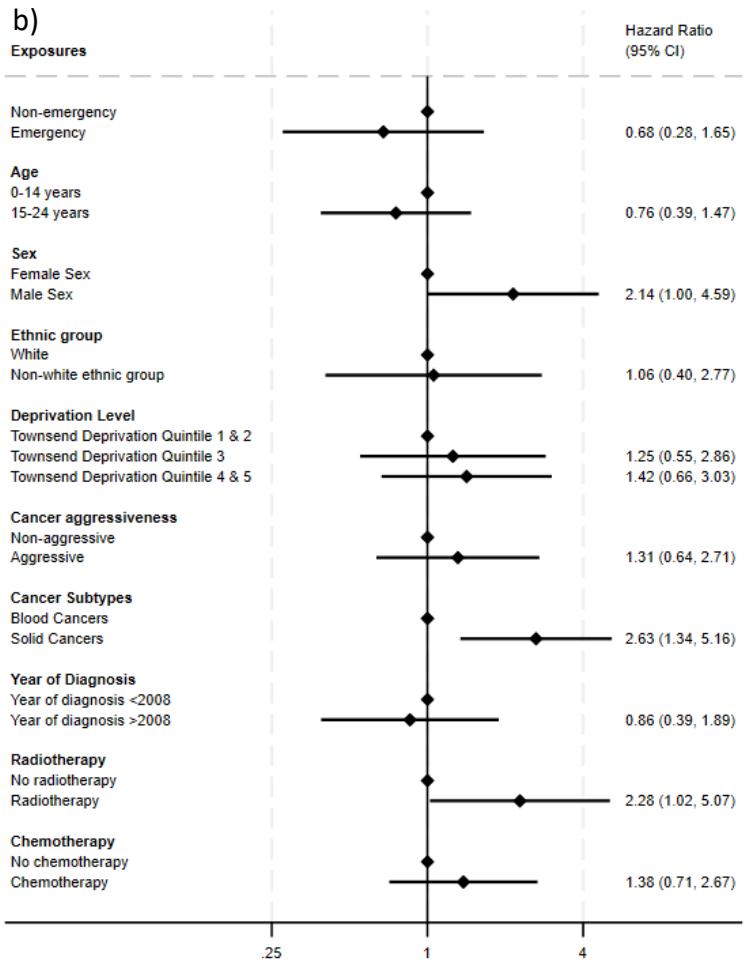
Supplementary Figure 9-9 *Forest plot demonstrating hazard ratios (with 95% confidence intervals) of anxiety-depressive disorders for cancer-route and adjusted factors following Cox proportional regression and multiple imputation in a) 1-year survivors and b) 5-year survivors.*





Supplementary Figure 9-10 *Forest plot demonstrating hazard ratios (with 95% confidence intervals) of psychotic disorders for cancer-route and adjusted factors following Cox proportional regression and multiple imputation in a) 1-year survivors and b) 5-year survivors.*





Chapter 10 Tables and Figures

Hyperlinks to Tables

Table 1-1 *Person and healthcare factors with prior evidence of contributing to delays in CTYA cancers (18, 23, 40, 47).*

Table 2-1 *Number of cancer diagnoses records in each dataset and Measures of Accuracy measures of SNOMED-CT and ICD-10 cancer codes using the National Cancer Registry as gold standard.*

Table 2-2 *Alignment of cancer diagnoses between SNOMED-CT/ICD-10 and cancer diagnoses in the National Cancer Registry.*

Table 2-3 *Evaluation of false positives by cancer subtype on GP and HES datasets compared to National Cancer Registry used as gold standard*

Table 2-4 *Evaluation of false negatives by cancer subtype on GP and HES datasets compared to National Cancer Registry used as gold standard*

Table 2-5 *Dates of cancer diagnosis by cancer subtype on GP and HES datasets compared to the gold-standard National Cancer Registry. IQR= interquartile range.*

Table 2-6 *Baseline characteristics of full population of 0-24 year olds and full cohort of cancer cases. All percentages are column percentages. Age in years is rounded to the nearest integer above.*

Table 2-7 *Baseline characteristics of cases by cancer subtype. All percentages except proportion of total cancer are column percentages. Age in years is rounded to the nearest integer above.*

Table 2-8 *Incidence rates per million population by sex in generated cohort and nationally reported world age-standardised incidence rates per million population by sex in England.*

Table 3-1 *Clinical features identified through literature search, patient input and national guidance, associated with CTYA cancer per cancer subtype*

Table 3-2 *Demographic characteristics of all cancers, all controls and individual cancer subtypes. Numbers are count (col%) unless otherwise specified. IQR= interquartile range. CI= confidence interval.*

Table 3-3 *Frequency of selected symptom/signs in cancer cases and controls. Association between individual clinical feature and a subsequent cancer diagnosis in univariable conditional logical regression analyses. Any counts <5 were suppressed for confidentiality. All percentages are column %.*

Table 3-4 *Frequency of selected clinical features in childhood cancer cases and controls. Association between individual clinical feature and a subsequent childhood cancer diagnosis in univariable conditional logical regression analyses. OR= Odds Ratio. CI=Confidence interval. All percentages are column %.*

Table 3-5 *Frequency of selected clinical features in TYA cancer cases and controls. Association between individual clinical feature and a subsequent TYA cancer diagnosis in univariable conditional logical regression analyses. OR= Odds Ratio. CI=Confidence interval. All percentages are column %.*

Table 3-6 *Median diagnostic intervals for each cancer subtype overall and across each age category. IQR= interquartile range*

Table 3-7 *Strength of association between any cancer diagnosis and early symptoms i) when considered in isolation, ii) when considered in combination with any of the other identified statistically significant symptoms within 6 months of diagnosis. aOR= adjusted Odds Ratio, PPV= Positive predictive value.*

Table 3-8 *Component Loading Matrix where loading scores indicate the correlation of the original variable with the principal component. Loading scores with significant contribution (above 0.3) shown only. PC= principal component.*

Table 3-9 *Association between principal components and CTYA cancers in univariable conditional logistic regression models, adjusted for ethnicity and deprivation level through the Townsend Quintile. aOR= adjusted Odds Ratio, CI= confidence interval, SD= standard deviation*

Table 3-10 *Proportion of cases and controls with the symptom combination for the principal component present and the corresponding positive predictive values. PC= principal component, PPV= Positive Predictive Value, CI= confidence interval*

Table 4-1 *Candidate predictor variables explored in model development*

Table 4-2 *Baseline characteristics of i) haematological cancer, ii) solid cancer and iii) full population cohorts used in internal-external cross validation. All data except age has been expressed in numbers (% column). Age in years is rounded to the nearest integer.*

Table 4-3 Baseline characteristics by blood test results. All data has been expressed in numbers (% row). Age in years is rounded to the nearest integer above. TQ = Townsend Deprivation Quantile 1-5.

