Investigating long-term outcomes and cure for children and young people diagnosed with cancer

Lesley Smith

Submitted in accordance with the requirement for the degree of

Doctor of Philosophy

The University Of Leeds

School of Medicine

October 2019

Intellectual Property and Publication Statement

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

Chapters 5, 6 and 8 contain work based on the following publications and conference presentations.

Publications

Long-term survival after childhood acute lymphoblastic leukaemia: population-based trends in cure and relapse by clinical characteristics.

Smith L, Glaser AW, Kinsey SE, Greenwood DC, Chilton L, Moorman AV, Feltbower RG.

British Journal of Haematology 2018; 182: 851-858 doi.org/10.1111/bjh.15424

Contributor's statement: LS, AG and RF designed the research study. LS analysed the data and drafted the manuscript. LS, AG, SK, DG, LC, AVM and RGF interpreted the results and critically reviewed the manuscript. All authors approved the final version of this paper.

Respiratory morbidity in young people surviving cancer: population-based study of hospital admissions, treatment-related risk factors and subsequent mortality.

Smith L, Glaser AW, Peckham D, Greenwood DC, Feltbower RG.

International Journal of Cancer 2019; 145: 20-28 doi.org/10.1002/ijc.32066

Contributor's statement: LS, AG and RF designed the study. LS analysed the data and drafted the manuscript. LS, AG, DG, DP and RF contributed to the in interpretation of the results and critical revision of the study. All authors approved the final study.

Cumulative burden of subsequent neoplasms, cardiovascular and respiratory morbidity in young people surviving cancer

Smith L, Glaser AW, Greenwood DC, Feltbower RG.

International Journal of Cancer 2019 (submitted September 2019)

Contributor's statement: LS AG and RF designed the study. LS analysed the data and drafted the manuscript. LS, AG, DG, and RF contributed to the in interpretation of the results and critical revision of the study. All authors approved the final study.

Conference Presentations

Cumulative burden of subsequent neoplasms, cardiovascular and respiratory morbidity in long-term survivors of childhood and young adult cancer

Smith L, Glaser AW, Greenwood DC, Feltbower RG

51st Congress of the International Society of Paediatric Oncology, October 2019, Lyon (Poster)

Risk and impact of respiratory hospitalisation among childhood and young adult cancer survivors

Smith L, Glaser AW, Peckham D, Greenwood DC, Feltbower RG

British Thoracic Society Winter Meeting 2018, December 2018, London (Poster)

Long-term respiratory disease among childhood and young adult cancer survivors

Smith L, Glaser AW, Peckham D, Greenwood DC, Feltbower RG

Public Health England Cancer Services, Data and Outcomes Conference 2018, June 2018, Manchester (Oral)

Trends in cure and relapse by clinical characteristics for children diagnosed with leukaemia aged 0-17 years in Yorkshire 1990-2009: a population-based study

Smith L, Glaser AW, Kinsey SE, Greenwood DC, Feltbower RG

Society for Social Medicine 61st Annual Scientific Meeting, September 2017, Manchester (Oral)

Trends in cure and relapse by clinical characteristics for children diagnosed with leukaemia aged 0-17 years in Yorkshire 1990-2009: a population-based study

Smith L, Glaser AW, Kinsey SE, Greenwood DC, Feltbower RG

Public Health England Cancer Data and Outcomes Conference 2017, June 2017, Manchester (Oral)

Acknowledgements

Firstly, this project uses data provided by patients and collected by the NHS as part of their care and support. This study would not have been possible without these data. I thank NHS Digital for the provision of HES data.

I would like to thank the Candlelighters Trust and the Laura Crane Youth Foundation Trust for funding my role as statistician on the Yorkshire Specialist Registers of Cancer in Children and Young People while conducting this PHD.

My supervisors, Richard Feltbower, Adam Glaser and Darren Greenwood provided invaluable support, guidance and feedback throughout my PhD.

I would also like to thank the other co-authors of the published papers arising from this thesis for their input and critical revisions of the manuscripts: Sally Kinsey, Anthony Moorman, Lucy Chilton and Daniel Peckham.

I wish to thank several of my colleagues for their support and friendship including Paula Feltbower, Ben Fox, Amanda Friend and Nicola Hughes. A special mention goes to Marlous Hall who has been a great friend always offering good advice and support.

Thanks to all my family for everything they do to support me. Thanks to my husband Paul, for keeping me grounded during this process and for your endless support, love and for listening to all my concerns, or at least pretending to! Final thanks go to my two amazing children, Adam and Callum, who constantly keep me on my toes but it’s worth every second.

Abstract

Survival for children and young adults (CYA) diagnosed with cancer has improved substantially over recent decades, with over 80% currently diagnosed expected to survive at least 5-years. However, survivors are at increased risk of the late effects of their treatment, with many reporting chronic health conditions in later life. The purpose of this project was to investigate cure and long-term health outcomes in CYA with cancer in Yorkshire using data from a population-based cancer registry. The study included 5471 patients diagnosed with a primary tumour in Yorkshire between 1990 and 2011 aged under 30.

Statistical cure models were utilised to describe survival trends. These models simultaneously estimate the percentage 'cured' and the survival of those 'uncured'. The percentage cured is a summary of long-term survival while the median survival time of the uncured provides important information on those who are not long-term survivors. Generally for most diagnostic groups there was an improvement in survival over time which was mainly driven by an increase in the proportion of patients cured rather than an increase in the survival of the uncured.

Long-term morbidity was assessed via linkage to hospital admission data for respiratory and cardiovascular disease and subsequent tumours obtained from cancer registrations. Long-term CYA had increased risk of each of these outcomes compared to the general population. Analysis incorporating the cumulative burden of all subsequent neoplasms and all respiratory and cardiovascular hospitalisations combined found that by age 40, an individual experienced an average of 2 of these events, mainly driven by hospitalisations for respiratory conditions.

Findings from this study provide an evidence base to aid risk-stratification for the long-term follow-up care for this high risk population.

Table of Contents

Intellectual Property and Publication Statement....i
Acknowledgements....v
Abstract....vii
Table of Contents....ix
List of Tables....xv
List of Figures....xix
List of Abbreviations....xxiii
Chapter 1 Introduction....1
1.1 Background....1
1.2 Study rationale....2
1.3 Thesis aims and objectives....3
1.4 Thesis outline....4
Chapter 2 Background and literature review....5
2.1 Introduction....5
2.2 Classification of cancers in children and young adults....5
2.2.1 Cancer registration....7
2.2.2 Specialist cancer registries....7
2.2.3 Yorkshire Specialist Register of Cancer in Children and Young People....8
2.3 Age definitions....9
2.4 Children and young adult cancer epidemiology studies....10
2.5 Cancer incidence....13
2.5.1 Distribution of tumours....13
2.5.2 Diagnostic subgroups....14
2.5.3 Incidence rates in children....16
2.5.4 Incidence rates in TYA....18
2.6 Cancer survival....20
2.6.1 Childhood cancer survival....21
2.6.2 TYA cancer survival....22
2.6.3 Childhood and TYA survival comparison....24
2.7 Statistical methods in survival analysis....25
2.7.1 Survival and hazard functions....26
2.7.2 Kaplan-Meier....26
2.7.3 Cox model....27

2.7.4 Parametric survival models 28
2.7.5 Relative and net survival 29
2.7.6 Relative survival models 32
2.7.7 Flexible parametric survival models 32
2.7.8 Restricted cubic splines 33
2.7.9 Flexible parametric models incorporating splines 34
2.7.10 Flexible parametric models in relative survival framework 35
2.8 Cure models 35
2.8.1 Mixture cure model 36
2.8.2 Non-mixture cure model 37
2.8.3 Parametric distributions and link functions 38
2.8.4 Flexible parametric cure model for relative survival 39
2.8.5 Model assumptions and checking model fit 41
2.8.6 Additional benefits of cure models 42
2.8.7 Estimating time to cure 43
2.9 Application of cure models in CYA 45
2.9.1 Sources of data 46
2.9.2 Types of cure model 46
2.9.3 Age of study participants 46
2.9.4 Diagnostic groups 47
2.9.5 Other risk factors included 48
2.9.6 Study sample size 48
2.9.7 Length of follow-up 48
2.9.8 Survival outcomes 49
2.9.9 Summary of key findings 49
2.10 Survivorship 52
2.10.1 National Cancer Survivorship Initiative 52
2.11 Late mortality 53
2.11.1 Cause of death 53
2.11.2 Variation by cancer type, sex and age and treatment 54
2.11.3 Changes over time 54
2.12 Late effects 55
2.13 Subsequent malignant neoplasms 56
2.14 Cardiovascular late effects 61
2.15 Respiratory late effects 64
2.16 Key gaps in the knowledge 68

Chapter 3 Data and methods 71
3.1 Data sources 71
3.1.1 Yorkshire Specialist Register of Cancer in Children and Young People 71
3.1.2 Hospital Episode Statistics 75
3.1.3 National Cancer Registration and Analysis Service 79
3.1.4 Cytogenetic risk group 79
3.2 Ethical approval and data security 80
3.3 Study variables 80
3.3.1 Diagnostic group 80
3.3.2 Age at diagnosis 81
3.3.3 Disease severity 81
3.3.4 Deprivation 81
3.3.5 Ethnicity 82
3.3.6 Treatment 83
3.4 Outcomes 84
3.4.1 Overall survival 84
3.4.2 Relative survival 84
3.4.3 Event-free survival 84
3.4.4 Subsequent malignant neoplasms 85
3.4.5 Late effects based on hospital admissions 86
3.5 Statistical Methods 88
3.5.1 Descriptive statistics 88
3.5.2 Cure models 89
3.5.3 Selection of confounders for adjustment 94
3.5.4 Competing risk regression models 96
3.5.5 Respiratory admissions 99
3.5.6 SMN analysis 100
3.5.7 Cardiovascular admissions 102
3.5.8 Cumulative burden 103
3.6 Overview of results chapters 106
Chapter 4 Descriptive analysis of the study population 107
4.1 Introduction 107
4.2 Cancer incidence 107
4.3 Survival 109
4.3.1 Comparison of overall and relative survival 111

4.4 Survival trends....113
4.4.1 All cancers combined....113
4.4.2 Leukaemia....113
4.4.3 Lymphoma....114
4.4.4 CNS tumours....115
4.4.5 Bone tumours....115
4.4.6 Soft tissue sarcoma....116
4.4.7 Germ cell tumours....117
4.4.8 Neuroblastoma....117
4.4.9 Renal tumours....118
4.4.10 Carcinomas....119
4.5 Demographic and clinical characteristics of the study population....119
4.6 HES linkage summary....123
4.7 Descriptive statistics of HES admission data....125
4.8 Discussion....125
4.8.1 Results in context....125
4.8.2 Strengths and limitations....127
4.8.3 Summary....128
Chapter 5 Application of cure models to children and young adults diagnosed with cancer....129
5.1 Introduction....129
5.2 Selecting the optimal flexible parametric cure model....129
5.2.1 Flexible Parametric model fit statistics....130
5.2.2 Flexible Parametric predicted survival plots....130
5.2.3 Optimal flexible parametric cure model....133
5.3 Comparison of flexible parametric, mixture and non-mixture cure models....133
5.3.1 Model fit statistics....133
5.3.2 Percentage cured....134
5.3.3 Survival of the uncured....136
5.3.4 Summary of cure model comparison....139
5.4 Cure by diagnostic group....139
5.4.1 Trends in cure by age group....139
5.4.2 Trends in cure by period of diagnosis....142
5.5 Trends in cure for childhood acute lymphoblastic leukaemia....144
5.5.1 Overall survival....146
5.5.2 Event-free survival....148

5.5.3 Cytogenetic risk groups....149
5.5.4 Risk of relapse....150
5.5.5 Checking model fit....151
5.6 Discussion....151
5.6.1 Results in context....151
5.6.2 Strengths and limitations....154
5.6.3 Summary....156
Chapter 6 Respiratory late effects....157
6.1 Introduction....157
6.2 Description of respiratory admissions....157
6.3 Comparison with general population....159
6.3.1 Comparison based on primary diagnosis only....162
6.4 Cumulative incidence....163
6.5 Association between treatment exposure and risk of admission....165
6.6 Checking proportionality assumption....169
6.7 Readmissions and subsequent mortality....170
6.8 Discussion....172
6.8.1 Results in context....172
6.8.2 Strengths and limitations....175
6.8.3 Summary....176
Chapter 7 Subsequent malignant neoplasms....177
7.1 Introduction....177
7.2 Description of SMNs....177
7.3 Cumulative incidence....179
7.4 Comparison with general population....180
7.4.1 Comparison with SMN development 5-years post diagnosis only....181
7.4.2 Comparison based on age at primary diagnosis....182
7.5 Latency for SMNs....184
7.6 Survival following SMN....186
7.7 Discussion....187
7.7.1 Results in context....187
7.7.2 Strengths and limitations....190
7.7.3 Summary....191
Chapter 8 Cumulative burden of respiratory and cardiovascular morbidity and subsequent tumours....193
8.1 Introduction....193

8.2 Description of cardiovascular admissions 193
8.2.1 Cumulative incidence 194
8.2.2 Comparison with general population 196
8.2.3 Comparison based on primary diagnosis only 198
8.3 Cumulative burden subsequent neoplasms, cardiovascular and respiratory admissions 200
8.3.1 Overall burden 202
8.3.2 Treatment specific risks 203
8.3.3 Differences by age 206
8.3.4 Differences by diagnostic group 207
8.3.5 General population rates 210
8.4 Discussion 211
8.4.1 Results in context 211
8.4.2 Strengths and limitations 214
8.4.3 Summary 216
Chapter 9 Discussion 217
9.1 Introduction 217
9.2 Originality of study findings 219
9.3 Clinical implications of study findings 220
9.4 Strengths and limitations 222
9.4.1 Data quality 222
9.4.2 Modelling issues 226
9.5 Recommendations for future research 229
9.6 Conclusions 230
References 231
Appendix A Summary of studies utilising cure modes in children and young adults with cancer 257

List of Tables

Table 2.1: Age standardised incidence rates (ASR) for male and female children, diagnosed 2001-2015 in England.....17

Table 2.2: Age standardised incidence rates (ASR) for male and female TYA (aged 13-24 years), diagnosed between 1979 and 2001 in England…19

Table 2.3: Terms and phrases used in literature search.....45

Table 3.1: Local Authority names and codes for Yorkshire.....74

Table 3.2: Non-malignant morphology codes included in analysis of subsequent tumours 86

Table 3.3: Flexible Parametric cure model knot positions.....91

Table 3.4: Age groups and time periods included in cure models for each diagnostic group 92

Table 3.5: Summary of study population included in each analysis chapter 106

Table 4.1: Number of diagnoses and age standardised incidence rate (ASR) per 1,000,000 person years.....108

Table 4.2: Number of deaths and 5-year overall survival by diagnostic group 110

Table 4.3: Comparison of 5-year overall survival and 5-year relative survival by age group and diagnostic group.....112

Table 4.4: Demographic and clinical characteristics of the study population 120

Table 4.5: Diagnostic groups for children and TYA.....121

Table 4.6: Stage and Grade distribution by diagnostic group.....122

Table 4.7: Comparison of those linked and not-linked to HES inpatient admissions (row %s).....124

Table 4.8: Summary of all HES admissions.....125

Table 4.9: Summary of HES admissions 5-years post diagnosis in 5-year survivors 125

Table 5.1: Comparison of model fit statistics for flexible parametric cure model with different knot points 130

Table 5.2: Comparison of model fit statistics from the flexible parametric, mixture and non-mixture cure model.....133

Table 5.3: Comparison of percentage cured from flexible parametric, mixture and non-mixture cure model 134

Table 5.4: Comparison of median survival time of the uncured and time when 90% uncured are dead from flexible parametric, mixture and non-mixture cure model 137

Table 5.5: Cure model results by age group and diagnostic group.....141

Table 5.6: Cure model results by period of diagnosis and diagnostic group 143

Table 5.7: Characteristics of ALL patients in Yorkshire aged 1-17 years ... 145

Table 5.8: Unadjusted and adjusted cure model results for overall survival for ALL patients 147

Table 5.9: Unadjusted and adjusted cure model results for event-free survival for ALL patients 148

Table 5.10: Comparison of patient characteristics for those linked and not linked to cytogenetic risk group data 149

Table 5.11: Cure model results for overall survival by cytogenetic risk group 150

Table 6.1: Patient characteristics overall and by respiratory admission 158

Table 6.2: Observed and expected respiratory admissions and hospitalisation rate ratio (HRR) by type of respiratory admission and age group 160

Table 6.3: Observed and expected respiratory admissions and hospitalisation rate ratio (HRR) by diagnostic group 161

Table 6.4: Observed and expected number of respiratory admissions and hospitalisation rate ratio (HRR) based on primary diagnosis for admissions only 163

Table 6.5: Association between treatment exposure and risk of admission for any respiratory disease, subdistribution Hazard Ratio (sHR) and 95% CI 166

Table 6.6: Association between treatment exposure and risk of admission for asthma, pneumonia and chronic respiratory diseases, subdistribution Hazard Ratio (sHR) and 95% CI 167

Table 6.7: Association between treatment exposure and risk of admission for respiratory disease including interaction with age group, adjusted subdistribution Hazard Ratio (sHR) and 95% CI 169

Table 6.8: Estimated subdistribution hazard ratio (sHR) including each variable as time varying coefficient in model for any respiratory admission 170

Table 6.9: Readmissions following initial hospitalisation for respiratory condition 171

Table 6.10: Mortality following first admission for respiratory disease for long-term survivors 172

Table 7.1: Patient characteristics of those with and without SMN diagnosed 178

Table 7.2: Diagnostic group of first and subsequent tumours 179

Table 7.3: Standardised incidence ratio (SIR) and absolute excess risk (AER) per 10,000 person years, for SMN by primary diagnostic group (all ages) 181

Table 7.4: Standardised incidence ratio (SIR) and absolute excess risk (AER) per 10,000 person years, by primary diagnostic group (all ages) for SMNs occurring 5-years post diagnosis.....182

Table 7.5: Latency period by age group.....184

Table 7.6: SMN types by latency period 185

Table 7.7: Association between latency period and subsequent mortality following SMN diagnosis, hazard ratio (HR), 95% CI.....187

Table 8.1: Number of survivors admitted for cardiovascular conditions.....194

Table 8.2: Observed and expected cardiovascular admissions and hospitalisation rate ratio (HRR).....197

Table 8.3: Observed and expected cardiovascular admissions and hospitalisation rate ratio (HRR) by age at diagnosis.....198

Table 8.4: Observed and expected number of admissions and hospitalisation rate ratio (HRR) based on primary diagnosis for admissions only.....199

Table 8.5: Summary of outcomes for cumulative burden analysis.....200

Table 8.6: Association between treatment exposures and risk of events, hazard ratios (HR) and 95% CI.....204

Table 8.7: Association between treatment exposure and time to first event for each outcome, sub-distribution hazard ratios (sHR) and 95% CI 205

Table 8.8: Association between treatment exposures and risk of events stratified by age at diagnosis, adjusted hazard ratios (HR) and 95% CI 207

Table 8.9: Association between treatment exposures and risk of events stratified by diagnostic group, adjusted hazard ratios (HR) and 95% CI 209

Table A.9.1: Summary of studies included in literature review of applications of cure models in children and young adults with cancer.....257

List of Figures

Figure 2.1: Distribution of the main diagnostic tumour groups by age at diagnosis....14

Figure 2.2: 10-year survival by diagnostic group, children (0-14 years) Great Britain....21

Figure 2.3: 5-year relative survival for TYA diagnosed 1991-2005 in UK....23

Figure 2.4: Weibull hazard and survival functions for various shape parameters....29

Figure 2.5: Mortality rate by age and sex, England, 2013-2015....31

Figure 2.6: Schematic representation of statistical cure....36

Figure 2.7: Temporal changes in cure model estimates measured at two time periods T_0 and T_1....43

Figure 2.8: Flow chart of literature search results....46

Figure 3.1: Hierarchical structure of CIP spells, Provider spells and Episodes in HES....77

Figure 3.2: DAG representing the relationship between treatment exposures and respiratory admissions....100

Figure 3.3: DAG representing the relationship between SMN latency period and survival....102

Figure 3.4: Diagrams of five hypothetical individuals with multiple and recurrent events....103

Figure 3.5: DAG showing relationship between treatment exposures and cumulative burden....105

Figure 4.1: Survival trend for all cancers combined....113

Figure 4.2: Survival trend for leukaemia....114

Figure 4.3: Survival trend for lymphoma....114

Figure 4.4: Survival trend for CNS tumours....115

Figure 4.5: Survival for bone tumours....116

Figure 4.6: Survival trend for soft tissue sarcomas....116

Figure 4.7: Survival trend for germ cell tumours....117

Figure 4.8: Survival trend for neuroblastoma....118

Figure 4.9: Survival trend for renal tumours....118

Figure 4.10: Survival trend for carcinomas....119

Figure 4.11: Summary of linkage of registry data to HES inpatient admissions....123

Figure 5.1: Plot of predicted survival from flexible parametric cure models with different degrees of freedom (df) and Ederer II lifetable estimates of relative survival (with 95% CI)....131

Figure 5.2: Plot of predicted survival of the uncured from flexible parametric cure models with different degrees of freedom (df).....132

Figure 5.3: Predicted relative survival from flexible parametric (FP), mixture and non-mixture cure model 135

Figure 5.4: Predicted survival of the uncured from FP, mixture and non-mixture cure model and observed survival for those patients that died 138

Figure 5.5: Flow chart of study sample included in ALL cure analysis.....144

Figure 5.6: Overall survival and event-free survival by period of diagnosis, ALL patients aged 1-17 years.....146

Figure 5.7: Cumulative incidence of relapse by diagnostic period 150

Figure 5.8: Plot of predicted relative survival from FP cure models for ALL patients and Ederer II lifetable estimates of relative survival 151

Figure 6.1: Percentage of survivors admitted for a respiratory condition by diagnostic group and age group 159

Figure 6.2: Box plot of age at first admission in cancer survivors compared to the general population 162

Figure 6.3: Cumulative incidence of respiratory admission by attained age164

Figure 6.4: Cumulative incidence of respiratory admission by attained age based on primary diagnoses only 165

Figure 6.5: Kaplan-Meier survival following first respiratory admission type 172

Figure 7.1: Cumulative incidence of SMN by age group.....180

Figure 7.2: Standardised incidence ratio (SIR) of SMN by age and primary diagnostic group 183

Figure 7.3: Absolute excess risk (AER) of SMN by age and primary diagnostic group 183

Figure 7.4: Box plot of latency time by SMN type 185

Figure 7.5: Kaplan-Meier survival from SMN by latency period 186

Figure 8.1: Cumulative incidence of cardiovascular admission by attained age.....195

Figure 8.2: Cumulative incidence of cardiovascular admission by attained age based on primary diagnoses only 196

Figure 8.3: Flow chart showing order of events and time between events.201

Figure 8.4: Cumulative incidence and cumulative burden of SMN, respiratory and cardiovascular admissions.....202

Figure 8.5: Mean cumulative count by treatment group.....203

Figure 8.6: Mean cumulative count by age group.....206

Figure 8.7: Mean cumulative count by diagnostic group.....208

Figure 8.8: Number of HES episodes for cardiovascular and respiratory disease by age group, England, 2017-2018. 210

Figure 8.9: Number of new cancer cases diagnosed by age group, UK, 2014-2016.....211

List of Abbreviations

A&E	Accident and Emergency
AAPC	Average Annual Percentage Change
ACCIS	Automated Childhood Cancer Information System
AER	Absolute Excess Risk
AIC	Akaike Information Criterion
ALL	Acute Lymphoblastic Leukaemia
AML	Acute Myeloiod Leukaemia
APC	Admitted Patient Care
ASR	Age Standardised Rate
AYA	Adolescents and Young Adults
BCCSS	British Childhood Cancer Survivor Study
BIC	Bayesian Information Criterion
CCSS	Childhood Cancer Survivors Study
CI	Confidence Interval
CIF	Cumulative Incidence Function
CIP	Continuous Inpatient Spell
CNS	Central Nervous System
COSD	Cancer Outcomes and Services Dataset
CYA	Children and Young Adults
DAG	Directed Acyclic Graph
df	Degrees of freedom
EFS	Event-Free Survival
FCE	Finished Consultant Episode
FP	Flexible Parametric
GP	General Practitioner
HES	Hospital Episode Statistics
HR	Hazard Ratio
HRR	Hospitalisation Rate Ratio
IACR	International Association of Cancer Registries
IARC	International Agency for Research on Cancer

ICCC-3	International Classification of Childhood Cancer, 3rd Edition
ICD-10	International Classification of Diseases Version 10
ICD-O3	International Classification of Diseases in Oncology, 3rd Edition
KM	Kaplan-Meier
MCC	Mean Cumulative Count
MST	Median Survival Time
NCI	National Cancer Institute
NCRAS	National Cancer Registration and Analysis Service
NCSI	National Cancer Survivorship Initiative
NEC	Not Elsewhere Classified
NHL	Non-Hodgkin Lymphoma
NHS	National Health Service
NOS	Not Otherwise Specified
NRCT	National Registry of Childhood Tumours
ONS	Office for National Statistics
OP	Outpatient
OPCS-4	Office of Population, Censuses and Surveys: Classification of Interventions and Procedures, 4th Revision
PBCR	Population-Based Cancer Registry
PFT	Pulmonary Function Test
PHE	Public Health England
PNET	Primitive Neuroectodermal Tumours
PTC	Principal Treatment Centre
PWP	Prentice, Williams and Peterson
PWT-TT	Prentice, Williams and Peterson Total Time
SEER	Surveillance, Epidemiology and End Results
sHR	Subdistribution Hazard Ratio
SIR	Standardised Incidence Ratios
SMN	Subsequent Malignant Neoplasm
SMR	Standardised Mortality Ratio
TYA	Teenagers and Young Adult

TYACSS	Teenage and Young Adult Cancer Survivor Study
UK	United Kingdom
USA	United States of America
WCC	White Cell Count
WHO	World Health Organisation
YSRCCYP	Yorkshire Specialist Register of Cancer in Children and Young People

Chapter 1 Introduction

1.1 Background

Cancer is major public health issue and accounts for a significant burden of disease with an estimated 18.1 million new cases of cancer diagnosed worldwide in 2018 [1]. Cancer affects people of all ages and is more common in older ages; on average half of all cancer diagnoses each year in the UK are in people aged 70 years and over [2]. Cancers in children and young adults (CYA) (aged 0-24 years) are rare and account for approximately 1% of all cancers diagnosed in the UK [2]. However, cancer is the one of the most common causes of death within this age range; accounting for 10-20% of all deaths in CYA [3].

Cancers diagnosed in CYA are a heterogeneous disease group and differ from those diagnosed in older adults [4]. The most common childhood cancers (0-14 years) are leukaemia, central nervous system (CNS) tumours and lymphomas; together these account for approximately two thirds of all childhood cancers [4, 5]. Carcinomas, which are the most common histological type in adults are very rare in childhood [5, 6]. Cancers in teenagers and young adults (TYA) have a distinct profile and can be grouped into three broad categories comprising of ‘late paediatric cancers’, ‘early onset adult cancers’ and certain diagnostic groups which have a peak incidence in this age group [7-9]. Lymphomas are the most common cancer in 15-24 years followed by carcinomas and germ cell tumours which together account for over half of all diagnoses in this age group [8].

Survival rates for children diagnosed with cancer have improved significantly over recent decades; 5-year survival for all cancers combined increased from 30% for children diagnosed in the 1960s [6, 10] to 82% for children diagnosed between 2006 and 2010 [11]. Predicted 5-year survival for those diagnosed in 2018 is estimated to be 85% [12]. These improvements in survival are due to several factors including advancements in treatments over time, including chemotherapy, surgical and radiotherapy techniques, participation in national

and international clinical trials, as well as centralisation of specialist centres and the development of supportive care packages [6, 10, 13]. Survival rates for TYA have also improved over time [14] and the latest overall 5-year survival rate for TYA diagnosed in the UK between 2001 and 2006 for all cancers combined was 82% [15, 16]. There is however, considerable variation in survival by diagnostic group. Childhood 5-year survival ranges from over 90% for children with retinoblastoma and Hodgkin lymphoma to around 60% for neuroblastoma and bone tumours [11, 17]. TYA 5-year survival ranges from over 80% for Hodgkin lymphoma, thyroid carcinoma, testicular and ovarian germ cell tumours and melanoma to around 60% for bone tumours and soft tissue sarcomas [15, 16]. For certain diagnostic groups such as leukaemia, bone and soft tissue sarcomas, TYA generally have poorer outcomes than children diagnosed with the same cancer [18-21].

1.2 Study rationale

The 5-year survival rate for patients diagnosed with cancer is frequently used as an indicator to monitor outcomes for cancer patients [22, 23] and it is often stated that patients who survive beyond this time period are said to be cured of their original cancer. In cancer epidemiology, cure models offer a statistical method to assess cure for long-term survivors [24-27]. Statistical cure is possible for a particular cancer if the survival curve flattens out and levels off after a sufficient length of time when the remaining patients have a similar death rate to the general population. Rather than model the survival of all patients as one, a cure model assumes there are two groups of patients, one who do not experience the outcome of interest and are ‘cured’ and the other who experience the outcome and their survival is estimated separately [24-27]. If assessing trends in survival over time, cure models can identify if survival has improved due to increasing the proportion of patients cured, improving the survival time of the uncured patients or a combination of both [28]. Cure models have been developed extensively in the statistical literature and applications to data from population-based cancer registrations are mainly to the most common adult cancer types, although two studies have specifically examined cure in childhood leukaemia using population-based cancer registrations [29, 30].

High survival rates for CYA cancers come at a cost and high intensity treatments make long-term cancer survivors (defined as those surviving beyond 5-years from diagnosis) at increased risk of premature mortality [31, 32], second

malignant neoplasms [33-35], and other morbidities [36-38] compared to the general population. These late health effects may not occur until several years after the end of treatment [39]. It is estimated that 62%-75% of childhood cancer survivors have at least one late side effects of treatment with many suffering from multiple conditions [36-38, 40]. There is also growing evidence that TYA are also at increased risk of these outcomes [41]. Data linkage of cancer registrations to hospital admissions provides an objective outcome of morbidity in which to study the late effects in long-term survivors of CYA cancers.

The purpose of this project was to examine population-based cure and long-term health outcomes in CYA with cancer in Yorkshire using objective outcome measures.

1.3 Thesis aims and objectives

This project had 3 key aims:

1. **To assess the feasibility of applying cure models to CYA diagnosed with cancer using data from a regional population-based specialist cancer register**

   Different methods for fitting cure models were researched and the most suitable statistical methods were applied to assess if cure, as defined by these statistical models, was a reasonable assumption. Cure models were investigated for a range of cancers which are most common in children and teenagers and young adults. Cure models incorporating clinical risk factors were investigated for children with leukaemia.

2. **To evaluate long-term health outcomes for children and young adult cancer survivors**

   Morbidity in long-term CYA cancer survivors was assessed using cancer registrations linked to hospital admission data for three specific outcomes:

   a) A detailed description of respiratory morbidity based upon hospital admissions was provided including risk of admissions related to earlier treatment.
   b) The risk of developing a secondary malignant neoplasm was assessed using cancer registration data, this included early onset tumours and the association with latency and subsequent mortality.
   c) Hospital admissions for cardiovascular disease were described.

These three outcomes were selected as they represent the most common causes of late mortality and morbidity in long-term childhood and young adult cancer survivors. For each outcome incidence rates in the cancer survivor cohort were compared to rates in the general population matched on age and sex.

3. **To assess the cumulative burden of subsequent tumours, cardiovascular and respiratory morbidity for children and young adult cancer survivors**

   Combining the three outcomes included in Aim 2 above, the cumulative burden of all three morbidities was described and the association between treatment exposure and cumulative burden investigated using novel statistical methodology to account for multiple and recurrent events.

1.4 Thesis outline

Chapter 2 provides a detailed background and evaluates the evidence base for this research, including a critical review of the current literature on CYA cancer epidemiology, cure models and studies on late effects. Key gaps in the current literature were identified. Chapter 3 introduces the data sources and describes the statistical methods used throughout the rest of the thesis. Chapter 4 includes a detailed descriptive analysis of the study population. The main results of the project relating to the three aims described above are presented in Chapters 5-8, including a detailed discussion at the end of each chapter summarising the results in the context of current research and identifying strengths and limitations of each analysis. Chapter 5 focusses on cure models (Aim 1), Chapter 6 on respiratory hospitalisations (Aim 2a), Chapter 7 on subsequent malignant neoplasms (Aim 2b) and Chapter 8 on the cumulative burden of subsequent tumours, cardiovascular and respiratory morbidity (Aim 3) including a detailed description of hospitalisations for cardiovascular disease (Aim 2c). Chapter 9 draws the final conclusions of this thesis together by proving a summary of the main findings in relation to the aims above, discusses the clinical implications, strengths and limitations of the work and identifies areas for further research.

Chapter 2 Background and literature review

2.1 Introduction

The aim of this chapter is to provide background information on the epidemiology of cancer in children and young people focusing on studies based on population-based cancer registry data. Statistical methods used to analyse survival data, including a description of cure models, are evaluated, followed by a literature review and critical appraisal of the application of cure models to children and young people with cancer. This chapter then goes on to critically review the current literature on the long-term health outcomes of survivors of cancer in childhood and young adulthood specifically focussing on three key areas: subsequent malignant neoplasms, cardiovascular disease and respiratory disease. Finally, an overall summary of the literature is provided and the key gaps in the knowledge are identified.

2.2 Classification of cancers in children and young adults

Cancers diagnosed in children and young adults (CYA) are a heterogeneous disease group and differ from those diagnosed in older adults, therefore age-specific classification and coding systems are used within this age range [4]. Generally, cancers are coded and classified according to the International Classification of Diseases in Oncology (ICD-O), currently in its third edition (ICD-O-3) [42]. ICD-O-3 describes tumours based on a topographical code, which describes the anatomical site of origin (or organ system) of the tumour and a morphological code, which describes the cell type (or histology) of the tumour and also the behaviour (malignant or benign). Cancers, particularly in adults, are often described and reported by primary site of origin based on topography alone, for example cancers of the breast, colon or lung. For children, the range and type of cancers diagnosed are different from adult cancers and classification for childhood tumours is based on morphology rather than topography alone [43]. The International Classification of Childhood Cancer, now in its third edition (ICCC-3) [43], is the current standard for presentation of data on childhood cancer incidence and survival based on the ICD-O-3. This classification system includes some non-malignant intracranial

and intraspinal tumours as these tumours present with similar clinical symptoms, prognosis and late effects as malignant tumours in childhood and therefore it is important to record the incidence of these tumours [43].

The ICCC-3 defines 12 main diagnostic group which are further split into 47 subgroups [43], the 12 main diagnostic groups are:

I. Leukaemias, myeloproliferative diseases and myelodysplastic diseases
II. Lymphomas and reticuloendothelial neoplasms
III. CNS and miscellaneous intracranial and intraspinal neoplasms
IV. Neuroblastoma and other peripheral nervous cell tumours
V. Retinoblastoma
VI. Renal tumours
VII. Hepatic tumours
VIII. Malignant bone tumours
IX. Soft tissue and other extraosseous sarcomas
X. Germ cell tumours, trophoblastic tumours and neoplasms of gonads
XI. Other malignant epithelial neoplasms and malignant melanomas
XII. Other and unspecified malignant neoplasms

The profile of cancers in teenagers and young adults (TYA) (see Section 2.3 for further discussion on age range) is distinct from those observed in children and adults. Cancers in TYA can be described by three broad categories: ‘late paediatric cancers’ such as Wilms tumours, rhabdomyosarcoma and neuroblastoma, those that arise as ‘early onset adult cancers’ such as melanoma and thyroid cancer and those which have a peak incidence in this age group such as Hodgkin lymphoma and gonadal germ cell tumours [9]. Therefore separate classification systems have also been defined for TYAs [8, 44]. In the UK, the Birch classification system is commonly used which groups diagnoses into one of 10 main groups with a further 32 subgroups within these [8]. This classification is also predominantly based on morphology. The 10 main diagnostic groups are:

1. Leukaemias
2. Lymphomas
3. CNS and other intracranial and intraspinal tumours
4. Osseous and chondromatous neoplasms, Ewings tumour and other neoplasms of the bone

5. Soft tissue sarcomas
6. Germ cell and trophoblastic neoplasms
7. Melanoma and skin carcinoma
8. Carcinomas (except of skin)
9. Miscellaneous specified neoplasms not elsewhere classified (NEC)
10. Unspecified malignant neoplasms NEC

The main differences between ICCC-3 and the Birch system is the classification of melanomas and carcinomas as separate groups in the Birch classification system, as these are more frequent in TYA than younger ages, while non-CNS embryonal tumours which are common in childhood and less frequent in TYA are grouped together in the Birch classification system [8, 45].

2.2.1 Cancer registration

Population-based cancer registries (PBCR) are responsible for recording all new cases of cancer in a defined population, usually defined by a geographic region [46], with a defined set of variables recorded for each case [47]. PBCRs play an important role in cancer control; monitoring trends in incidence, mortality, survival and prevalence as well as supporting and planning service and care for cancer patients [47]. The advantage of using data from a PBCR rather than a single institution data set or a clinical trial is that PBCRs cover the whole population of cancer patients and are not limited to a self-selected, often atypical, subgroups of patients [47].

Cancer registration in the UK started in the 1920s with regional coverage in some areas. Since 2013 registrations have been recorded centrally for the whole of England via the National Cancer Registration and Analysis Service (NCRAS) within Public Health England (PHE) [48]. Data are collected in accordance with the Cancer Outcomes and Services Dataset (COSD) which defines a general core dataset for all cancers in addition to extra key clinical and pathological data items for specific tumour types [49].

2.2.2 Specialist cancer registries

A general cancer register refers to one that records all new cancer registrations for all ages and all cancer types generally within a pre-defined geographic region. Specialist registers refer to those that only cover a specified age range

at diagnosis, such as paediatric registers for 0-14 year olds, or one disease type, for example haematological tumours only [50]. Across Europe eight countries (including the UK) have a national register for paediatric cancers and another three countries have large regional paediatric registers [51]. Specialist registers for children and young adults (CYA) are important due to differences in coding and types of diagnoses in this age range compared to cancers in older ages. In the UK national coverage of childhood cancers were recorded in the National Registry of Childhood Tumours (NRCT) which included all diagnoses of cancer in 0-14 year olds from 1962 onwards for England, Scotland and Wales and from 1993 onwards included Northern Ireland [10]. Since 2013, in line with changes to cancer registrations in England this registry was dissolved and registrations are now recorded as part of the NCRAS within PHE for England and the national general cancer registries of Scotland, Wales and Northern Ireland [51]. In England in addition to the NRCT there are several specialist childhood and young adult cancer registries including the Manchester Children's Tumour Registry [52], the Northern Region Young Persons Malignant Disease Registry [53], the West Midlands Regional Children's Tumour Registry [54] and the Yorkshire Specialist Register of Cancer in Children and Young People [55]. These regional registers collect more detailed information regarding key clinical prognostic risk factors at diagnosis such as stage and grade of tumours as well as detailed treatment information compared to data collected nationally. National and regional registries regularly exchange data to ensure completeness of ascertainment of cases.

2.2.3 Yorkshire Specialist Register of Cancer in Children and Young People

The Yorkshire Specialist Register of Cancer in Children and Young People (YSRCCYP) was used for the analysis presented in this thesis and is described in further detail in Chapter 3. Briefly it includes all diagnoses of cancer in children age 0-14 years from 1974 onwards and diagnoses in TYA aged 15-29 years from 1990 onwards for patients resident in the former Yorkshire Health Authority. Data on patients resident in the South Yorkshire area are included from 1998 onwards only.

Advantages of using data for a regional register include less geographic variation in data quality; errors may be easier to check and update as the register may only include a limited number of hospitals making it easier to check

and verify the patient's hospital notes for data quality errors. The YSRCCYP incudes more detailed information than is available in the NRCT, for example information on stage and treatment which have limited availability nationally. The limitations are smaller sample sizes especially for very rare childhood cancers where only a few cases are diagnosed per year. If patients live in one area but are treated in another then they may be counted more than once in different regional registers which may result in them being counted more than once in national figures. A change of address outside the region may also increase the chances of loss to follow-up.

2.3 Age definitions

In terms of cancer epidemiology, childhood cancers refer to those diagnosed before a person's 15th birthday so includes cancers diagnosed between 0-14 years of age. This age range has been used extensively in published research worldwide [56]. Adolescents generally refer to the age group 15-19 years [20, 57-59], however the definition of the age range for teenagers and young adults, or adolescents and young adults (AYA) as frequently used in literature from the USA and Canada, is less clearly defined and varies between studies. One commonly used age range is 15-24 years including much of the published work in the UK [8, 15, 16, 60] and in the EUROCARE- 4 study [19]. Others have used the age range 13-24 years [14, 18, 61-63]. An upper age limit of 29 years is often included in studies from the USA and Canada [44, 45, 64] and this is the age range included in the YSRCCYP [65]. More recently it has been recommended to use the age group 15-39 years [7, 9, 58, 66-68]. Often upper age limits are chosen to reflect national clinical treatment practices [7]. Using different age ranges has implications for comparison between studies as the types of cancers diagnosed and incidence rates vary with age. There are 2.7 times more patients diagnosed aged 15-29 years old compared to 0-14 years old and approximately half of the 15-29 year old patients diagnosed with cancer are 25-29 years at diagnosis [64]. The male to female cancer incidence ratio also varies by age group due to differences in cancer type by age and sex; around two thirds of cancers diagnosed in 15-39 years are in women, compared with a slight excess of childhood cancers in males [68, 69].

The YSRCCYP includes cancers diagnosed in patients aged 0-29 years at diagnosis, therefore for the basis of my analysis and the rest of this background literature review, children and young adults (CYA) will refer to those aged

between 0-29 years at diagnosis, with a specific focus of teenagers and young adults (TYA) to mean those aged 15-29 years unless otherwise stated.

TYA diagnosed with cancer represent a specific group of interest for several reasons. The teenage years span an age of transition from childhood to adulthood. It is an important time developmentally as individuals gain more control and independence in many aspects of their lives including in terms of education, relationships and sexuality [70]. A diagnosis of cancer and its treatment may adversely affect this transition period, for example prolonged absence from school while undergoing cancer treatment may affect not only educational outcomes but also friendships with peers [70]. TYA cancer patients fall between two groups in terms of cancer care, often too old to be treated in paediatric wards but too young to be on adult wards [66, 71]. TYA patients may want to be involved in the decision making process regarding treatment choices [70] and may prefer to be treated alongside their peers rather than with younger children or older adults [71]. Recent cancer policy has addressed this issue by recommending specialist provision of care for this age group including age appropriate services as well as clinical expertise [72]. This group is also an understudied group in term of late effects compared to childhood cancer survivors and further research is needed for this unique group [67, 73].

2.4 Children and young adult cancer epidemiology studies

There are several key epidemiological studies that are frequently referenced throughout this thesis when reviewing the literature on CYA cancer, particularly studies on long-term survivors. A brief overview of these studies including their main strengths and limitations is given below.

Automated Childhood Cancer Information System

The Automated Childhood Cancer Information System (ACCIS) project was a European project which included cancer registrations for children (0-14 years) and adolescents (15-19 years) from 80 PBCRs in 35 countries. The projects aims were to collect and report data on cancer incidence and survival in Europe and initially included over 160,000 cases of cancers diagnosed from 1970 to 2001 [20] and was subsequently updated to include all diagnoses between 1991 and 2010 [74] including 180,000 cases in 19 European countries. This was a large, population-based study therefore enabling incidence and survival

trends of relatively rare cancers within 0-19 year olds to be assessed. The limitations are that data are not available for older teenagers and young adults and there is a lack of prognostic information such as stage, to assess variation in survival.

EUROCARE

The EUROCARE project started in the late 1980s and aimed to provide analysis on survival rates in cancer patients of all ages across Europe. The latest EUROCARE-5 project included 22 million patients diagnosed from 117 PBCRs in 31 countries diagnosed between 1978 and 2007 [75]. Specific subgroup analysis of the EUROCARE database have included children or children and adolescents [19, 21, 76]. Again the main strengths of this study are that it is large and population-based and therefore able to assess survival trends by diagnostic groups across Europe. Data are available for all ages and survival estimates have been compared for children (0-14 years) and adolescents and young adults (15-39 years) and older adults (40-69 years) [21]. The main limitation of this study is a lack of data on stage at diagnosis and treatment.

Surveillance, Epidemiology and End Results program

In the USA, population-based data on the epidemiology of malignant diseases in CYA are maintained by the Surveillance, Epidemiology and End Results program (SEER) which is run by the National Cancer Institute [77]. This registry includes a representative sample of 26% of the US population, and comprehensive reports on incidence and survival for approximately 30,000 patients aged 0-19 years diagnosed from 1975-1995 [59] and approximately 60,000 patients 15-29 years diagnosed with cancer between 1975 and 2000 [64] have been published. Key strengths are again that studies based on SEER data are large and population-based covering all ages; data on treatment and stage are also available. However, it does only cover 26% of the US population.

Childhood Cancers Survivor Study

The Childhood Cancer Survivors Study (CCSS), is a multi-institution study in the USA and Canada of over 14,000 5-year cancer survivors diagnosed between 1970 and 1986, aged 0-20 years at diagnosis and a similarly aged cohort of 3600 non-cancer siblings [78]. It has been recently expanded to include additional diagnoses between 1987 and 1999 [79] so now includes approximately 34,000 5-year survivors from 31 institutions. The CCSS is not

population-based and only includes patients diagnosed with leukaemia, CNS tumours, Hodgkin lymphoma, non-Hodgkin lymphoma, Wilms tumours, neuroblastoma, soft tissue sarcomas and bone tumours so is not truly representative of all childhood cancer survivors. Strengths are the large study sample with comprehensive follow-up, detailed treatment information and a large sibling comparison group. The main limitations are that outcomes are self-reported by the survivors and only subsequent tumours are validated. There are also limitations of using sibling controls as both the cancer survivor and sibling control will be exposed to the same genetic and environmental risk factors.

British Childhood Cancer Survivor Study

The British Childhood Cancer Survivor Study (BCCSS) includes 17,000 5-year cancer survivors diagnosed in Great Britain between 1940 and 1991 aged 0-14 years at diagnosis [80]. It has also recently been extended to include patients diagnosed between 1992 and 2006 so now includes approximately 35,000 childhood cancer survivors [81]. The BCCSS was ascertained from the National Registry of Childhood tumours therefore its main strengths are the large sample size and that it is population-based. Limitations are that follow-up is via self-report obtained from completion of questionnaires therefore may be prone to recall and selection bias. Furthermore long-term outcomes are not validated. There is limited treatment information available; only binary indicators of receiving surgery, chemotherapy or radiotherapy are available and this information is missing in around a third of the study population [32].

Teenage and Young Adult Cancer Survivor Study

The Teenage and Young Adult Cancer Survivor Study (TYACSS) includes approximately 200,000 survivors of cancer diagnosed between the ages of 15-39 years, between 1971 and 2006 in England and Wales [82, 83]. This study was established based on national cancer registrations, with long-term outcomes based upon linkage to Hospital Episode Statistics in England and Patient Episode Database for Wales in Wales. Key strengths of this study are the large population-based sample size allowing detailed examination of outcomes by diagnostic groups specific to this age group and objective outcomes. However, limited treatment information is available.

Other large survivor cohort studies exist in several European countries [84] including the Nordic countries, the Netherlands and Switzerland which have

been used to study late effects of childhood cancer survivors which will be described later in this chapter (Sections 2.11-2.15)

2.5 Cancer incidence

2.5.1 Distribution of tumours

Cancers in CYA are a heterogeneous group and the distribution of diagnostic groups varies with age across the age spectrum of CYAs. In children, leukaemia, lymphoma and CNS tumours make up around two thirds of all cancers diagnosed [11]. Embryonal tumours are made up of undifferentiated cells similar to ones in a developing embryo and mainly occur in children accounting for about 20% of childhood cancers [85]. For TYA, lymphomas, carcinomas and germ cell tumours account for over half of all cancers [15]. Figure 2.1 shows the distribution of diagnostic groups by age for 0-19 year olds [20]. Many of the embryonal tumours, such as retinoblastoma, neuroblastoma, nephroblastoma and medulloblastoma, diagnosed in infancy (<1 year) and early childhood (1-9 years) are very rare in older childhood and adolescents. For infants sympathetic nervous system tumours (neuroblastoma) are the most common tumour diagnosed while for children aged 1-4 years leukaemias are most frequent and CNS tumours are the most common diagnosis in 5-9 year olds. After age 10 years, lymphomas, carcinomas, germ cell tumours and bone tumours become more frequent [5, 20]. Cancers in TYA have a distinct profile [7-9]. Lymphomas are the most common diagnoses in both 15-19 year olds and 20-24 year olds [8, 20, 64]. Leukaemias and bone tumours are less common in 20-24 year olds compared to younger ages while germ cell tumours, melanomas and carcinomas are more common [8]. In the USA melanoma increases from the 5th most common cancer in 15-19 year old to the most common cancer in 25-29 years [64].

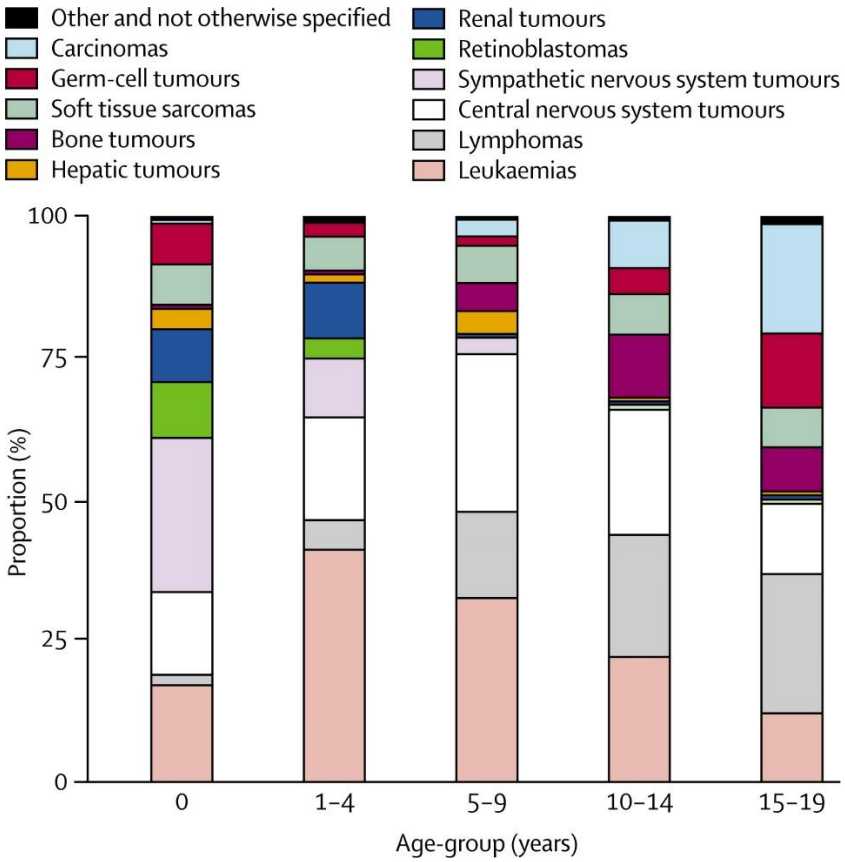


Figure 2.1: Distribution of the main diagnostic tumour groups by age at diagnosis.

Source: [20]

2.5.2 Diagnostic subgroups

In children, acute lymphoblastic leukaemia (ALL) is the most common diagnostic subgroup not only among all leukaemias but for all cancers combined and accounts for around 79% of all leukaemias and 25% of all childhood cancers [10]. Acute myeloid leukaemia (AML) accounts for 15% of leukaemias and 5% of all childhood cancers [10]. In 15-24 year olds ALL accounts for 46% of all leukaemias and 5% of all cancers diagnosed, while AML accounts for 37% of all leukaemias and 4% of all cancers in this age group [8].

Hodgkin lymphomas account for 41% of lymphomas diagnosed in children and are mainly diagnosed in older children, non-Hodgkin lymphoma (NHL) account for about 57% of lymphomas, these are very rare in infancy and increase steadily through childhood [10]. In TYA, Hodgkin lymphomas account for 72% of all lymphomas and overall 19% of all cancers diagnosed in this age range while NHL account for 28% of lymphomas [8].

Astrocytomas are the most common CNS tumours in children and TYA accounting for 43% and 51% of all CNS tumours in children and TYA respectively [8, 10]. In children the next most common CNS tumours are embryonal tumours (including medulloblastoma and primitive neuroectodermal tumours (PNET)) accounting for 19% of CNS tumours [10]; these are less frequent in TYA, accounting for 9% of all CNS tumours [8].

Neuroblastoma is the most common sympathetic nervous system tumour in childhood and the most common embryonal tumour of childhood and accounts for 6% of all childhood cancers [10]. The majority of renal tumours in children are nephroblastoma, also known as Wilms tumours, accounting for 90% of all renal tumours [10]. Neuroblastoma, retinoblastoma, Wilms tumours and hepatoblastoma are very rare over the age of 15 and in the Birch classification system for TYA cancers are not included as separate groups but are grouped together in group 9 miscellaneous specified neoplasms [8].

The two most frequent subgroups of bone tumours are osteosarcomas, accounting for 55% and 51% of childhood and TYA bone tumours respectively, and Ewings sarcoma, accounting for 38% and 30% of childhood and TYA bone tumours respectively [8, 10].

Rhabdomyosarcoma is the most common soft tissue sarcoma in childhood accounting for 53% of all soft tissue sarcomas and 4% of all cancers in childhood [10]. For TYA, rhabdomyosarcoma account for 22% and fibrosarcomas account for 24% of all soft tissues sarcomas [8].

The distribution of germ cell tumour subgroups varies markedly by age and sex. In children malignant gonadal germ cell tumours were the most common (42%) followed by intracranial and intraspinal tumours (35%) and malignant extracranial and extragonadal sites (22%) [10]. For 15-24 year olds germ cell tumours make up 14% of all cancers diagnosed of which 93% are gonadal germ cell tumours [8].

Similarly the distribution of carcinomas varies with age. For children carcinomas represent 3% of all diagnoses and within this the most common subgroups were

malignant melanoma (32%), skin carcinoma (17%) and thyroid carcinoma (15%) [10]. For TYA, melanoma is classified as a separate diagnostic group in the Birch classification representing 8% of all cancers, while thyroid carcinoma account for 3% of all cancers diagnosed in TYA [8].

In addition to differences in the types of cancers diagnosed in children and TYA, significant biological differences exist between TYA and younger and older patients with the same histological tumour type. For example the biologic characteristics of ALL change in post pubertal patients towards subtypes with worse prognosis [66].

2.5.3 Incidence rates in children

In children for all cancers combined, age-specific incidence rates are highest in the first five years of life, then reduce for those aged 5-9 years, before increasing again for ages 10-14 years. This increase in age-specific incidence then continues across the life-course [10].

Table 2.1 shows the age standardised incidence rates (ASR) per 1,000,000 persons per year, for childhood cancers diagnosed in England between 2001 and 2015, for the 12 main groups of the ICCC-3 [69]. Between 2001 and 2015, for all cancers combined the ASR was 166.9 per million for boys and 145.6 per million for girls. ASRs ranged from around 52 per million for leukaemias in boys to 1 per million for the other and not otherwise specified. Incidence patterns varied by sex with the highest sex ratio for lymphomas.

Table 2.1: Age standardised incidence rates (ASR) for male and female children, diagnosed 2001-2015 in England.

Diagnostic group (ICCC-3)	Male	Female	M:F ratio
I. Leukaemia	51.7	44.3	1.17
II. Lymphoma	20.9	9.9	2.11
III. CNS tumours	40.4	36.4	1.11
IV. Neuroblastoma	10.4	9.9	1.05
V. Retinoblastoma	4.4	4.8	0.92
VI. Renal tumours	8.7	10.3	0.84
VII. Hepatic tumours	2.2	2.0	1.10
VIII. Bone tumours	6.4	5.6	1.14
IX. Soft tissue sarcoma	11.5	8.7	1.32
X. Germ cell tumours	4.7	6.3	0.75
XI. Carcinomas	4.4	6.2	0.71
XII. Other and not otherwise specified	1.3	1.2	1.08
All cancers combined	166.9	145.6	1.15

Source: [69]

ASR per 1,000,000 persons per year standardised to World Standard Population

For all cancers combined the incidence rate across Europe increased by 1% per year in children from 118 per million for children diagnosed during the 1970s to 139 per million for children diagnosed in the 1990s, this increase was observed in all age groups [20]. Increasing incidence trends were observed for the majority of diagnostic groups; between 1978-82 and 1993-1997 the highest increases in the average annual percentage change (AAPC) in incidence rates were observed for soft tissue sarcoma (1.8%), CNS tumours (1.7%) and germ cell tumours (1.6%) [86]. Between 1978 and 1997 for all cancers combined incidence increased for both sexes but the AAPC was slightly higher in girls (1.4%) than in boys (0.9%) [86]. Between 1991 and 2010 incidence continued to increase by 0.5% per year in European children with increases observed for leukaemias (0.7%), lymphomas (0.3%), CNS tumours (0.5%) and other tumours (0.6%) with further temporal variation in rates also observed by age group and geographical region [74]. In the UK the incidence rate for all cancers combined increased by 15% from 1993 to 2016 [11].

2.5.4 Incidence rates in TYA

The incidence rate for all cancers combined is significantly higher in 20-24 year olds compared to 15-19 years, with ASR estimated at 157.7 per million for 15-19 years and 248.1 per million for 24-24 years for those diagnosed in England between 1979 and 2001 [62]. Table 2.2 shows the ASR per million for TYA cancers (aged 13-24 years) diagnosed in England between 1979 and 2001, for the 10 main groups of the Birch Classification [62]. Overall, for all cancers combined the ASR was 188.7 per million, again this was higher in males (201.5) than females (174.5). The ASRs by diagnostic group ranged from 45.3 per million for lymphoma to 1.2 per million for unclassified tumours. The ASR for germ cell tumours was over 7 times higher in males compared to females. There was a male excess also for leukaemia, lymphoma, CNS tumours, bone and soft tissue sarcomas, while rates were higher in females for melanoma and carcinomas. National ASRs by Birch Classification group are not publically available for more recently diagnosed cases, however, a 2018 report published by Public Health England and the Teenage Cancer Trust reported crude incidence rates for the 10 most common cancer subgroups separately for males and females for those aged 13-24 years diagnosed 2013-2015 [87]. In males testicular germ cell tumours were the most common subgroup and in females Hodgkin lymphoma was the most common subgroup.

Table 2.2: Age standardised incidence rates (ASR) for male and female TYA (aged 13-24 years), diagnosed between 1979 and 2001 in England.

Diagnostic group (Birch)	Male	Female	All	M:F ratio
Leukaemia	25.2	17.1	21.2	1.47
Lymphoma	50.5	40.0	45.3	1.26
CNS tumours	28.2	25.9	27.1	1.09
Bone tumours	13.9	9.8	11.9	1.42
Soft tissue sarcoma	9.8	8.4	9.1	1.17
Germ cell tumours	43.3	6.0	24.8	7.22
Melanoma	10.1	19.4	14.7	0.52
Carcinoma	17.4	44.0	30.6	0.40
Miscellaneous NEC	2.1	2.6	2.3	0.81
Unclassified	1.0	1.4	1.2	0.71
Total	201.5	174.5	188.7	1.15

Source: [62]

ASR per 1,000,000 persons per year standardised to the European standard population

The incidence for all cancers combined for adolescents (15-19 years) has also been increasing since the 1970s although at a faster rate than in children with estimates of the AAPC of 2% per year for 15-19 year olds in Europe between 1978 and 1997 [88], and 1% per year between 1991 and 2010 [74]. Variation in temporal trends has been reported by diagnostic group with significant increases observed for leukaemias, lymphomas, CNS tumours, germ cell tumours and epithelial tumours, specifically thyroid, melanoma and skin carcinoma [20, 57, 74].

In England between 1979 and 1997 incidence increased by 1.5% per year for 15-24 year olds [8] with significant increases observed for NHL, astrocytoma, germ cell tumours, melanoma and thyroid carcinoma [8]. There was a larger overall increase for females compared to males: between 1993 and 2016 the ASR for all cancers combined increased by 22% for males compared to 45% for females [15]. The latest incidence rates for all cancers combined for TYA (15-24

years) diagnosed in the UK between 2014 and 2016 were 300 per million for males and 341 per million for females and 320 per million for both sexes combined [15].

Some caution should be taken interpreting these trends and comparing rates from different sources and populations. Changes in trends over time can be difficult to interpret due to changes in coding and classification and changes to diagnostic procedures, resulting in improved classification of some tumours and a decrease in less specific groups [10]. This is particularly important for large European studies which may have variable coverage in some regions. Different standard reference populations have been used to calculate the ASRs in different studies so rates may not be directly comparable across studies. Different classification systems and different age ranges used for TYA may limit direct comparability between studies also. Generally, studies have shown that: 1) incidence trends are increasing over time for both children and TYA; 2) rates are increasing at a faster rate in females compared to males and 3) the increases are across a range of diagnostic groups. In children, incidence for all cancers combined is greater for males compared to females, while for TYA incidence rates are higher in females, however incidence by sex varies by diagnostic group.

2.6 Cancer survival

The improvement in survival for children diagnosed with cancer since the 1960s has been substantial. The 5-year survival estimates for all cancers combined for children has increased from 30% for children diagnosed in the 1960s [6, 10] to 82% for children diagnosed between 2006 and 2010 [11]. Survival has continued to increase in the 2010s, with predicted 5-year survival for those diagnosed in 2018 estimated to be 85% [12]. There are many factors that have influenced these rates including advancements to treatment protocols, participation in national and international clinical trials, as well as centralisation of specialist centres and the development of supportive care packages [6, 10, 13]. Survival rates for TYA have also improved over time [14] and the overall 5-year survival estimate for TYA diagnosed in the UK between 2001 and 2006 for all cancers combined was also 82% [15]. However in the USA, the rate of increase in survival observed for adolescents and young adults has not been as great as that for older and younger ages and reasons for the lack of progress

for this age group has been attributable to the lack of participation in clinical trials and lack of health insurance among this age group [66].

2.6.1 Childhood cancer survival

There is considerable variation in childhood cancer survival by diagnostic group [11, 17]; Figure 2.2 shows the time trends in 10-years survival by diagnostic group for children in Great Britain [11]. Improvements in survival were observed for all diagnostic groups and for the most recent time period 10-year survival rates vary from 99% for retinoblastoma to 57% for bone tumours [11].

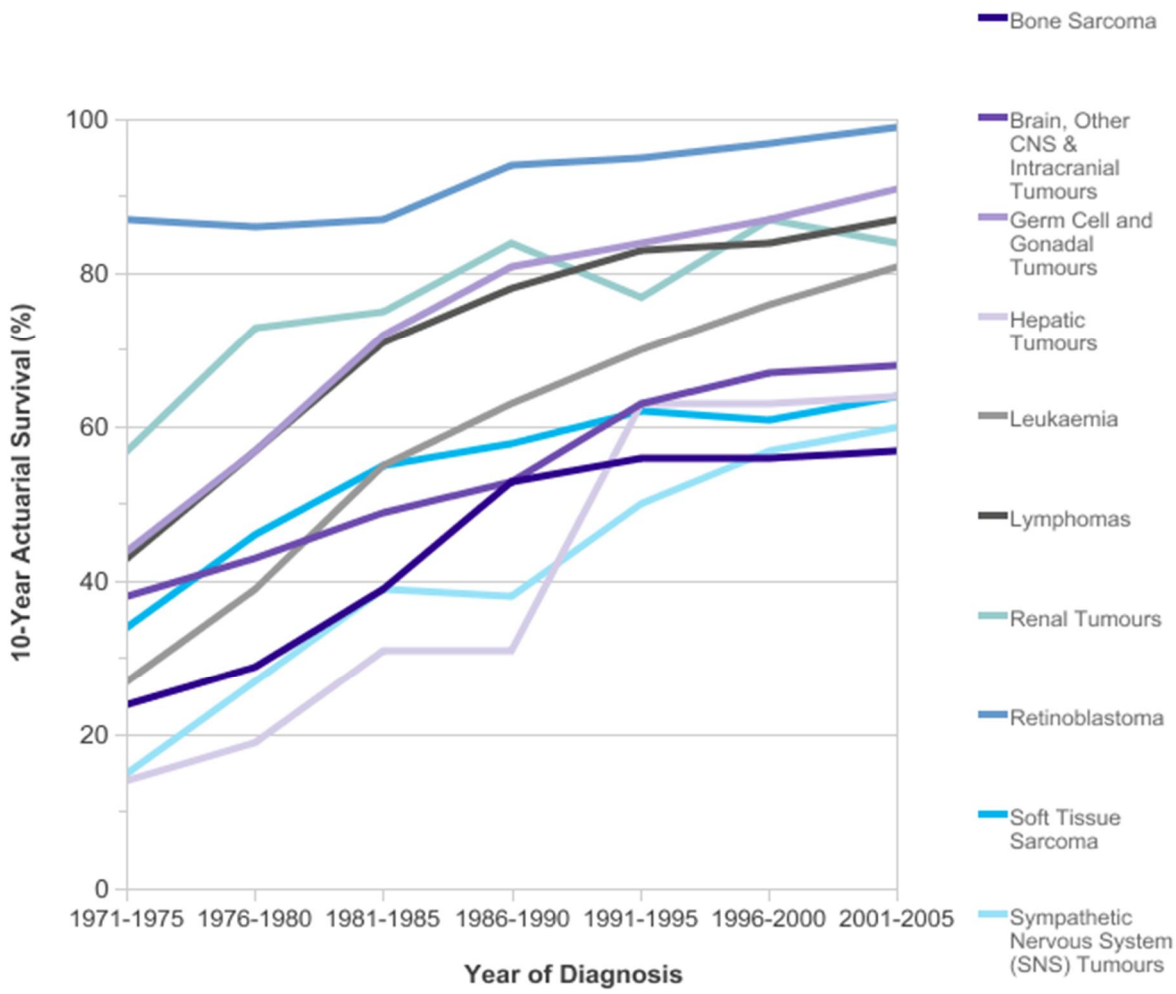


Figure 2.2: 10-year survival by diagnostic group, children (0-14 years) Great Britain.

Source: [11]

There is also significant variation in survival between diagnostic subgroups for the main cancer types. For haematological cancers 5-year survival rates are high ranging from 84% to 95% for ALL, NHL, Burkitt lymphoma and Hodgkin lymphoma, but are lower for AML where 5-year survival was estimated at 63% [19, 76]. 5-year survival for all CNS tumours is around 60%, however survival rates are lower for those diagnosed with embryonal CNS tumours (5-year survival 57%) compared to ependymoma (5-year survival 63%) [76]. Survival rates are also good for nephroblastoma (5-year survival 89%) but lower for other solid tumours, osteosarcoma (69-77%), neuroblastoma (71-72%), rhabdomyosarcoma (68-69%) and Ewing sarcoma (67-68%) [19, 76].

In the EUROCARE-4 study, for all cancers combined the risk of death was 9% higher in males compared to females [19], however in the more recent EUROCARE-5 study there were no difference by sex in survival for all cancers combined but there were sex differences for specific diagnostic groups, for example survival from ALL was higher in girls than in boys [76]. Age at diagnosis is another prognostic risk factor for survival, differences in survival by age at diagnosis vary by diagnostic group. Infants (<1 year) had poorer survival for ALL, AML, NHL and CNS tumours, whereas infants had the highest survival for neuroblastoma compared to older ages. Children aged 10-14 years had the poorest survival for astrocytoma, nephroblastoma and Ewing's sarcoma [76].

Results from the ACCIS and EUROCARE studies have shown that there is geographical variation in cancer survival for children with survival rates lowest in Eastern Europe and the largest survival differences for cancers with the poorest prognosis [20, 76]. There were also inequalities between countries in CNS survival but this may be attributable to differences in registration and classification of these tumours between countries [76].

2.6.2 TYA cancer survival

For TYA aged 15-24 in the UK, there was a significant improvement in survival for all cancers combined and several cancer types, however, there is still significant variation by cancer type [15, 16] (Figure 2.3). Cancer types with high survival (5-year survival >80%) include Hodgkin lymphoma, ovarian germ cell tumours, testicular germ cell tumours, non-gonadal germ cell tumours, melanoma, thyroid cancer, breast carcinoma and cervical cancers [15, 16, 58]. Diagnostic types with moderate survival (5-year survival rates between 50-80%)

include ALL, AML, NHL, CNS tumours, osteosarcoma, non-rhabdomyosarcoma soft tissue sarcomas, ovarian carcinoma, and colorectal carcinoma. Two cancer types, Ewings sarcoma and rhabdomyosarcoma, have poorer prognosis with estimated 5-year survival less than 50% [15, 16, 19].

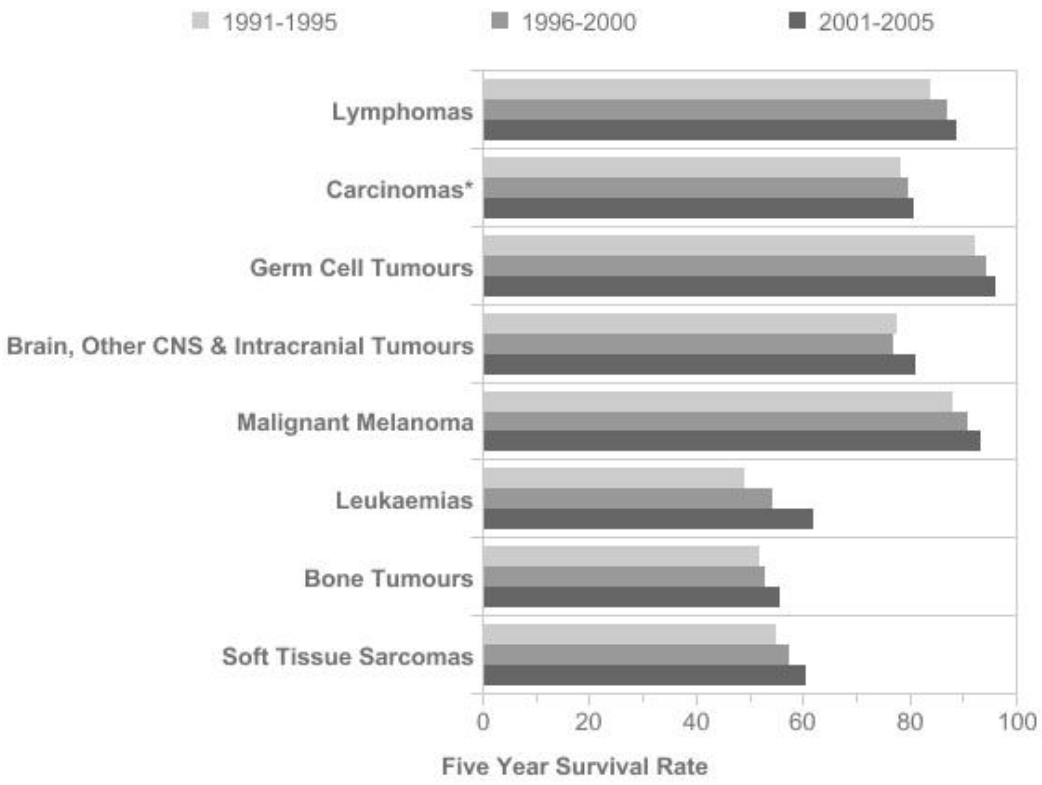


Figure 2.3: 5-year relative survival for TYA diagnosed 1991-2005 in UK.

Source: [15]

A sex difference in TYA cancer survival has been shown for all cancers combined both in the UK and in European studies with females having a survival advantage over males [16, 19, 21]. In the UK between 2002 and 2006, 5-year survival for females was 84% compared to 81% for males [16]. Results from EUROCARE-4 also showed that the risk of death was 18% higher in males compared to females [19]. Sex differences in survival vary by diagnostic group. In the UK, sex differences in survival diminished over time for several cancer types and for patients diagnosed 2002-2006 only remained for melanoma, with estimated 5-year survival of 96% for females and 84% for males, and non-rhabdomyosarcoma soft tissue sarcomas, with estimated 5-year survival of 73% for females and 62% for males [16]. Results from the EUROCARE-5 study

showed that survival was higher in females compared to males for AML, Hodgkin lymphoma, NHL, CNS tumours, soft tissue sarcomas, melanoma, thyroid carcinoma, breast carcinoma, head and neck carcinoma, lung and tracheal carcinoma while survival was higher for males for urinary tract carcinomas and gonadal germ-cell tumours, although it should be noted that this study included all patients aged 15-39 years old at diagnosis [21].

Within the TYA age range there are differences in survival by age. Birch et al compared survival between 13-16 years, 17-20 years and 21-24 years and found that the pattern of survival with age varied with diagnostic group: for leukaemia and CNS tumours survival was better in the younger age group but for germ cell tumours survival was highest for the older age group [14]. The EUROCARE-5 study compared survival by 5-year age bands for the 15-39 year age group and reported the 5-year survival for ALL was higher for 15-19 years (62%) compared to 20-24 years (46%) and 25-29 years (48%), but for all other haematological cancers the survival rates were similar in all age groups; approximately 90% for Hodgkin lymphoma, 77% for NHL and 50% for AML [21]. Survival for soft tissue sarcomas was lower in the 15-19 year age group (63%) compared to the 25-29 year age group (69%) [21].

There was also regional variations in cancer survival for TYA by country within Europe and survival was generally highest in Northern Europe and lowest in Eastern Europe [19, 88].

2.6.3 Childhood and TYA survival comparison

For certain diagnostic groups, TYA generally have poorer outcomes than children diagnosed with the same cancer [18-21]. Adolescents, aged 15-19 years, had worse survival than children for ALL, AML, Hodgkin lymphoma, NHL, astrocytoma, Ewing's sarcoma, rhabdomyosarcoma and osteosarcoma, however this age group had better survival than children for medulloblastoma and germ-cell tumours [21]. Survival in TYA aged 15-24 years was lower than in children for lymphoid leukaemia and osteosarcoma for patients diagnosed across Europe from 2000-2002 [19].

Comparing survival in TYA aged 15-39 years to that in older adults aged 40-69 years, found that survival for most carcinomas was higher in TYA with the

exception of colorectal, breast and prostate carcinomas; colorectal survival was similar in both age groups, while survival for breast and prostate carcinomas was significantly lower in TYA [21].

Similarly to the comparison of incidence rates between studies, some caution should be taken when directly comparing different survival rates between studies. There may be differences in inclusion and exclusion criteria, for example different age ranges, different geographic regions which may register cancers differently, particularly regarding those of non-malignant or border line malignancy. For example, pilocytic astrocytoma which has a borderline malignant code in ICD-O3 but in previous editions of ICD-O was classified as a malignant tumour type. Studies also differ in statistical methodology such as the use of relative survival, which accounts for the background mortality rate and is important to include when assessing survival for cancer diagnosed in older ages, but not always used when estimating survival for children as the background mortality rate is low. For example Gatta et al in the Eurocare-4 study estimated both the observed and relative survival but only present observed survival as deaths due to other causes are rare in this age group [19]. This issue will be discussed in detail later in Section 2.7.5.

To summarise, trends have shown significant improvement over time in survival but there are still some specific diagnostic groups with poor prognosis. Age and sex are important prognostic factors to consider when assessing survival. Although overall survival rates are similar for children and TYA there are certain diagnostic groups where survival for TYA is poorer than that of children.