Table 4-4 *Performance measures per age, sex and ethnic subgroups for i) symptom-only and ii) symptom + blood test solid cancer model. AUC= area under the curve, 95% CI = 95% confidence interval*

Table 4-5 *Measures of Accuracy of each model, following validation, based on the top 0.1%, 0.5%, 1%, 2%, 5%, 10%, 20%, 40%, 60%, 80%, 90% and 95% at highest risk in symptom-only and symptom + blood test solid cancer models*

Table 4-6 *Performance measures per age, sex and ethnic subgroups for i) symptom-only and ii) symptom + blood test haematological cancer models. AUC= area under the curve, 95% CI = 95% confidence interval*

Table 4-7 *Measures of Accuracy of each model, following validation, based on the top 0.1%, 0.5%, 1%, 2%, 5%, 10%, 20%, 40%, 60%, 80%, 90% and 95% at highest risk in symptom-only and symptom + blood test haematological cancer models*

Table 4-8 *Performance measures per age, sex and ethnic subgroups for i) symptom-only and ii) symptom + blood test leukaemia models. AUC= area under the curve, 95% CI = 95% confidence interval*

Table 4-9 *Measures of Accuracy of each model, following validation, based on the top 0.1%, 0.5%, 1%, 2%, 5%, 10%, 20%, 40%, 60%, 80%, 90% and 95% at highest risk in symptom-only and symptom + blood test leukaemia models*

Table 4-10 *Performance measures per age, sex and ethnic subgroups for i) symptom-only and ii) symptom + blood test lymphoma models. AUC= area under the curve, 95% CI = 95% confidence interval*

Table 4-11 *Measures of Accuracy of each model, following validation, based on the top 0.1%, 0.5%, 1%, 2%, 5%, 10%, 20%, 40%, 60%, 80%, 90% and 95% at highest risk in symptom-only and symptom + blood test lymphoma models*

Table 4-12 *Performance measures per age, sex and ethnic subgroups CNS tumour model. AUC= area under the curve, 95% CI = 95% confidence interval*

Table 4-13 *Measures of Accuracy, following validation, based on the top 0.1%, 0.5%, 1%, 2%, 5%, 10%, 20%, 40%, 60%, 80%, 90% and 95% at highest risk in CNS tumour model*

Table 4-14 Clinical Scenarios and predicted risks using developed models for a) haematological (leukaemia and lymphoma individually), b) solid cancers (CNS tumour individually).

Table 5-1 *Route to cancer diagnosis (adapted National Cancer Registry(14))*

Table 5-2 *Factors associated with diagnostic delays and emergency presentations*

Table 5-3 *Exposures-of-interest in those with and without a recorded diagnostic interval. All percentages are column percentages (%). Age is in integers and rounded up to the nearest integer. IQR= interquartile range.*

Table 5-4 *The association between age-at-diagnosis and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for sex, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Table 5-5 *The association between sex and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Table 5-6 *The association between ethnic group and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, sex, cancer subtype and cancer aggressiveness. 95% CI = 95% confidence interval.*

Table 5-7 *The association between deprivation level, as assessed by the Townsend Deprivation Score, and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, sex, cancer subtype and cancer aggressiveness. Baseline quintile is Quintile 1 (i.e., least deprived). 95%CI = 95% confidence interval.*

Table 5-8 *The association between cancer types and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Baseline cancer type is leukaemia. Adjusted for age and sex. 95%CI = 95% confidence interval.*

Table 5-9 *The association between cancer biology (as defined by histological aggressiveness) and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted*

quantile regression models. Adjusted for age, sex and cancer subtype. 95%CI = 95th confidence interval.

Table 5-10 *The association between presence of red flag symptoms (as defined by NICE, Grace Kelly and HEADSMART guidance) and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, , cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Table 5-11 *The association between symptoms/signs and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models for leukaemias, sarcomas and abdominal tumours. Adjusted for age and sex. 95%CI = 95% confidence interval*

Table 5-12 *The association between blood test request and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, sex, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Table 5-13 *The association between imaging test request and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Adjusted for age, sex, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Table 5-14 *The association between regions and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Baseline region selected was South of England. Adjusted for age, sex, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval*

Table 5-15 *The association between the three COVID-19 pandemic periods and 25th centile, median and 75th centile diagnostic intervals in unadjusted and adjusted quantile regression models. Baseline period was selected pre-pandemic period (2017-2019). Adjusted for age, sex, cancer subtype and cancer aggressiveness. 95%CI = 95% confidence interval.*

Table 5-16 *Exposures-of-interest in those with and emergency, non-emergency and those without a recorded presentation. All percentages are column percentages (%). Age is in integers and rounded up to the nearest integer. IQR= interquartile range. Note: p-value for analyses between those with and without a presentation.*

Table 6-1 *Baseline characteristics of young people in the cohorts (cohort 1 for 1-year follow-up, cohort 2 for 5-year follow-up) who had or did not have death recorded. All percentages are column percentages (%). Age is in integers and rounded up to the nearest integer. IQR= interquartile range.*

Table 7-1 *Description of diagnoses and drug names used to define each outcome in this Chapter.*

Table 7-2 *Minimum detectable hazard ratios in this study given sample at 80% power and significance threshold of 0.05. Comparison of previously reported hazard ratios (163).*

Table 7-3 *Baseline characteristics of cancer cases and matched controls. All percentages are column percentages (%). Age is in integers and rounded up to the nearest integer. IQR= interquartile range.*

Table 7-4 *Mental health disorders in 1-year cancer survivors and controls.*

Table 7-5 *Mental health disorder in 5-year cancer survivors and controls*

Supplementary Table 9-1 *ICD-10 codes used to define CTYA cancers in HES Dataset*

Supplementary Table 9-2 SNOMED-CT codes used to define CTYA cancers in GP Dataset

Supplementary Table 9-3 *National Cancer Registry Morphology Codes based on ICCC-3 (adapted from Public Health England, National Cancer Registry Analysis Services(1))*

Supplementary Table 9-4 Measures of Accuracy measures of SNOMED-CT/ICD-10 cancer codes using the National Cancer Registry as gold standard.

Table 9-5 Diagnostic codes of false positives identified in GP and HES datasets when National Cancer Registry is used as gold standard

Table 9-6. *Definition of clinical features using SNOMED-CT clinical codes. Grouped into gastrointestinal, genitourinary, cardiorespiratory, neurological, systemic, haematological and musculoskeletal symptom groups.*

Table 9-7 Minimum sample size required to detect an odds ratio of 2 for a single feature separately for childhood (0-14 years) and TYA (15-24 years). 95% CI= 95% Confidence Interval

Table 9-8 Minimum sample size required to detect odds ratio of 2 for two interacting dependent variables x_1 and x_2 (where x_1 is the clinical feature detailed in Supplementary Table 6.3-3)

Table 9-9 Frequency of selected clinical features in cancer cases and controls per cancer subtype. Association between individual clinical feature and a subsequent cancer diagnosis in univariable conditional logical regression analyses. OR= Odds Ratio. CI=Confidence interval.

Table 9-10 *Frequency of selected clinical features in cancer cases and controls per cancer subtype in children. Association between individual clinical feature and a subsequent cancer diagnosis in univariable conditional logical regression analyses. OR= Odds Ratio. CI=Confidence interval.*

Table 9-11 Frequency of selected clinical features in cancer cases and controls per cancer subtype in TYA. Association between individual clinical feature and a subsequent cancer diagnosis in univariable conditional logical regression analyses. OR= Odds Ratio. CI=Confidence interval.

Table 9-12 *Proportional hazard test using Schoenfeld residuals demonstrating that proportionally assumption is violated ($p<0.05$)*

Table 9-13 *Sample size requirements for clinical prediction model development using Riley et al., 2020(128)*

Table 9-14 Meta-regression analysis for age-at-diagnosis, ethnicity, Townsend deprivation score. 95%CI = 95% confidence interval.

Table 9-15 Search Strategies for Literature Review in Chapter 5

Table 9-16 Clinical and histological classification of high risk cancer biology.

Table 9-17 Baseline Characteristics of study population for time period 1st January – 1st May 2017-2021. All % are column percentages. Age rounded up to the nearest integer.

Table 9-18 Baseline Characteristics of study population for time period 2nd May- 31st December 2017-2021. All % are column percentages. Age rounded up to the nearest integer.

Table 9-19 List of pre-defined clinical features associated with a diagnosis of CTYA cancer (adapted from NICE Guideline [NG12] – Suspected Cancer: Recognition and Referral)

Table 9-20 The association between exposures-of-interest and 25th centile, median and 75th centile diagnostic intervals in adjusted quantile regression models following multiple imputation. Adjusted for age, sex, cancer subtype and biology. When variable of interest was an adjusting factor, that factor was excluded from regression models. 95%CI = 95% confidence interval.

Table 9-21 Forest plot demonstrating adjusted odds of emergency presentation compared to non-emergency presentation in CTYA cancers for each factor/exposure in multiple imputation models. 95% CI = 95% confidence interval.