2.7 Statistical methods in survival analysis

Survival analysis studies are based on time to event data, where the main outcome is the time to an event of interest. Often this event is death but may be any event of interest such as relapse or disease progression. Individuals are followed up from a defined starting time, such as the date of cancer diagnosis, and only some individuals may experience the event of interest by the end of the follow-up period, therefore the survival times for all individuals may not be know, this is known as censoring. Censoring may occur for several reasons such as; the individual may not have experienced the event by the end of the study period, the individual may be lost to follow-up at some point during the study period or the individual may experience another outcome and therefore

follow-up for the event of interest is not feasible, for example if the event of interest is death from cancer but the individual dies from another cause they would be censored at time of death [89]. Survival times are often skewed and may have many events near the start of follow-up and fewer events later on [89].

2.7.1 Survival and hazard functions

Three key mathematical functions in survival analysis are the survival function, $S(t)$; the hazard function, $h(t)$; and the cumulative hazard function, $H(t)$. These three functions can be defined as transformations of each other as defined below [90]. Let the random variable T be the survival time since the origin of the study ($t = 0$). T can be any non-negative value. The survivor function or probability, $S(t)$, is the probability that an individual survives from the time origin to a specified time t

$$S(t) = \Pr(T > t)$$

The hazard function, $h(t)$, is the rate that an individual who is under observation at time t has an event at that time. It is the instantaneous event rate for an individual who has already survived to time t

$$h(t) = \lim_{\delta t \to 0} \frac{\Pr(t \leq T < t + \delta t | T > t)}{\delta t}$$

The cumulative hazard function, $H(t)$, is the integral of the hazard function and is the accumulation of risk by time t

$$H(t) = \int_0^t h(u)du$$

There are important relationships between these functions

$$S(t) = \exp\{-H(t)\}$$

$$S(t) = 1 - F(t)$$

$$H(t) = -\ln\{S(t)\}$$

Where $F(t)$ is the cumulative distribution function of T.

2.7.2 Kaplan-Meier

The survival probability can be estimated from the observed survival times using the Kaplan-Meier (KM) method [91]. Suppose that k patients have events

in the period of follow-up at distinct times $t_1 < t_2 < t_3 < \cdots < t_k$. Events are assumed to be independent, therefore the probabilities of surviving from one interval to the next can be multiplied together to give the cumulative survival probability. The probability of being alive at time t_j, $S(t_j)$, is calculated from , $S(t_{j-1})$, the probability of being alive at time t_{j-1}, n_j the number of patients alive just before t_j, *and* d_j, the number of events at t_j

$$S(t_j) = S(t_{j-1})\left(1 - \frac{d_j}{n_j}\right)$$

Where $t_0 = 0$ and $S(0) = 1$. The value of $S(t)$ is constant between times of events and the estimated probability is a step function that changes value at the time of each event. Each individual contributes information to the survival probability for as long as they are event-free. The survival curve or Kaplan-Meier curve is a plot of the KM survival probability against follow-up time and is frequently used and useful nonparametric measure to describe the survival function. It is plotted as a step function with joins at each time of death or event of interest. The graph will reach 0 if the individual with the longest follow-up time dies, otherwise it will plateau at the time of the last death and continue until the censored survival time of the longest surviving individual [92]. Visual emphasis may be placed on the right hand tail of the curve which may be unreliable and unstable as the number of individuals at risk becomes smaller over time and this estimate may be based on only a few cases or events [92, 93].

2.7.3 Cox model

The most commonly used regression model for analysis of survival data is the Cox proportional hazards model [94] which takes the form

$$h_i(t) = h_0(t)\exp(\boldsymbol{x_i}\beta)$$

Where the hazard function of the *i*th individual, $h_i(t)$, is conditional on covariates x_i, where $\beta = \beta_1, \ldots, \beta_k$ is the vector of regression coefficients and the baseline hazard function is $h_0(t)$. Hazard ratios, estimated from the model, describe the multiplicative effect of covariates on the event of interest. The main assumption of the Cox model is that the estimated parameters are not associated with time and are proportional over time, (also known as a proportional hazards model). This means that the estimated hazard ratio is the same regardless of the length of follow-up. The baseline hazard function is not specified and the Cox model is known as a semi-parametric model, and parametric assumptions are only made about the effects of the covariates on the hazard function but not about the shape of the hazard function. The

baseline hazard function is estimated non-parametrically and not assumed to follow a particular statistical distribution. A major advantage of the Cox model is that the baseline hazard does not need to be defined, it can take any shape, constant, increasing, decreasing or a combination of these, but assumes that is baseline hazard is the same for everyone [95]. A limitation of the Cox model, in addition to the assumption of proportional hazards, is that since the baseline hazard is not estimated we can only estimate the relative differences between groups and not absolute differences in risk or survival [93].

2.7.4 Parametric survival models

Another approach to estimating the hazard function is to assume that the survival times follow a specific mathematical distribution, known as a parametric model [96]. Parametric models generally provide smooth estimates of the hazard and survival functions for combinations of covariates [93]. Many different functions can be used with the most commonly being the Exponential, Weibull, log-normal, log-logistic, Gompertz and Gamma distributions, these models can be easily implemented in statistical software packages for example the `streg` command in Stata [95]. These distributions vary in complexity and the shape of the estimated hazard. The exponential distribution corresponds to a constant hazard rate over time. The Weibull distribution assumes either a strictly increasing or decreasing rate over time. The log-normal distribution corresponds to a combination of increasing and decreasing hazard rate over time. The log-logistic model is similar to the log-normal model, however the log-logistic model has simpler mathematical expressions of the hazard and survivor functions (that do not include the cumulative distribution function). The Gompertz model represents a hazards that either increase or decrease exponentially with time. The Gamma distribution is a highly flexible hazard function that allows for many possible shapes and includes the Weibull and exponential distributions as special cases [95]. These models still assume the proportionality of hazards, and hazard ratios are interpreted in the same way as the Cox model [96], however it is easier to estimate predicted survival from a parametric survival model compared to the Cox model and the parametric model is slightly more efficient resulting in more precise estimates [96].

In parametric cure models (described in Section 2.8) the Weibull distribution is commonly used due to the flexibility of the function and model fit [26, 28]. The distributional functions for this model are described below as these will be

refererred to later in this chapter. The Weibull hazard function takes the form [92, 95]

$$H(t) = \lambda t^{\gamma}$$

$$h(t) = \lambda\gamma t^{\gamma-1}$$

$$S(t) = \exp(-\lambda t^{\gamma})$$

Where γ is the shape parameter which is >0 and λ is the scale parameter. If $\gamma = 1$ then $h(t) = \lambda$, which is equivalent to the exponential distribution, if $\gamma < 1$ then the hazard is monotonic decreasing and if if $\gamma > 1$ the hazard is monotonic increasing. Figure 2.4 shows the hazard and corresponding survival functions for different values of the shape parameter γ.

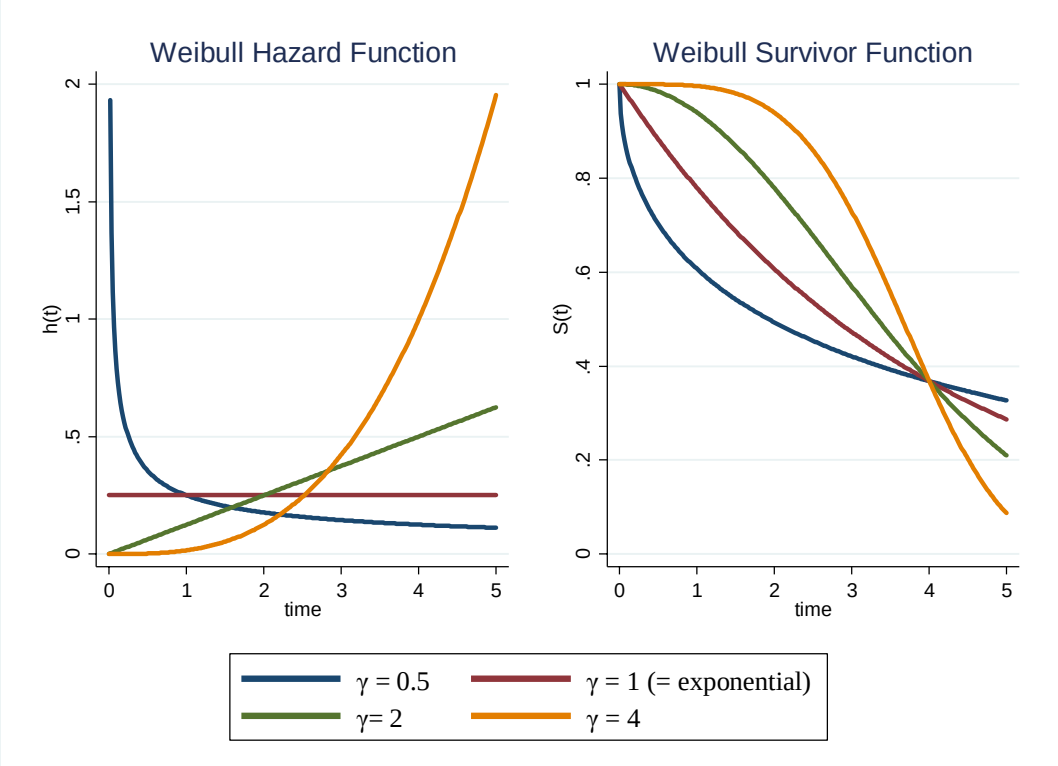


Figure 2.4: Weibull hazard and survival functions for various shape parameters

2.7.5 Relative and net survival

In population-based cancer studies relative and net survival are commonly used methods to estimate survival. Net survival is defined as the survival which might occur if all risks of dying from other causes were removed [97]. It is a theoretical measure to obtain the proportion of patients dying from the direct or indirect consequences of cancer. Two approaches to estimating net survival are used: cause-specific survival and relative survival [22, 98]. Cause-specific survival relies upon not only cause of death being recorded but also to be able to

classify deaths into one of two groups: those directly related to cancer and those unrelated to the cancer under study, this is not always possible and may result in misclassification of deaths [22, 98]. Misclassification of cause of death has a greater impact on survival for cancers with poorer prognosis than cancers with good prognosis [98]. Relative survival is defined as the ratio of observed survival in a group of cancer patients compared to that expected in the general population with similar characteristics with respect to age, sex, calendar period and possibly other factors such as deprivation or geographic region [99]. Relative survival is a standard method of estimating survival for population-based cancer registry data and estimates the excess mortality associated with a diagnosis of cancer regardless of whether the death is directly or indirectly attributable to cancer [22]. Lifetables are used to calculate the expected mortality rate in the general population as it is difficult to obtain a cohort of cancer free-patients, with the assumption that the cancer deaths are a negligible proportion of all deaths [99].

Relative survival as a function of time, $R(t)$, can be defined as

$$R(t) = \frac{S(t)}{S^*(t)}$$

Where $S^*(t)$ is the expected survival and $S(t)$ is the observed (all-cause) survival for the cancer patients. The hazard analogue of relative survival is the excess hazard rate

$$h(t) = h^*(t) + \lambda(t)$$

Where $h(t)$ is the observed mortality rate amongst the cancer patients and $h^*(t)$ is the background mortality rate in the general populations (matched for age, sex and possibly other covariates) and $\lambda(t)$ is the excess mortality associated with a diagnosis of cancer (or other disease of interest).

Different methods exist for measuring the expected survival including life table approaches, with the three most common estimates being the Ederer I method, the Ederer II method and the Hakulinen method [100-102]. These methods differ in the in the calculation of the expected survival, due to differences in the length of follow-up the matched individuals are at risk for. There is little difference in these estimates for five-year survival estimates, however the Ederer II method is preferred when considering longer-term survival [100, 102, 103].

In childhood cancer studies relative survival is not often used as the mortality rates in the background population from causes other than cancer are low, for example in the EUROCARE studies on childhood cancer survival only overall survival was reported [19, 76]. However, survival estimates based on the TYA age range are frequently estimated within the relative survival framework (for example studies in the UK [14, 16], Europe [21] and USA [64]). Figure 2.5 shows the mortality rate for England by age and sex for the years 2013-2015 separately for males and females [104]. It can be seen that apart from in the first year of life where there is a slight increase, the mortality rate for both sexes is low in those aged under 30 and starts to increase from about age 50 onwards.

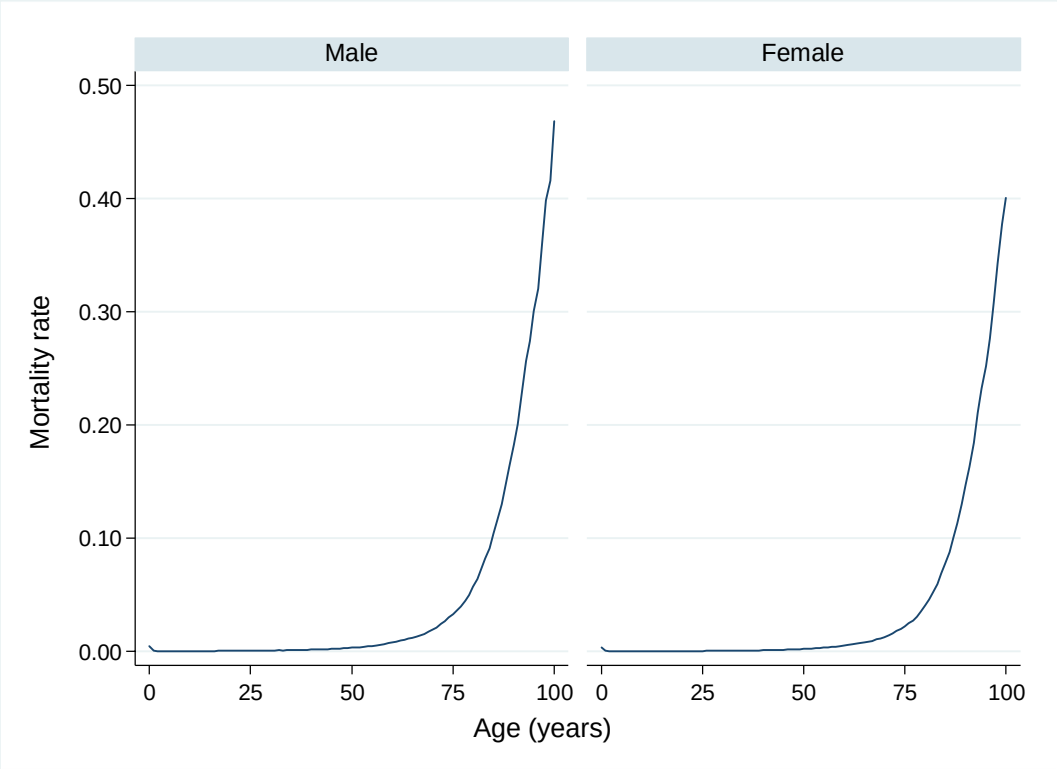


Figure 2.5: Mortality rate by age and sex, England, 2013-2015.

Source: [104]

For relative survival to give an estimate of net survival the mortality due to cancer and due to other causes should be independent [101]. A further method to estimate net survival was suggested by Pohar-Perme et al in 2012 [105] where each observation is weighted by the inverse of the expected survival to obtain an estimated of the net survival [105]. This estimator accounts for the

‘informative censoring’ bias where some groups of patients are less likely to be observed for the full duration of follow-up than others and is an unbiased estimator or net survival in a relative survival framework. When comparing survival across countries, for example in the EUROCARE studies, survival estimates are often age-standardised to account for differences in survival by age and to provide one singe measure to compare survival [101]. It has been shown that when calculating age-standardised survival the Ederer II and Pohar-Perme estimates are similar and the bias is small [101, 106]. The Pohar-Perme estimate of net survival is the preferred method to estimate net survival as recommended by the United Kingdom and Ireland Association of Cancer Registries [107].

2.7.6 Relative survival models

A Poisson modelling approach can be used to model relative survival, where the follow-up timescale is split into a number of intervals and the excess mortality rate within each follow-up interval is calculated [108]. The model assumes the excess hazard rate is constant within each interval which may not be appropriate depending on the length of the intervals. Yearly intervals are commonly used but this may be inappropriate particularly in the first year of follow-up where there is a large change in the hazard and the timescale may need to be split into narrower time bands, for example monthly intervals.

2.7.7 Flexible parametric survival models

Flexible parametric models, also known as Royston-Parmer models, have been proposed as an alternative to the Cox model and other parametric models for the analysis of survival data [93, 109]. These models treat time continuously rather than splitting the follow-up time (as in Poisson models). Restricted cubic splines are used to estimate the shape of the baseline hazard and this modelling approach is useful for the incorporation of time-dependent effects and easily extended into the relative survival framework [93, 101]. These models are parametric although by using splines to model the survival curve they are more flexible that standard parametric survival models [110].

If we consider the Weibull model which is a proportional hazards model with the limitation that the shape of the baseline hazard is either monotonic increasing or decreasing [110], the survival function is

$$S(t) = \exp(-\lambda t^{\gamma})$$

Transforming this to the log cumulative hazard scale, gives

$$ln\{H(t)\} = ln[-ln\{S(t)\}] = ln(\lambda) + \gamma ln(t)$$

Which is a linear function of log time on the log cumulative hazard scale. Adding covariates gives

$$ln\{H(t|x_i)\} = ln(\lambda) + \gamma\, ln(t) + x_i\beta$$

The baseline log cumulative hazard function is $ln(\lambda) + \gamma\, ln(t)$, with covariates additive on this scale. Flexible parametric models relax the assumption of linearity of log time by using restricted cubic splines. Under the proportional hazards assumption, the covariates can be interpreted as (log) hazard ratios because proportional hazards also imply proportional cumulative hazards. The cumulative hazard as a function of log time is generally a stable function [93, 110].

2.7.8 Restricted cubic splines

Splines are flexible mathematical functions defined by piecewise polynomials with some constraints to ensure the overall curve is smooth [93]. The points at which these polynomials are joined are called knot points. The fitted function is forced to have continuous 0th, 1st and 2nd derivatives. Cubic splines are often used in practice. Restricted cubic splines restrict the fitted function to be linear before the first knot and after the final knot. Restricted cubic splines with K knots can be fit by creating k-1 derived variables. For knots $k_1, \ldots, k_K$ a restricted cubic spline can be written as

$$s(x) = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \cdots + \gamma_{K-1} z_{K-1}$$

The derived variables z_j are calculated as

$$z_1 = x$$

$$z_j = (x - k_j)_+^3 + \phi_j (x - k_1)_+^3 - \left(1 - \phi_j\right)(x - k_K)_+^3 \qquad j = 2, \ldots, K-1$$

Where

$$\phi_j = (k_K - k_j)/(k_K - k_1)$$

The derived variables can be highly correlated and orthogonalization can lead to more stable parameter estimates and quicker convergence [93]. Restricted cubic splines give more realistic estimates in the tails of the distribution where the data are often sparse compared to standard splines [93].

2.7.9 Flexible parametric models incorporating splines

Flexible parametric models are fitted on the log cumulative hazards scale, a proportional hazards models can be written as

$$ln\{H(t|x_i)\} = ln\{H_0(t)\} + x_i\beta$$

A restricted cubic spline function of $ln(t)$, with knots k_0, can be written as $s\{ln(t)\,|\gamma, k_0\}$. This is then used for the baseline log cumulative hazard in a proportional hazards model

$$ln\{H(t|x_i)\} = \eta_i = s\{ln(t)\,|\gamma, k_0\} + x_i\beta$$

For example, with four knots

$$ln\{H(t|x_i)\} = \eta_i = \gamma_0 + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \gamma_3 z_{3i} + x_i\beta$$

Which can be transformed to the survival and hazard scale

$$S(t|x_i) = exp\{-\exp(\eta_i)\}$$

$$h(t|x_i) = \frac{ds\{ln(t)\,|\gamma, k_0\}}{dt}\exp(\eta_i)$$

The hazard function is calculated from the derivatives of the restricted cubic spline function. Covariate effects can be interpreted as log hazard ratios under the assumption of proportional hazards [110].

The fitted model depends on the number and location of knot points. These models can be implemented in Stata using the `stpm2` command with the knot points placed at the centiles of the distributions of uncensored log event times [110]. Better fitting models are obtained using the log of follow-up time, given the generally positively skewed distribution observed when analysing survival time [93]. In Stata the default number of knots is 2 (3 degrees of freedom), with knot points placed at the 33rd and 67th centiles of the uncensored log survival times [110]. Studies have shown that hazard ratios are generally insensitive to the number and location of knots and the choice of knots is not crucial [93]. Too many knots will overfit the baseline hazard and too few knots with underfit it, usually between 1 and 5 will be sufficient but this may depend on the size of the dataset [93]. The AIC and BIC can be used to compare models to select the optimal number of knots to use, however the default knot positions generally work well [93].

Flexible parametric models are easily extended to include time-dependent effects by including interactions with spline terms and the covariates of interest. For each time-dependent effect there is an interaction between the covariate and the spline variables. The number of spline variables for a particular time-dependent effect will depend on the number of knot points. The model allows for non-proportional cumulative hazards, and if the hazard ratio is a function of time the best way to estimate this is to plot it as a function of time with 95% confidence intervals [110].

2.7.10 Flexible parametric models in relative survival framework

Flexible parametric models also easily extend into the relative survival framework [93]. The all-cause hazard (mortality) rate, $h(t)$, can be defined as the sum of two components: the background mortality rate, $h^*(t)$, and the excess mortality rate associated with the disease of interest defined by $\lambda(t)$. The background mortality rate is assumed known and usually estimated from national or regional life tables, often stratified by age, sex, calendar year and other covariates as appropriate (e.g. deprivation). The flexible parametric model defined above can be adapted to relative survival and models the cumulative excess hazard on the log scale using restricted cubic splines [93]. An advantage of using this type of model over a Poisson model for relative survival is that the flexible parametric model uses the continuous timescale and does not require splitting the follow-up time which may be computationally intensive for large data sets. It is also easy to incorporate time-dependent effects into the flexible parametric model [93, 110].

2.8 Cure models

Cure models offer an alternative approach to standard survival methods to model long-term outcomes when some of the individuals will not experience the event of interest and can be defined as being ‘cured’ of their cancer [25-27, 111]. Plots of survival curves can be used to identify if a particular group of patients can be defined as being cured. When working in the relative survival framework, for most cancers the relative survival will reach a plateau some years after diagnosis indicating that the mortality among the group of patients still alive is the same as in the general population [26]. Likewise when assessing overall survival if the Kaplan-Meier curve flattens out and plateaus

then a cure model may be appropriate [111]. Figure 2.6 shows a schematic representation of statistical cure when the survival curves flattens out and plateaus during the follow-up period, with the dashed line showing the proportion of patients cured. In these situations a cure model may be appropriate and useful to describe the survival.

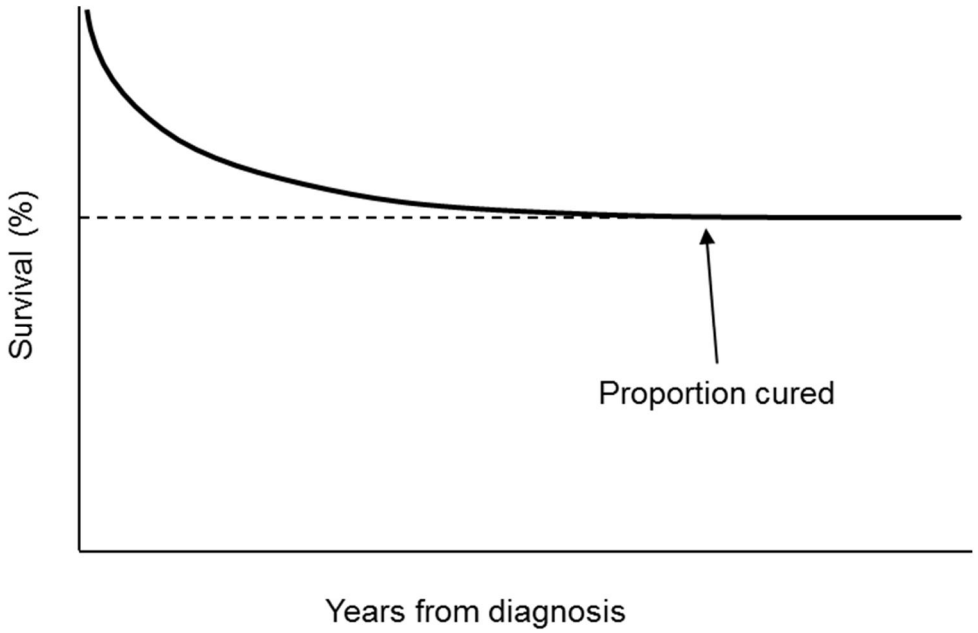


Figure 2.6: Schematic representation of statistical cure

Rather than assuming all patients follow a single survival distribution, cure models assume patients can be split into one of two groups; those who are ‘cured’ and those who are not with a separate survival curve modelled for the ‘uncured. Cure models were first proposed by Boag in 1949 [112] and Berkson and Gage in 1952 [113]. Both studies included the long-term outcomes of cancer patients where a proportion of the patients were estimated to be cured of their disease. Two main types of cure models have been extensively described and developed: the mixture and the non-mixture model. Both of these models have also been extended to the relative survival framework. More recently a flexible parametric cure model has been proposed and also developed in the relative survival framework. These three different types of models are described below.

2.8.1 Mixture cure model

A mixture cure model assumes there are two groups of patients; those who are cured, and therefore have the same mortality rate as the general population, and those who are not cured [25, 26].

Assume that a fraction π (between 0 and 1) of patients is cured and a fraction $(1-\pi)$ is not. Standard mixture cure models are of the form

$$S(t) = \pi + (1-\pi)S_u(t)$$

Where π is the proportion cured and $S_u(t)$ is the survival function for the uncured individuals which can be estimated parametrically or non-parametrically. The hazard function of this model is

$$h(t) = \frac{(1-\pi)f_u(t)}{S(t)}$$

Where $f_u(t)$ is the probability density function associated with $S_u(t)$.

This model extends to the relative survival framework [26]

$$S(t) = S^*(t)\{\pi + (1-\pi)S_u(t)\}$$

Where $S^*(t)$ is the expected survival function. On the hazard scale this becomes

$$h(t) = h^*(t) + \frac{(1-\pi)f_u(t)}{\pi + (1-\pi)S_u(t)}$$

Where $h^*(t)$ is the expected hazard (mortality) rate.

2.8.2 Non-mixture cure model

The non-mixture cure model was developed to model cancer recurrence and defines an asymptote to estimate the cure proportion [25]. The overall survival function is

$$S(t) = \pi^{F(t)}$$

The hazard function is

$$h(t) = -\ln(\pi)\, f(t)$$

where π is the proportion cured and $F(t)$ is a cumulative distribution function generally chosen to be $1 - S(t)$, where $S(t)$ is the survival function and $f(t)$ is

the probability density function for $F(t)$. The survival function has an asymptote at the cure fraction π, and the cumulative hazard has an asymptote at $-\ln(\pi)$.

This model may also be expressed as

$$S(t) = \exp\big(\ln(\pi)\, F(t)\big)$$

$$S(t) = \exp(B(t) \exp\big(a_1 z_1 + a_2 z_2 + \cdots + a_p z_p\big))$$

Where $B(t) = \exp(\alpha_0)\, F(t)$. This corresponds to the integrated baseline hazard function of the Cox regression model. If the parameters $F(t)$ do not vary by covariates, then the above is a proportional hazards model. This is an advantage of the non-mixture cure model over the mixture model, as the non-mixture model has the proportional hazards model as a special case [25, 26].

The non-mixture model can be re-written as a mixture model

$$S(t) = \pi + (1-\pi)\left(\frac{\pi^{F(t)} - \pi}{1-\pi}\right)$$

Therefore the survival distribution of the uncured can also be obtained from the non-mixture model by a transformation of the model parameters [26].

2.8.3 Parametric distributions and link functions

Various parametric distributions can be used to model the survival functions in both the mixture and non-mixture models. The Weibull, lognormal and gamma distributions can all be implemented in Stata commands written to fit cure models, `strsmix` and `strsnmix` [114]. The Weibull distribution is often used as it provides a flexible function and fits well in many situations except when there is a high cure fraction (>80%) or a high excess mortality rate in the first few weeks of follow-up [24, 26]. The log-normal distribution may not fit well as it has a long tail and an imposed rise and fall of the hazard function leading to the estimated cure fraction being based on extrapolation past the end of the follow-up period. The gamma distribution may be useful as it has the Weibull, exponential, lognormal and standard gamma distributions as special cases, however, there may be problems with convergence and this is likely to occur when the Weibull distribution does not provide a good estimate of the cure fraction [114].

The survival function of the Weibull distribution is

$$S(t) = \exp(-\lambda t^{\gamma})$$

Both the scale, λ, and shape, γ, parameters can vary by covariates or remain constant. If these parameters do not vary by covariates then it is assumed that the survival of the uncured is the same for all subgroups of patients which is probably an unrealistic assumption. Therefore, in most cases it is best to allow both the scale and shape parameter to vary by covariates [114].

The cure fraction, π, can also vary by covariates and the dependence of this modelled by different link functions. Assume X is the covariate matrix

1. The identity link $\pi_i = \beta' X$
2. The logistic link $\log(\pi_i/(1-\pi_i) = \beta' X$
3. The log(-log) link $\log(-\log(\pi_i)) = \beta' X$

The identity link has the advantage that is it relatively easy to interpret as it is measured in the units of the proportion cured, however, there may be boundary problems for low or high cure proportions. The logistic link function expressed the covariate effects as (log) odds ratios and has similar interpretation as in logistic regression. The log(-log) link is useful for the non-mixture model as covariate effects are expressed as (log) excess hazard ratios, if the parameters within the distribution do not vary by covariates then proportional excess hazards can be assumed [25, 26].

2.8.4 Flexible parametric cure model for relative survival

Flexible parametric survival models in the relative survival framework described in Section 2.7.10 have been extended to incorporate cure as a special case and allow estimation of the cure proportion and the survival of the uncured [115]. When cure is reached the excess hazard rate is zero and the cumulative hazard excess hazard will be constant after this time. The cure proportion is estimated by forcing the log cumulative excess hazard in the flexible parametric survival model to be linear and have zero slope after the last knot. This is done by calculating the spline variables “backwards”, treating the knots in reverse order and then restricting the linear spline variable to be zero. The spline basis functions, $v_j(x)$, are then defined as

$$v_j(x) = (k_{K-j+1} - x)_+^3 - \lambda_j (k_{max} - x)_+^3 - (1-\lambda_j)(k_{min} - x)_+^3$$

For $j = 2, \ldots K-1$, and $\lambda_j = (k_{K-j+1} - k_{min})/(k_{max} - k_{min})$. The relative survival function for the flexible parametric survival model, with splines calculated backward and with restriction for the linear spline variable is defined as

$$R(t) = exp[-exp\{\gamma_{00} + \gamma_{02}v_2(x) + \cdots + \gamma_{0K-1}v_{K-1}(x)\}]$$

Which can be written as

$$R(t) = \pi^{exp\{\gamma_{02}v_2(x)+\cdots+\gamma_{0K-1}v_{K-1}(x)\}}$$

Where $\pi = \exp(-\exp(\gamma_{00}))$. This is a special case of a non-mixture cure model: the cure proportion is

$$\pi = \exp(-\exp(\gamma_{00}))$$

The distribution function is

$$F_z(t) = \{\gamma_{02}v_2(x) + \cdots + \gamma_{0K-1}v_{K-1}(x\}$$

Covariates can be included

$$R(t;\boldsymbol{z}) = exp\left[-\exp(\gamma_{00} + \beta^T\boldsymbol{z})\, exp\left\{\gamma_{02}v_2(x) + \cdots + \gamma_{0K-1}v_{K-1}(x) + \sum_{i=1}^{D} s(x;\gamma_i)z_i\right\}\right]$$

The constant parameters γ_{00} and β are used to model the cure proportion and the time dependent parameters are used to model the distributional function $F_z(t)$. All spline variables take the value 0 from the point of the last knot, which means that the constant parameter, γ_{00}, is the log cumulative excess hazard at and beyond the last knot for the reference group and can therefore be used to predict cure. The survival of the uncured can be predicted in the same way as the survival of the uncured in the non-mixture cure model [115].

Flexible parametric cure models are fairly robust to the number and location of the knots points, but some caution need to be taken regarding the location of the last knot [115]. It is important it is not placed too early and recommended to be placed at the last observed death or later. It is also important to distribute the knots along the whole follow-up time as the model needs to fit well at the end of the follow-up even when most of the events may be near the beginning [115].

Flexible parametric cure models give similar results to the mixture and non-mixture cure model when cure is a reasonable assumption [28, 115]. A Weibull non-mixture cure model may give biased results in situations when the Weibull model is not flexible enough to capture the shape of the survival function, for example when mortality is high is early follow-up time, however the flexible parametric models offers an alternative approach that gives a better model fit [28, 115]. Flexible parametric cure model have also been shown to fit well in situations when mixture and non-mixture models fit poorly or do not converge for example when survival is relatively high or relatively low [28].

2.8.5 Model assumptions and checking model fit

There are two key assumptions underlying cure models. The first is that cure is a reasonable assumption. This can be assessed graphically from Kaplan-Meier curves or plot of the relative survival over time. If the survival curves plateaus after sufficient follow-up time then cure may be a reasonable assumption, if not then a cure model should not be applied [28]. Breast cancer is an example where the excess mortality does not plateau even many years after diagnosis and cure is not a reasonable assumption [116, 117].

The second assumption is that the survival distribution of the uncured can be described appropriately. The mixture, non-mixture and flexible parametric models described above offer alternative ways to describe the survival of the uncured. The Weibull distribution has been used in many examples and been shown to fit reasonably well. Flexible parametric models use splines to model the underlying survival and may be suitable in situations when the mixture or non-mixture models do not fit the data well [115].

Other factors to consider when modelling cure include the size of the study and the length of follow-up. The proportion cured is based on the point the survival curve levels out which may be many years after diagnosis and the data at this point may be sparser. The length of follow-up must be sufficiently long enough to have observed cure, this will depend on the cancer type and may vary by other covariates.

There are currently no diagnostic tools to directly assess model fit for cure models. Akaike Information Criterion (AIC) has been used to select the best

fitting model, however this compares model fit over the whole timescale and for cure models the interest is in the model fit towards the end of follow-up where data are more sparse, therefore this evaluation method should be interpreted with caution [28, 115, 116]. Graphical assessment of the model fit is recommended by comparing the predicted relative survival from the cure models with empirical life table estimates, for example using the Ederer II method [28].

2.8.6 Additional benefits of cure models

Cure models offer additional information to standard survival estimates that may be of interest to clinicians, epidemiologists and patients. While 5-year (or 10-year) survival estimates are often produced these rates are not synonymous with cure (in this case statistical cure where the mortality rate in a group of cancer survivors is equivalent to that in the general population). The information obtained from a cure model on the survival trends of the uncured, in addition to the cure proportion, are informative particularly when assessing temporal trends in survival. For example cure models can identify if survival is improving because more patients are cured, or if the survival of patients who ultimately die has increased or a combination of both. Figure 2.7 shows four different situations in improved survival that may be identified via a cure model which are measured in terms of the proportion of patients cured and the median survival time of the uncured at two time periods T_0 and T_1: a) an improvement in both; b) an improvement in the proportion of patients cured but a decrease in the median survival time of the uncured; c) no change in the proportion of patients cured but an increase in the median survival time of fatal cases and d) an increase in the cured proportion but no change in the median survival time of the uncured [118].

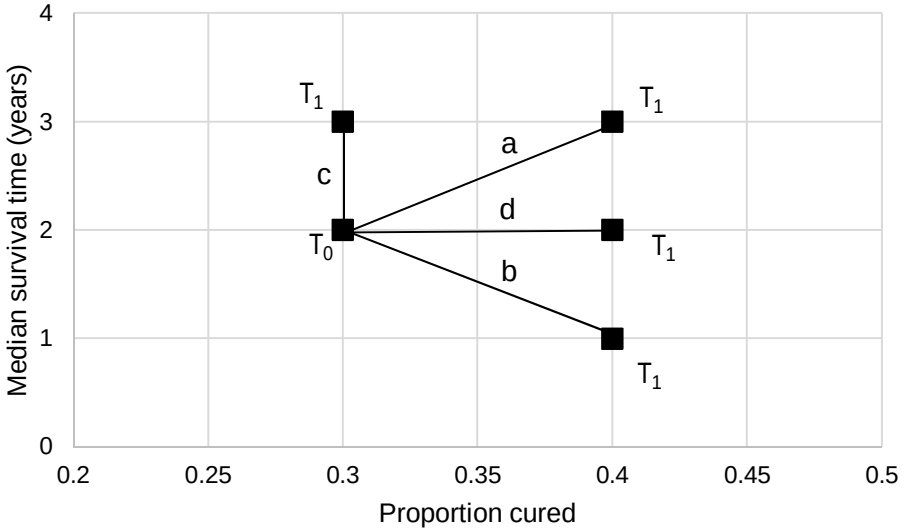


Figure 2.7: Temporal changes in cure model estimates measured at two time periods T_0 and T_1.

Source: [118]

In a cure model covariates may have different associations with the proportion cured and the survival of the uncured. Therefore this methodology provides a single analytic method to study a patient's survival and assess both the long-term outcomes and short-term effects [25, 27].

2.8.7 Estimating time to cure

Standard estimates obtained from a cure model are the proportion of patients cured and the median survival time of the uncured. Some authors have also defined and estimated the time to cure using different methodology as described below.

Time to cure can be defined as the time when an arbitrary but small proportion of fatal cases are still alive [24]. A 1% threshold has been used to estimate the average time to cure for childhood leukaemia patients [29] and glioblastoma patients [119]. However, this arbitrary threshold is sensitive to both the sample size and the length of follow-up.

An alternative approach to measure the time to cure can be estimated from the number of years needed so that conditional relative survival in the following five-years exceeds 95% [120]. Conditional relative survival is the cumulative survival

in the following X-years given that the patient has already survived a certain number of years. When the relative survival curve reaches a plateau, and therefore statistical cure can be defined, the conditional relative survival approaches 100%. Although the 95% threshold is an arbitrary cut point, it reflects the time when the mortality rate in the group of cancer patients is very similar to that of the general population. This method has been applied to data from Italy [120] and has been further extended by Dal Maso et al [121] who report the time when the 5-year conditional survival exceeds 95% and also when 5-year conditional relative survival reaches 90% and when the 10-year conditional survival exceeds 95%. These different estimates provide sensitivity analysis around the estimate of the time to cure. These three estimates of the time to cure showed consistency for some cancer sites but also variation for other cancers [121], highlighting the difficulty identifying the time to cure and the need to ensure sufficient follow-up of patients.

The minimum follow-up period required to allow the estimation of statistical cure varies by cancer site. A study based on SEER data estimated the threshold year, which is the minimum years of follow-up needed to estimate statistical cure [122]. This varied from 2.6 years for pancreatic cancer to 25 years for cancer of the salivary gland. These estimates were based on cancer-specific survival rates and there are issues with the coding and classification of death due to cancer. However, this study does highlight that the minimum follow-up period required depends on the cancer under study and this is independent of the proportion of patients cured. For thyroid and breast cancer even after 27 years follow-up cure could not be defined due to increased excess mortality many years after diagnosis [122], other studies on breast cancer have also found this [116, 117].

It is not recommended that the time to cure is estimated from the flexible parametric cure model [123]. Although flexible parametric cure models are robust to the number and placement of the knot points, in a cure model the knots should be placed over the whole follow-up period and the last knot should be positioned at the last observed death time or possibly later. This last knot point is used to estimate the cure proportion. Flexible parametric cure models with different knot points can be compared formally using model fit statistics such as the AIC. By comparing models where the last knot point is placed at different time points it may be possible to predict the estimated time to cure, however this approach is not recommended as the comparison between models

mainly relies upon differences at the beginning of follow-up where most of the data are [123].

2.9 Application of cure models in CYA

To critically review the applications of cure models for children and young people with cancer a literature review was carried out in Medline, Embase and Web of Science to identify all published papers up to 2016.

Table 2.3 shows the terms used in the search strategy covering three topics: cure models, cancer and children and young people and Figure 2.8 shows the resulting papers identified from the search and included in the review. A total of 46 studies were identified where cure models had been applied to exclusively, or included children or young adults with cancer, a summary table of the included studies is included in Appendix A.

Table 2.3: Terms and phrases used in literature search

Cure models	**Cancer**	**Children and young adults**
Cure* adj2 model*	Neoplasm/	Child*
statistical adj cure*	neoplasm*	paediatric
proportion adj1	cancer*	pediatric
cure* fraction adj1	cancer regist*	adolescen*
cure*	tumour*	teenage*
	tumor*	young adult*
	malignanc*	CYA
	oncology	TYA
		AYA

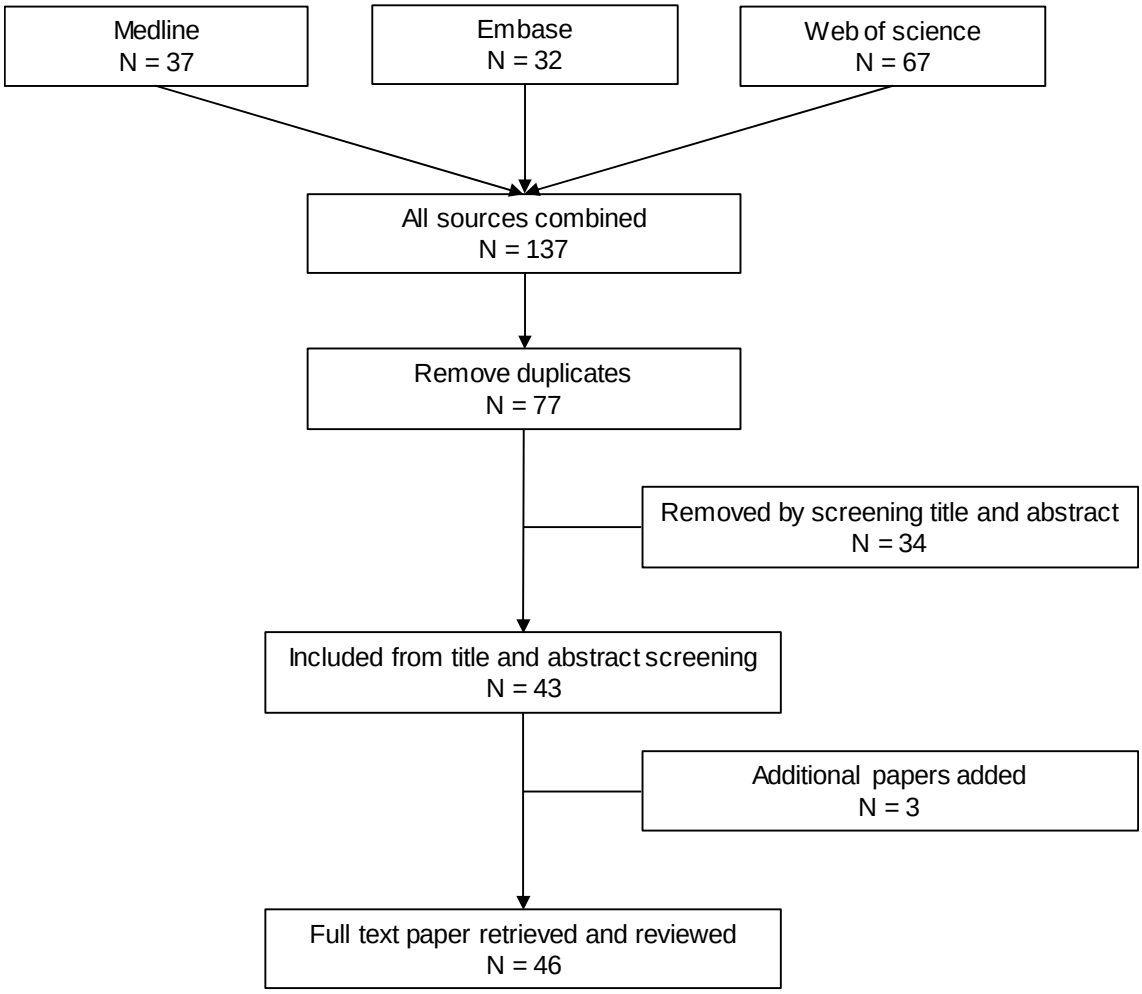


Figure 2.8: Flow chart of literature search results

2.9.1 Sources of data

In total 31 studies were based on data from population-based cancer registries, 12 on clinical trial data and the remaining 3 studies were based on clinical hospital data on patients undergoing surgery for specific cancers.

2.9.2 Types of cure model

Mixture cure models were used in 29 studies, 10 studies used a non-mixture cure model, 2 studies used both, flexible parametric cure models were used in 4 studies and 1 study did not explicitly define the type of cure model used.

2.9.3 Age of study participants

Thirteen studies were based on the childhood and teenager and young adult (TYA) age range (i.e. <30 years at diagnosis) exclusively. The majority of these were studies based upon clinical trials (11 studies); these studies were mainly in children, however the age range was not always explicitly defined in the papers. The remaining 2 studies were based on population-based cancer registry data, both on leukaemia; one on 0-14 year olds in Great Britain on all leukaemias with further analysis by subgroup [29] and the other study included ALL patients only diagnosed in children (0-14 years) and TYA (15-24 years) in Europe using EUROCARE data [30].

The other studies either included all ages (13 studies) or included patients within a specified age range mainly including patients aged 15 and over (11 studies). Within these studies the cure proportion was often presented by age group which included a younger age group, such as 0-44 years or 15-44 years, however the age groupings were not consistent across studies (see Appendix A).

2.9.4 Diagnostic groups

The studies span a range of types of cancer. Four studies estimated cure rates for all cancers combined and the most frequently diagnosed 20 or more cancer sites in adults [120, 121, 124, 125]. Considering studies based on individual cancer types, colorectal cancer was the most frequently included cancer type (9 studies [24, 118, 126-132]). Colorectal cancer is a good example of a cancer where the application of the cure model is appropriate as the relative survival curve plateaus after about 8-10 years follow-up and it is one of the most common cancers diagnosed in adults, however it is not commonly diagnosed in children and young people.

Haematological cancers were one of the most studied diagnostic groups, included in 13 studies although some papers included more than one subtype; ALL was included in 5 studies [25, 29, 30, 133, 134], AML in 4 [127, 135-137], Hodgkin lymphoma in 4 [25, 138-140], chronic myeloid leukaemia in 1 [141] and NHL in 1 [25]. It is important to consider these subgroups separately as the survival patterns vary by subtype, however some of these applications were on clinical trial data, therefore the results may not be generalizable to the wider population. Seven studies based on cancer registry data [29, 30, 127, 135, 137-139] are particularly relevant for children and young people with cancer and the

key results of these papers are synthesised and critically evaluated in Section 2.9.9.

Three studies have used data from the RARECARE study, which is a project looking at the incidence, survival and prevalence of rare cancers (defined as incidence lower than 6 per 100,000) across Europe, also using data from several cancer registries [142], including CNS glial and non-glial tumours [143], embryonal tumours [144] and germ cell tumours [145]. Although these tumour types are relevant for the childhood and young adult age range, the results from these studies are limited as the authors only present the overall cure proportion and do not provide a further breakdown by age group or report results on the survival of the uncured.

2.9.5 Other risk factors included

Other variables included, either as covariables in the cure model or presented as stratified analysis, were age, sex, diagnostic sub group, study (for studies including more than one study), period of diagnosis, region of diagnosis, metastases, biological markers, stage, mode of detection, treatment, socio-economic status. Detailed information on clinical patient characteristics were more common on clinical trial studies compared to population-based cancer registry studies.

2.9.6 Study sample size

The study sample sizes varied markedly from approximately 100 in some studies (mainly trial data) to over 6 million for studies based on EUROCARE data. One study utilising data from a regional population-based cancer registry in Austria showed it is possible to model cure in smaller populations (as opposed to studies on large national datasets) [125], this was also observed in the study by Bouliotis and Bessel who modelled cure for patients with Hodgkin lymphoma based on a regional registry in Nottingham on 768 patients [138]. The main reasons for non-convergence of models was that long-term survival did not plateau rather than lack of sample size and power.

2.9.7 Length of follow-up

The length of follow-up varied from at least one year to up to 40 years. Studies based on population-based cancer registry data generally had at least 5-years follow-up for all patients (as this is a commonly used metric to measure survival in these cohorts). The length of follow-up needed will depend on the cancer type as it is essential to have enough follow-up time to observe a flattening out of the survival curve, therefore studies based on shorter follow-up periods may provide biased estimates of the proportion cured or may not converge. For example, there are issues modelling cure for breast cancer patients as the excess mortality risk continues many years after follow-up. Woods et al follow-up breast cancer patients up for up 23 years after diagnosis, and found that many of their cure models did not converge as the main assumption of cure was not met [117].

2.9.8 Survival outcomes

Studies using cancer registry data were generally analysed within the relative survival framework. Studies in clinical trials used event-free survival which was defined from time of diagnosis or treatment to relapse, disease progression, secondary tumour or death. Two studies have looked at cure models in childhood leukaemia, one used relative survival [29] while the other did not use relative survival for children but did for the analysis of 15-24 year olds [30].

2.9.9 Summary of key findings

Eight of the papers in this review were methodological papers and found in statistical journals and the emphasis is on the application of the methods rather than interpretation of the model results [25, 133, 134, 146-150]. Studies that looked at all ages and included a separate age group for younger ages generally found cure rates were higher in younger ages. Key studies that included children and young people are described below.

Studies on childhood leukaemia show that the proportion of cured cases has increased over time. In Great Britain for children diagnosed with leukaemia the percentage cured increased from 25% for those diagnosed during 1971-1975 to 73% for those diagnosed during 1996-2000 [29]. The average time to cure was also estimated to increase from 11.0 years to 15.9 years over the same period; this increase was only observed for children with ALL, from 12 years to 19 years, whilst for acute non-lymphoblastic leukaemia the average time to cure

remained about 5 years. In Europe for children diagnosed with ALL, the percentage cured was estimated by age group and was poorest in infants (under 1 year) and highest for children aged 1-4 years, the percentage of adolescents and young adults (15-24 years) cured was similar to that of infants [30]. The percentage cured increased in all age groups with the greatest improvement in infants (under 1 year) from 26% in 1982-84 to 58% in 2000-02. The percentage cured for children aged 1-4 years increased from 70% to 90% [30]. Both these studies included patients diagnosed up to 2002, and no estimates of the percentage cured are available for more recently diagnosed leukaemia patients. Both studies used a mixture cure model, however, neither of them reported results on the survival of the uncured. Other than age and period of diagnosis no other prognostic risk factors were included in the cure models. A key strength of the Shah paper is the estimation of the time to cure [29]. Both studies are based on large populations (the whole of England and Europe) with sufficient follow-up to ensure reliability and precision of reported results.

National studies on temporal trends in AML survival and cure have been conducted in England, including patients aged 15-99 years diagnosed between 1971 and 2006 [137], and Sweden, for patients aged 19-80 years diagnosed between 1973 and 2001 [135]. Estimates were reported for 15-24 years and 25-39 years in the English study and for the 19-40 year age group for the Swedish study. In England the percentage cured increased between 1975 and 2006 for ages 15-69 with the greatest improvement in those aged 15-24 years from 8% to 48%, while for 25-39 years it increased from 6% to 44%. The median survival time of the uncured cases for 15-24 years increased from 5 months to 15 months between 1975 and 2006, and for 25-39 years it increased from 5 months to 11 months [137]. In Sweden for AML patients the largest increase in the cure percentage was also observed in the youngest age group; increasing from 4% in 1975 to 68% in 2000 for those aged 19-40 years. Over this same time period the median survival time of the uncured increased until the beginning of the 1990s, then decreased again and was estimated to be around 9 months for those diagnosed in 2000 [135].

Eloranta et al applied a cure model in a competing risk framework to account for cancer and non-cancer causes of death with an application to AML using national cancer registry data from Sweden [127]. Patients under 50 were included as a separate age group and it was estimated that patients diagnosed

at age 50 the percentage cured (40% for males and 44% for females) was similar to the percentage of patients who were predicted to die from causes other than AML (44% for males and 45% for females). The probability of dying from AML was strongly associated with age, with worse prognosis for older patients.

These three studies on AML were large national population-based studies but two of them did not include children. There are some limitations to the available data, for example no national data on stage was available in Sweden or England, and results were only presented by age at diagnosis and time period.

The percentage cured for Hodgkin lymphoma, estimated from a regional cancer registry in Nottingham, increased from 45% for patients diagnosed 1973-1982 to 77% for those diagnosed 1993-2002 [138]. Over the same time period the median survival time of the uncured increased from 1.5 years to 4.0 years. Although this study collected detailed information on the patients including treatment, stage and death information, these variables were not included in the cure models. A further study on young and middle aged Hodgkin lymphoma patients in Sweden found selective improvement in survival in the 18-29 year age group between 1992 and 1999; the percentage cured increased but the median survival time of the uncured decreased. However between 2000 and 2009 both the percentage cured and the median survival time of the uncured increased with the percentage cured approaching 1 for patients diagnosed in 2009 [139]. This study included detailed clinical information including treatment, stage and relapse information which was incorporated in the cure modelling. However, it was based on patients aged 18-59 years so did not include children or adolescents.

The percentage cured was estimated for patients aged 16-39 years in a study of glioblastoma multiforme using SEER data; the percentage of cases cured in 16-39 year olds was 12% and the time to cure was estimated at 9.8 years [119]. Cure was not estimated for other age groups as population-based cure was unlikely for these groups based on results from the Kaplan-Meier survival curves. The cure model did not include any other prognostic factors and was only estimated for patients diagnosed in one time period (2001-06).

2.10 Survivorship

The previous sections describe methods and studies relating to cancer survival. However, increases in cancer survival rates over times have led to the development of research into the long-term health and late effects of treatment for cancer survivors known as cancer survivorship [151]. This is particularly important for children and young people with cancer with many published studies examining the long-term health of CYA cancer survivors (discussed in Sections 2.11 to 2.15). Currently for many children and TYA diagnosed with cancer the focus is on reducing treatment intensity while maintaining survival rates to reduce treatment-related morbidity and mortality [152].

2.10.1 National Cancer Survivorship Initiative

The 2007 Cancer Reform Strategy highlighted the need for a greater focus on cancer survivorship [153]. In the UK, the National Cancer Survivorship Initiative (NCSI) was launched in 2010 to understand the needs of those living with cancer and develop models of care that meet their needs, and to help support cancer survivors to live as healthy and active a life as possible for as long as possible [154]. In UK in 2008 there were an estimated 2 million people living with a previous diagnosis of cancer and this was predicted to increase by 3% per year as more people are diagnosed with cancer, treatment becomes more effective and people live longer after cancer [155]. Survivors of childhood and young adult cancers are a particular group with special needs. Health services are needed to monitor and support this group who may have long-term consequences from their earlier cancer treatment [156]. Therefore life-long follow-up of childhood and adolescent cancer survivors is recommended [72].

The NCSI also highlighted the need for linkage between routine data sources such as cancer registry data and primary and secondary care data to measure health outcomes for cancer survivors [154]. One of the key factors in studying long-term outcomes for cancer survivors is having detailed information on treatment received as part of their care. The YSRCCYP records detailed treatment information for cancer patients therefore with linkage to hospital data is a very valuable source of data to explore these long-term health effects and their associations with initial treatment.

2.11 Late mortality

Five-year survival is often used as a benchmark to measure cancer survival. Outcomes are measured from the date of diagnosis and those surviving beyond five years from this date are considered to be a long-term survivor. Late mortality of childhood cancer survivors refers to mortality occurring beyond 5-years from diagnosis [79].

Long-term childhood cancer survivors are at an increased risk of excess late mortality compared to the general population and several studies have described these risks. The overall late mortality rate in 5-year childhood cancer survivors has been estimated to be between 8 and 13 times higher compared to the general population [31, 32, 79, 81, 157-161]. The age range of patients included in the studies varies as does the period of diagnosis therefore results may not be directly comparable. These estimates are based the CCSS including those diagnosed aged 0-19 years [31, 79, 157, 159], the BCCSS which includes those at 0-14 years at diagnosis [32, 81] and population-based cancer registry studies in the Nordic region including those aged 0-19 years at diagnosis [158, 160]. In Scotland, the standardised mortality ratio (SMR) was 6.1 for 0-24 year olds, which was 11.0 for children and 4.7 for 15-24 year olds [161].

2.11.1 Cause of death

The main causes of death 5-years post diagnosis are recurrence and/or progression of original cancer, subsequent neoplasms, diseases of the circulatory system and diseases of the respiratory system [31, 32, 79, 81, 160, 161]. Recurrence or progression of the original disease is the leading cause of late mortality accounting for between 58% and 66% of deaths in long-term childhood and young adult cancer survivors [31, 81, 160, 161]. Second or subsequent tumours were the next most common cause of death accounting for between 11% and 19% of all deaths [31, 81, 160, 161]. Diseases of the circulatory system account for 5-7% of deaths [31, 81, 161] and diseases of the respiratory system account for 2-4% of deaths [31, 81, 161].

The leading causes of death change over follow-up period. Recurrence or progression is the leading cause of death in the period 5-10 years post diagnosis. The cumulative mortality of death due to recurrence increases rapidly

with time from diagnosis to about 15 years from diagnosis and then levels off during further follow-up when mortality due to second cancers and other causes of death start to increase [31, 35, 160]. For example in the BCCSS, 97% of deaths are attributable to recurrence in the period 5-14 years post diagnosis and this reduces to 8% of deaths more than 45 years post diagnosis, while over the same time period, deaths due to second primary tumours account for 8% of the excess deaths increasing to 58% [32]. Among survivors aged over 60 circulatory disease overtakes subsequent primary neoplasms as the leading cause of death [81].

2.11.2 Variation by cancer type, sex and age and treatment

The overall mortality rate varies by type of primary cancer, age at diagnosis, sex and treatment received. Significant increased mortality were observed for all diagnostic groups with SMRs highest for leukaemia patients and CNS tumours, in particular medulloblastoma and PNET, and Ewing sarcoma patients [31, 32, 81]. The mortality rate and absolute excess risk of death were higher in males, however the SMR was higher in females compared to males due to lower background mortality rates in females [31, 79, 81]. Children diagnosed between 0-4 years and 5-9 years had higher SMRs compared to those diagnosed at older ages [31, 32, 81, 161]. Reulen et al found that the SMR for all causes was 11.8 for patients treated with radiotherapy compared to 7.2 for those not treated by radiotherapy while patients who received chemotherapy had an SMR of 15.8 compared to 8.1 for those not receiving chemotherapy, however no further details of treatment information were given [32]. Mertens et al used the detailed treatment information recorded in the CCSS and investigated the associations with cause of death and found that exposure to radiation, alkylating agents, and epipodophyllotoxins were associated with an increased risk of mortality from a subsequent tumour; exposure to cardiac radiation and anthracyclines were associated with an increased risk of death from cardiac disease; and exposure to radiation was associated with an increased risk of death due to causes other than recurrence, external causes, subsequent malignancy, cardiac or respiratory diseases [31].

2.11.3 Changes over time

Both the CCSS and BCCSS published studies in 2016 examining late mortality [79, 81] including additional patients diagnosed more recently than those included in previous studies published in 2008 [31] and 2010 [32] respectively.

These studies have shown that late mortality for long-term survivors has reduced over time. In the CCSS, 15 years post diagnosis the cumulative mortality decreased from 10.7% for patients diagnosed in the 1970s to 5.8% for those diagnosed in 1990s [79]. While in the BCCSS, patients diagnosed and treated between 1990 and 2006 experienced 30% of the excess number of deaths experienced by those diagnosed and treated before 1970 [81]. Both studies found reductions in deaths from recurrence and progression and also in health related or non-neoplastic causes of death. Garwic et al also found in the Nordic countries that the SMR in the time period 5-9 years after diagnosis decreased from 30.2 for patients diagnosed in the 1960s to 18.6 for patients diagnosed in the 1990s [160]. In Scotland 0-24 year olds diagnosed 1998-2003 had 46% decreased risk of death compared to those diagnosed 1981-85 [161]. These results provide evidence that reducing treatment exposure in order to decrease the frequency of late effects is translating to a significant reduction in late mortality.

To summarise, childhood cancer survivors have an increased risk of late mortality that remains elevated throughout the life course. The leading causes of death are recurrence or progression of primary cancer, subsequent tumours, cardiac and respiratory diseases. Type of primary tumour, age, sex and treatment modality are all associated with late mortality. There is evidence that late mortality rates are decreasing over time reflecting changes in treatment protocols to reduce exposure to toxic therapies.

2.12 Late effects

As well as having an increased risk of late mortality childhood cancer survivors experience a range of adverse health outcomes. Complications and side effects of treatment can arise anytime following treatment and can generally be split into three phases: acute (during treatment), early (within months after treatment ends) and late effects. Late effect is the term used to describe a long-term adverse health outcome that persists or develops several years after cancer is diagnosed and treated [162]. Several studies have been used to observe and describe late effects in long-term childhood cancer survivors including studies based on clinical assessments which may also pick up sub-clinical disorders, those based on patient self-reported outcomes, as well as studies based on linkage to hospital admissions and other sources of routine data [162]. Late effects in TYA have also been examined but to a lesser extent than for children

and many of the late effects in TYA have been derived from studies of childhood cancer survivors [67].

Based on the CCSS with a mean follow-up of 18 years, it was estimated that: 62% of childhood cancer survivors had at least one chronic condition, which was 3.3 times higher compared to matched sibling controls; 28% had a severe or life threating chronic condition which was 8.2 times higher than in the sibling controls; and 38% of survivors reported multiple chronic conditions which was 4.9 times higher compared to siblings [37]. Survivors of bone tumours, CNS tumours and Hodgkin lymphoma were at highest risk of subsequent chronic conditions and significant associations between treatment received for childhood cancer, sex and age at diagnosis and adverse health conditions were reported [37]. A study from the Netherlands also reported that 75% of long-term childhood cancer survivors had one or more adverse health outcomes and 40% had a severe or life-threatening condition [36]. Results from the St Jude Lifetime Cohort Study in the USA, which included 10-year childhood cancer survivors, reported that 98% of this cohort had 1 or more chronic health conditions and 68% had a severe or disabling or life threatening condition [163]. By age 50, a survivor experienced on average 5 severe or disabling or life threatening conditions compared to an average of 2 conditions in matched controls [40]. Findings from the CCSS have shown that the incidence of serious chronic health conditions in long-term survivors has decreased for those diagnosed more recently; the 20-year cumulative incidence of severe or disabling or life threatening conditions decreased from 33% for those diagnosed 1970-79 to 28% for those diagnosed 1990-99 [164].

The next section of this chapter focuses on three specific outcomes: 1) subsequent malignant neoplasms (Section 2.13), 2) cardiovascular late effects (Section 2.14) and 3) respiratory late effects (Section 2.15). These were selected for inclusion as these are the leading causes of late mortality and morbidity in long-term CYA cancer survivors and were included as outcomes for analysis in this thesis.

2.13 Subsequent malignant neoplasms

It is well recognised that the intense treatments used to treat paediatric cancers increase the patient’s risk of developing another cancer in later life. Deaths from

subsequent malignant neoplasms are one of the leading causes of late mortality in childhood cancer survivors [31, 81, 160, 161].

A primary cancer is one that originates in a primary site or tissue and is not an extension, nor a recurrence, nor a metastasis [165]. A second cancer is defined as a new primary cancer that occurs in a person who has had cancer in the past, this may also be described as a subsequent primary cancer, second primary cancer, subsequent malignant neoplasm or a second malignant neoplasm [166]. Throughout the rest of this thesis the abbreviation SMN is used to define subsequent malignant neoplasm unless otherwise stated. The International Agency for Research on Cancer (IARC), the International Association of Cancer Registries (IACR), World Health Organisation (WHO) and the European Network of Cancer Registries have published rules for defining and recording multiple primary tumours in cancer registries where the recognition and existence of two or more primary cancers does not depend on time [165]. Neoplasms of different morphology should be regarded as multiple cancers even if diagnosed simultaneously in the same site [165]. Studies may use different follow-up periods and eligibility criteria to examine SMNs.

Studies based on the CCSS, BCCSS and TYACSS only include 5-year survivors of childhood cancer and examine all new neoplasms recorded after this date [35, 82, 167, 168]. Several studies have based their definition of SMNs on the IARC and IACR rules for defining multiple primaries irrespective of the time elapsed since the primary neoplasm was diagnosed; including in a large study of SMNs after childhood non-CNS solid tumours based on data from 13 cancer registries [169], a Canadian study focussing on SMNs developing in the first 5 years following diagnosis [170], and a large study based in the Nordic countries [34]. While other studies have examined SMNs in patients who survived a minimum of two months [171], three months [172], three years [173] or five years [174] following diagnosis of a primary cancer.

For childhood cancer survivors the risk of developing a SMN was between 3 – 10 times higher than that of the general population [34, 35, 167, 170, 171]. These studies differed in terms of inclusion criteria of childhood cancer survivors and the definitions used to define the SMN making direct comparison between studies difficult. However, all found a substantial increased risk compared to the general population. Pole et al showed that 40% of SMNs in childhood cancer survivors occurred in the first 5 years following diagnosis [170]

therefore comparisons between studies including all cancer survivors and those only containing 5-year survivors may vary substantially in the number of SMNs reported. The magnitude of excess risks and specific types of second cancer vary widely with type of first cancer and also in length of follow-up. In the CCSS the median time to first occurrence of subsequent malignant neoplasm was 18 years and was shortest for development of leukaemia at 9 years and longest for small intestine and colorectal cancer at 23 years [167].

The cumulative incidence for a SMN continues to increase across the life-course; 25-30 years after diagnosis of a childhood cancer the cumulative incidence of developing a SMN ranged from 4-8% [167, 170, 171, 173, 174]. Long-term childhood cancer survivors from the CCSS had a cumulative incidence of a SMN by age 55 of 16% [168], similar results were also found in the BCCSS where the cumulative incidence by age 55 was 14% [35]. Findings based on the CCSS have shown that the cumulative incidence of SMNs decreased for those diagnosed in the 1990s compared to those diagnosed 1970s [164, 175].

All types of childhood cancer are associated with an increased risk of SMN however several studies have shown that the standardised incidence ratios (SIR) are highest for a primary diagnosis of Hodgkin lymphoma (SIRs range from 6-16 [35, 167, 170, 171, 173]), retinoblastoma (SIRs range from 13-15 [35, 171]) and bone tumours (SIR range from 4-18 [35, 170, 173]) and in particular Ewings sarcoma (SIRs range from 9-13 [167, 171]). The highest cumulative incidence of second malignancy following a non-CNS solid primary tumour occurred after retinoblastoma reaching 18% 50 years after the diagnosis of the primary cancer [169].

A range of SMNs are diagnosed in long-term childhood cancer survivors, studies reporting these have used different coding and groupings of SMN and have different periods of follow-up but the most common sites and types of SMN include female breast cancer, CNS tumours, bone tumours, soft tissue sarcomas, melanoma, thyroid, digestive tumours, genitourinary tumours and endocrine tumours [35, 167-170, 173]. The incidence and range of subsequent neoplasms change over follow-up. Pole et al examined SMN in the first 5-years from diagnosis of a childhood cancer and found early SMNs (those developing within 5 years of diagnosis) were more likely to be leukaemia, lymphoma or sympathetic nervous system tumours and around one third of early SMNs were

solid tumours [170]. The site distribution of second cancers changes over the life-course. In the Nordic countries CNS tumours accounted for 39% of SMN diagnosed in 0-14 years olds but only 9% of SMNs in 60+ years where breast cancers were the most common SMN comprising 32% of all SMNs [34]. In the BCCSS bone tumours and glioma accounted for 50% of the excess risk for patients aged <20 years at diagnosis of a SMN, whereas digestive and genitourinary tumours accounted for 36% of the excess risk in those aged over 40 [35].

An increased risk of SMNs in those who received radiotherapy or chemotherapy has been reported in several studies [35, 167, 168, 170, 173], with the greatest excess risk for those that received both treatment modalities [173]. The risk of developing a colorectal cancer for childhood cancer survivors treated with abdominopelvic radiation is similar to that of individuals with a strong family history of colorectal cancer; cumulative incidence by age 50 is 1.4% for childhood cancer survivors compared to 1.2% those with a family history of colorectal cancer [35].

Other risk factors associated with an increased risk of SMN were female sex [167, 168], age at diagnosis, although Friedman et al [167] reported that older age at diagnosis increased the risk of SMN while Pole et al [170] reported younger age at diagnosis increased the risk, attained age [35] and treatment era [34, 167].

Several studies based on pooled European data on almost 70,000 5-year survivors of childhood cancer have examined the risk of diagnosis of specific subsequent tumours including leukaemia and soft-tissue sarcomas and bone tumours [176-178]. Compared to the general population childhood cancer survivors had 4-times the expected risk of leukaemia [176], 16-times the expected risk of soft tissue sarcoma [177] and 22-times the expected risk of bone tumours [178]. The large sample size of these studies enabled a detailed examination of risks for specific subtypes of these diagnostic groups.

Studies on SMNs in TYA are limited. The TYACSS study examined SMNs in a cohort of 200,000 5-year survivors diagnosed aged 15-39 years in England and Wales and reported the risk of SMNs after each specific AYA cancer type [82]. SMNs were most frequently diagnosed in survivors of breast cancer, cervical

cancer, testicular cancer and Hodgkin lymphoma with the cumulative incidence 35-years post-diagnosis of all SMNs ranging from 12% for breast cancer survivors, to 27% in female survivors of Hodgkin Lymphoma. Lung cancers accounted for a substantial proportion of the excess number of SMNs diagnosed within these groups [82]. A study based on SEER data reported on SMNs diagnosed in a cohort of 150,000 AYAs (aged 15-39 years) in the US after specific AYA cancer types and reported an SIR of 1.6, compared to an SIR of 4.3 or children and 1.1 for older adults [174]. Higher risks of SMNs were observed for patients with a primary diagnosis of AML, Hodgkin lymphoma, NHL, testicular cancer, melanoma, breast cancer and sarcoma, with the highest SIRs for survivors of Hodgkin lymphoma (SIR=3) [174]. For those aged 15-39 years at diagnosis of primary cancer the 30-year cumulative incidence was 18% for those who received radiation compared to 12% for those that did not [174]. A further US study based on two-year survivors aged 15-39 years at diagnosis estimated the incidence of SMNs were 2.6 time higher than matched controls with a cumulative incidence of 13% twenty-years post diagnosis [179]. Older age at primary diagnosis, female sex, ethnic group, advanced stage of disease and radiotherapy exposure were all associated with an increased risk of SMN, although these risk factors varied by first cancer type [179].

In childhood cancer survivors it is estimated that 40% of SMNs are diagnosed within 5-years from primary tumour [170] and in the US, a study of 15-39 year olds estimated that 73% of SMNs were diagnosed 1-5 years from primary diagnosis [180]. There are differences in the types of SMNs diagnosed by latency period: early onset SMNs are more likely to be leukaemias and lymphomas [181]. Few studies have assessed the impact of latency on survival. In AYA aged 15-39 years, the risk of death doubled for those with a latency period of 1-5 years compared to those with a latency period of 6 years or more [180]. While in Canada childhood cancer survivors with early onset SMNs were 1.8 times more likely to die that those who developed an SMN after 5-years [170]. Another US study of SMNs developing before the age of 20 found those with a latency of less than 5-years had lower survival, but this study only included primary solid tumours [182].

Survival from SMNs is lower in children and AYAs compared to survival rates for the same type of primary tumour; 5-year survival in children was 80% for primary tumours compared to 47% for SMNs and in AYA 81% for primary tumours compared with 60% for SMNs [183]. Another study of childhood cancer

survivors reported an increased risk of death for those with a SMN compared to those diagnosed with a first cancer after adjustment for potential confounders including sex, age at diagnosis, decade of diagnosis, ethnicity and diagnostic group [184]. An increased risk of death was found across several diagnostic groups: breast cancer, thyroid cancer, AML, CNS tumours, melanoma, bone cancers and soft tissue sarcomas [184]. A Dutch study found survival from sarcoma SMNs was worse than for patients with a first primary sarcoma, and there were no survival differences between primary tumours and SMNs for breast cancer or melanoma. However, this study was based on a small number of long-term survivors (45 sarcoma, 41 breast cancer and 17 melanoma survivors) [185].

To summarise, the risk of developing a SMN continues to increase throughout follow-up and varies by type of primary cancer and other risk factors such as sex, age at diagnosis, length of follow-up and treatment received for primary cancer. A range of different types of SMN are diagnosed and these vary across the life course. Studies based on the TYA age range are more limited. Prognosis following SMN diagnosis is an important area for further research.

2.14 Cardiovascular late effects

Improvements in survival for childhood cancer are due to advancements in treatments such as chemotherapy and radiotherapy, however, these treatments are cardiotoxic and can cause persistent and progressive damage to the cardiovascular system [186, 187]. As childhood cancer survivors age they may experience impaired myocardial growth as a consequence of earlier cardiotoxic treatment [186]. In particular, exposure to anthracyclines and radiation to the chest have been shown to increase the risk of cardiovascular late effects [186-188], with a dose response relationship [189, 190]. It is important to monitor patients to identify early signs of cardiac disease before it progresses to clinical presentation as if left undetected and untreated it may lead to heart failure and death [186-188].

Cardiovascular disease is one of the leading causes of non-cancer related late mortality in childhood cancers survivors [79, 81]. The standardised mortality ratio (SMR) for cardiovascular disease is between 3 and 7 times higher for childhood cancer survivors compared to the general population [31, 32, 81, 189, 191] and 13% of all excess deaths 45 years after diagnosis are attributable to

cardiac causes [32]. SMRs are highest for survivors diagnosed with Hodgkin lymphoma and renal tumours [31]. In the CCSS deaths from cardiac related events have decreased over time [79]. However, findings based on the BCCSS showed a quadratic relationship with treatment area with the greatest risk of cardiac mortality observed for those treated in the 1980s, and a subsequent decline for those treated more recently [191]. The cumulative incidence for cardiac mortality in long-term childhood cancer survivors was 2% 35-years after diagnosis [189] increasing to 5% at 60-years post diagnosis [191]. The TYACSS study reported an overall SMR for cardiac mortality of 1.4, which was highest for those aged 15-19 year at diagnosis (SMR=4.2) and decreased with increasing age to and SMR of 1.2 for those aged 35-39 years at diagnosis [192].

The late effects of cardiovascular diseases are one of the most studied late effects in childhood cancer survivors. A systematic review published in 2017 identified 64 papers which included all cardiovascular clinical and subclinical outcomes occurring at least 1-year post diagnosis [193]. The authors found the definitions of outcomes used varied by study therefore the estimated prevalence of the study endpoints varied considerably. A range of disease groups were included, the most frequent end point was heart failure, included in 41 studies with the prevalence ranging from 0.1% to 54%. There was large variation in the size and design of the included studies as well as the differences in the definition of the endpoints and methods of reporting and measurement of the outcomes, for example some studies relied upon self-report whereas other included clinical measurements. A meta-analysis was conducted for hypertension and stroke only and pooled estimates of the mean prevalence were 20% for hypertension and 2% for stroke, however there was considerable statistical heterogeneity for the hypertension estimate. Pooled estimates for other endpoints could not be calculated due to lack of consistency in the definitions of outcomes [193].

Significant risk factors for cardiovascular late effects are female sex, and younger age at diagnosis and longer duration of follow-up [186, 187, 190]. An increased risk of cardiac events are found for most diagnostic groups, with particularly high risks for survivors of Hodgkin lymphoma [190]. Many long-term survivors will also develop modifiable risk factors related to cardiovascular disease (such as hypertension or obesity) which may further compound their risk of a cardiac event in later life [186, 187]. For example, survivors who received chest radiation or anthracycline chemotherapy in combination with

hypertension had a significantly increased risk of coronary artery disease, heart failure and valvular disease beyond that expected under an additive assumption [194].

Clinical outcomes have been identified by various different approaches each with their own strengths and limitations. Studies based on the CCSS [190, 194] use patient self-report which may suffer from recall bias, however this study has very detailed treatment information. Compared to sibling controls cancer survivors were more likely to report heart failure, myocardial infarction, pericardial disease and valvular disease with a significant increased risk for those exposed to anthracyclines and cardiac radiation [190].The cumulative incidence continued to increase throughout follow-up; by age 45 the cumulative incidence of coronary artery disease was 5.3%, heart failure was 4.8%, valvular disease was 1.5% and arrhythmia was 1.3% [194]. By age 50 years, the cumulative incidence of ischemic heart disease was 7.7% and for stroke was 6.3% [195].