Table 9-22 Forest plot demonstrating adjusted odds of emergency presentation compared to non-emergency presentation in CTYA cancers for symptoms only in multiple imputation models. 95% CI = 95% confidence interval.

Table 9-23 SNOMED-CT and ICD-10 Codes for each mental health outcome

Hyperlinks to Figures

Figure 1-1 *Risk of cancer by age, adapted Walker et al., 2021(7).*

Figure 1-2 *Annual average number of cases within the twelve broad diagnostic groups of CTYA cancers in the United Kingdom, adapted Cancer Research UK Cancer Statistics(6).*

Figure 1-3 *Survival following CTYA cancer from 1980 to 2010, adapted Cancer Research UK Cancer Statistics (1).*

Figure 1-4 *An illustration of a cancer diagnostic pathway, starting from symptom onset to confirmed diagnosis.*

Figure 2-1 *Flowchart for Cancer Diagnosis Identification and Classification. Only identification and merging steps are used for validation study described below.*

Figure 2-2 *Validation study for cancer diagnoses with definitions for sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). SNOMED-CT/ICD-10 codes compared against National Cancer Registry.*

Figure 2-3 *Plot depicting number of cancer diagnoses identified in intersecting datasets. Dots represent diagnoses present in individual dataset, lines between dots represent diagnoses present in multiple datasets (74). Total number of cancer diagnoses in each dataset summarised on bar chart on bottom left. GP= General Practice Dataset, HES= HES Hospital Dataset, NCR= National Cancer Registry Dataset*

Figure 2-4 *Proportion (%) by 4-year age band and sex for i) QResearch CTYA cohort, ii) ONS population estimates.*

Figure 2-5 *Proportion (%) of ethnic categories for i) QResearch CTYA cohort, ii) 2021 census population estimates(75).*

Figure 2-6 Summary flow diagram depicting the variations of the CTYA and cancer cohorts used in each chapter of the Thesis.

Figure 3-1 *Clinical features encountered in primary care associated with any cancer in CTYA in multivariable conditional logistic regression models. Early symptoms denoted with . OR= Odds ratio, CI= Confidence interval, PPV=Positive Predictive Value.*

Figure 3-2 *Clinical features encountered in primary care associated with any childhood cancer in multivariable conditional logistic regression models. Early symptoms denoted with . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value. All percentages are column %.*

Figure 3-3 *Clinical features encountered in primary care associated with any TYA cancer in multivariable conditional logistic regression models. Early symptoms denoted with . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value.*

Figure 3-4 *Clinical features encountered in primary care associated with CTYA leukaemia in multivariable conditional logistic regression models. Early symptoms denoted with . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value.*

Figure 3-5 *Clinical features encountered in primary care associated with CTYA lymphoma in multivariable conditional logistic regression models. Early symptoms denoted with . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value.*

Figure 3-6 *Clinical features encountered in primary care associated with CTYA CNS tumours in multivariable conditional logistic regression models. Early symptoms denoted with . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value.*

Figure 3-7 *Clinical features encountered in primary care associated with CTYA sarcomas in multivariable conditional logistic regression models. Early symptoms denoted with . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value.*

Figure 3-8 *Clinical features encountered in primary care associated with CTYA abdominal tumours in multivariable conditional logistic regression models. Early symptoms denoted with . OR= Odds ratio, CI= Confidence interval PPV=Positive Predictive Value.*

Figure 4-1 *Schematic overview of the internal-external cross-validation process*

Figure 4-2 *Final logistic regression model predicting odds of solid cancer 6 months following study entry/symptom onset in a) symptom-only and b) symptom + blood test model. Models*

contain fractional polynomials for age (-2,-2). Adjusted OR= Adjusted Odds ratio, CI= Confidence interval, Hb= haemoglobin, CRP=C-reactive protein.

***Figure 4-3** Calibration of the a) symptom-only and b) symptom + blood test solid cancer models tested demonstrated through calibration plots which show the alignment between predicted and observed risks across 10 risk deciles/groups.*

***Figure 4-4** Region-level and pooled performance estimates following internal-external cross validation for a) & d) area-under-the-curve (AUC), b) & e) calibration slope and, c) & f) calibration-in-the-large for symptom-only (a-c) and symptom + blood test solid cancer models (d-f). Effect Size=adjusted Odds ratio, CI= Confidence interval.*

***Figure 4-5** Decision curve to assess net benefit of symptom only and symptom + blood test models across a range of threshold probabilities.*

***Figure 4-6** Final logistic regression model predicting odds of haematological cancer 6 months following study entry/symptom onset in a) symptom-only and b) symptom + blood test model. Models contain fractional polynomials for age (-2,-2). Adjusted OR= Adjusted Odds ratio, CI= Confidence interval, Hb= haemoglobin, MCV= mean cell volume, ESR= Erythrocyte sedimentation rate, CRP=C-reactive protein.*

***Figure 4-7** Calibration of the a) symptom-only and b) symptom + blood test haematological cancer models tested demonstrated through calibration plots which show the alignment between predicted and observed risks across 10 risk deciles/groups.*

***Figure 4-8** Region-level and pooled performance estimates following internal-external cross validation for a) & d) area-under-the-curve (AUC), b) & e) calibration slope and, c) & f) calibration-in-the-large for symptom-only (a-c) and symptom + blood test haematological cancer models (d-f). Effect Size=adjusted Odds ratio, CI= Confidence interval.*

***Figure 4-9** Decision curve to assess net benefit of symptom only and symptom + blood test models across a range of threshold probabilities.*

Figure 4-10 *Final logistic regression model predicting odds of leukaemia 6 months following study entry/symptom onset in a) symptom-only and b) symptom + blood test model. Models contain fractional polynomials for age (-2,-2). Adjusted OR= Adjusted Odds ratio, CI= Confidence interval, Hb= haemoglobin.*

Figure 4-11 *Calibration of the a) symptom-only and b) symptom + blood test leukaemia models, demonstrated through calibration plots which show the alignment between predicted and observed risks across 10 risk deciles/groups.*

Figure 4-12 *Region-level and pooled performance estimates following internal-external cross validation for a) & d) area-under-the-curve (AUC), b) & e) calibration slope and, c) & f) calibration-in-the-large for symptom-only (a-c) and symptom + blood test leukaemia models (d-f). Effect Size=adjusted Odds ratio, CI= Confidence interval.*

Figure 4-13 *Decision curve to assess net benefit of symptom-only and symptom + blood test models across a range of threshold probabilities*

Figure 4-14 *Final logistic regression model predicting odds of lymphoma 6 months following study entry/symptom onset in a) symptom-only and b) symptom + blood test model. Models contain fractional polynomials (FP) for age (-2,-2), CI= Confidence interval, Hb= haemoglobin, CRP = C-reactive protein, ESR = Erythrocyte Sedimentation Rate.*

Figure 4-15 *Calibration of the a) symptom-only and b) symptom + blood test lymphoma models, demonstrated through calibration plots which show the alignment between predicted and observed risks across 10 risk deciles/groups.*

Figure 4-16 *Region-level and pooled performance estimates following internal-external cross validation for a) & d) area-under-the-curve (AUC), b) & e) calibration slope and, c) & f) calibration-in-the-large for symptom-only (a-c) and symptom + blood test lymphoma models (d-f). Effect Size=adjusted Odds ratio, CI= Confidence interval.*

Figure 4-17 *Decision curve to assess net benefit of models across a range of threshold probabilities for lymphoma models*

Figure 4-18 *Final logistic regression model predicting odds of CNS tumours 6 months following study entry/symptom onset. Models contain fractional polynomials (FP) for age (-2,-2), CI= Confidence interval*

Figure 4-19 *Calibration of CNS tumour model, demonstrated through a calibration plot which shows the alignment between predicted and observed risks across 10 risk deciles/groups.*