Other studies have used linkage of cancer registrations to hospital admissions [196-198], which provide an objective measure of cardiovascular disease morbidity, however less severe cases of cardiovascular complications which may be treated in primary care will not be included. These studies had varying amounts of information on treatment available. Admission rates for cardiovascular disease in long-term survivors of childhood and young adult cancers are estimated to be between 2-5 times higher than in general population controls [196, 197, 199-203]. Survivors of Hodgkin lymphoma had the highest hospitalisation rates for cardiovascular disease in the CCSS and in Scandinavia [197, 204]. Studies examining specific cardiovascular diseases showed an increased risk of hospitalisations for several cardiovascular disorders including: hypertension, ischemic heart disease, pulmonary heart disease, heart failure, valvular disease, conduction disorders, pericardial disease and cerebrovascular disease [196, 197, 199, 203]. Based on clinical follow-up of 10-year survivors, the St Jude Lifetime Cohort Study reported a cumulative incidence of severe chronic cardiovascular events of 17% by age 40 increasing to 35% by age 50 [40]. Further examination of this cohort restricted to survivors of Hodgkin lymphoma only, reported a cumulative incidence of 46% by age 50 of a severe chronic cardiovascular event with myocardial infarction and structural heart defects making the largest contribution to this [205].

Exposure to cardiac radiation was associated with an increased risk but no association with anthracycline dose was found [205] .

In Yorkshire cardiac late effects were examined in 0-29 year olds using linkage of YSRCCYP data to Hospital Episode Statistics, including over 3000 long-term survivors with up to 20 years follow-up [196]. Overall 4% survivors had at least one hospitalisation for a cardiovascular event and compared to the general population, hospitalisations were increased for childhood survivors but not for TYA survivors. However, there were increased rates for TYA for specific outcomes including: pericardial disease, cardiomyopathy and heart failure, pulmonary heart disease and hypertension. This study included patients diagnosed up to 2006 and included hospital admissions up to 2011, further information is now available for patients diagnosed since 2007 and hospital admission data are now available up to 2017. As the risk of cardiovascular events continues to increase across the life course it is important to continue to monitor and report these adverse outcomes to understand more about the burden of late effects in childhood and young adult cancer survivors. Eight European countries, including the UK, are contributing data to establish a large cohort to study symptomatic cardiac late events in 5-year childhood cancer survivors. Ascertainment of outcomes will be obtained through linkage to hospital data and medical records, patient and GP questionnaires and visits to follow-up clinics. The cohort will include approximately 60,000 cancer survivors and this large sample size will allow detailed examination of outcomes in relation to several risk factors [206].

To summarise, the risk of cardiovascular late effects increases throughout follow-up. Age at primary diagnosis, type of cancer, sex, anthracycline use and chest radiation are all risk factors for cardiovascular late effects. Hospitalisations for cardiovascular conditions have been examined in several studies including one study based on the YSRCCYP.

2.15 Respiratory late effects

Primary lung tumours are very rare in children, however, the lung is a common site for metastases, even several years after treatment [207]. The lung is one of the most radiation sensitive structures in the body and radiotherapy and chemotherapy used to treat childhood cancer can cause permanent lung damage. In addition lung function diminishes over time as a function of normal

aging and the effects of early lung damage for cancer treatments may compound these developments. Individual and combined exposure to chest radiation, certain chemotherapy drugs, hematopoietic stem cell transplant and thoracic surgery increase the risk of respiratory late effects among childhood cancer survivors [207]. Specific chemotherapy agents identified as increasing the risk of respiratory late effects are: bleomycin, alkylating agents such as busulfan and cyclophosphamide, and nitrosureas such as carmustine (BCNU), lomusine (CCNU) [207, 208]. Respiratory conditions are a common cause of morbidity in long-term survivors and can have a major impact on quality of life [208]. Respiratory complications range in severity from subclinical to severe and life threatening complications. Children with cancer may have other risk factors that predispose them to long-term respiratory problems including: genetic susceptibility to chemotherapy or radiotherapy, underlying asthma or chronic obstructive lung disease, infection, cigarette use and exposure to environmental respiratory toxins [207].

Late mortality from respiratory disease is between 3 and 9 times higher for childhood cancer survivors compared to the general population [31, 32, 81, 83, 157, 159, 161]. However, a decrease in the risk of death from respiratory conditions has been observed for children diagnosed more recently as cancer treatments have been modified to reduce long-term side effects [79, 83]. A UK study investigating respiratory mortality in survivors of cancer diagnosed before age 40 years reported SMRs of 6.8 for children and 1.7 for TYA (aged 15-39 years) with pneumonia being the most common cause of respiratory death in both age groups [83]. The most recent figures from the BCCSS estimated an absolute excess risk of 2.3 deaths per 10,000 person years from respiratory diseases [81]. In the CCSS survivors of AML and neuroblastoma had the highest risk of respiratory deaths [159].

In studies of hospital admissions for long-term survivors of CYA cancer, there was an increased risk of admissions for respiratory conditions in survivors compared to the general population with rates in cancer survivors 2-6 times higher than in general population controls [199-204]. These studies of hospital admissions were population-based and used an objective measure of hospital admission as a proxy indicator for morbidity, however only one study included admissions for specific respiratory conditions [203]. Compared to the general population childhood cancer survivor had an increased risk of admissions for pneumonia, acute upper respiratory infections, bronchitis and emphysema and

respiratory failure [203]. In Scotland the admissions for respiratory diseases were associated with area level deprivation with the cumulative incidence significantly higher among the most deprived compared to the least deprived which may reflect differences in smoking prevalence by deprivation [199]. Many respiratory conditions may be diagnosed and treated by GPs (primary care) and patients may not present at hospital with these conditions therefore hospital admissions may only capture the more extreme morbidity within these patients. Results from the CCSS based on self-reported hospital admissions showed that admissions for respiratory conditions were higher in all diagnostic groups compared to the general population and were highest for survivors of Hodgkin lymphoma with a SIR of 4.5 (95%3.8, 5.4) [204].

Three studies based on the CCSS have investigated self-reported respiratory outcomes, two studies included patients diagnosed with all cancers [209, 210] while one was restricted to patients diagnosed with CNS tumours only [211]. The first study by Mertens et al in 2002 [209] included 15 different respiratory outcomes. Analysis was based on three defined time periods (diagnosis to end of treatment, end of treatment to 5 years post diagnosis and >5 years post diagnosis) and the patients report of the earliest age at onset of disease. Therefore if a patient had a condition that first occurred during treatment but was still present 5-years post diagnosis this would not be included in the later time period. The authors found that survivors had an increased risk of lung fibrosis, recurrent pneumonia, chronic cough, pleurisy, use of supplemental oxygen, abnormal chest wall, exercise induced shortness of breath, bronchitis, recurrent sinuthres infection and tonsillitis during follow-up compared to matched sibling controls [209]. Treatment factors were found to be significantly associated with respiratory conditions 5-years post diagnosis, particularly chest radiation; the cumulative incidence of lung fibrosis for those treated with chest radiation was 3.5% at 20 years post diagnosis [209]. Sex and age at diagnosis were also associated with worse outcomes; females had increased risk of chronic cough, shortness of breath and exercise induced shortness of breath, those aged less than 5 years at diagnosis had increased risk of exercise induced shortness of breath, while those aged ≥15 years at diagnosis had an increased risk of pleurisy [209]. A more recent study of the CCSS with a longer follow-up time period was published in 2016 and included 6 respiratory conditions [210]. By age 45, cumulative incidence of any respiratory condition was 29.6% (95% CI 29.1% to 30.0%) for childhood cancer survivors compared to 26.5% (95% CI 24.9 to 28.0%) for sibling controls [210]. Survivors had increased risks of chronic cough, oxygen need, lung fibrosis and recurrent

pneumonia, again these were strongly associated with treatment modality [210]. Huang et al included 7 respiratory outcomes in CNS childhood cancer survivors and found there was an increased risk of lung fibrosis, chest wall abnormality, chronic cough and supplemental oxygen use in survivors compared to matched siblings [211]. Craniospinal radiation was associated with a 10 fold increased risk of chest wall deformity (RR = 10.4, 95% CI 7.6 to 14.4) [211], but not with increased risk of other respiratory outcomes reported. Based on clinical follow-up of 10-year survivors, the St Jude Lifetime Cohort Study reported a cumulative incidence of severe chronic respiratory conditions of 13% by age 40 increasing to 22% by age 50 [40]. Within this cohort those with impaired lung function had reduced physical function [212]. A study of long-term childhood cancer survivors in Switzerland based on patient reported outcomes with sibling comparisons reported a cumulative incidence of any respiratory disease 35-years post diagnosis of 21% [213]. Cumulative incidence of pneumonia was highest for those treated with both pulmonary toxic chemotherapy and chest radiation [213].

Other studies based on long-term cancer survivors include results from pulmonary function test (PFT) mainly carried out when survivors attend long-term follow-up clinics and limited to single centre studies which are not population-based with smaller sample sizes [163, 214-217], including several studies identified in a systematic review published in 2011 [207]. PFT provide objective measures of lung diseases including obstructive lung disease, restrictive lung disease, diffusion capacity and overall respiratory dysfunction, however the definitions of these can vary between studies making direct comparisons difficult. These adverse respiratory outcomes may also include undiagnosed conditions in survivors who are not presenting with symptoms. The prevalence of respiratory abnormalities based on PFT results ranged from 44% to 65%, however the percentage of patients with symptoms was estimated at between 19% and 29% [163, 214-217]. At age 50 the cumulative prevalence of abnormal respiratory function was 81% [163].

Several smaller studies have examined the association between treatment and respiratory abnormalities. Mulder et al included 193 PFTs in patients who received pulmotoxic therapy for a primary cancer in childhood and found that chest radiation particularly in combination with bleomycin or lung surgery was the most important risk factor for impaired pulmonary function up to 18 years post diagnosis [215]. In a study of 121 childhood cancer survivors who

underwent a baseline PFT a median of 12 years from diagnosis and a further follow-up 5 years later, found that those treated with chest radiation had a significant decline in lung diffusion function over time [214]. A further study of 143 childhood cancer survivors found that patients who underwent chest radiation or surgery were more likely to have a respiratory abnormality when assessed at long-term follow-up clinics [217].

Lung cancer is rarely reported as a second malignancy after childhood cancer [208], for example, amongst 14,000 survivors in the CCSS only 11 cases of subsequent lung cancer were observed with a median time to diagnosis of 20 years [167].In the BCCSS, there were 36 respiratory SMNs reported in 17981 survivors, with respiratory SMNs accounting for 9% of the total absolute excess risk in survivors aged over 40 years [35]. However, subsequent lung cancers were more frequently diagnosed in long-term survivors in the TYACSS, where it was observed that lung cancers accounted for a substantial proportion of the excess number of subsequent tumours across a range of AYA cancers 35 years post diagnosis including breast (2.9%), cervical (3.6%), testicular (2.7%) and Hodgkin lymphoma (5.1% in males and 3.8% in females) [82]. Given the rarity of subsequent lung tumours in childhood cancer survivors, treatment associated risk factors have not been identified. However with longer follow-up of childhood cancer survivors it is likely that subsequent lung cancers will increase and studies investigating the role of treatment exposures can be conducted [208].

In summary, many studies have shown that treatment is a major risk factor for respiratory late effects as are sex and age at diagnosis. Studies looking at specific causes of respiratory conditions were either based on self-reported outcomes or abnormalities based upon lung function tests which may also detect asymptomatic disease. Analyses of hospital admissions have only considered the broad category of all respiratory conditions, only one study has examined specific respiratory conditions. No studies have examined hospital admissions by treatment modality.

2.16 Key gaps in the knowledge

Cancer incidence and survival trends in children and young adults have been studied extensively, however it is important to continue to monitor these trends in population-based studies to assess the impact of changes to treatment and health policies and variation in outcomes by patient characteristics.

Furthermore, many studies have limited clinical prognostic information such as stage at presentation and treatment. The research presented in this thesis aims to explore alternative statistical methods to gain further insights into survival trends by utilising the statistical cure model. These models are becoming more frequently used and reported for a range of adult cancers, however there are limited applications to population-based studies in children and young people with cancer. Cure models are based on mortality outcomes and while survival rates for children and young people with cancer are high overall there is wide variation by diagnostic group, therefore it is important to assess cure separately by diagnostic group. The application of cure models to data from population-based cancer registries in children and young adults are limited to mainly studies of leukaemia. No studies have considered the application of cure models to a range of diagnostic groups in children and young adults as has been conducted for adult cancers. The use of other prognostic risk factors in the cure models (such as age, sex, stage/grade, treatment for studies based on cancer registry data) are limited. This evidence gap is addressed by Aim 1 of this thesis with the results presented in Chapter 5.

There is a growing body of literature reporting on the long-term late health effects of treatment for childhood and young adult cancers. However, much of late effects literature for TYAs is based upon extrapolation of findings from childhood studies despite these being distinct populations. More studies are needed to explore risks specifically in the TYA population. TYAs are included in analysis in this study to address this knowledge gap.

The prevalence of chronic health conditions in childhood cancers survivors is high and the incidence of chronic conditions increases over time with increasing follow-up and does not appear to plateau. Studies based on large cohorts with sibling controls are limited by self-reported outcomes whereas studies based on clinical assessments are often based on single-institutions and therefore not population-based and also do not compare with the health of the general population. Important prognostic risk factors for adverse health outcomes in later life include treatment factors, diagnostic group, age and sex.

The YSRCCYP linked to hospital admission data provides an important source of data to enable the examination of morbidity requiring hospitalisation for the Yorkshire population, this work has previously been conducted for cardiovascular disease [196] but not for other disease areas. Updated

estimated of cardiovascular admissions are included in Aim 2c of this thesis (Chapter 8). Respiratory disease is one of the most common causes of late mortality and morbidity and is associated with exposure to treatment for childhood cancers including specific chemotherapy drugs and radiation to the chest. Therefore it is important to estimate the prevalence of hospital admissions for respiratory conditions and associated risk factors to identify the groups of survivors at greatest risk. This is addressed by Aim 2a of this thesis with results included in Chapter 6.

Childhood cancer survivors are at increased risk of subsequent malignant neoplasm, however this risk has not been quantified in the Yorkshire region and while many studies have been conducted in children, studies including the TYA age range are more limited. Several studies have examined SMN occurrence but mainly focussed on SMNs after 5-years and few examined the impact of latency on survival. This is addressed by Aim 2b with results presented in Chapter 7.

Finally, many studies focus on one outcome only but it is also important to consider multiple morbidities, as cancer survivors are at increased risk of an array of late effects as a consequence of their earlier treatment with over one third of long-term survivors reporting multiple chronic conditions [37]. By utilising several different sources of routinely collected follow-up data this project aimed to evaluate the cumulative burden of morbidity from cardiovascular and respiratory diseases and subsequent tumours in long-term survivors (Aim 3, Chapter 8). Details of the specific data sets and methodology used in this thesis are described in Chapter 3 and Chapter 4 includes a detailed descriptive analysis of the study population.

Chapter 3 Data and methods

This chapter outlines the datasets included in the analysis presented in this thesis and details the statistical methods applied to address the thesis aims and objectives listed in Section 1.3. The project is based upon data from the Yorkshire Specialist Register of Cancer in Children and Young People (YSRCCYP), which is described in detail at the start of this chapter. Data linkage to other datasets, including Hospital Episode Statistics (HES), National Cancer Registration and Analysis Service (NCRAS) and cytogenetic data, were performed to obtain additional outcomes and other clinical data. For each additional data source a description is included detailing the key variables and all outcomes included in analysis. This study utilised individual level identifiable and sensitive data and the ethical implications and details of ethical approvals are included. Finally the statistical methods used in analysis are described.

3.1 Data sources

The main source of data used in this study was the YSRCCYP. Further linkage to data from other sources was carried out to produce the results presented in this thesis. All sources of data are detailed below including details of the key variables extracted from each source.

3.1.1 Yorkshire Specialist Register of Cancer in Children and Young People

The YSRCCYP is a population-based cancer register which records all cancers diagnosed in children and young people aged 0-29 years resident in Yorkshire. The register was established in 1974, originally only recording cancers in children aged 0-14 years, but from 1990 onwards included teenagers and young adults aged 15-29 years. When the register began, the NHS in England was administered by regional Health Authorities so the region covered by the register was that of the Yorkshire Regional Health authority which included West Yorkshire, North Yorkshire and Humberside, but excluded South Yorkshire. From 1998-2009 cancers diagnosed in 0-29 year olds from South

Yorkshire were included in the register to ascertain complete coverage of the Yorkshire and Humber region for this time period.

The total population of the Yorkshire and Humber region from the 2011 census was 5.3 million, which included 2 million individuals aged 0-29 years. The region is broadly representative of the age and sex structure of the rest of England [218, 219]. The region covers both urban and rural communities with a significant ethnic minority population in parts of West Yorkshire, mainly those of South Asian origin and in particular of Pakistani origin; 4% of the Yorkshire and Humber population was Pakistani compared with 2% in England and Wales [219]. The Yorkshire and Humber region has a different deprivation profile compared to England with twice as many areas of high deprivation (based on the Index of Multiple Deprivation) [220].

The register receives notifications of cancers diagnosed from several sources including directly from the Principal Treatment Centres (PTC) in Leeds and Sheffield, which treat the majority of children and teenagers with cancer in the region. Secondary sources of ascertainment include neuropathology reports, hospital admissions and other regional and national cancer registries. The register regularly receives extracts from Public Health England on all patients diagnosed in Yorkshire aged 0-29 years from the National Cancer Registration and Analysis Service (NCRAS). This is the main source of notifications for patients aged 25-29 years. Notifications of cancer include basic personal demographic information, date of diagnosis and details of morphology and topography. Validation checks are conducted on NHS number, postcode, age at diagnosis, date of birth, date of death, morphology and topography codes when these data are entered into the database. Further fields that must be completed during the registration process are date of diagnosis and gender. Validation reports are conducted quarterly to monitor data quality and completeness of these data fields.

A dedicated data collection manager then extracts further information from the patient medical records from the relevant hospitals in the region, including details on clinical factors including stage at diagnosis and treatment. Notifications of relapse are also received directly from the PTCs and this information, including the date of relapse, is recorded in the database. Each patient is followed up every two years via hospital consultants and GP to ascertain their vital status and any further information on relapses. Further

information on date and cause of death are obtained from NCRAS. All data received from NCRAS is cross-checked against the data held on YSRCCYP and additional registrations added. Completeness of registrations each year are monitored by examination of annual incidence trends for each diagnostic group.

When treatment data is received from NCRAS, this is cross-checked against the register to check, for example, surgery and chemotherapy start dates are consistent. Any further anomalies with data that are identified, for example while undertaking specific data analysis, are flagged with the data collection manager and further verification from patient notes may be undertaken, and data updated on the database.

This study included all patients aged 0-29 years diagnosed between 1990 and 2011 resident in the Former Yorkshire Health Authority region only. Table 3.1 lists the local authority areas included in the NHS Yorkshire and Humber Region and those included in the Former Yorkshire Health Authority (when the register was established). These local authority areas are used to derive population statistics for the region covered by the registry. Patients resident in South Yorkshire were not included in this study these patients were only included in the YSRCCYP database from 1998-2009 and not the full time period of interest.

Table 3.1: Local Authority names and codes for Yorkshire.

Area code	Area name	Included in Former Yorkshire Health Authority
E06000010	Kingston upon Hull, City of	Yes
E06000011	East Riding of Yorkshire	Yes
E06000012	North East Lincolnshire	Yes
E06000013	North Lincolnshire	Yes
E06000014	York	Yes
E07000163	Craven	Yes
E07000164	Hambleton	Yes
E07000165	Harrogate	Yes
E07000166	Richmondshire	Yes
E07000167	Ryedale	Yes
E07000168	Scarborough	Yes
E07000169	Selby	Yes
E08000016	Barnsley	No
E08000017	Doncaster	No
E08000018	Rotherham	No
E08000019	Sheffield	No
E08000032	Bradford	Yes
E08000033	Calderdale	Yes
E08000034	Kirklees	Yes
E08000035	Leeds	Yes
E08000036	Wakefield	Yes

Source: [221]

The aims of this study were to examine long-term outcomes; therefore patients were included from 1990 onwards for several reasons. The latest follow-up information on death at the time analysis started was to the end of 2016; therefore all patients were followed up for at least 5 years with the potential to have up to 26 years follow-up for those diagnosed in 1990. To ensure consistency over time in the age range of patients only those diagnosed from 1990 onwards were included, as prior to this diagnoses in the 15-29 year age group were not registered. Furthermore, prior to 1990 NHS number was not as well recorded and this is a key personal identifier used for linkage to other data

sources (see next sections) so to ensure the maximum potential to link to other sources this cut off was also chosen.

Diagnoses are defined based on ICD-O3 topography and morphology codes. These are then mapped to the ICCC-3 codes [43] which were used to define the diagnostic groups used in this study. The ICD-O3 codes are also mapped to the Birch classification [8], which is commonly used for TYA cancer diagnosis. ICCC-3 and Birch are generally similar, however the Birch classification system separates out the carcinomas into separate groups as these are more common in TYA compared to children, where in the ICCC-3 classification they are all grouped together. (See 3.3.1 for more details on diagnostic groups used in analysis).

Other data items extracted from the register included: patient's date of birth, date of diagnosis, sex, stage or grade of cancer at diagnosis, treatment received, which included binary indicators of receiving surgery, chemotherapy or radiotherapy as well as further details of the treatment including surgery codes (OPCS-4), dates of surgery and outcome of surgery, information on chemotherapy regimens and drugs and radiotherapy site. Postcode at diagnosis was extracted and used as a personal identifier (along with others) for linkage to other data sources and postcode was also mapped to area level deprivation indices (see Section 3.3.4).

3.1.2 Hospital Episode Statistics

The Hospital Episode Statistics (HES) database includes records of all NHS hospital activity for individuals resident in England and is managed by NHS Digital (formally known and the Health and Social Care Information Centre) [222]. It includes several different datasets including admitted patient care (APC, inpatient), outpatient care (OP) and accident and emergency care (A&E) [223]. HES data are not collected for research purposes but are collected to allow each hospital to be reimbursed for the care they deliver, so called 'payment by results'. In order to be paid correctly hospital providers need to accurately record all procedures carried out and clinically code all details of the patient's conditions and treatments. HES data are collated for each financial year and stored as a series of separate records for each period of care rather than at the patient level as patients may have more than one admission each year. For each patient a unique identifier, called the HESID, is created by NHS

Digital, which uniquely identifies each patient and provides a method to track them in HES data. An alternative method would be to use the patients unique NHS number, however NHS number was not recorded in all records, particularly for historic records, and also the NHS number is an identifiable field [224]. HES is commonly used for secondary data analysis and research [225]. Furthermore, by linking to a cohort of cancer patients, provides a rich and powerful source of data to examine hospitalisations in long-term cancer survivors.

There have been issues with the quality of data recorded in HES although the quality of data is improving year on year. Inpatient data have been recorded since 1989, however prior to 1997 the data quality was not regarded as of sufficient quality for research purposes: the HESID was introduced in 1996 to allow the tracking of individuals across multiple admissions [224]. Outpatient data are only available from 2003/04 onwards and A&E data from 2007/08 onwards, there are also concerns regarding the data quality of these datasets, for example examination of the outpatient HES linked to the YSRCCYP from 2003-2016 identified that 99.4% of records had the main diagnostic code recorded as “unknown and unspecified cause of morbidity”. Therefore in order to estimate the prevalence of late effects in a cohort of cancer patients, only the inpatient (APC) data from 1997 onwards was used in this study.

Inpatient HES are recorded as a series of Finished Consultant Episodes (FCE) which represent a period of care (known as an episode) under a particular consultant speciality at a single hospital provider. A patient’s whole stay in hospital is known as a spell and a spell may contain more than one episode if a patient is treated under more than one consultant during their admission. There are two main types of spells: provider spells and continuous inpatient (CIP) spells. Provider spells are the time a patient stays in one hospital and the provider spell ends when the patient is transferred to another hospital, dies or is discharged. A CIP is a continuous period of care within the NHS, regardless of any transfers between hospital providers and the spell ends when the patient dies or is discharged from hospital [226]. The hierarchical structure of spells and episodes is shown in Figure 3.1.

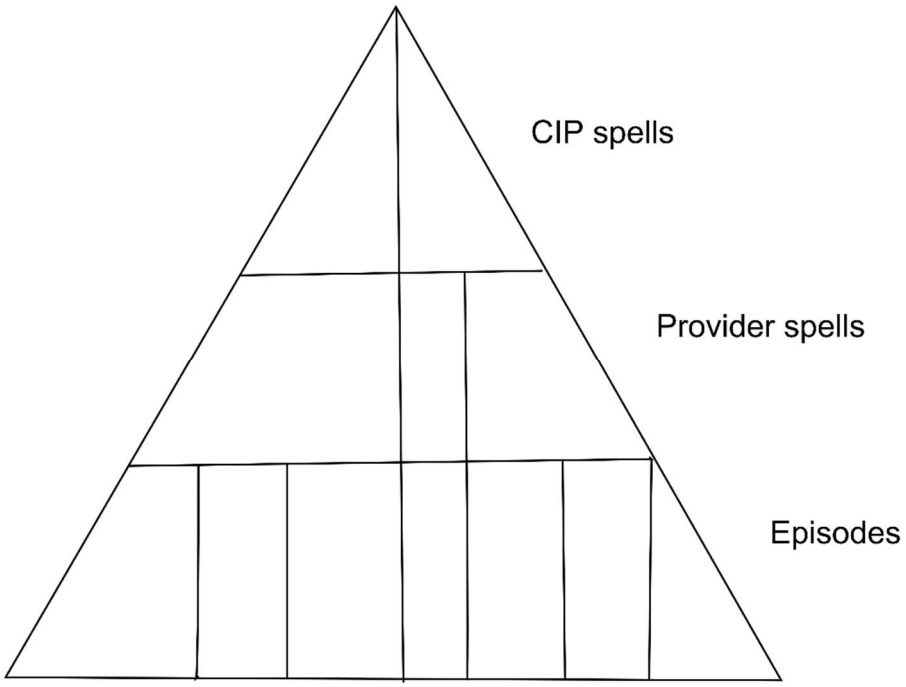


Figure 3.1: Hierarchical structure of CIP spells, Provider spells and Episodes in HES.

Source: [226]

Identifiers from patients from the YSRCCYP were sent to NHS Digital for linkage to HES on NHS number, date of birth, sex and postcode [222]. Separate data extracts were sent on 4 different occasions, the first extract was received in 2012 and included patients diagnosed between 1974 (the start of the register) and 02/02/2012 who were alive on 01/04/1996 and HES records from April 1996 to March 2011. Since then three further extracts have been received to allow further follow-up of the cohort and HES data to be obtained for patients recorded on the register since 2012 with follow-up including admissions up to 31st March 2017. In this study patients diagnosed between 1990 and 2011 were included with corresponding hospital admission data available from April 1997 to March 2017.

3.1.2.1 Diagnostic codes and operation codes

HES uses the WHO International Classification of Diseases 10 (ICD-10) coding system to record all diagnostic information. In the APC episodes this is recorded as a primary diagnosis, the main reason the patient was admitted, and up to 19 other diagnostic codes can be recorded, these can be used to record other important co-morbidities related to the admission. All procedures or

interventions are recorded using the Office of Population, Censuses and Surveys: Classification of Interventions and Procedures, 4th Revision (OPCS-4) and up to 20 of these can be recorded alongside the date of the procedure.

The analysis in this thesis considers all diagnostic codes listed in any of the 20 fields to ensure the whole admission patterns of the patient were included. Information on cardiovascular and respiratory admissions were extracted from these data (see Section 3.4.5). Within HES data the outcome of interest was the date of first admission and also any subsequent admissions. CIPs were created for each patient within the cohort using the methodology described by NHS Digital [226], to ensure that all episodes within the same admission for each individual were captured as one event. For the rest of this thesis a CIP will be described as an admission. The respiratory or cardiovascular diagnosis may occur in any episode of an admission for each individual. Previous research in the Yorkshire register examining cardiovascular outcomes using HES data was also based on CIPs [196].

3.1.2.2 Comparison group

The rate of hospitalisations in the cancer survivor cohort were compared to age and sex matched admissions in the general population of Yorkshire and Humber. These data were requested and received from NHS Digital in addition to the linked data for the cancer patients. This consisted of an extract of all HES episodes in Yorkshire between 1997 and 2017 for those aged <56 years including age, sex, date of admission and all diagnostic codes associated with each episode along with a unique ID (derived from the HESID) to link individuals between episodes over the study period. From this, an age-matched cohort was constructed for each calendar year. For example, the cancer survivor cohort was aged 5 to 34, so for the background Yorkshire population all admissions in those age 5 to 34 years were retained, and in 2016 the cancer survivor cohort was aged 5 to 56 years therefore all admission in this age range were retained. Then for each calendar year the number of admissions for each diagnostic code of interest, the total number of admissions by single year of age and sex were tabulated and used to create admission rates in the general population by dividing the number of admissions by mid-year population estimates obtained from the Office for National Statistics [227]. This then provided an external reference data set to calculate indirect standardised rates.

All diagnostic and procedure fields were included and the same codes considered to identify population-based rates of admissions for respiratory and cardiovascular morbidity (see Section 3.4.5).

3.1.3 National Cancer Registration and Analysis Service

Data on all subsequent malignant neoplasms (SMN) were obtained from the YSRCCYP database and the National Cancer Registration and Analysis Service (NCRAS) [48]. If a patient was diagnosed with a SMN while under the age of 30 and resident in Yorkshire then this is recorded in the YSRCCYP database. However, if the patient is 30 or older when diagnosed or is no longer resident in Yorkshire this information was obtained from NCRAS. An application to obtain these data was approved through Public Health England (PHE) Office for Data Release (ODR1516_163 A1). Patient identifiers including NHS Number, Date of Birth, Postcode and Sex were sent to PHE for linkage to identify SMNs for patients within the Yorkshire register.

Data received back from NCRAS included date of subsequent tumour diagnosis and diagnostic group, coded to ICD-O2 or ICD-O3 (depending on date of diagnosis) for all tumours diagnosed up to 31st December 2015.

3.1.4 Cytogenetic risk group

Analysis presented in Chapter 5 focussed on acute lymphoblastic leukaemia (ALL) patients only. A key prognostic indicator for ALL is cytogenetic risk group. Cytogenetic information is important not only for predicting survival but also to identify patients at increased risk of relapse and those less likely to respond to treatment after relapse [228]. Cancer registries do not routinely collect this information therefore this information was obtained via linkage to the Leukaemia Research Cytogenetics Group database (held and managed at Newcastle University) for a subset of patients recruited into clinical trials. Patients were matched on personal identifiers including NHS number, patient names, date of birth and sex. Cytogenetic risk group was coded as good, intermediate or poor for B-cell precursor ALL and all T-cell precursor ALL were included in one group [229]. Some patients after linkage, were unable to be assigned a risk group and

were categorised as “Unknown”, with the “not linked” group included as a separate category.

3.2 Ethical approval and data security

The YSRCCYP has received ethical approval from the Northern and Yorkshire Multi Centre Research Ethics Committee (MREC/00/3/001) and approval from the Health Research Authority Confidential Advisory Group (CAG 1-07(b)/2014) which permits the processing of identifiable cancer registration data without the need for informed patient consent. An amendment to ethics was made to allow the collection of long-term follow-up data on SMNs as part of this project. Favourable ethical approval for this amendment was granted in January 2017. Approval for the updated HES data was granted by NHS Digital Data Access Request Service in April 2017 and approval from PHE Office for Data Release in November 2017 for SMN information.

The YSRCCYP database contains personal, identifiable and sensitive data including names, dates of birth, addresses, NHS numbers and detailed clinical information. All data are held within secure networks within the University of Leeds and can only be accessed by authorised members. No information is ever published in which individuals can be identified. To ensure no patient identifiable data are disclosed numbers may be suppressed where there are fewer than 5 cases, other cells in tables may also be suppressed to avoid disclosure by difference.

3.3 Study variables

3.3.1 Diagnostic group

Diagnostic group was based on the ICCC-3 as described in Section 3.1.1 above. As well as all cancers combined analyses were conducted for the main diagnostic groups and some specific subgroups. ICCC-3 was used when considering all ages so direct comparisons between children and TYA for specific diagnostic groups could be made. If only considering TYA then Birch classification was used as this system more appropriate for this age group – in particular the carcinoma groups. Analysis based on Birch only have been identified in the relevant sections of this thesis. Analysis based on cure models

including clinical risk factors was based on those diagnosed with ALL only (ICCC-3 group 1a) (see Section 3.5.2.3).

3.3.2 Age at diagnosis

For most analysis the whole 0-29 year age range was included, mainly categorised into two groups: children (0-14 years) and TYA (15-29 years). However the cure model ALL analysis was based on those aged 1-17 years only (see Section 3.5.2.3 for further justification and details).

3.3.3 Disease severity

Stage or grade of disease was available for some tumours but the levels of completeness varied by diagnostic group [230]. White cell count (WCC) was used a proxy for stage for leukaemia patients; the Ann Arbour staging system was used for lymphoma; the Royal Marsden or TNM stage was used for testicular germ cell tumours and FIGO stage for ovarian germ cell tumours. There was insufficient stage information recorded for bone tumours and soft tissue sarcomas. CNS tumours were categorised according to WHO grade. Stage or grade of disease was included in descriptive analysis including the reporting of levels of missing data. Methods to deal with missing data, such as multiple imputation [231, 232] were not considered as this was out with the scope of this project but it is an important analytical issue to address in future studies (see Chapter 9 Discussion). Section 3.5.3 describes the methods used to select potential confounders for adjustment in statistical models including identifying a minimum adjustment variable set, where possible variables other than stage were used in adjusted models to avoid excluding cases with incomplete data.

3.3.4 Deprivation

Postcode at diagnosis was linked to 2001 Census wards to obtain the Townsend Deprivation score [233]. Townsend deprivation scores are area based measures derived from Census data on the percentage of homes that are not owner-occupied, the percentage of economically active residents who are unemployed, the percentage of households that do not have access to a car and the percentage of households with more than one person per room. Each of these variables are standardised using z-scores and then a composite score

is created to give an overall score for each area [234]. The areas used throughout this thesis were wards and the scores for each ward in England were then ranked and split into fifths. The scores and deprivation fifths were then assigned to each individual within the study population by mapping their postcode at diagnosis to a 2001 electoral ward. For patients in the register no individual level social class variables are available therefore an area based measure of deprivation had to be used, although this may lead to problems with ecological fallacy where area-level associations may not infer associations at the individual level [235]. Other area level measures of deprivation are available, such as Carstairs and the Index of Multiple Deprivation, however to ensure comparability with other studies based on the Yorkshire register, Townsend was used to this has been used previously within the register.

3.3.5 Ethnicity

Ethnicity was obtained from a combination of ethnicity recorded in HES and the naming algorithm program Onomap [236]. Since 1995 in England it has been mandatory to collect self-reported ethnic group data in HES data [237]. Ethnicity recorded in HES is based on ethnic groups used in the Census and different ethnic groups were recorded in HES from 2001 onwards to reflect changes to the Census ethnic group categories [223]. Within HES data ethnicity is recorded for each episode, therefore each patient may have more than one episode and potentially multiple ethnic codes may be recorded. For patients with multiple admissions and more than one ethnic groups recorded the most common ethnic group was assigned to each individual, as recommended by several authors [238-240]. Due to small numbers in the Black and Chinese groups, ethnicity was categorised as White, South Asian or Other.

Onomap is a name recognition program which was developed based on surnames and forenames from public name registers from over 26 countries and classifies individual names into cultural ethnic or linguistic groups [241, 242]. It includes all ethnic minority groups in the UK and unlike other naming algorithms such as Nam Pehchan [243] and SANGRA [244], is not limited to the South Asian group only. The Onomap taxonomy classifies names into one of 185 different types, which are nested within 66 subgroups, which are then nested within 16 larger groups. The surname and forename recorded in the YSRCCYP database for each person was matched to an Onomap type which is the lowest level in the Onomap classification system. As above three ethnic

groups were included: White, South Asian or Other to be consistent with the ethnicity groups from HES.

A combined HES and Onomap ethnic group was assigned based on the previous reported methodology [236]. (Lesley Smith was lead author of this publication and contributed to the study design, conducted the statistical analysis and drafted the manuscript). If both sources agreed then this group was assigned to each patient. If either source was missing but the other was not then the non-missing group was used. A further Onomap type is the Muslim group which is defined as having origins in the Middle East and this group cannot be directly mapped to the South Asian group [245]. Where Onomap assigned Muslim and HES was recorded as South Asian, then the individual was assigned to the South Asian group. There were still patients with discrepancies between the two classifications therefore two further ethnic group indicators were created: one where these patients were assigned the HES ethnic group and another where they were assigned the Onomap ethnic group. Analysis of incidence trends within this population found that the two combined indicators showed a similar trend in results [236], therefore the combined indicator which prioritises HES was chosen to categorise patients into one of three ethnic groups; White, South Asian or Other. Despite using these two sources ethnicity information was still missing for some individuals. The Onomap program licence expired in 2016, with no further updates available, and at this stage an ethnic group had not been assigned using this program to a small number of individuals included in the study population.

3.3.6 Treatment

Treatment data were extracted from the register. Three separate binary indicators for receiving surgery, radiotherapy or chemotherapy were created. From these a combined treatment modality indicator was created within each diagnostic group where patients with no treatment recorded were included as a separate category.