Figure 4-20 *Region-level and pooled performance estimates following internal-external cross validation for a) area-under-the-curve (AUC), b) calibration slope and, c) calibration-in-the-large for symptom-only CNS tumour models. Effect Size=adjusted Odds ratio, CI= Confidence interval.*

Figure 4-21 *Decision curve to assess net benefit of the CNS tumour model across a range of threshold probabilities*

Figure 5-1 *Forest plot demonstrating adjusted odds of emergency presentation compared to non-emergency presentation in CTYA cancers for each factor/exposure. 95% CI = 95% confidence interval.*

Figure 5-2 *Forest plot demonstrating adjusted odds of emergency presentation compared to non-emergency presentation in CTYA cancers for symptoms only.*

Figure 6-1 *Direct acyclic graph demonstrating causal pathways between diagnostic pathway (DI/Emergency) and mortality, with biasing paths through age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressive cancer, deprivation level, region and ethnic group. DI= diagnostic interval.*

Figure 6-2 *Forest plots demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and diagnostic intervals following Cox proportional regression using Approach 1.*

Figure 6-3 *Predicted probability of a) 1- and b) 5-year mortality across the diagnostic interval modelled using restricted cubic splines with 4 knots.*

Figure 6-4 *Forest plots demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and diagnostic intervals following Cox proportional regression using Approach 2.*

Figure 6-5 *Predicted probability of 1-year mortality across the restricted diagnostic intervals, a) 0-6 days; b) 7-25 days; c) 26-67 days (top panel). Predicted probability of 5-year mortality across the restricted diagnostic intervals, d) 0-6 days; e) 7-25 days; f) 26-67 days (bottom panel).*

Figure 6-6 *Forest plot demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and emergency presentations following Cox proportional regression.*

Figure 7-1 *Updated direct acyclic graph demonstrating causal pathways between diagnostic pathway (DI/Emergency) and outcomes-of-interest, with biasing paths through age-at-diagnosis, year-of-diagnosis, cancer subtype, aggressive cancer, cancer treatment, deprivation level, region and ethnic group. DI= diagnostic interval.*

Figure 7-2 *Forest plot showing risk (measured using adjusted hazard ratio) of psychotic and anxiety-depressive disorders in 1 year CTYA cancer survivors. 95% CI = 95% confidence interval*

Figure 7-3 *Forest plot showing risk (measured using adjusted hazard ratio) of psychotic and anxiety-depressive disorders in 5- year CTYA cancer survivors. 95% CI = 95% confidence interval*

Figure 7-4 *Forest plot demonstrating hazard ratios (with 95% confidence intervals) for diagnostic interval and adjusted factors following Cox proportional regression using Approach 1 in a) 1-year survivors and b) 5-year survivors.*

Figure 7-5 *Predicted probability of mental health disorder in 1-year and 5-year survivors of cancer across full diagnostic interval*

Figure 7-6 Forest plot demonstrating hazard ratios (with 95% confidence intervals) for diagnostic interval and adjusted factors following Cox proportional regression using Approach 2 in a) 1-year survivors and b) 5-year survivors.

Figure 7-7 *Predicted probability of 1-year anxiety-depressive disorders across the restricted diagnostic intervals, a) 0-5 days; b) 6-22 days; c) 23-62 days (top panel). Predicted probability of 5-year anxiety-depressive disorders across the restricted diagnostic intervals, d) 0-5 days; e) 6-22 days; f) 23-62 days (bottom panel).*

Figure 7-8 *Forest plot demonstrating hazard ratios (with 95% confidence intervals) of anxiety-depressive disorders for cancer-route and adjusted factors following Cox proportional regression in a) 1-year survivors and b) 5-year survivors.*

Figure 7-9 *Forest plot demonstrating hazard ratios (with 95% confidence intervals) of psychotic disorders for cancer-route and its association with psychotic disorders following Cox proportional regression in a) 1-year survivors and b) 5-year survivors.*

Supplementary Figure 9-1 *Venn diagrams depicting number of cancer diagnoses identified in each dataset for each cancer subtype.*

Figure 9-2 Proportion (%) of 4 ethnic categories for i) 2021 census population estimates ii) QResearch CTYA cohort across 9 geographical regions in England (75).

Supplementary Figure 9-3 Clinical features encountered in primary care associated with a) leukaemia, b) lymphoma, c) CNS tumours, d) sarcomas in children (0-14 years) in multivariable conditional logistic regression models. Early symptoms denoted with *. OR= Odds ratio CI= Confidence interval PPV=Positive Predictive Value.

Figure 9-4 Fractional polynomial terms for age for a) the logistic model for solid cancers and b) logistic model for haematological cancers c) logistic model for leukaemias, d) logistic model for lymphomas, e) logistic model for CNS tumours

Figure 9-5 Forest plots demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and diagnostic intervals following Cox proportional regression and multiple imputation using Approach 1.

Figure 9-6 Forest plots demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and diagnostic intervals following Cox proportional regression and multiple imputation using Approach 2.

Figure 9-7 Forest plot demonstrating adjusted hazard ratios (with 95% confidence intervals) for a) 1- and b) 5-year mortality and emergency presentations following Cox proportional regression.

Figure 9-8 Forest plot showing risk (measured using adjusted hazard ratio) of psychotic and anxiety-depressive disorders in a) 1 year and b) 5-year CTYA cancer survivors. 95% CI = 95% confidence interval

Figure 9-9 *Forest plot demonstrating hazard ratios (with 95% confidence intervals) of anxiety-depressive disorders for cancer-route and adjusted factors following Cox proportional regression and multiple imputation in a) 1-year survivors and b) 5-year survivors.*

Figure 9-10 *Forest plot demonstrating hazard ratios (with 95% confidence intervals) of psychotic disorders for cancer-route and adjusted factors following Cox proportional regression and multiple imputation in a) 1-year survivors and b) 5-year survivors.*

Chapter 11 Glossary

ACE	Accelerate, Coordinate Evaluate
AUC	Area under the curve
CCLG	Childhood Cancer and Leukaemia Group
CDT	Clinical Decision Tool
CI	Confidence Interval
CNS	Central nervous system
COVID-19	Coronavirus-2019
CPRD	Clinical Practice Research Datalink
CRUK	Cancer Research UK
CTYA	Children, Teenager and Young Adult
EHR	Electronic healthcare record
FHIR	Fast Healthcare Interoperability Resources
GP	General practice
HES	Hospital Episode Statistics
HR	Hazard Ratio
ICCC-3	International Classification of Childhood Cancers – 3
ICD-10	International Classification of Diseases- 10
ICD-O-3	International Classification of Diseases for Oncology – 3
KMO	Kaiser Meyer Olkin
NCR	National Cancer Registry
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NPV	Negative predictive value
ONS	Office for National Statistics
OPCS-4	Classification of Interventions and Procedures - 4
OR	Odds ratio
PC	Principal component

PCA	Principal Component Analysis
PPV	Positive predictive value
RECORD	REporting of studies Conducted using Observational Routinely-collected health Data
SACT	Systemic Anti-Cancer Therapy
SNOMED-CT	Systematized Nomenclature of Medicine -- Clinical Terms
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
TWW	Two week wait
TYA	Teenager, Young Adult
UK	United Kingdom

Chapter 12 References

1. Public Health England. *Children, teenagers and young adults UK cancer statistics report* 2021 [cited 2024 4th August]; Available from: https://digital.nhs.uk/ndrs/data/data-outputs/cancer-publications-and-tools/ctya-uk-cancer-statistics-report-2021#chapter-index.
2. Steliarova-Foucher, E., et al., *International Classification of Childhood Cancer, third edition.* Cancer, 2005. **103**(7): p. 1457-67.
3. Stiller, C.A., *Global burden of childhood cancer: growing, but controllable.* Lancet Oncol, 2019. **20**(9): p. 1184-1185.
4. Bhakta, N., et al., *Childhood cancer burden: a review of global estimates.* Lancet Oncol, 2019. **20**(1): p. e42-e53.
5. Botta, L., et al., *Long-term survival and cure fraction estimates for childhood cancer in Europe (EUROCARE-6): results from a population-based study.* Lancet Oncol, 2022. **23**(12): p. 1525-1536.
6. Cancer Research UK. 2024; Available from: https://crukcancerintelligence.shinyapps.io/CYPCancerData/.
7. Walker, D.A., *Helping GPs to diagnose children's cancer.* Br J Gen Pract, 2021. **71**(705): p. 151-152.
8. Force, L.M., et al., *The global burden of childhood and adolescent cancer in 2017: an analysis of the Global Burden of Disease Study 2017.* The Lancet Oncology, 2019. **20**(9): p. 1211-1225.
9. Royal College of Paediatrics and Child Health. *State of Child Health: Child Mortality*. 2020; Available from: https://stateofchildhealth.rcpch.ac.uk/evidence/mortality/child-mortality/.
10. Fidler-Benaoudia, M.M., et al., *A Comparison of Late Mortality Among Survivors of Childhood Cancer in the United States and United Kingdom.* J Natl Cancer Inst, 2021. **113**(5): p. 562-571.
11. Aldiss, S., et al., *Research priorities for young people with cancer: a UK priority setting partnership with the James Lind Alliance.* BMJ Open, 2019. **9**(8): p. e028119.
12. Hanna, T.P., et al., *Mortality due to cancer treatment delay: systematic review and meta-analysis.* BMJ, 2020. **371**: p. m4087.
13. Torring, M.L., et al., *Evidence of increasing mortality with longer diagnostic intervals for five common cancers: a cohort study in primary care.* Eur J Cancer, 2013. **49**(9): p. 2187-98.
14. Love, R.R., *The Natural History of Cancer in Humans*, in *Manual of Clinical Oncology*, C.D. Sherman, et al., Editors. 1990, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 20-30.
15. Brasme, J.F., et al., *Delays in diagnosis of paediatric cancers: a systematic review and comparison with expert testimony in lawsuits.* Lancet Oncol, 2012. **13**(10): p. e445-59.
16. Forster, A.S., et al., *Associations between diagnostic time intervals and health-related quality of life, clinical anxiety and depression in adolescents and young adults with cancer: cross-sectional analysis of the BRIGHTLIGHT cohort.* Br J Cancer, 2022. **126**(12): p. 1725-1734.
17. World Health Organisation. *Guide to Cancer Early Diagnosis*. 2017; Available from: https://www.who.int/publications/i/item/9789241511940.

18. National Health Service Long Term Plan. *NHS Long Term Plan*. Available from: https://www.longtermplan.nhs.uk/online-version/chapter-3-further-progress-on-care-quality-and-outcomes/a-strong-start-in-life-for-children-and-young-people/children-and-young-people-with-cancer/.
19. Cancer Strategy Taskforce. *Statement of Intent: Cancer Strategy for England: 2015 – 2020.* . 2015; Available from: https://www.cancerresearchuk.org/sites/default/files/statement_of_intent_-final_0.pdf
20. Childhood Cancer and Leukaemia Group. *Strategic Plan 2018-2025*. 2018; Available from: https://www.cclg.org.uk/write/MediaUploads/About%20CCLG/CCLG_Strategic_Plan_2018-2025.pdf
21. Herbert, A., et al., *Diagnostic timeliness in adolescents and young adults with cancer: a cross-sectional analysis of the BRIGHTLIGHT cohort.* Lancet Child Adolesc Health, 2018. **2**(3): p. 180-190.
22. Ahrensberg, J.M., et al., *The initial cancer pathway for children - one-fourth wait more than 3 months.* Acta Paediatr, 2012. **101**(6): p. 655-62.
23. Dang-Tan, T. and E.L. Franco, *Diagnosis delays in childhood cancer: a review.* Cancer, 2007. **110**(4): p. 703-13.
24. Ahrensberg, J.M., M. Fenger-Gron, and P. Vedsted, *Use of primary care during the year before childhood cancer diagnosis: a nationwide population-based matched comparative study.* PLoS One, 2013. **8**(3): p. e59098.
25. Dommett, R.M., et al., *Features of childhood cancer in primary care: a population-based nested case-control study.* Br J Cancer, 2012. **106**(5): p. 982-7.
26. Dommett, R.M., et al., *Features of cancer in teenagers and young adults in primary care: a population-based nested case-control study.* Br J Cancer, 2013. **108**(11): p. 2329-33.
27. Pritchard-Jones, K., et al., *Evidence for a delay in diagnosis of Wilms' tumour in the UK compared with Germany: implications for primary care for children.* Arch Dis Child, 2016. **101**(5): p. 417-20.
28. HeadSmart Be Brain Tumour, A., *A new clinical guideline from the Royal College of Paediatrics and Child Health with a national awareness campaign accelerates brain tumor diagnosis in UK children--"HeadSmart: Be Brain Tumour Aware".* Neuro Oncol, 2016. **18**(3): p. 445-54.
29. Walker, D., et al., *Strategies to accelerate diagnosis of primary brain tumors at the primary-secondary care interface in children and adults.* CNS Oncol, 2013. **2**(5): p. 447-62.
30. The Health Foundation. *HeadSmart Be Brain Tumour Aware: Final Report for Closing the Gap through Clinical Communities* 2013; Available from: https://www.health.org.uk/sites/default/files/CtGtCC_HeadSmart_report.pdf.
31. Whitfield, E., et al., *A taxonomy of early diagnosis research to guide study design and funding prioritisation.* Br J Cancer, 2023. **129**(10): p. 1527-1534.
32. Weller, D., et al., *The Aarhus statement: improving design and reporting of studies on early cancer diagnosis.* Br J Cancer, 2012. **106**(7): p. 1262-7.
33. Petrova, D., et al., *The patient, diagnostic, and treatment intervals in adult patients with cancer from high- and lower-income countries: A systematic review and meta-analysis.* PLoS Med, 2022. **19**(10): p. e1004110.

34. NHS Digital National Disease Registration Service. *Routes to Diagnosis*. 2022; Available from: https://digital.nhs.uk/ndrs/our-work/ncras-work-programme/routes-to-diagnosis.
35. Wilne, S., et al., *Presentation of childhood CNS tumours: a systematic review and meta-analysis.* Lancet Oncol, 2007. **8**(8): p. 685-95.
36. Saatci, D., et al., *Presentation of B-cell lymphoma in childhood and adolescence: a systematic review and meta-analysis.* BMC Cancer, 2024. **24**(1): p. 718.
37. Ni Cheallaigh L, L.J., Fern L, et al, *Clinical presentation of childhood soft tissue sarcomas: a systematic review and meta-analysis.* 2023.
38. Clarke, R.T., et al., *Clinical presentation of childhood leukaemia: a systematic review and meta-analysis.* Arch Dis Child, 2016. **101**(10): p. 894-901.
39. Ahrensberg, J.M., et al., *Childhood cancer and factors related to prolonged diagnostic intervals: a Danish population-based study.* Br J Cancer, 2013. **108**(6): p. 1280-7.
40. Rafiq, M., et al., *Predictive value of abnormal blood tests for detecting cancer in primary care patients with nonspecific abdominal symptoms: A population-based cohort study of 477,870 patients in England.* PLoS Med, 2024. **21**(7): p. e1004426.
41. Cranfield, B.M., et al., *Primary care blood tests before cancer diagnosis: National Cancer Diagnosis Audit data.* Br J Gen Pract, 2023. **73**(727): p. e95-e103.
42. Koo, M.M., et al., *Symptom Signatures and Diagnostic Timeliness in Cancer Patients: A Review of Current Evidence.* Neoplasia, 2018. **20**(2): p. 165-174.
43. Kawamoto, K., et al., *Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success.* BMJ, 2005. **330**(7494): p. 765.
44. Kawamoto, K., et al., *System-agnostic clinical decision support services: benefits and challenges for scalable decision support.* Open Med Inform J, 2010. **4**: p. 245-54.
45. McDonald, K.M., et al., *Patient safety strategies targeted at diagnostic errors: a systematic review.* Ann Intern Med, 2013. **158**(5 Pt 2): p. 381-9.
46. Usher-Smith, J., et al., *Risk prediction tools for cancer in primary care.* Br J Cancer, 2015. **113**(12): p. 1645-50.
47. Friis Abrahamsen, C., J.M. Ahrensberg, and P. Vedsted, *Utilisation of primary care before a childhood cancer diagnosis: do socioeconomic factors matter?: A Danish nationwide population-based matched cohort study.* BMJ Open, 2018. **8**(8): p. e023569.
48. Martins, T., et al., *Assessing Ethnic Inequalities in Diagnostic Interval of Common Cancers: A Population-Based UK Cohort Study.* Cancers (Basel), 2022. **14**(13).
49. Renzi, C. and G. Lyratzopoulos, *Comorbidity and the diagnosis of symptomatic-but-as-yet-undiagnosed cancer.* Br J Gen Pract, 2020. **70**(698): p. e598-e599.
50. Nicholson, B.D., et al., *Consultations for clinical features of possible cancer and associated urgent referrals before and during the COVID-19 pandemic: an observational cohort study from English primary care.* Br J Cancer, 2022. **126**(6): p. 948-956.
51. Quinn-Scoggins, H.D., et al., *Cancer symptom experience and help-seeking behaviour during the COVID-19 pandemic in the UK: a cross-sectional population survey.* BMJ Open, 2021. **11**(9): p. e053095.
52. Shanmugavadivel, D., et al., *Accelerating diagnosis for childhood brain tumours: an analysis of the HeadSmart UK population data.* Arch Dis Child, 2020. **105**(4): p. 355-362.