Further details of surgery were extracted based on OPCS-4 codes. Details on chemotherapy drugs were extracted. Details on radiotherapy site were also obtained from the register database. For analysis on respiratory late effects (Section 3.5.5) three treatment exposures were of interest: 1) patients who had received any chemotherapy drugs with known lung toxicity which included

bleomycin, busulphan, carmustine, cyclophosphamide and lomustine, 2) those who had received radiotherapy to the chest which included radiotherapy to the lungs, heart and mediastinum as well as total body irradiation and 3) those who had received thoracic surgery including operations of the chest wall, lobectomy and other operations on the lung. For the analysis involving the cumulative burden of admissions for respiratory and cardiovascular disease and SMNs (Section 3.5.8) any patients who had received anthracyclines were also identified. Accurate dose information for radiotherapy or chemotherapy was not available, therefore these treatment exposures were included as binary variables only.

3.4 Outcomes

3.4.1 Overall survival

Overall survival was defined from date of diagnosis to date of death or censoring. All patients were followed-up to 31st December 2016 to ensure at least 5 years follow-up for each patient.

3.4.2 Relative survival

Relative survival is described in detail in the background chapter (Section 2.7.5). Briefly, it is defined as the observed survival divided by the expected survival where the expected survival is obtained from national life tables stratified by age, sex and calendar year. Mortality rates in the general population were obtained by age, sex and year from national lifetables for England published by the Office for National Statistics [104].

3.4.3 Event-free survival

In addition to overall survival there is interest in event-free survival (EFS) which incorporates relapse, in addition to death and censoring. This outcome is commonly reported from clinical trial data but population-based information on relapse are limited. Relapse information is collected with the YSRCCYP. Relapse was defined as recurrent disease either occurring locally at the same site as the initial diagnosis and/or elsewhere [246]. The exact date of relapse was extracted for analysis. EFS was defined from date of diagnosis to date of

relapse or date of death, whichever occurred first. EFS was examined on a subset of children with ALL and the results are included in Chapter 5.

3.4.4 Subsequent malignant neoplasms

Ascertainment of subsequent malignant neoplasms (SMN) were obtained from the YSRCCYP and linkage to the National Cancer Registration and Analysis Service (NCRAS) (Section 3.1.3). All tumours were registered following the coding of multiple primary cancer rules recommended by the International Association of Cancer Registries (IACR) and the International Agency for Research on Cancer (IARC) [165]. These rules state that a primary cancer is one that originates in a primary site of tissue and is not an extension, a recurrence or a metastases. All subsequent tumours, regardless of time since diagnosis of first tumour, were included.

A SMN was defined as a malignant neoplasm of any site with a different morphology from that of the primary tumour. Subsequent diagnoses were coded according the ICD-O3, tumours with a behaviour code of 0 (benign) and 1 (in-situ) were excluded with a few exceptions (Table 3.2). CNS tumours with benign or in-situ behaviour that are included in the ICCC-3 were included (for example pilocytic astrocytomas and meningiomas) [43]. Unlike tumours located elsewhere in the body, benign CNS tumours in children present with similar clinical symptoms, prognosis and late effects as malignant tumours including serious neurologic morbidity. All bladder cancers were included regardless of behaviour type as there are known difficulties classifying the malignant potential of tumours of benign, in-situ or uncertain behaviour [35, 247]. Non-melanoma skin cancers (ICD-O3 codes C44) were excluded as there is known variation in registration of these tumours in England and not all tumours are reported resulting in an underestimation in the general population [248].

Table 3.2: Non-malignant morphology codes included in analysis of subsequent tumours

Topography	Morphology	Description
C70.0	M9530/0	Meningioma, not otherwise specified (NOS)
C70.9	M9530/0	Meningioma, NOS
C71.2	M9530/0	Meningioma, NOS
C71.9	M9413/0	Dysembryoplastic neuroepithelial tumor
C71.9	M9421/1	Pilocytic astrocytoma
D32.0	M9530/0	Meningioma, NOS
D32.9	M9530/0	Meningioma, NOS
D32.9	M9532/0	Fibrous meningioma
D42.0	M9530/1	Meningioma, NOS
D43.2	M9505/1	Ganglioglioma, NOS
D43.4	M9505/1	Ganglioglioma, NOS
D46.2	M9983/1	Refractory anemia with excess blasts

The ICCC-3 or Birch classification system used to categorise tumours in children and TYA is not appropriate to classify SMNs within this cohort as follow-up for SMN includes those diagnosed up to age 53 years therefore ICDO2/03 was used as this is used to classify adult cancers. SMNs were categorised into a broad diagnostic groups based on ICDO2/O3 codes. These included: 1) Leukaemia, 2) Lymphoma, 3) CNS tumours, 4) Digestive, 5) Respiratory, 6) Breast, 7) Testicular, 8) Thyroid, 9) Soft tissue and 10) Other. The specific groups of SMN tumours were selected based on the numbers observed within the cohort and those used in other published studies [35].

3.4.5 Late effects based on hospital admissions

The linked cancer registry-HES data were used to identify late effects based on hospital admissions in 5-year survivors. Only admissions occurring at least 5-years post diagnosis were included. Whilst this is an arbitrary cut point it is the standard convention in studies based on long-term survivors. A five-year period is chosen to ensure sufficient time has passed since the end of treatment (treatment for leukaemia may last up to 3 years) so that any conditions diagnosed may be defined as a late effect of treatment and not a consequence of short term toxicity. Additionally 5-year survival is frequently used as a bench

mark and patients surviving beyond this period as assumed to be cured of their cancer. Therefore there is great interest in understanding and quantifying health issues after this period. By using this cut-off it also allows comparison with other published literature.

A long-term cancer survivor cohort was established from the YSRCCYP database including all children and young people, aged 0-29 years, diagnosed between 1990 and 2011 and surviving at least 5-years from diagnosis. HES admissions occurring between April 1997 and March 2017 were used to identify late effects for respiratory and cardiovascular conditions as described below. HES data were available from 1997 onwards, therefore in this study patients who were diagnosed in 1990-1991 did not start follow-up 5-years from diagnosis but shortly after when admission data were available.

For both respiratory and cardiovascular conditions admissions were based on the primary diagnosis and all secondary diagnostic fields within each HES episode to ensure all conditions were identified. Sensitivity analysis was also carried out based on identifying conditions based on the primary diagnosis field only.

3.4.5.1 Respiratory late effects based on hospital admissions

Respiratory admissions were classified based on ICD-10 codes and included:

1. Any respiratory conditions (J00-J99)
2. Asthma (J45-J46)
3. Pneumonia (J10.0, J11.0, J12-J18)
4. Chronic lower respiratory disease (J40-J44, J47)
5. Lung fibrosis (J84.1)
6. Respiratory conditions due to other external agents (J70) (this includes radiation and drug-induced lung disorders).

To quantify hospitalisations due to respiratory conditions the first admission for each disease type was included. Readmissions following first admissions were also identified. Time to first admission for each condition was calculated starting from 5-years after the diagnosis date.

3.4.5.2 Cardiovascular late effects based on hospital admissions

Cardiovascular admissions were classified based on ICD-10 codes and included:

1. All cardiovascular conditions (I00-I99, G45)
2. Hypertension (I10-I15)
3. Coronary heart disease (I20-I25)
4. Cardiomyopathy and heart failure (I42, I43, I50, I51)
5. Valvular heart disease (I34-I39)
6. Pericardial disease (I30-I32)
7. Conduction disorders (arrhymias) (I44-I49)
8. Cerebrovascular disease (I60-I69, G45)

These are the same cardiovascular included in previous analysis based on the YSRCCYP [196]. To quantify hospitalisations due to cardiovascular conditions the first admission for each disease type was included. Time to first admission for each condition was calculated starting from 5-years after the diagnosis date.

3.4.5.3 Admission in the general population

In order to establish if the admission rates for respiratory and cardiovascular disease were higher in the cancer survivor cohort compared to the general population, individual-level inpatient admission data for the whole Yorkshire and Humber region were obtained matching the cancer survivor cohort in terms of age and sex over the same time period. These data were used to estimate admission rates in the general population for each specified admission type using population denominator data based on single-year of age, sex and calendar year for the Yorkshire and Humber region obtained from the Office for National Statistics.[227]

3.5 Statistical Methods

3.5.1 Descriptive statistics

The number of patients and incidence rates by diagnostic group were calculated, and described for all patients and for two broad age groups: children (0-14 years at diagnosis) and TYA (15-29 years at diagnosis). Age standardised incidence rates (ASR) were calculated using the direct method and the European standard population, to allowed valid comparisons to be made over

time [249]. ASRs and 95% confidence intervals (95% CI) are presented per 1,000,000 persons per year.

Overall survival by diagnostic group and age group, were calculated using the Kaplan-Meier survival estimate. Survival estimates were only calculated for groups including at least 50 patients following the methodology of PHE [250]. Retinoblastoma, hepatic tumours and Other specified and unspecified tumours were excluded from survival analysis due to the small numbers in these groups. Kaplan-Meier survival curves were used to graphically assess trends in survival over time, initially using three time periods 1990-1996, 1997-2003, 2004-2011, however if there was fewer than 50 cases in each time period then two time periods were considered 1990-2000 and 2001-2011. These plots were used to assess if statistical cure was a reasonable assumption. In addition to overall survival plots, relative survival was also estimated using Ederer II method [99]. For children and young adults diagnosed with cancer, deaths due to other causes within this age range are rare, and the estimates of overall survival and relative survival are very similar; however much of the methodology and statistical programs written to model cure were developed in the relative survival framework therefore relative survival is presented also.

3.5.2 Cure models

As described in detail in Chapter 2 different types of cure models are available to model statistical cure including the flexible parametric, the mixture and the non-mixture cure model. As recommended, the flexible parametric cure model is the most suitable when survival is relative high, as is the case with survival from childhood and young adult cancers, and these models perform as well as the mixture and non-mixture models in other situations [28, 115]. The first aim of this thesis was to assess the feasibility of applying cure models to children and young people diagnosed with cancer and this was done in three stages; 1) comparison of different types of cure models, 2) estimation of trends in cure by diagnostic group and 3) examination of association between prognostic risk factors and cure for childhood leukaemia. Each of these stages is described in detail below.

3.5.2.1 Comparison of cure models

For all cancers combined different cure models were run and the results obtained compared between models. The models included were: 1) flexible parametric (FP) cure model, 2) mixture cure model with Weibull distribution and 3) non-mixture cure model with Weibull distribution. The models were run on all cancers combined to ensure the largest sample size as possible rather than stratifying by diagnostic group and reducing statistical power. For each model covariates included were time period, using three periods of diagnosis (1990-1996, 1997-2003, 2004-2011) and age group, using two broad age groups (0-14 years and 15-29 years). The resulting estimates (and 95% CI) obtained for each model were the percentage cured, the median survival time (MST) of the uncured and the time at which 90% of the uncured had died. These metrics are commonly reported and clinically informative in addition to the 5-year survival rate [28, 115, 251].

The FP cure model uses splines to model the underlying survival curve and it has been shown that the FP cure models are fairly robust to the number and position of the knots used as long as the last knot is placed at the last observed death or later and the other knots are distributed along the full follow-up time [115]. FP cure models were run in Stata using the `stpm2` command with the cure option as described by Andersson et al [115]. These models place boundary knots at the minimum and maximum of the uncensored survival times and an additional knot is placed at the 95% centile of the uncensored survival time. The degrees of freedom can be specified to alter the number of internal knot points; the default is 5 degrees of freedom. Sensitivity analysis was conducted to evaluate the optimum number of knots to ensure the best model fit. Models with between 4 and 8 degrees of freedom (between 3 and 7 internal knots) were fitted using the centiles of the uncensored survival times shown in Table 3.3 [115]. Models were compared using Akaike information criterion (AIC) and Bayesian information criterion (BIC) and graphical assessments of the model fit. Models fit statistics such as AIC and BIC may be of limited value when comparing cure models as they measure model fit over the full follow-up time and cure model needs to fit well at end of follow-up where data may be more sparse [28, 115]. However, they are still useful to give some indication of the best model fit but should not be used alone but in combination with other assessments, such as graphical plots of model fit. Diagnosis period and age group were included as time varying coefficients so that the percentage cured and the survival function of the uncured varied by these covariates.

Table 3.3: Flexible Parametric cure model knot positions

Degrees of freedom	Number of internal knots	Centile positions
4	3	33, 67, 95
5	4	25, 50, 75, 95
6	5	20, 40, 60, 80, 95
7	6	17, 33, 50, 67, 83, 95
8	7	14, 29, 43, 57, 71, 86, 95

Source: [115]

Mixture and non-mixture cure models were run in Stata using the `strsmix` and `strsnmix` commands using the Weibull distribution to model the survival of the uncured [114]. The percentage cured, the MST of the uncured and the time which 90% of the uncured had died as well as model fit statistics (AIC and BIC) from these models were estimated and compared with results from the FP cure model.

Cure models provide estimates separately for each combination of covariates in the model, therefore to make comparisons between levels of each covariate while adjusting for the other covariates standardised estimates were calculated [251, 252]. For example, the cure percentage for each age group was estimated assuming that the distribution of the other covariates (in this example time period) were the same as the whole study population. Standardised estimates were calculated for all reported outcomes.

For each model, the model fit was assessed by comparing the survival estimates from the cure model with empirical life table estimates of relative survival. Finally to investigate differences in the survival curves of the uncured the observed Kaplan-Meier survival for patients who died was estimated and plotted alongside the survival curve of the uncured estimated from the cure models.

3.5.2.2 Estimation of cure by diagnostic group

For a range of diagnostic groups the FP cure model was run to describe trends in the percentage cured and MST of the uncured over time and differences by age group between children and TYA. Only groups with at least 50 patients in each time period and age group were considered further in the cure models. Initially three time periods (1990-1996, 1997-2003 and 2004-2011) were chosen

but for diagnostic groups with less than 50 patients in each time period (and age group) two time periods (1990-2000, 2001-2011) were then considered to ensure enough patients in each group. For certain diagnostic groups models were only considered for children only or TYA only depending on the numbers in each group. Table 3.4 shows the age groups and time periods included for each diagnostic group. For bone tumours all years were considered together as there were less than 50 children diagnosed in the earlier time period. For germ cell tumours and other epithelial tumours, due to the small number of children diagnosed only models for TYA were included.

Table 3.4: Age groups and time periods included in cure models for each diagnostic group

Diagnostic group	Age groups	Time periods
Leukaemia	Children, TYA	1990-1996, 1997-2003, 2004-2011
Lymphoma	Children, TYA	1990-1996, 1997-2003, 2004-2011
CNS tumours	Children, TYA	1990-1996, 1997-2003, 2004-2011
Bone tumours	Children, TYA	All years combined
Soft tissue sarcoma	Children, TYA	1990-2000, 2001-2011
Germ cell tumours	TYA only	1990-1996, 1997-2003, 2004-2011
Neuroblastoma	Children only	1990-2000, 2001-2011
Renal tumours	Children only	1990-2000, 2001-2011
Other epithelial	TYA only	1990-1996, 1997-2003, 2004-2011

Each covariable (age group and period of diagnosis) was included in a univariate model and also an adjusted model including both. Results presented for each diagnostic group and time period include the standardised percentage cured and the standardised median survival time of the uncured, adjusting for the other variable in the model.

3.5.2.3 Association between prognostic risk factors and cure for childhood leukaemia

A detailed exploration of the association between prognostic clinical risk factors and statistical cure was assessed for patients with acute lymphoblastic leukaemia (ALL). ALL was chosen to examine in detail as it is the most commonly diagnosed cancer in childhood accounting for 25% of all childhood cancers [10] and has clinically defined prognostic risk factors [229, 253]. All

patients diagnosed with ALL between October 1990 and June 2011 aged 1-17 years were included. This age range was included rather than the commonly used childhood age range 0-14 years, as it reflects the age range treated in clinical practice at the hospitals in the study region. This also was the upper age limit of the UKALL 2003 trial which started in 2003 although this increased to 20 years in 2006 and to 24 years from 2007 onwards [253], 18-24 year olds were not included in this analysis, as prior to 2006 they would have been treated on different protocols. Infants (aged <1 year) were excluded from this analysis also as they comprised a small group with poor prognosis.

Overall survival and event-free survival (EFS) were examined by prognostic risk factors graphically by Kaplan-Meier survival curves. The prognostic risk factors included were period of diagnosis, age, sex and white cell count (WCC). These risk factors were chosen as these are used in clinical practice for risk stratification [253]. Trends over time were assessed using three time periods corresponding to the recruitment periods of three main trials for ALL in the UK: UKALL XI from October 1990 to March 1997 [254], ALL97 and ALL97/99 from April 1997 to September 2003 [255-257] and UKALL2003 from October 2003 to June 2011 [253]. Within the ALL 97 trial, the duration and treatment intensity changed in November 1999 (with this phase known as ALL97/99) [256], however, due to small sample size it was not possible to consider these two separate sub-periods. Age was categorised as 1-9 years and 10-17 years and WCC was used a proxy for disease severity and categorised as $<50 \times 10^9$/L and $\geq 50 \times 10^9$/L, with those with higher WCC having worse prognosis.

Flexible parametric cure models were used to estimate the percentage cured and the median survival time of the uncured for both overall survival and EFS. Models based on overall survival were modelled in the relative survival framework using the background mortality rate from national lifetables. No such estimates can be used when modelling EFS so these models were run in a non-relative survival framework. In the models based on EFS the uncured group includes a mixture of those who died and those who relapsed therefore interpretation of the survival time of the uncured is difficult. For analysis of cure based on EFS only the percentage cured is presented.

Univariable and multivariable models were included as described above with the multivariable model including period of diagnosis, age, sex and WCC. From the multivariable model standardised cure proportion and MST of the uncured

were calculated to allow comparison between different levels of each risk factor. Excess mortality rate ratios (EMRR), which are equivalent to the hazard ratio from a Cox model, were also estimated from the cure model to allow the examination of the association of covariates on survival and cure for the overall survival model only. A separate cure model including cytogenetic risk group was also included.

Finally, the cumulative incidence of relapse was estimated by time period with death as a competing risk [258] to examine trends in the risk of relapse over time.

3.5.3 Selection of confounders for adjustment

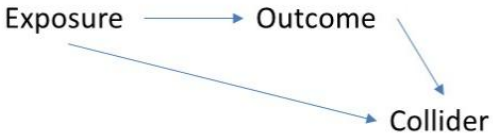
In several of the analyses described in Sections 3.5.5 to 3.5.8 regression models were used to estimate the association between an exposure and the outcome of interest. These analyses have been undertaken within a causal inference framework where identification and selection of confounders were based on graphical causal diagrams. Directed acyclic graphs (DAGs) can be used to visually represent theoretical causal relationships between a series of variables and identify confounders, mediators and colliders in relation to the exposure and outcome in the research question of interest [259-266].

In a DAG, arrows (also known as arcs or edges) connect nodes which represent variables in a causal relationship. A path is a set of arrows connecting any two variables. Arrows may only be uni-directional and relationships must be acyclic so that no paths can lead back to a node. Ideally the DAG should include all variables that influence two or more other variables in the DAG, even if these are unmeasured within the dataset. The causal path of interest is the hypothesized association between exposure and outcome. A “back-door” path is an alternative path between the exposure and outcome. Confounding is the presence of at least one open back-door path between exposure and outcome.

A confounder is a variable which is associated with the exposure and outcome of interest and does not lie on the causal pathway between exposure and outcome. Confounders should be identified from prior knowledge and not based upon available data. When a confounder is controlled for the back-door path is closed.

A mediator is a variable that lies on the causal path between exposure and outcome. A path that includes a mediator is also known as an indirect causal path. Mediators leave the indirect causal path open and controlling for mediators will close the indirect causal path preventing or limiting the ability to observe an association between exposure and outcome. Mediators should not be treated as confounders and should not be adjusted in the model of interest.

Another important variable to consider in the context of DAGs are collider variables [264]. A collider can be identified on a DAG when two arrows along a path point to a variable. Controlling for a collider will open the back door path introducing confounding, therefore colliders should not be adjusted for.



Relationships between exposure and outcomes are often complex involving many potential confounders and mediators therefore causal diagrams are a useful tool to explore such relationships. When more than one variable lies on the back door path the adjustment of one variable is sufficient to close this path. In a DAG with many paths, control of a small number of variables (a minimum set of confounders) will close all back-door paths. Dagitty in an online tool (http://www.dagitty.net/) that can be used to visual DAGs and identify minimum adjustment variable sets [267]. DAGitty was used to plot all DAGs used throughout this thesis and to obtain a minimum sufficient adjustment set for each model of interest. The details for each included DAG are included in the relevant sections relating to each analysis.

The advantages of using DAGs in this context are that the assumptions underlying the models are explicit which aids transparency and reproducibility of the research [264, 266]. The limitations of using DAGs are that they are non-parametric and the direction of the arrows is not always known. In the DAGs presented in this thesis all variables have been drawn in temporal order to help show the assumptions being made in terms of direction of causality. Another limitation is that decisions on which relationships exists can be difficult to decide. Indeed the lack of paths between two variables is as strong an assumption as the inclusion of a path [264]. The DAGS developed in this thesis were developed with discussion with the PhD supervisors and based on previous published literature and clinical expertise. It is acknowledged that the misspecification if a DAG may lead to incorrect adjustment and statistical inferences.

3.5.4 Competing risk regression models

A competing risk in survival analysis is defined as an event which prevents the occurrence of the primary event of interest [268-271]. In the analyses of hospital admissions and SMNs (as measures of late effects) presented in this thesis death was considered as a competing risk, since if a patient died they were no longer at risk for hospitalisation or SMN.

3.5.4.1 Cumulative incidence function (CIF)

If competing risks are not present then the complement of the Kaplan-Meier (KM) function can be used to estimate the incidence of an outcome over time. Estimating the cumulative incidence using the KM method in the presence of competing risks, by treating those who experience a competing event as censored, is not appropriate as this will produce biased estimates [270]. The KM survival function will overestimate the incidence in the presence of competing risks and it is recommended that the cumulative incidence function should be used to estimate incidence which describes the absolute risk of the event of interest over time [270-272]. The cumulative incidence function (CIF) was calculated to estimate the probability of late effects over time including death as a competing risk using the `stcompet` command in Stata [258].

3.5.4.2 Competing risk regression models

In standard survival analysis (with no competing risks) Cox proportional hazards models can be used to estimate the relative effect of covariates on the hazard function. There is a direct correspondence between the effect of a covariate on the hazard of the outcome and the effect of a covariate on the incidence of the outcomes. If a covariate increases the hazard of the occurrence of the outcome it will also increase the incidence of the outcome [273]. In the presence of competing risks there is no longer a direct relationship between the hazard and the risk. The way in which covariates are associated with cause-specific hazards may not be the same as the way they are associated with the cumulative incidence. Two different hazard based regression models have been described and used to deal with competing risks: 1) estimating the effect of covariates on the cause-specific hazard function and 2) estimating the effects of covariates on the subdistribution hazard function (or the CIF) [269-273].

These two methods differ in their use and interpretation and the method chosen should depend on the specific research question. Details of the two approaches are given below.

3.5.4.3 Cause-specific regression models

Cause-specific models can be used to estimate the association between covariates and the rate of occurrence of the event of interest (the hazard). In these models subjects who experience a competing event are treated as censored subjects and removed from the risk set for calculation of the hazard [269, 271, 272]. This model can be implemented using, for example, a Cox model. The cause-specific hazard ratio provides a summary of the relationship between a covariate and the rate of occurrence in subjects who are currently event-free without considering the effect of the competing risk. These models are best suited to address aetiological research questions [269, 271, 272].

3.5.4.4 Fine-Gray subdistribution hazard model

Fine and Gray [274] defined a regression model to directly estimate the relationship between covariates and the cumulative incidence function, or the probability of the occurrence of the event of interest. These models are known as Fine-Gray models, subdistribution hazard models or CIF regression models.

The subdistribution hazard is the probability of failure due to an event at that moment in time, given that this event has not already occurred. The risk set includes all subjects who have not yet experienced the outcome of interest, so includes those who are event-free and also those who have experienced a competing event [269, 273]. Subjects who experienced the competing event are included in the risk set so that they can be counted in the proportion of the population that cannot have the event of interest. These models are recommended if the research question is focussed on estimating incidence and predicting prognosis [269, 271, 272]. Therefore, this model was chosen and used to investigate the relationship between patient risk factors and the incidence of respiratory late effects where death was considered a competing risk.

The interpretation of the coefficients from the Fine-Gray model is not straightforward [272, 273]. Exponentiated regression coefficients denote the subdistribution hazard ratio (sHR) and can be used to describe the direction of the observed association but cannot be used to directly quantify the magnitude of the association since the magnitude of the relative effect of the covariate on the subdistribution hazard function is different from the magnitude of the effect of the covariate on the CIF [273]. A sHR=1 implies no association between the covariate and the CIF, while if the sHR>1 then this implies than a 1-unit increase in the covariate is associated with an increased incidence of the event of interest and if the sHR<1 then the covariate is associated with a decreased incidence [272].

The magnitude of the regression coefficients do not provide information of the magnitude of the covariate on the incidence, however, the magnitude of coefficients from the same model may be compared [273]. For example, if one covariate has a larger regression coefficient than a second covariate then the magnitude of the first covariate on the incidence of the outcome will be greater than the magnitude of the second covariate. Limitations of these models are the sHRs cannot be directly compared from different models with different outcomes, or from different studies since the CIF will not be the same for different types of events [269, 273].

The Fine-Gray model is a semi-parametric model. Similar to the Cox model, the baseline subhazard does not need to be specified and the model assumes that the subdistribution hazards are proportional [274]. This assumption can be

checked graphically by plotting non-parametric cumulative incidence functions by covariates to check for crossing incidence curves or including time varying coefficients where the assumption is violated if there is a significant interaction between the covariate and time [95, 270].

3.5.5 Respiratory admissions

The cumulative incidence for each respiratory condition based on attained age was calculated, treating death as a competing risk [258]. This was based on all diagnosis codes in HES and those based on the primary diagnostic code only.

Admission rates in the general population were used to calculate the expected number of admissions in the cancer survivor cohort and indirect standardisation techniques were used to obtain the hospitalisation rate ratio (HRR) standardised to the general population by age, sex and year [275]. HRRs compare the ratio of the observed to expected admission counts and were calculated for all ages and separately for children and TYA.

Fine-Gray competing risk regression models (as described in 3.5.4.4) were used to examine the association between three treatment exposures (pulmonary toxic chemotherapy, radiation to the chest and thoracic surgery) and the risk of admission for a respiratory condition. Models were included for: 1) any respiratory admission, 2) asthma, 3) pneumonia and 4) chronic lower respiratory disease. Models were not included for lung fibrosis and conditions due to other external agents because of a small number of observed admissions (fewer than 10 admissions). The proportional subdistribution hazards assumption was tested by including each covariable as a time varying coefficient and assessing its statistical significance [95].

Further models were examined including two-way interactions between age group and treatment. Models included an interaction term between age group (children and TYA) and pulmonary toxic chemotherapy and an interaction term between age group and chest radiation to determine whether the association between treatment and risk of admission differed by age. No interaction models were included for thoracic surgery due to small numbers.

The DAG shown in Figure 3.2 shows the potential causal relationships between treatment exposures (in this case pulmonary toxic chemotherapy as the main exposure of interest) and respiratory admissions. From this a minimal sufficient adjustment set was derived and included deprivation, diagnosis age, diagnosis year, diagnostic group, ethnicity and the other treatment exposures. These covariates were all included in an adjusted model. Similarly for the other treatment exposures (chest radiotherapy and thoracic surgery) the same adjustment set was identified.

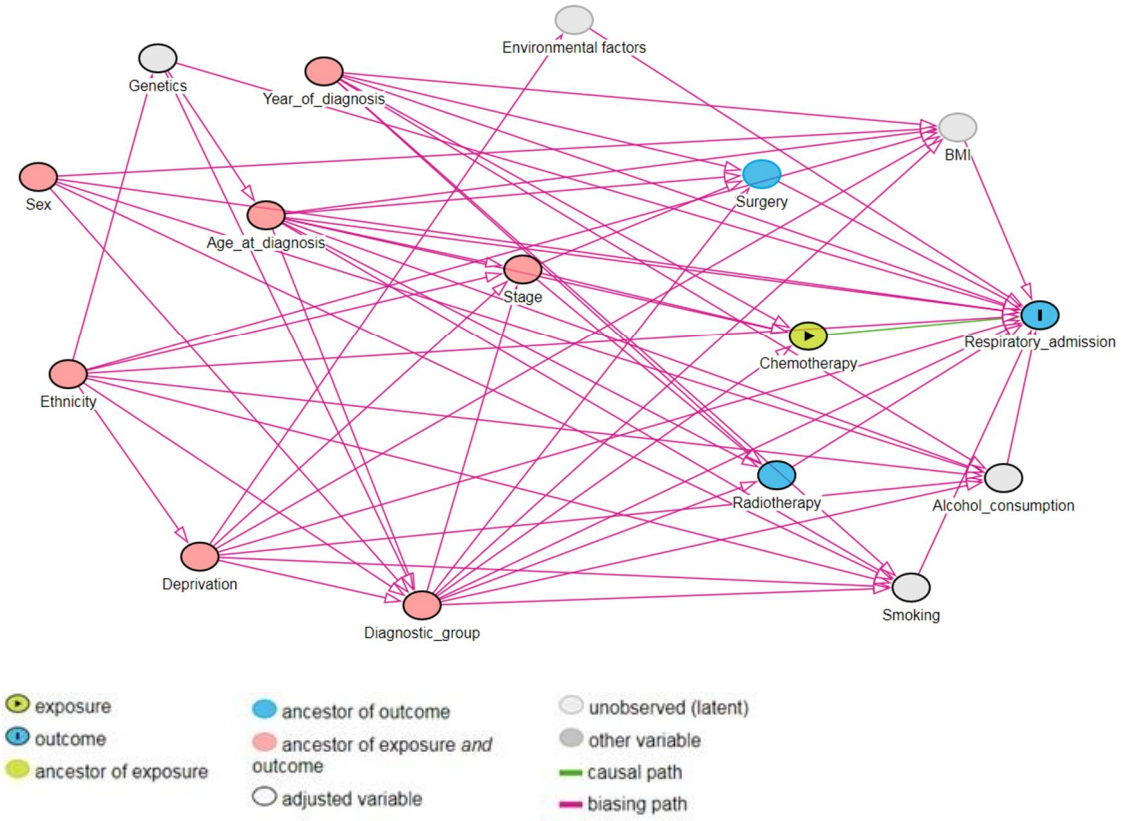


Figure 3.2: DAG representing the relationship between treatment exposures and respiratory admissions

Subsequent admissions and mortality were examined for those admitted for at least one respiratory condition. Flexible parametric survival models were used to estimate the risk of subsequent mortality comparing those whose first admission (five-years post diagnosis) was for pneumonia compared to those admitted for other respiratory conditions.

3.5.6 SMN analysis

The cumulative incidence of developing a SMN was calculated treating death as a competing risk [258]. This was calculated overall and by age at primary diagnosis for children (0-14 years) and TYA (15-29 years). Time at risk for developing a SMN for each person was calculated from date of diagnosis until the earliest of first SMN diagnosed, death or the end of follow-up period (31/12/2015).

To compare SMN rates in cancer survivors with rates in the general population, standardised incidence ratios (SIRs) and absolute excess risks (AER) were calculated. SIRs were obtained by calculating the ratio of the observed number of SMNs to the expected number of incidence cancers based on general population data. National incidence rates by 5-year age band, sex and 1-year calendar period were obtained from the Office of National Statistics [276]. The AER was calculated as the difference between the observed and expected number of SMNs, divided by the total number of person years at risk, reported per 10,000 person years. It can be interpreted as the excess number of subsequent tumours observed per 10,000 survivors per year. The SIR provides relative excess risk, while the AER is a measure of the absolute excess risk. The SIR and AER were calculated for all cancers combined, by primary diagnostic group and by age at primary tumour diagnosis (children and TYA). Further sensitivity analysis was conducted, to allow comparison with the published literature, by estimating the SIR and AER for tumours occurring 5-years post diagnosis only in 5-year survivors only.

The latency time, the time between diagnosis of the first and subsequent tumour, was grouped into time periods (<5 years, 5 to <10 years, 10+ years). Descriptive analyses are presented by latency periods, including the median time between primary and subsequent tumours, by age and SMN type.

Survival analysis was conducted for those who developed an SMN to investigate the relationship between latency time and the impact on survival including Kaplan-Meier plots and flexible parametric survival models. Follow-up time started on the date of diagnosis of SMN and follow-up for death or censoring until 31st December 2016 to allow at least one year follow-up for all. The main exposure of interest was latency period, the association between this and survival was modelled based in the DAG presented in Figure 3.3. From this DAG the minimal adjustment set identified included primary tumour treatment, year of diagnosis and SMN type.

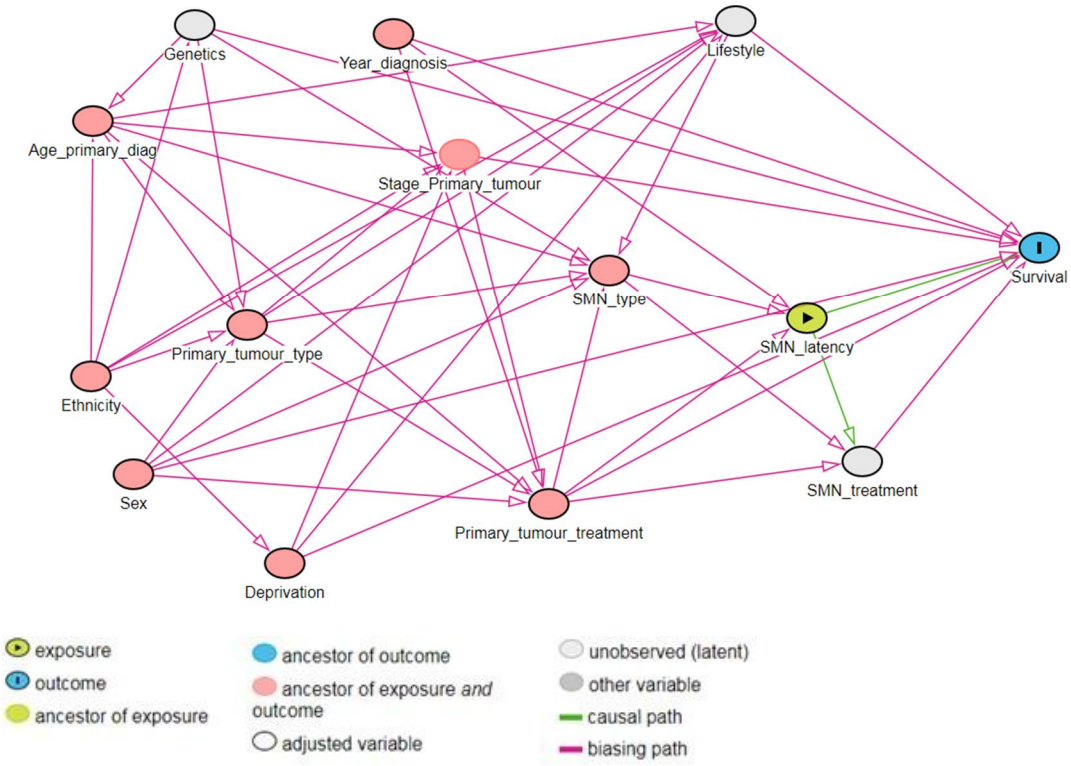


Figure 3.3: DAG representing the relationship between SMN latency period and survival

3.5.7 Cardiovascular admissions

Hospitalisations for cardiovascular admissions for survivors of childhood and young adult cancer have been examined in detail previously using the Yorkshire register [196], therefore a detailed analysis of risk factors for cardiovascular admissions is out with the scope of this thesis. However, the previous work only included HES admissions up to 2011, therefore incidence of admissions up to 2017 for Yorkshire patients are included in Chapter 8. It is also important to include cardiovascular admissions in this analysis as they represent one of the most common causes of late morbidity and mortality in childhood cancer survivors and are included as an event in the cumulative burden analysis (methods described in Section 3.5.8).

The cumulative incidence for each cardiovascular condition based on attained age was calculated, treating death as a competing risk [258]. This was based on all diagnosis codes in HES and those based on the primary diagnostic code only.

Admission rates in the general population were used to calculate the expected number of admissions in the cancer survivor cohort and indirect standardisation techniques were used to obtain the hospitalisation rate ratio (HRR) standardised to the general population by age, sex and year [275]. HRRs were calculated for all ages and separately for children and TYA.

3.5.8 Cumulative burden

In addition to focussing on each late effect (respiratory admission, cardiovascular admission and SMN) independently the cumulative burden of all three events was examined. This also included multiple hospital admissions for respiratory and cardiovascular conditions and multiple SMN diagnoses. Figure 3.4 shows hypothetical examples for 5 individuals with different patterns of events. For example subject A had two respiratory admissions before censoring, while subject B had 3 cardiovascular admissions then died.

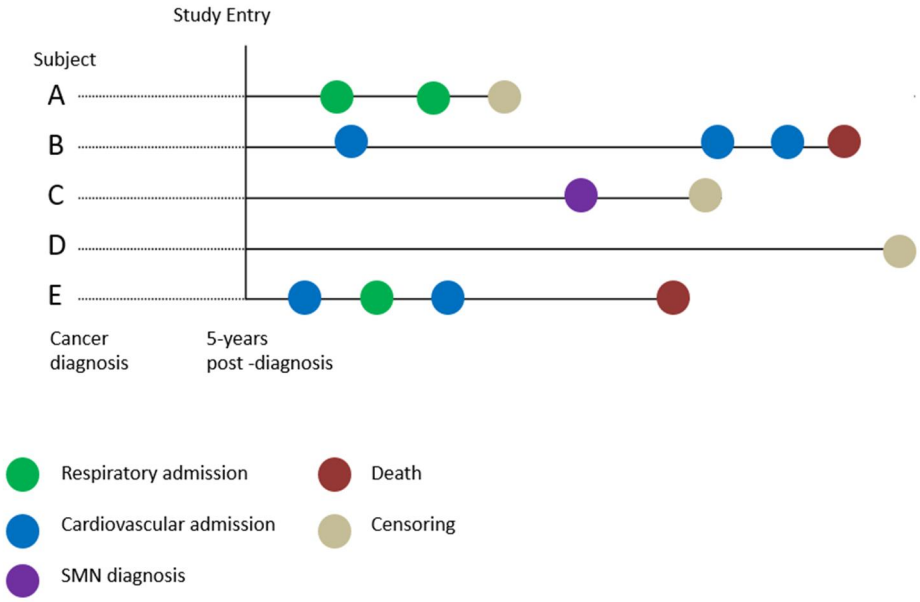


Figure 3.4: Diagrams of five hypothetical individuals with multiple and recurrent events

The cumulative incidence for each event type (respiratory admission, cardiovascular admission or SMN diagnosis) were examined, where follow-up time started 5-years post diagnosis and ended at date of event or date of death

or end of follow-up (31st December 2015), which ever came first [258]. The cumulative incidence of combined outcomes were also estimated where the time at risk ended at the date of the last event. The cumulative incidence was calculated on the attained age scale.