53. Taylor, R.M., et al., *Involving young people in BRIGHTLIGHT from study inception to secondary data analysis: insights from 10 years of user involvement.* Res Involv Engagem, 2018. **4**: p. 50.
54. Dixon-Woods, M., et al., *Parents' accounts of obtaining a diagnosis of childhood cancer.* Lancet, 2001. **357**(9257): p. 670-4.
55. Clarke, R.T., et al., *'Shouting from the roof tops': a qualitative study of how children with leukaemia are diagnosed in primary care.* BMJ Open, 2014. **4**(2): p. e004640.
56. Dommett, R.M., et al., *Risk of childhood cancer with symptoms in primary care: a population-based case-control study.* Br J Gen Pract, 2013. **63**(606): p. e22-9.
57. Vezyridis, P. and S. Timmons, *Evolution of primary care databases in UK: a scientometric analysis of research output.* BMJ Open, 2016. **6**(10): p. e012785.
58. QResearch Database. *The QResearch Database*. Available from: www.qresearch.org.
59. NHS Digital. 2024; Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics.
60. Office for National Statistics. *Population Estimates for England and Wales, Scotland and Northern Ireland*. [cited 2024 18/07]; Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland.
61. Bright, C.J., et al., *Data Resource Profile: The Systemic Anti-Cancer Therapy (SACT) dataset.* Int J Epidemiol, 2020. **49**(1): p. 15-15l.
62. Henson, K.E., et al., *Data Resource Profile: National Cancer Registration Dataset in England.* Int J Epidemiol, 2020. **49**(1): p. 16-16h.
63. Office for National Statistics. *Cancer registration statistics*. 2016; Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/cancerregistrationstatisticsqmi.
64. Office for National Statistics. *Ethnic group, England and Wales Census 2021*. 2021; Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021.
65. Townsend P, D.N., *Inequalities in Health: The Black Report*. 1982: Dept of Health and Social Security.
66. Wardle, M. and A. Spencer, *Implementation of SNOMED CT in an online clinical database.* Future Healthc J, 2017. **4**(2): p. 126-130.
67. Vorisek, C.N., et al., *Fast Healthcare Interoperability Resources (FHIR) for Interoperability in Health Research: Systematic Review.* JMIR Med Inform, 2022. **10**(7): p. e35724.
68. Public Health England. *NIPE newborn eye screening: screen positive pathway*. 2024; Available from: https://www.gov.uk/government/publications/newborn-and-infant-physical-examination-programme-handbook/nipe-newborn-eye-screening-screen-positive-pathway.
69. StataCorp. *Stptime*. 2023 [cited 2024 18/07]; Available from: https://www.stata.com/manuals/ststptime.pdf.
70. Stiller, C.A., *International patterns of cancer incidence in adolescents.* Cancer Treat Rev, 2007. **33**(7): p. 631-45.
71. Kroll, M.E., et al., *Childhood leukaemia and socioeconomic status in England and Wales 1976-2005: evidence of higher incidence in relatively affluent communities persists over time.* Br J Cancer, 2011. **105**(11): p. 1783-7.

72. Hippisley-Cox, J. *Validation of the UK QResearch Database*. 2007; Available from: https://www.qresearch.org/about/ethics-and-confidentiality/validation-and-initial-reports/.
73. Andrews, C., et al., *OpenSAFELY: Representativeness of electronic health record platform OpenSAFELY-TPP data compared to the population of England.* Wellcome Open Res, 2022. **7**: p. 191.
74. Herrett, E., et al., *Data Resource Profile: Clinical Practice Research Datalink (CPRD).* Int J Epidemiol, 2015. **44**(3): p. 827-36.
75. Shiekh, S.I., et al., *Completeness, agreement, and representativeness of ethnicity recording in the United Kingdom's Clinical Practice Research Datalink (CPRD) and linked Hospital Episode Statistics (HES).* Popul Health Metr, 2023. **21**(1): p. 3.
76. Kaseb, H., M.A. Tariq, and G. Gupta, *Lymphoblastic Lymphoma*, in *StatPearls*. 2024: Treasure Island (FL) ineligible companies. Disclosure: Muhammad Ali Tariq declares no relevant financial relationships with ineligible companies. Disclosure: Gunjan Gupta declares no relevant financial relationships with ineligible companies.
77. Murray, M.J., et al., *Delphi method to identify expert opinion to support children's cancer referral guidelines.* Arch Dis Child, 2020. **105**(3): p. 241-246.
78. National Institute for Health and Care Excellence. *NICE Suspected Cancer Guidelines*. Available from: https://www.nice.org.uk/guidance/ng12/resources/research-recommendations-from-an-individual-piece-of-guidance.
79. Richardson, D.B., *An incidence density sampling program for nested case-control analyses.* Occup Environ Med, 2004. **61**(12): p. e59.
80. Grace Kelly Trust, *Regional Red Flags Childhood Cancer.*
81. Sayeed, S., I. Barnes, and R. Ali, *Childhood cancer incidence by ethnic group in England, 2001-2007: a descriptive epidemiological study.* BMC Cancer, 2017. **17**(1): p. 570.
82. Kehm, R.D., et al., *Socioeconomic Status and Childhood Cancer Incidence: A Population-Based Multilevel Analysis.* Am J Epidemiol, 2018. **187**(5): p. 982-991.
83. Chowdhury, M.Z.I. and T.C. Turin, *Variable selection strategies and its importance in clinical prediction modelling.* Fam Med Community Health, 2020. **8**(1): p. e000262.
84. van Zaane, B., et al., *Comparison of approaches to estimate confidence intervals of post-test probabilities of diagnostic test results in a nested case-control study.* BMC Med Res Methodol, 2012. **12**: p. 166.
85. Mercaldo, N.D., K.F. Lau, and X.H. Zhou, *Confidence intervals for predictive values with an emphasis to case-control studies.* Stat Med, 2007. **26**(10): p. 2170-83.
86. Jolliffe, I.T. and J. Cadima, *Principal component analysis: a review and recent developments.* Philos Trans A Math Phys Eng Sci, 2016. **374**(2065): p. 20150202.
87. University of California Los Angeles Advanced Research Computing. *Principal Components Analysis*. Available from: https://stats.oarc.ucla.edu/spss/output/principal_components/.
88. Clift, A.K., et al., *Sickle Cell Disorders and Severe COVID-19 Outcomes: A Cohort Study.* Ann Intern Med, 2021. **174**(10): p. 1483-1487.
89. Saatci, D., et al., *Association Between Race and COVID-19 Outcomes Among 2.6 Million Children in England.* JAMA Pediatr, 2021. **175**(9): p. 928-938.
90. Sterne, J.A., et al., *Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls.* BMJ, 2009. **338**: p. b2393.
91. White, I.R., P. Royston, and A.M. Wood, *Multiple imputation using chained equations: Issues and guidance for practice.* Stat Med, 2011. **30**(4): p. 377-99.
92. StataCorp., *Stata Statistical Software: Release 17*. 2023, College Station, TX: StataCorp LLC.

93. von Elm, E., et al., *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies.* Lancet, 2007. **370**(9596): p. 1453-7.
94. Benchimol, E.I., et al., *[The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement].* Z Evid Fortbild Qual Gesundhwes, 2016. **115-116**: p. 33-48.
95. HealthTalk. *Signs and Symptoms for Cancer in Young People* [cited 2023 15/04]; Available from: ttps://healthtalk.org/cancer/signs-and-symptoms-of-cancer-in-young-people.
96. Koo, M.M., et al., *Association of Self-reported Presenting Symptoms With Timeliness of Help-Seeking Among Adolescents and Young Adults With Cancer in the BRIGHTLIGHT Study.* JAMA Netw Open, 2020. **3**(9): p. e2015437.
97. Rafiq, M., et al., *Clinical activity in general practice before sarcoma diagnosis: an Australian cohort study.* Br J Gen Pract, 2024. **74**(745): p. e508-e516.
98. Hippisley-Cox, J. and C. Coupland, *Symptoms and risk factors to identify people with suspected cancer in primary care.* Br J Gen Pract, 2013. **63**(608): p. 125-6.
99. Scales, S. *Science Surgery: 'How are children's cancers different from adults' cancers?'*. 2020; Available from: https://news.cancerresearchuk.org/2020/09/28/science-surgery-how-are-childrens-cancers-different-from-adults-cancers/.
100. Haj-Hassan, T.A., et al., *Which early 'red flag' symptoms identify children with meningococcal disease in primary care?* Br J Gen Pract, 2011. **61**(584): p. e97-104.
101. Holdsworth, F., et al., *A guide to Hodgkin and non-Hodgkin lymphomas: similarities and differences.* Br J Nurs, 2021. **30**(17): p. S16-S22.
102. Dommett, R.M., et al., *Achieving a timely diagnosis for teenagers and young adults with cancer: the ACE "too young to get cancer?" study.* BMC Cancer, 2019. **19**(1): p. 616.
103. Sutton, R.T., et al., *An overview of clinical decision support systems: benefits, risks, and strategies for success.* NPJ Digit Med, 2020. **3**: p. 17.
104. Chima, S., et al., *Decision support tools to improve cancer diagnostic decision making in primary care: a systematic review.* Br J Gen Pract, 2019. **69**(689): p. e809-e818.
105. Segal, I., et al., *Abnormal liver transaminases and conjugated hyperbilirubinemia at presentation of acute lymphoblastic leukemia.* Pediatr Blood Cancer, 2010. **55**(3): p. 434-9.
106. Ruka, W., et al., *Alterations of routine blood tests in adult patients with soft tissue sarcomas: relationships to cytokine serum levels and prognostic significance.* Ann Oncol, 2001. **12**(10): p. 1423-32.
107. Rafiq, M., et al., *Inflammatory marker testing in primary care in the year before Hodgkin lymphoma diagnosis: a UK population-based case-control study in patients aged </=50 years.* Br J Gen Pract, 2022. **72**(721): p. e546-e555.
108. Kakaje, A., et al., *Rates and trends of childhood acute lymphoblastic leukaemia: an epidemiology study.* Sci Rep, 2020. **10**(1): p. 6756.
109. Royal College of Paediatrics and Child Health. *RCPCH Laboratory Reference Ranges*. 2016 [cited 2024 31/05]; Available from: https://www.rcpch.ac.uk/sites/default/files/rcpch/HTWQv8.7/Reference%20ranges%20Feb%2018%20FINAL.pdf.
110. Watson, J., et al., *Predictive value of inflammatory markers for cancer diagnosis in primary care: a prospective cohort study using electronic health records.* Br J Cancer, 2019. **120**(11): p. 1045-1051.

111. Hippisley-Cox, J., et al., *Development and validation of a new algorithm for improved cardiovascular risk prediction.* Nat Med, 2024. **30**(5): p. 1440-1447.
112. Hippisley-Cox, J., C. Coupland, and P. Brindle, *Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study.* BMJ, 2017. **357**: p. j2099.
113. Steyerberg, E.W. and F.E. Harrell, Jr., *Prediction models need appropriate internal, internal-external, and external validation.* J Clin Epidemiol, 2016. **69**: p. 245-7.
114. Collins, G.S., et al., *Evaluation of clinical prediction models (part 1): from development to external validation.* BMJ, 2024. **384**: p. e074819.
115. IntHout, J., J.P. Ioannidis, and G.F. Borm, *The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method.* BMC Med Res Methodol, 2014. **14**: p. 25.
116. Van Calster, B., et al., *Calibration: the Achilles heel of predictive analytics.* BMC Med, 2019. **17**(1): p. 230.
117. de Hond, A.A.H., E.W. Steyerberg, and B. van Calster, *Interpreting area under the receiver operating characteristic curve.* Lancet Digit Health, 2022. **4**(12): p. e853-e855.
118. Riley, R.D., et al., *A guide to systematic review and meta-analysis of prognostic factor studies.* BMJ, 2019. **364**: p. k4597.
119. Higgins, J.P., et al., *Measuring inconsistency in meta-analyses.* BMJ, 2003. **327**(7414): p. 557-60.
120. Vickers, A.J., B. Van Calster, and E.W. Steyerberg, *Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests.* BMJ, 2016. **352**: p. i6.
121. Collins, G.S., et al., *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD).* Ann Intern Med, 2015. **162**(10): p. 735-6.
122. Riley, R.D., et al., *Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes.* Stat Med, 2019. **38**(7): p. 1276-1296.
123. Saatci, D., et al., *Childhood, teenage and young adult cancer diagnosis during the first wave of the COVID-19 pandemic: a population-based observational cohort study in England.* Arch Dis Child, 2022. **107**(8): p. 740-746.
124. Accelerate Coordinate Evaluate Programme. *Using Cancer Decision Support Tools to support the early diagnosis of cancer*. 2017 [cited 2024 31/07]; Available from: https://www.cancerresearchuk.org/sites/default/files/ace_cancer_decision_support_tools_final_report_v1.1_080517.pdf.
125. Huang, C., et al., *Bone and bone marrow involvement in neuroblastoma: A case report.* Medicine (Baltimore), 2020. **99**(40): p. e22505.
126. Hart, P.C., et al., *C-Reactive Protein and Cancer-Diagnostic and Therapeutic Insights.* Front Immunol, 2020. **11**: p. 595835.
127. Barlow, P., et al., *Area-level deprivation and geographic factors influencing utilisation of General Practitioner services.* SSM Popul Health, 2021. **15**: p. 100870.
128. NICE. *2024 exceptional surveillance of suspected cancer: recognition and referral (NICE guideline NG12)* 2024; Available from: https://www.nice.org.uk/guidance/ng12/resources/2024-exceptional-surveillance-of-suspected-cancer-recognition-and-referral-nice-guideline-ng12-pdf-17407864344517.
129. Rask, O., et al., *Prospective registration of symptoms and times to diagnosis in children and adolescents with central nervous system tumors: A study of the Swedish Childhood Cancer Registry.* Pediatr Blood Cancer, 2022. **69**(11): p. e29850.