The mean cumulative count (MCC) was used to estimate the total burden of all hospitalisations for respiratory and cardiovascular disease and all SMNs. The MCC estimates the mean number of multiple and recurrent events per individual in the population within a given time period in a competing risk framework [277]. Patients who experience an event are kept in the risk set until they experience a competing risk or are censored. The MCC estimates are presented as the average number of events per survivor. The MCC was calculated for the full cohort and three treatment groups: anthracyclines, pulmonary toxic chemotherapy and radiation to the chest.

3.5.8.1 Multiple failure time models

To estimate the association between previous cancer treatment and cumulative burden, adjusting for potential confounders, the Prentice, Williams and Peterson total time (PWP-TT) survival model for multiple-failure times was used [278]. This model is an extension of the Cox model which allows recurrent events for each individual. The PWP-TT model is a stratified model for ordered multiple events, where all individuals are at risk for the first stratum but only those with an event in the previous stratum are at risk for the successive one. Robust standard errors to account for correlations within individuals. The total time model was used which measures time to events from the start of follow-up [279]. The PWP-TT model may fail to converge if the risk sets for recurrent events are too small, therefore for respiratory and cardiovascular admissions number of events was limited to the first 10 admissions. These models have not been developed within a competing risk framework therefore deaths were also considered as a failure event in these models using the approach described by Westbury et al [280]. An alternative would be to code each death as a censored event, however this would not be appropriate as censoring should be uninformative and this is not the case if the individual has died (as they are then unable to experience the events of interest). Alternative methods would be to use multi-state modelling and this is discussed further in Chapters 8 and 9. Sensitivity analysis was conducted looking at time to first admission for each outcome using standard competing risk regression (where death was treated as a competing risk).

The DAG shown in Figure 3.5 shows the relationships between patient and tumour related variables and cumulative burden (with anthracycline as the main exposure of interest). From this the minimal sufficient adjustment set included diagnostic group, age at cancer diagnosis, year of diagnosis, deprivation, ethnicity and the other treatment exposures (pulmonary toxic chemotherapy and chest radiation). Similarly for the other treatment exposures (pulmonary toxic chemotherapy and chest radiation) the same adjustment set was identified. These variables were included in the adjusted model.

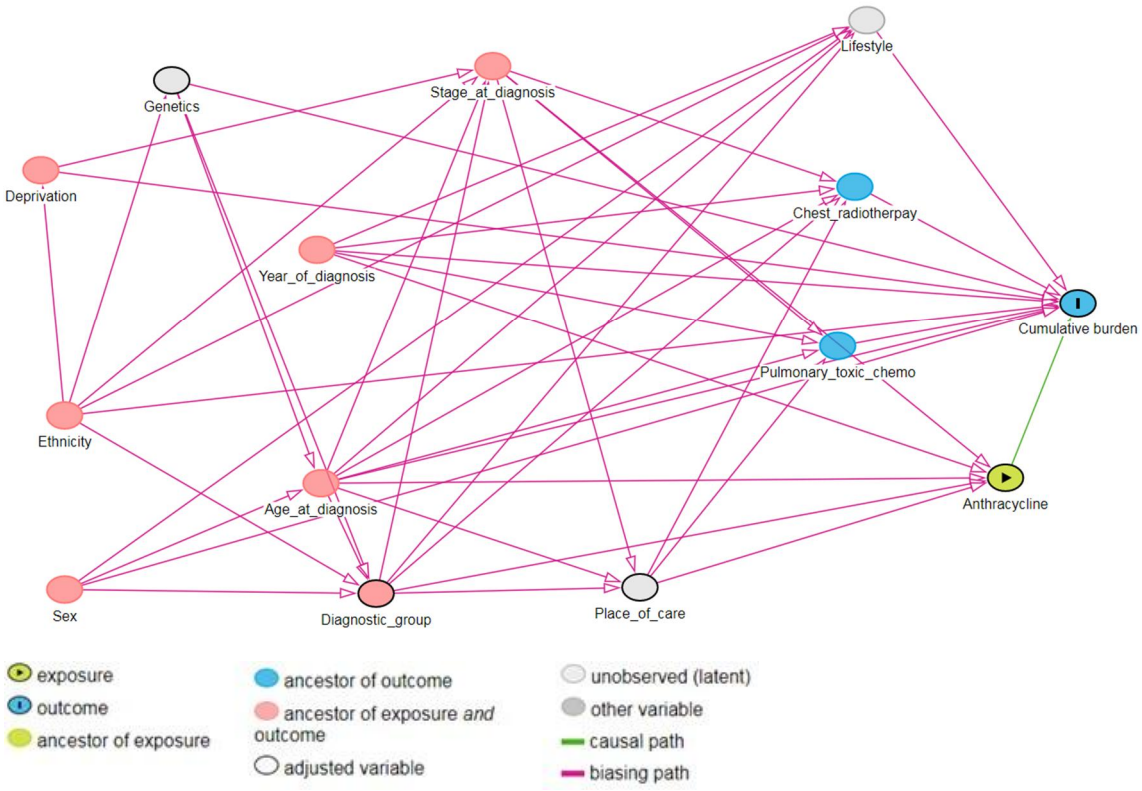


Figure 3.5: DAG showing relationship between treatment exposures and cumulative burden

Further subgroup analysis was conducted by estimating the MCC by age at diagnosis and by diagnostic group. PWP-TT models with two-way interactions between age group at diagnosis and each treatment exposure and further models with two-way interactions between diagnostic group and each treatment exposure were tested to determine whether the association between treatment and cumulative burden differed by age at diagnosis or diagnostic group.

3.6 Overview of results chapters

The following four chapters of this thesis include the main results from the statistical analysis described above. Slightly different follow-up time periods were available for outcomes based on different data sources therefore the total number of individuals included in each chapter differs (in addition some chapters are based on 5-year survivors only). Table 3.5 details the study populations included in the following chapters.

Table 3.5: Summary of study population included in each analysis chapter

Analysis (Chapter)	**Diagnosis period**	**End of follow-up**
Descriptive analysis (Chapter 4)	1990-2011	End of 2016
Cure models (Chapter 5)	1990-2011	End 2016
ALL cure (Chapter 5)	Oct 1990-June 2011	End 2016
Respiratory admissions (Chapter 6)	1990-2011, 5-year survivors only	March 2017 (admissions start 1997)
SMN (Chapter 7)	1990-2010	End 2015
SMN survival (Chapter 7)	1990-2010 (SMN diagnosed up to 2015)	End 2016 (at least 1-year for all)
Cardiovascular admissions (Chapter 8)	1990-2011, 5-year survivors only	March 2017 (admissions start 1997)
Cumulative burden (Chapter 8)	1992-2010, 5-year survivors only	End 2015 (follow-up starts 1997)

Chapter 4 Descriptive analysis of the study population

4.1 Introduction

This chapter describes the study population used in this thesis. Firstly incidence rates are described, followed by trends in survival including survival curves for each diagnostic group for children and TYA. Examination of the survival curves is the first step in assessing if statistical cure is a valid assumption. A description of the demographic and clinical characteristics of the study population is presented. Results from linkage to HES data are reported including a comparison of the characteristics of those matched and unmatched. Finally descriptive statistics are provided for general admissions to hospital within the study population.

4.2 Cancer incidence

The full study sample comprised a total of 5471 patients diagnosed with a primary tumour in Yorkshire between 1990 and 2011 aged under 30, this included 2109 children aged 0-14 years and 3362 TYA aged 15-29 years. The (European) age standardised incidence rate (ASR) was 169 per 1,000,000 persons per year (per million) (95% CI 165 to 174). The ASR was higher in TYA (202 per million, 95% CI 195 to 209) compared to children (138 per million), 95% CI 132 to144). In children the most commonly diagnosed cancers were leukaemias, CNS tumours and lymphomas while for TYA it was lymphomas, germ cell tumours, other epithelial tumours and CNS tumours. (Table 4.1). For TYAs the other epithelial group consisted of 50% thyroid carcinomas and 45% other and unspecified carcinomas. Further breakdown of the TYA tumours by Birch classification are provided in Table 4.5.

Table 4.1: Number of diagnoses and age standardised incidence rate (ASR) per 1,000,000 person years

	All ages (0-29 years)		Children (0-14 years)		TYA (15-29 years)	
Diagnostic group*	**N (%)**	**ASR (95% CI)**	**N (%)**	**ASR (95% CI)**	**N (%)**	**ASR (95% CI)**
I Leukaemia	994 (18.2)	32.2 (30.2, 34.2)	644 (30.6)	42.8 (39.4, 46.1)	350 (10.6)	21.1 (18.9, 23.3)
II Lymphoma	1108 (20.3)	33.1 (31.2, 35.1)	239 (11.3)	15.0 (13.1, 16.9)	869 (25.9)	52.1 (48.7, 55.6)
III CNS tumours	899 (16.4)	28.1 (26.3, 30.0)	483 (22.9)	31.0 (28.2, 33.8)	416 (12.4)	25.1 (22.7, 27.5)
IV Neuroblastoma	163 (3.0)	5.7 (4.8, 6.6)	150 (7.1)	10.4 (8.7, 12.0)	13 (0.4)	0.8 (0.4, 1.2)
V Retinoblastoma	66 (1.2)	2.4 (1.8, 2.9)	66 (3.1)	4.6 (3.5, 5.7)	-	-
VI Renal tumours	159 (2.9)	5.4 (4.5, 6.2)	121 (5.7)	8.3 (6.8, 9.8)	38 (1.1)	2.3 (1.6, 3.0)
VII Hepatic tumours	41 (0.8)	1.4 (0.9, 1.8)	24 (1.1)	1.7 (1.0, 2.3)	17 (0.5)	1.0 (0.5, 1.5)
VIII Bone tumours	232 (4.2)	7.1 (6.2, 8.0)	89 (4.2)	5.5 (4.4, 6.7)	143 (4.3)	8.7 (7.3, 10.2)
IX Soft tissue sarcomas	370 (6.8)	11.5 (10.3, 12.7)	166 (7.9)	10.7 (9.0, 12.3)	204 (6.1)	12.4 (10.7, 14.1)
X Germ cell tumours	928 (17.0)	27.5 (25.8, 29.3)	85 (4.0)	5.6 (4.4, 6.8)	843 (25.1)	50.5 (47.1, 54.0)
XI Other Epithelial tumours	502 (9.2)	14.8 (13.5, 16.1)	40 (1.9)	2.5 (1.7, 3.3)	462 (13.7)	27.7 (25.1, 30.2)
XII Other tumours	9 (0.2)	0.3 (0.1, 0.5)	2 (0.1)	0.1 (0, 0.3)	7 (0.2)	0.4 (0.1, 0.8)
All cancers	5471	169.4 (164.9, 173.9)	2109	138.1 (132.2, 144.0)	3362	202.2 (195.3 209.0)

*Diagnostic group based on ICCC-3.

4.3 Survival

Table 4.2 shows the number of deaths and 5-year survival estimates by age group and diagnostic group. Survival estimates were only calculated for groups including at least 50 cases (therefore hepatic tumours and other tumours were excluded), and no estimates for retinoblastoma were calculated as there was only 1 death within this group. For all patients the median follow-up was 11.2 years (IQR 5.5 to 18.1 years). This was similar for children (11.6 years (IQR 5.3 to 18.7 years)) and TYA (11.0 years (IQR 5.6 to 17.8 years)).

For all cancers combined, for those diagnosed 0-29 years the 5-year survival was 78% (95% CI 76 to 79), this varied from 55% for neuroblastoma (95% CI 47 to 62) to 94% for germ cell tumours(95% CI 92 to 95). Overall 5-year survival was 76% (95% CI 75 to 78) for children ranging from 55% for neuroblastoma (95% CI 46 to 62) to 93% for germ cell tumours (95% CI 85 to 97). For TYA 5-year survival for all cancers combined was 78% (95% CI 77 to 80), which ranged from 58% for leukaemia (95% CI 53 to 63) to 94% for germ cell tumours (95% CI 92 to 95). Compared to children survival was lower for TYA for leukaemia (81% for children (95% CI 78 to 84) and 58% for TYA (95% CI 53 to 63) and CNS tumours (72% for children (95% CI 68 to 76) and 67% for TYA (95% CI 62 to 71)).

Table 4.2: Number of deaths and 5-year overall survival by diagnostic group

	All ages (0-29 years)			Children (0-14 years)			TYA (15-29 years)		
Diagnostic group	**N**	**No deaths* (%)**	**5-year survival (95% CI)**	**N**	**No deaths* (%)**	**5-year survival (95% CI)**	**N**	**No deaths* (%)**	**5-year survival (95% CI)**
I Leukaemia	994	305 (30.7)	73 (70, 76)	644	148 (23.0)	81 (78, 84)	350	157 (44.9)	58 (53, 63)
II Lymphoma	1108	199 (18.0)	86 (84, 88)	239	42 (17.6)	85 (79, 89)	869	157 (18.1)	86 (84, 88)
III CNS tumours	899	361 (40.2)	69 (66, 72)	483	177 (36.7)	72 (68, 76)	416	184 (44.2)	67 (62, 71)
IV Neuroblastoma	163	76 (46..6)	55 (47, 62)	150	70 (46.7)	55 (46, 62)	13	6 (46)	-
V Retinoblastoma	66	1 (2)	-	66	1 (2)	-	-	-	-
VI Renal tumours	159	31 (19.5)	85 (78, 90)	121	22 (18.2)	85 (77, 90)	38	9 (23.7)	-
VII Hepatic tumours	41	20 (49)	-	24	7 (29)	-	17	13 (76.5)	-
VIII Bone tumours	232	106 (45.7)	61 (55, 67)	89	40 (45)	61 (50, 70)	143	66 (46.2)	62 (53, 69)
IX Soft tissue sarcomas	370	157 (42.4)	61 (56, 66)	166	68 (41.0)	63 (55, 70)	204	89 (43.6)	60 (53, 66)
X Germ cell tumours	928	78 (8.4)	94 (92, 95)	85	8 (9)	93 (85, 97)	843	70 (8.3)	94 (92, 95)
XI Other Epithelial tumours	502	133 (26.5)	77 (73, 81)	40	6 (15)	-	462	127 (27.5)	76 (72, 79)
XII Other tumours	9	3 (38)	-	2	0	-	7	3 (43)	-
All cancers	5471	1470 (26.9)	78 (76, 79)	2109	589 (27.9)	76 (75, 78)	3362	881 (26.2)	78 (77, 80)

*percentage within each diagnostic group
Survival estimates only provided for diagnostic groups including at least 50 patients
Survival estimates not provided for retinoblastoma due to small number of deaths within this group

4.3.1 Comparison of overall and relative survival

Table 4.3 compares the 5-year overall survival with 5-year relative survival estimates. For all ages and diagnostic group the estimates and 95% confidence intervals are very similar, only differing to the first decimal place. The Kaplan-Meier survival curves by diagnostic group and age group presented in the next section, (Section 4.4) are based on overall survival.

Table 4.3: Comparison of 5-year overall survival and 5-year relative survival by age group and diagnostic group

	All ages (0-29 years)		Children (0-14 years)		TYA (15-29 years)	
Diagnostic group	**Overall survival (95% CI)**	**Relative survival* (95% CI)**	**Overall survival (95% CI)**	**Relative survival* (95% CI)**	**Overall survival (95% CI)**	**Relative survival* (95% CI)**
I Leukaemia	73.1 (70.3, 75.6)	73.3 (70.4, 75.9)	81.4 (78.1, 84.2)	81.5 (78.2, 84.3)	58.0 (52.7, 63.0)	58.2 (52.8, 63.1)
II Lymphoma	85.8 (83.6, 87.8)	86.0 (83.8, 88.0)	84.5 (79.3, 88.5)	84.6 (79.4, 88.6)	86.2 (83.7, 88.3)	86.4 (84.0, 88.6)
III CNS tumours	69.4 (66.3, 72.3)	69.5 (66.4, 72.4)	71.6 (67.4, 75.4)	71.7 (67.5, 75.5)	66.8 (62.0, 71.1)	66.9 (62.2, 71.3)
IV Neuroblastoma	54.6 (46.6, 61.9)	54.8 (46.8, 62.0)	54.7 (46.4, 62.2)	54.8 (46.5, 62.4)	-	-
VI Renal tumours	84.9 (78.3, 89.6)	85.1 (78.5, 89.8)	85.1 (77.4, 90.4)	85.3 (77.6, 90.5)	-	-
VIII Bone tumours	61.2 (54.6, 67.1)	61.3 (54.7, 67.3)	60.7 (49.7, 69.9)	60.7 (49.8, 70.0)	61.5 (53.1, 69.0)	61.7 (53.2, 69.2)
IX Soft tissue sarcomas	61.1 (55.9, 65.8)	61.2 (56.0, 66.0)	62.7 (54.8, 69.5)	62.7 (54.9, 69.6)	59.8 (52.7, 66.2)	60.0 (52.9, 66.4)
X Germ cell tumours	93.6 (91.9, 95.0)	94.0 (92.2, 95.4)	92.9 (85.0, 96.8)	93.2 (85.2, 97.0)	93.7 (91.9, 95.2)	94.1 (92.2, 95.5)
XI Other Epithelial tumours	76.5 (72.5, 80.0)	76.7 (72.7, 80.1)	-	-	75.8 (71.6, 79.4)	75.9 (71.8, 79.6)
All cancers	77.6 (76.4, 78.7)	77.8 (76.6, 78.8)	76.4 (74.6, 78.2)	76.5 (74.7, 78.3)	78.3 (76.9, 79.6)	78.5 (77.1, 80.0)

*Relative survival based on Ederer II estimates

4.4 Survival trends

This section describes trends in survival by time period and age group, for each diagnostic group. The included time periods were chosen to include at least 50 patients. These Kaplan-Meier plots were also used to graphically assess if statistical cure is a reasonable assumption for analysis in Chapter 5.

4.4.1 All cancers combined

For both children and TYA survival for all cancers combined has steadily improved over time, with similar survival curves for children and TYA diagnosed between 2004 and 2011 (Figure 4.1). The survival curves flatten out after about 10 years follow-up, but in earlier time periods, particularly for TYA the curves do not level off until about 15-20 years after diagnosis. Statistical cure would seem a reasonable assumption for all cancers combined.

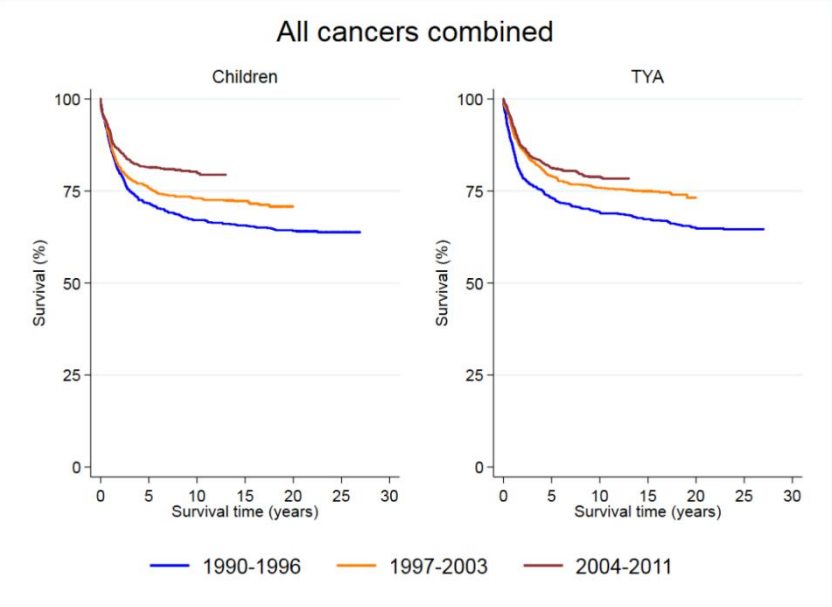


Figure 4.1: Survival trend for all cancers combined

4.4.2 Leukaemia

For both children and TYA survival has improved over time but at different rates (Figure 4.2). In each time period survival was higher in children compared to TYA. For TYA, there was only a small increase in survival between the first and second time period but then a substantial improvement in the latest time period. All curves tended to level off and flatten out between 8-10 years from diagnosis, suggesting it is appropriate to model statistical cure for leukaemia.

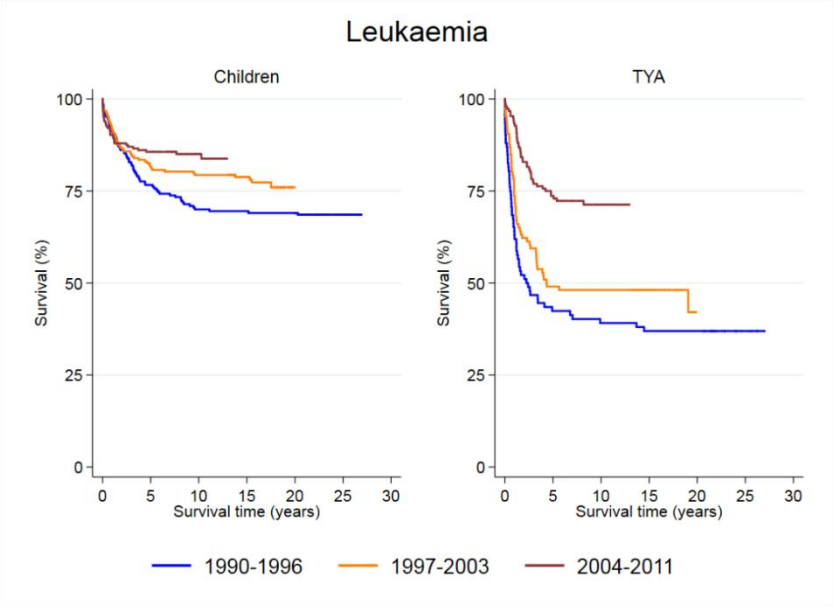


Figure 4.2: Survival trend for leukaemia

4.4.3 Lymphoma

Lymphoma survival has improved over time and is relatively high and similar for both children and TYA diagnosed in the latest time period (Figure 4.3). Survival curves for children levelled off after around 10 years follow-up but for TYA it was slightly later around 10-15 years post diagnosis.

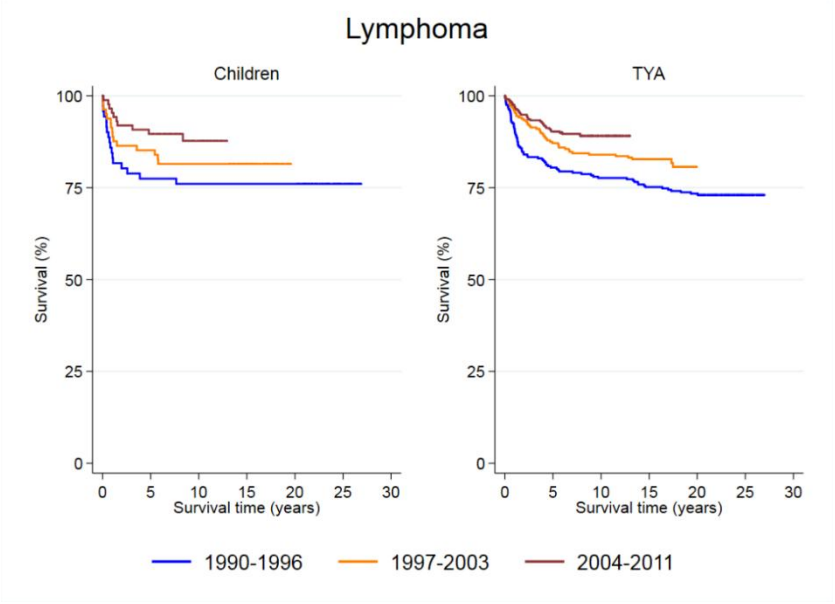


Figure 4.3: Survival trend for lymphoma.

4.4.4 CNS tumours

For children CNS survival improved from 1990-96 to 1997-2003 but no further increase was observed in 2004-2011 (Figure 4.4). For TYA survival improved in each time period but was slightly lower than that for children. The survival curves tended to continue to decrease over time and only showed evidence of plateauing around 15-20 years post diagnosis. Statistical cure may not be feasible for this diagnostic group without further follow-up.

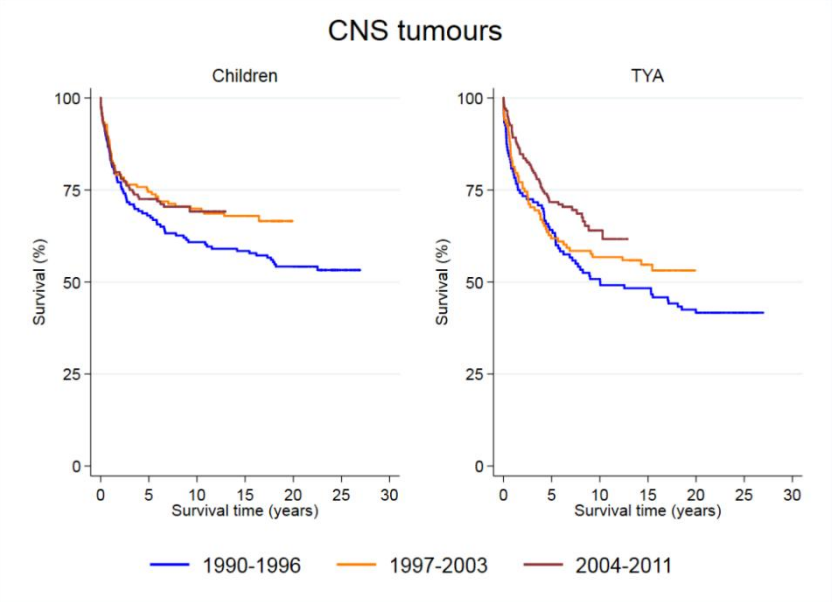


Figure 4.4: Survival trend for CNS tumours.

4.4.5 Bone tumours

Temporal trends in survival for bone tumours could not be assessed due to small numbers, even when two time periods were considered there were still fewer than 50 children in each time period. Therefore survival was caluclated for all years combined for children and TYA separately (Figure 4.5). The survival curves are similar for both children and TYA, the curves tend to flatten out after around 10-15 years suggesting that statistical cure may be an appropriate assumption for this diagnostic group.

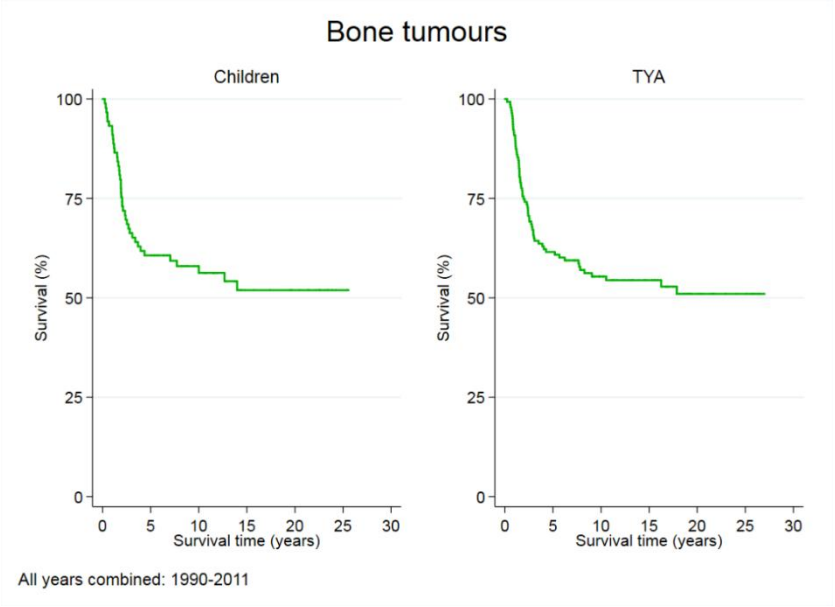


Figure 4.5: Survival for bone tumours

4.4.6 Soft tissue sarcoma

Temporal trends for soft tissue sarcoma were analysed using two time periods to ensure sufficient numbers in each group. For children survival improved slightly for patients diagnosed 2001-2011 compared to those diagnosed 1990-2000, however for TYA survival was slightly lower for those diagnosed more recently (Figure 4.6). For both age groups and time periods the curves plateau about 5-10 years from diagnosis.

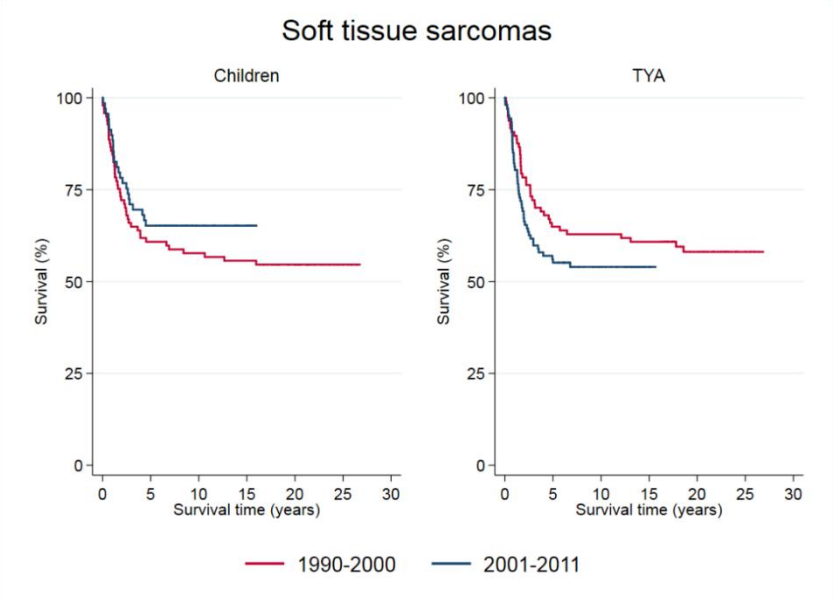


Figure 4.6: Survival trend for soft tissue sarcomas

4.4.7 Germ cell tumours

Temporal trends for germ cell tumours could not be estimated for children due to insuffient sample size, therefore temporal trends in survival were only estimated for TYA using three time periods. Survival for TYA were high in each time period and curves were relatively flat, in the earliest time period there was some suggestion that the survival curve continued to decrease over time, because of this and the relatively high survival statistical cure maybe more difficult to model for this group (Figure 4.7).

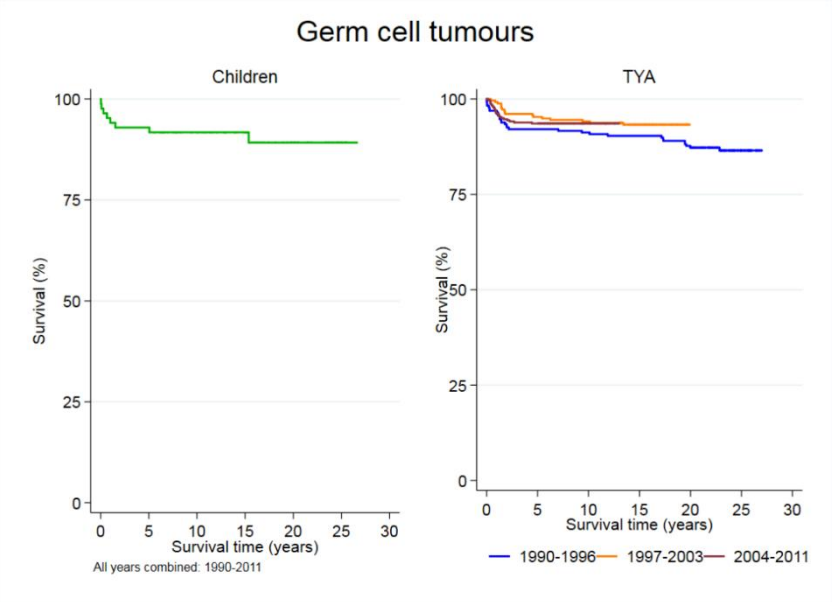


Figure 4.7: Survival trend for germ cell tumours

4.4.8 Neuroblastoma

Neuroblastoma survival was only estimated for children using two time periods. Survival improved for those diagnosed between 2001 and 2011 compared to those diagnosed from 1990-2000 (Figure 4.8). The survival curves flatten out around 5-10 years after diagnosis. Cure would seem a reasonable assumption for children with neuroblastoma.

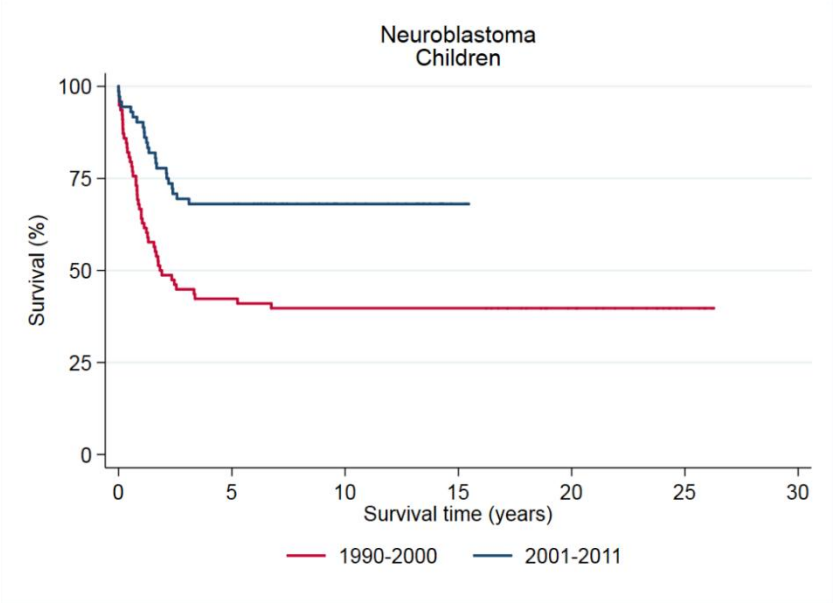


Figure 4.8: Survival trend for neuroblastoma

4.4.9 Renal tumours

Again survival trends for renal tumours were only estimated for children using two time periods. There was a slight increase in survival over time. Survival curves plateaued around 5 years, suggesting statistcal cure would seem a reasonable assumption for children with renal tumours (Figure 4.9).

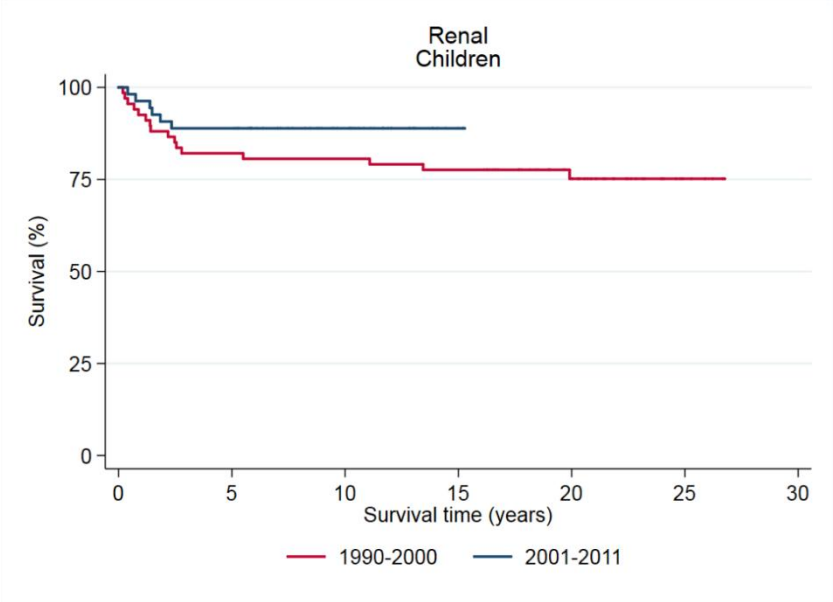


Figure 4.9: Survival trend for renal tumours

4.4.10 Carcinomas

Temporal trends for carcinomas were only estimated for TYA (based on Birch classification) using three time periods. Survival improved from 1990-1996 to 1997-2003 and remained the same in 2004-2011 (Figure 4.10). The curves remained stable after about 10 years from diagnosis suggesting statistical cure may be feasible.

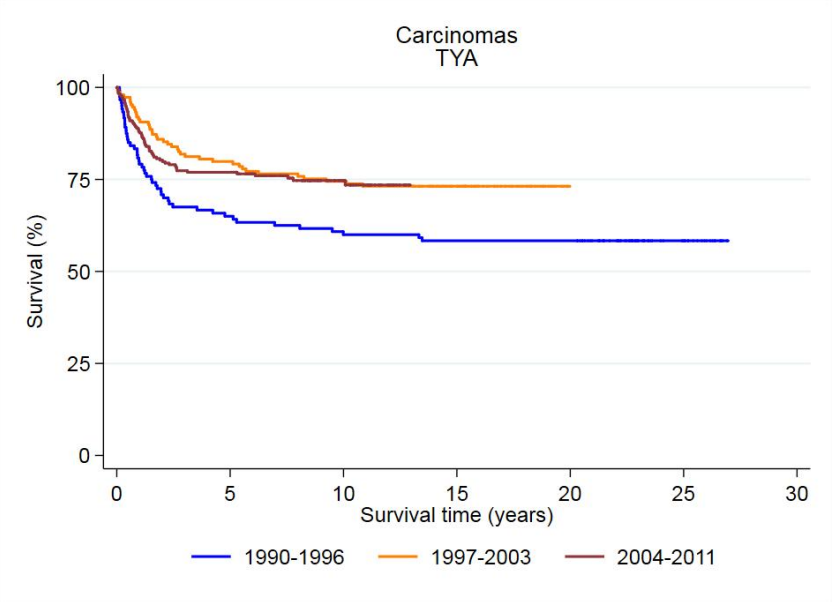


Figure 4.10: Survival trend for carcinomas

4.5 Demographic and clinical characteristics of the study population

The demographic and clinical characteristics of the study population are described in Table 4.4 and Table 4.5. Nearly one third (27%) of the study population were aged 25-29 years at diagnosis and 60% were male. Chemotherapy was the most common treatment received with 64% received some chemotherapy, 49% had surgery and 29% radiotherapy and 15% relapsed. Ethnicity was assigned based on a combination of HES and Onomap as described in Chapter 3, however ethnicity was still not available for 2% of the study population. Yorkshire has greater levels of deprivation compared to England as whole; based on the Townsend indicator 24% of all wards in Yorkshire are in the most deprived fifth. However, 41% of the study population were resident in the most deprived deprivation quintile at the time of diagnosis.