130. Benitez Majano, S., et al., *Mental Health Morbidities and Time to Cancer Diagnosis Among Adults With Colon Cancer in England.* JAMA Netw Open, 2022. **5**(10): p. e2238569.
131. Pearson C, W.L. *Diagnostic intervals for cancer, analysis by Route to Diagnosis in England, 2014 to 2015*. 2020; Available from: https://www.gov.uk/government/statistics/cancer-patients-in-england-diagnostic-intervals/diagnostic-intervals-for-cancer-analysis-by-route-to-diagnosis-in-england-2014-to-2015.
132. Martins, T., et al., *Ethnic inequalities in routes to diagnosis of cancer: a population-based UK cohort study.* Br J Cancer, 2022. **127**(5): p. 863-871.
133. Kehoe, K., et al., *Delay in the diagnosis of paediatric brain tumours: a systematic review.* Childs Nerv Syst, 2023. **39**(8): p. 2053-2063.
134. Azizi, A.A., et al., *From Symptom to Diagnosis-The Prediagnostic Symptomatic Interval of Pediatric Central Nervous System Tumors in Austria.* Pediatr Neurol, 2017. **76**: p. 27-36.
135. Ruco, A., et al., *Factors Associated with the Breast Cancer Diagnostic Interval across Five Canadian Provinces: A CanIMPACT Retrospective Cohort Study.* Cancers (Basel), 2023. **15**(2).
136. Petrovici, I., M. Ionica, and O.C. Neagoe, *Economic Crisis: A Factor for the Delayed Diagnosis of Breast Cancer.* Int J Environ Res Public Health, 2021. **18**(8).
137. de Aguirre-Neto, J.C., et al., *International Comparisons of Clinical Demographics and Outcomes in the International Society of Pediatric Oncology Wilms Tumor 2001 Trial and Study.* JCO Glob Oncol, 2022. **8**: p. e2100425.
138. McPhail, S., et al., *Risk factors and prognostic implications of diagnosis of cancer within 30 days after an emergency hospital admission (emergency presentation): an International Cancer Benchmarking Partnership (ICBP) population-based study.* Lancet Oncol, 2022. **23**(5): p. 587-600.
139. Roskin, J., J. Diviney, and V. Nanduri, *Presentation of childhood cancers to a paediatric shared care unit.* Arch Dis Child, 2015. **100**(12): p. 1131-5.
140. Le Cook, B. and W.G. Manning, *Thinking beyond the mean: a practical guide for using quantile regression methods for health services research.* Shanghai Arch Psychiatry, 2013. **25**(1): p. 55-9.
141. Nikitina, L., R. Paidi, and F. Furuoka, *Using bootstrapped quantile regression analysis for small sample research in applied linguistics: Some methodological considerations.* PLoS One, 2019. **14**(1): p. e0210668.
142. Hughes, R.A., et al., *Accounting for missing data in statistical analyses: multiple imputation is not always the answer.* Int J Epidemiol, 2019. **48**(4): p. 1294-1304.
143. Loh, A.H., et al., *Diagnostic delay in pediatric solid tumors: a population based study on determinants and impact on outcomes.* Pediatr Blood Cancer, 2012. **58**(4): p. 561-5.
144. Halperin, E.C. and H.S. Friedman, *Is there a correlation between duration of presenting symptoms and stage of medulloblastoma at the time of diagnosis?* Cancer, 1996. **78**(4): p. 874-80.
145. NHS Digital National Disease Registration Service, *Emergency Presentations.*
146. Lethaby, C.D., et al., *A systematic review of time to diagnosis in children and young adults with cancer.* Arch Dis Child, 2013. **98**(5): p. 349-55.
147. Harrell, F.E., Jr., *Regression Modeling Strategies*. 2001.
148. Von Hippel, P., *HOW TO IMPUTE INTERACTIONS, SQUARES, AND OTHER TRANSFORMED VARIABLES.* Sociological Methodology, 2009. **39**(1): p. 265-291.

149. StataCorp. *stpower*. 2024; Available from: https://www.stata.com/manuals13/ststpowerlogrank.pdf.
150. Fidler, M.M., et al., *Long term cause specific mortality among 34 489 five year survivors of childhood cancer in Great Britain: population based cohort study.* BMJ, 2016. **354**: p. i4351.
151. Oeffinger, K.C., et al., *Chronic health conditions in adult survivors of childhood cancer.* N Engl J Med, 2006. **355**(15): p. 1572-82.
152. Erdmann, F., et al., *Childhood cancer: Survival, treatment modalities, late effects and improvements over time.* Cancer Epidemiol, 2021. **71**(Pt B): p. 101733.
153. Wang, Y., et al., *Subsequent female breast cancer risk associated with anthracycline chemotherapy for childhood cancer.* Nat Med, 2023. **29**(9): p. 2268-2277.
154. Veiga, L.H., et al., *Association of Breast Cancer Risk After Childhood Cancer With Radiation Dose to the Breast and Anthracycline Use: A Report From the Childhood Cancer Survivor Study.* JAMA Pediatr, 2019. **173**(12): p. 1171-1179.
155. Mueller, S., et al., *Radiation, atherosclerotic risk factors, and stroke risk in survivors of pediatric cancer: a report from the Childhood Cancer Survivor Study.* Int J Radiat Oncol Biol Phys, 2013. **86**(4): p. 649-55.
156. Lee, A., et al., *Lifetime Burden of Psychological Symptoms, Disorders, and Suicide Due to Cancer in Childhood, Adolescent, and Young Adult Years: A Systematic Review and Meta-analysis.* JAMA Pediatr, 2023. **177**(8): p. 790-799.
157. Mind. *How common are mental health problems?* 2024; Available from: https://www.mind.org.uk/information-support/types-of-mental-health-problems/statistics-and-facts-about-mental-health/how-common-are-mental-health-problems/.
158. Royal College of Paediatrics and Child Health, *Prevalence of mental health conditions.* 2020.
159. NHS Digital. *OPCS Classification of Interventions and Procedures version 4*. 2024 [cited 2024 03/02]; Available from: https://classbrowser.nhs.uk/ref_books/OPCS-4.9_NCCS-2021.pdf.
160. Frederiksen, L.E., et al., *Employment status and occupational positions of childhood cancer survivors from Denmark, Finland and Sweden: A Nordic register-based cohort study from the SALiCCS research programme.* Lancet Reg Health Eur, 2022. **12**: p. 100258.
161. National Institute for Health and Care Excellence Clinical Knowledge Summaries. *Risk factors for depression*. 2024; Available from: https://cks.nice.org.uk/topics/depression/background-information/risk-factors/.
162. Liu, J.F., et al., *Public awareness of childhood, teenager and young adult cancer signs and symptoms in Great Britain: a cross-sectional survey.* Arch Dis Child, 2023. **108**(12): p. 987-993.
163. Nunan, D., J. Aronson, and C. Bankhead, *Catalogue of bias: attrition bias.* BMJ Evid Based Med, 2018. **23**(1): p. 21-22.
164. Coughlin, S.S., *Recall bias in epidemiologic studies.* J Clin Epidemiol, 1990. **43**(1): p. 87-91.
165. Mahtani, K., et al., *Catalogue of bias: observer bias.* BMJ Evid Based Med, 2018. **23**(1): p. 23-24.
166. Hofler, M., *The effect of misclassification on the estimation of association: a review.* Int J Methods Psychiatr Res, 2005. **14**(2): p. 92-101.
167. Office for National Statistics. *Cancer survival by stage at diagnosis for England*. 2020 [cited 2024 05/08]; Available from:

https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalbystageatdiagnosisforenglandexperimentalstatistics/adultsdiagnosed20122013and2014andfollowedupto2015.

168. American Cancer Society. *Prognostic factors in childhood leukemia*. 2020 [cited 2024 05/08]; Available from: https://www.cancer.org/cancer/types/leukemia-in-children/detection-diagnosis-staging/prognostic-factors.html.