Table 4.4: Demographic and clinical characteristics of the study population

Variable	Category	N (N=5471)	%
Age group	0-4 years	1001	18.3
	5-9 years	544	9.9
	10-14 years	564	10.3
	15-19 years	803	14.7
	20-24 years	1071	19.6
	25-29 years	1488	27.2
Sex	Males	3287	60.1
	Females	2184	39.9
Period of diagnosis	1990-1996	1645	30.1
	1997-2003	1685	30.8
	2004-2011	2141	39.1
Surgery	No	2781	50.8
	Yes	2690	49.2
Chemotherapy	No	1986	36.3
	Yes	3485	63.7
Radiotherapy	No	3880	70.9
	Yes	1591	29.1
Relapse	No	4658	85.1
	Yes	813	14.9
Deprivation quintile	1 (least deprived)	341	6.2
	2	646	11.8
	3	1112	20.3
	4	1132	20.7
	5 (most deprived)	2240	40.9
Ethnicity*	White	4729	86.4
	South Asian	409	7.5
	Other	233	4.3
	Missing	100	1.8

* Ethnicity combination of HES, and Onomap (see Section 3.3.5 for methods).

Around one third of children were diagnosed with leukaemia and 26% of TYAs were diagnosed with lymphoma, 26% with germ cell tumours and 15% with carcinomas. (Table 4.5)

Table 4.5: Diagnostic groups for children and TYA

Age group	Diagnostic group	N	%
Children (0-14 years)	**ICCC-3**	**N=2109**	
	Leukaemia	644	30.5
	Lymphoma	239	11.3
	CNS tumours	483	22.9
	Neuroblastoma	150	7.1
	Soft tissue sarcoma	166	7.9
	Retinoblastoma	66	3.1
	Renal tumours	121	5.7
	Hepatic tumours	24	1.1
	Bone tumours	89	4.2
	Germ cell tumours	85	4.0
	Other tumours	42	2.0
TYA (15-29 years)	**Birch classification**	**N=3362**	
	Leukaemia	347	10.3
	Lymphoma	863	25.7
	CNS tumours	412	12.3
	Bone tumours	170	5.1
	Soft tissue sarcoma	180	5.4
	Germ cell tumours	828	24.6
	Melanoma and skin cancer	14	0.4
	Carcinomas	512	15.2
	Other	36	1.1

A further key prognostic indicator for cancer survival is severity of disease at diagnosis, generally measured by stage or grade at presentation. For leukaemia patients white cell count (WCC) is used as a proxy for stage and used in clinical practice for risk stratification [253]. Stage is not well recorded for all diagnostic groups and Table 4.6 includes the stage details, including the number with missing stage for specific diagnostic groups where previous analysis of the Yorkshire Register data has indicated the staging data are of sufficient quality for inclusion in analysis [281]. Previous research based on the YSRCCYP

explored the missing data mechanisms for stage in detail and used multiple imputation to impute stage under the missing at random assumption [281].

The range of missing stage/grade data ranged from 20% for leukaemia to 43% for both lymphoma and germ cell tumours (Table 4.6). After excluding those with missing stage, 75% of leukaemia patients had low white cell count (<50 x 10^9/L), over two thirds of lymphoma patients presented with stage I and II disease. 40% of CNS tumours were grade I while 27% were grade IV. Only 7% of patients with germ cell tumours presented with advanced stage disease (stages III and IV).

Table 4.6: Stage and Grade distribution by diagnostic group

Diagnostic group and staging system	**Stage category**	**N**	**%**	**% excluding missing**
Leukaemia	<50 x 10^9/L,	595	59.9	75.2
White Cell count	≥50 x 10^9/L	196	19.7	24.8
	Missing	203	20.4	
Lymphoma	I	125	11.3	19.9
Ann Arbor stage	II	297	26.8	47.4
	III	116	10.5	18.5
	IV	89	8.0	14.2
	Missing	481	43.4	
CNS tumours	I	243	27.0	40.2
WHO grade	II	138	15.4	22.8
	III	63	7.0	10.4
	IV	161	17.9	26.6
	Missing	294	32.7	
Germ cell tumours (TYA only)	I	345	40.9	71.3
Royal Marsden, TNM and FIGO combined	II	101	12.0	20.9
	III	18	2.1	3.7
	IV	20	2.4	4.1
	Missing	359	42.6	

4.6 HES linkage summary

HES data were available for admissions between 1st April 1997 and 31st March 2017. The flowchart in Figure 4.11 shows the results from the linkage of the cancer registry data to HES. The linkage was to any record in the HES database, not just restricted to inpatient admissions only. This include linkage to any HES records in the full time period so may relate to episodes around the time of cancer diagnosis and treatment. Overall 9.8% of eligible patients did not link to an inpatient HES episode.

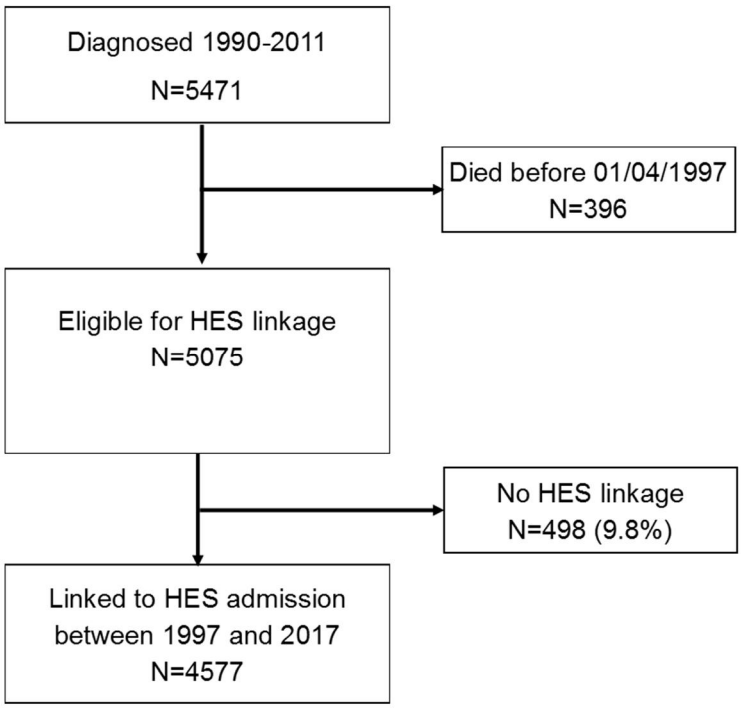


Figure 4.11: Summary of linkage of registry data to HES inpatient admissions

A comparison between the characteristics of those linked and not linked to HES admission data are shown in Table 4.7. There were significant differences between linked and not linked individuals by diagnostic group, age group, sex, year of diagnosis. There was no evidence of a difference by deprivation.

Table 4.7: Comparison of those linked and not-linked to HES inpatient admissions (row %s)

Characteristic	Total n (%)	Linked n (%)	Not linked n (%)	
ALL	5075	4577 (90.2)	498 (9.8)	
Diagnostic group				<0.0001
Leukaemia	903	858 (95.0)	45 (5.0)	
Lymphoma	1043	932 (89.4)	111 (10.6)	
CNS tumours	814	771 (94.7)	43 (5.3)	
Neuroblastoma	132	126 (95.5)	6 (4.5)	
Bone tumours	210	204 (97.1)	6 (2.9)	
Soft tissue sarcoma	336	317 (94.3)	19 (5.7)	
Germ cell tumours	910	783 (86.0)	127 (14.0)	
Other	727	586 (80.6)	141 (19.4)	
Age group				<0.0001
0-4 years	916	872 (95.2)	44 (4.8)	
5-9 years	449	477 (95.1)	22 (4.9)	
10-14 years	525	506 (96.4)	19 (3.6)	
15-19 years	764	704 (92.1)	60 (7.9)	
20-24 years	987	862 (87.3)	125 (12.7)	
25-29 years	1384	1156 (83.5)	228 (16.5)	
Sex				<0.0001
Males	3052	2707 (88.7)	345 (11.3)	
Females	2023	1870 (92.4)	153 (7.6)	
Year of diagnosis				<0.0001
1990-1996	1250	1014 (81.1)	236 (18.9)	
1997-2003	1684	1637 (97.2)	47 (2.8)	
2004-2011	2141	1926 (90.0)	215 (10.0)	
Deprivation quintile				0.05
1 (least deprived)	309	272 (88.0)	37 (12.0)	
2	595	519 (87.2)	76 (12.8)	
3	1028	938 (91.2)	90 (8.8)	
4	1066	964 (90.4)	102 (9.6)	
5 (most deprived)	2077	1884 (90.7)	193 (9.3)	

4.7 Descriptive statistics of HES admission data

A total of 4577 patients linked to inpatient HES data, of these there were a total of 101104 episodes and 94603 CIPS. The median number of episodes per individual was 13 (Inter quartile range (IQR) 5, 13) (Table 4.8).

Table 4.8: Summary of all HES admissions

	Mean	Median (IQR)
Number of episodes per patient	22	13 (5, 31)
Number of spells per patient	21	12 (4, 30)
Number of CIPS per patient	21	12 (4, 29)

In Chapters 6 and 8, HES data were used to estimate late effects in 5-year survivors incorporating admissions 5-years post diagnosis only. Out of 4235 5-year survivors, 2649 (62.6%) had at least one admission 5-years post diagnosis. There were a total of 20019 episodes and 18299 CIPs. The median number of episodes per individual was 3 (IQR 1, 7).

Table 4.9: Summary of HES admissions 5-years post diagnosis in 5-year survivors

	Mean	Median (IQR)
Number of episodes per patient	8	3 (1, 7)
Number of spells per patient	7	3 (1, 7)
Number of CIPS per patient	7	3 (1, 7)

4.8 Discussion

4.8.1 Results in context

Cancers in children and young adults are rare. In Yorkshire, as reported elsewhere, incidence is higher in TYA compared to children. Since the early 1990s cancer incidence in children in the UK has increased by 15%, the incidence rate for cases diagnosed 2010-2012 was 157 per million [11]. This compared with an incidence rate of 138 per million in this study based on all cases diagnosed from 1990-2011. Nationally for TYA (aged 15-24 years) incidence rates have increased by 33% since the early 1990s. National rates

have been estimated to be 298 per million in 2010-2012 [15]. For England, crude incidence rates for the 13-24 year age group were estimated to be 298 per million for 2013-2015 [87]. Neither of these rates included the 25-29 year age group that are included in this thesis, therefore direct comparison between rates is difficult. In Yorkshire for 15-29 year olds the incidence rate from 1990-2011 was 202 per million.

For the study cohort overall 5-year survival was 76% for children and 78% for TYA, with both age groups showing significant increases in survival over time. National 5-year survival for children diagnosed 2006-2010 was 82% [11] and for TYA (15-24 years) diagnosed 2001-2005 was 81% [15]. In general, there is no evidence to suggest that incidence and survival rates and trends are different in Yorkshire compared to national data.

Survival varies by diagnostic group and from the study data included in this thesis for most diagnostic groups there were sufficient numbers to examine survival trends. Although for some diagnostic groups, bone tumours, soft tissue sarcoma, neuroblastoma and renal tumours, several years' data had to be aggregated and limited temporal trends analysis could be conducted.

The first step in determining if statistical cure is an appropriate assumption is to graphically check if the survival curves level off and plateau during follow-up [28]. For most diagnostic groups this seems a reasonable assumption, although the length of follow-up needed may vary by diagnostic group, for example longer follow-up may be needed for CNS tumours. Further examination of cure models by diagnostic group are presented in Chapter 5. Leukaemia is the most commonly diagnosed cancer within children and acute lymphoblastic leukaemia (ALL) accounts for 80% of all leukaemias [10]. This subgroup is included in a detailed analysis examining statistical cure and relapse by clinical characteristics, including further linkage to cytogenetic risk factor data (results also presented in Chapter 5).

Linkage to HES admission data were available for 90% of the study cohort. This is comparable to other cancer registry linked HES admission studies [240, 281, 282]. There were some differences in individuals linked and not linked to HES and results based on hospital admissions need to be interpreted with this in mind (See Section 4.8.2 below and further discussion in Chapter 9). HES

admissions will be used as a proxy for long-term morbidity in Chapters 6 and 8 to quantify and assess the late effects of respiratory and cardiovascular disease in childhood and young adult cancer survivors.

4.8.2 Strengths and limitations

Key strength of the data used throughout this thesis are that the data are population-based and includes detailed patient, tumour and treatment related factors. The full 0-29 year age range were included, compared to other studies that may not include this full age range. No data were available for those aged 30-39 years at diagnosis so limited comparisons with studies based on the TYA age range of 15-39 years (such as TYACCS) were possible. However this thesis does provide essential intelligence on long-term outcomes for this understudied TYA group.

The late effects for this cohort were based upon hospital admission data providing an objective outcome measure compared to other studies based on self-reported questionnaires, which may suffer from recall bias and non-response, or studies based on clinical assessments, which may pick up non-symptomatic conditions and are generally single-centred.

There are several limitations to be acknowledged. Firstly given the rarity of certain diagnostic groups, limited analysis is possible due to small numbers. For example examining trends in survival for each diagnostic group a threshold of a minimum of 50 cases was used to ensure sufficient number to enable robust estimation of survival by age groups and time period. While this is an arbitrary threshold it is in line with national recommendations from PHE [250].

Stage data were not available for all diagnostic groups and even for the diagnostic groups with sufficient stage information it was still missing for up to 40% of cases. This has implications for inclusion in statistical models as a potential confounder. However, as discussed in the methods (Section 3.5.3), variables other than stage were selected based on the relevant DAGs.

Linkage to HES was only available for 90% of the study population and analysis of those linked and not linked showed difference by patient characteristics. This means that certain groups may be under-represented in the analysis of late

effects, including those diagnosed with germ cell tumours, who also have high survival rates, and those aged 25-29 years at diagnosis. This is a limitation of this study. However, there were no differences in the linkage rate by deprivation. The linkage rate was lower for those diagnosed in the earlier time period as HES data were only available from 1997 onwards so for these patients no admissions around the time of diagnosis and treatment were available, only longer-term admissions. Hospital admissions were used as the basis to evaluate long-term morbidity and for those not linked to any HES admissions it is unclear if no late effects were observed in this group as a result of actually having no long-term admissions or as a results of linkage errors and therefore these admissions were not captured. This means that the estimates of long-term morbidity may be an underestimation of the true burden of disease. Further issues around potential linkage errors and the implications of this are discussed further in Chapter 9, Section 9.4.1. It was observed that 60% of 5-year survivors had at least one admission 5-years post diagnosis. Reassuringly the linkage rate for these data are similar to other national studies based on cancer registry linked HES data [240, 281, 282] and compare favourably with studies based on questionnaire responses which typically have lower response rates, for example the BBCSS had a response rate of 70% [80].

4.8.3 Summary

This chapter provides a detailed description of the registry and linked hospital admission data that were used in the analysis presented in Chapters 5-8. Chapter 5 includes further detailed modelling of statistical cure. HES admissions will be used as a proxy for morbidity in Chapters 6 and 8 to quantify and assess the late effects of respiratory and cardiovascular disease in childhood and young adult cancer survivors. Chapter 7 includes investigation of subsequent tumours based on cancer registration data. Table 3.6 showed the different study populations included in each chapter and the descriptive characteristics of these groups are provided at the start of each relevant chapter in relation to the outcome of interest.

Chapter 5 Application of cure models to children and young adults diagnosed with cancer

5.1 Introduction

The results presented in this chapter address Aim 1 of this thesis and a key gap in the literature which was to assess the feasibility of applying cure models to CYA diagnosed with cancer using data from a regional population-based specialist cancer register. Three different cure models were compared: 1) the flexible parametric (FP) cure model, 2) mixture cure model and 3) non-mixture cure model. In order to fit the FP cure model the optimal model had to be determined first, which was achieved by comparing models with different internal knot points and comparing model fit statistics to select the best fitting FP cure model (Section 5.2). Then a comparison between the three different types of cure models was carried out based on all cancers combined (Section 5.3). All cancers combined were used for methodological comparisons to maximise the study sample, however there are wide variations in survival by diagnostic group so trends in cure by diagnostic group were estimated using the FP cure model, comparing differences by age group and time period (Section 5.4). Full details of the statistical methods were included in Chapter 3 and descriptive statistics of the study population were included in Chapter 4 including Kaplan-Meier survival plots. The final section of this chapter focusses on patients with ALL and incorporates clinical risk factors (Section 5.5). Full details of the justification of this specific study population and the statistical methods were detailed in Chapter 3 (Section 3.5.2.3).

5.2 Selecting the optimal flexible parametric cure model

To identify the best fitting FP cure model several models with different degrees of freedom and internal knots were compared based on all cancers combined (see methods in Section 3.5.2). Model fit statistics, the AIC and BIC, for each model were compared. Predicted survival curves were plotted graphically and compared to the life table Ederer II estimated at yearly intervals.

5.2.1 Flexible Parametric model fit statistics

Five different models were compared, with knot points placed at different times throughout follow-up, model fit statistics are presented in Table 5.1. Based on the AIC the models with 7 or 8 degrees of freedom were similar with the lowest AIC and based on the BIC models with 5-8 degrees of freedom were similar and had the lowest BIC (Table 5.1), suggesting a model with at least 5 degrees of freedom (4 internal knot points) would be optimal.

Table 5.1: Comparison of model fit statistics for flexible parametric cure model with different knot points

Degrees of freedom	Number of internal knots	Centile positions of knot points	AIC	BIC
4	3	33, 67, 95	11774.9	11860.8
5	4	25, 50, 75, 95	11744.5	11837.0
6	5	20, 40, 60, 80, 95	11743.0	11842.1
7	6	17, 33, 50, 67, 83, 95	11732.8	11838.6
8	7	14, 29, 43, 57, 71, 86, 95	11728.0	11840.3

5.2.2 Flexible Parametric predicted survival plots

Predicted survival was plotted for the full cohort for each model with different knot points and compared to yearly Ederer II life table estimates of relative survival stratified by age group and time period to assess differences in predicted survival from the models. Figure 5.1 shows that all models provided similar model fit in terms of estimating survival. For nearly all age groups and time periods the predicted survival from the cure model was very similar to the lifetable estimates with the exception of the 0-14 year age group diagnosed between 1997 and 2003 and 15-29 year age group diagnosed between 1990 and 1996. For these groups the predicted survival from the cure model was slightly higher than the lifetable estimates, however the data are sparse and the predicted survival curve was within the 95% confidence limits of the Ederer II estimates indicating adequate model fit.

Figure 5.2 shows the estimated survival curves of the uncured from the different models. Again all models showed similar predicted survival of the uncured with the exception of the model based on 4 degrees of freedom.

Figure 5.1: Plot of predicted survival from flexible parametric cure models with different degrees of freedom (df) and Ederer II lifetable estimates of relative survival (with 95% CI)

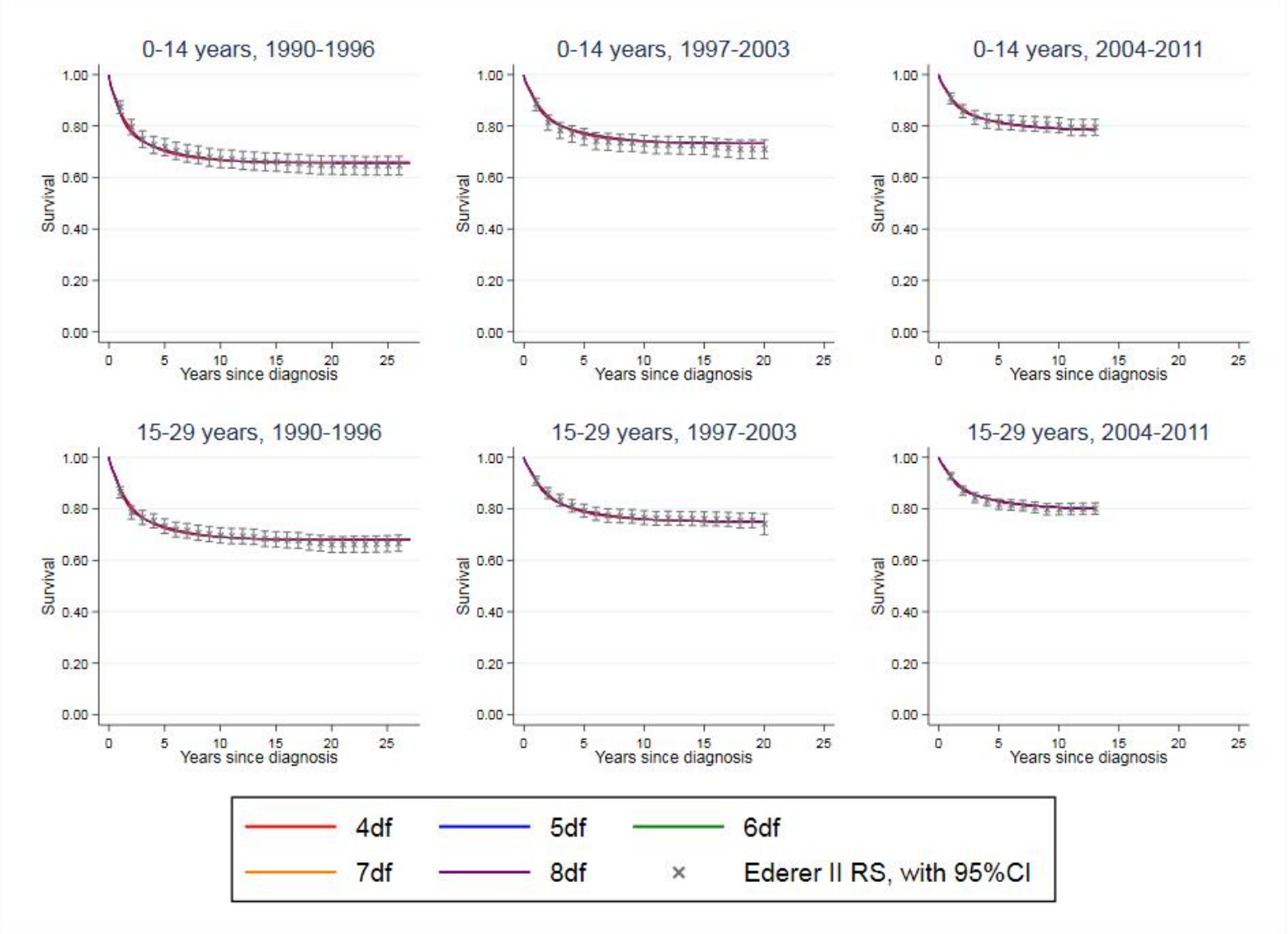
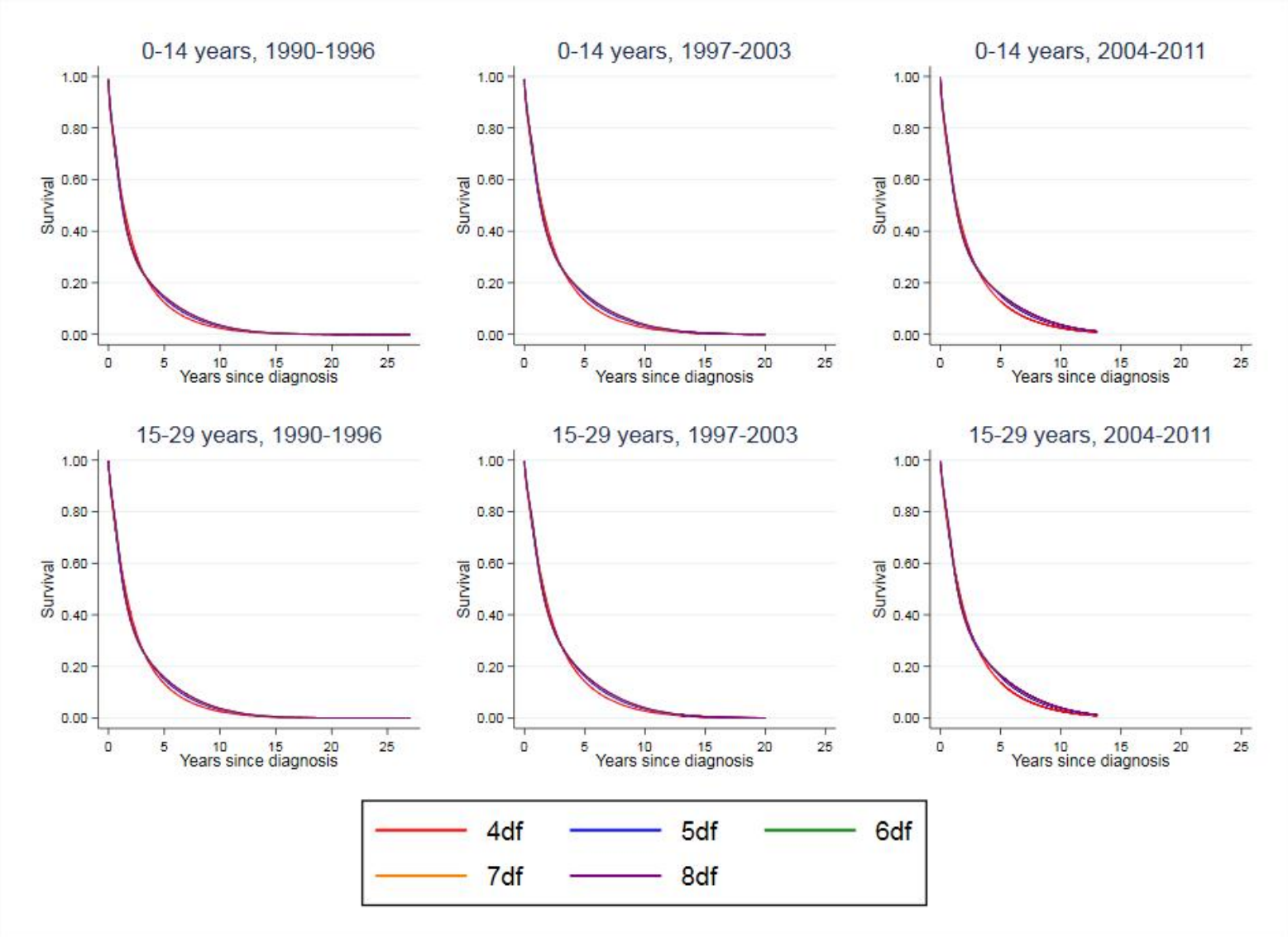


Figure 5.2: Plot of predicted survival of the uncured from flexible parametric cure models with different degrees of freedom (df)



5.2.3 Optimal flexible parametric cure model

The model with 5 degrees of freedom (4 internal knots placed at the 25th, 50th, 75th, and 95th centiles of the uncensored death times) was chosen as the most parsimonious model. This model had the lowest BIC, similar AIC to the model with 6 degrees of freedom and showed similar predicted survival estimates compared to models with more degrees of freedom. This model was used in further comparisons with the mixture and non-mixture cure models presented in the next section.

5.3 Comparison of flexible parametric, mixture and non-mixture cure models

Three different types of cure model were compared: flexible parametric (FP), mixture and non-mixture. These models were compared in terms of model fit statistics and model parameters including the percentage cured, the median survival time (MST) of the uncured and the time when 90% of the uncured had died. Plots of predicted survival curves and the survival curves of the uncured group were also compared between models.

5.3.1 Model fit statistics

Based on model fit statistics (Table 5.2), the mixture and non-mixture model showed very similar values, which were slightly higher than the AIC for the FP model, and slightly lower than the BIC for the FP model.

Table 5.2: Comparison of model fit statistics from the flexible parametric, mixture and non-mixture cure model

	Flexible parametric	**Mixture**	**Non-Mixture**
AIC	11744.46	11747.49	11746.26
BIC	11836.96	11826.77	11825.55

5.3.2 Percentage cured

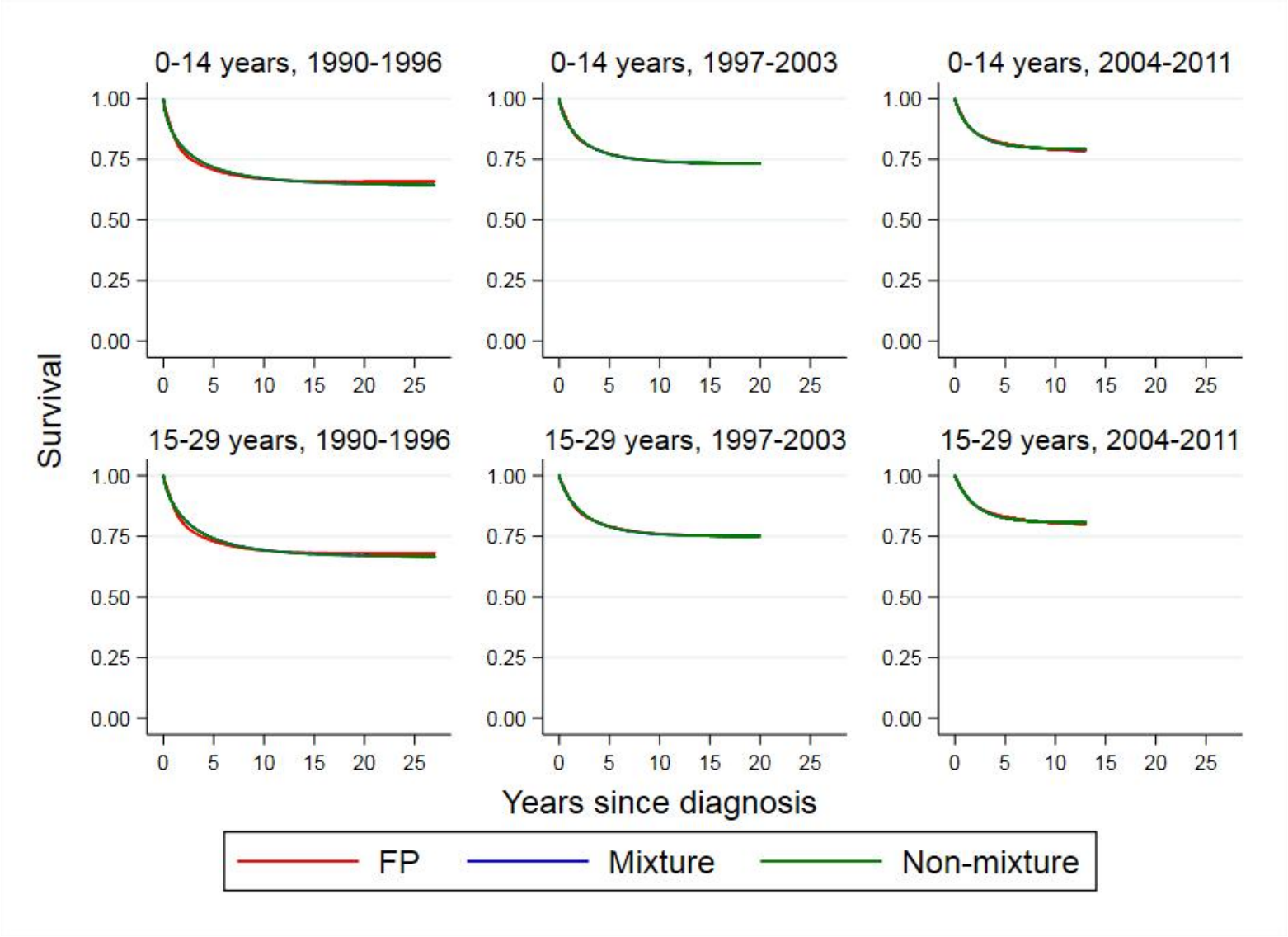
The three different cure models gave very similar estimates of the percentage of patients cured for each age group and period of diagnosis; 73% for children, 75% for TYA, and all showed that the percentage of patients cured increased over time. Based on the FP model 67% of patients diagnosed between 1990 and 1996 were cured increasing to 79% for patients diagnosed between 2004 and 2011 (Table 5.3).

Table 5.3: Comparison of percentage cured from flexible parametric, mixture and non-mixture cure model

	Percentage cured (95% CI)		
	Flexible Parametric	**Mixture**	**Non-mixture**
Age group			
0-14 years	73 (71, 75)	73 (71, 75)	73 (71, 75)
15-29 years	75 (73, 76)	75 (73, 76)	75 (73, 76)
Diagnosis period			
1990-1996	67 (65, 70)	66 (63, 68)	66 (63, 68)
1997-2003	74 (72, 77)	74 (72, 76)	74 (72, 76)
2004-2011	79 (77, 81)	80 (78, 82)	80 (78, 82)

The predicted survival curves for the three models were similar for all age groups and time periods (Figure 5.3).

Figure 5.3: Predicted relative survival from flexible parametric (FP), mixture and non-mixture cure model



5.3.3 Survival of the uncured

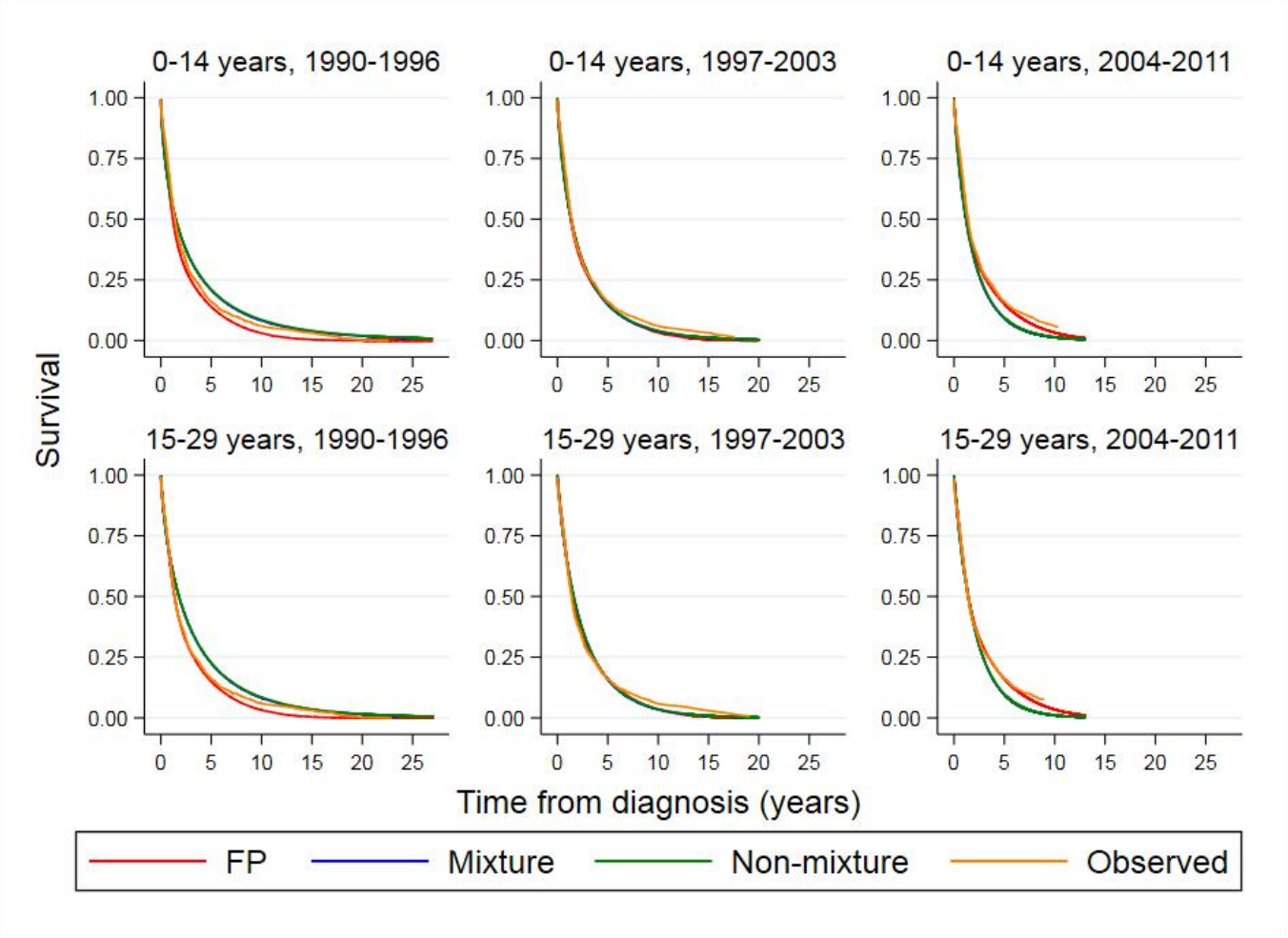
The median survival time (MST) of the uncured estimated from the mixture and non-mixture model were very similar (Table 5.4), however these varied slightly from those obtained from the FP model, which were generally lower. For example, for 15-29 year olds the MST from the FP model was 1.4 years (95% CI 1.3 to 1.6) compared to 1.6 years (95% CI 1.4 to 1.8) from the mixture model and 1.6 years (95% CI 1.4 to 1.7) from the non-mixture model. The largest difference in estimates of the MST was for those diagnosed between 1990 and 1996; the MST was 1.3 years (95% CI 1.1 to 1.4) from the FP model, 1.7 years (95% CI 1.4 to 1.9) from the mixture model and 1.7 years (95% CI 1.4 to 1.9) from the non-mixture model. Based on the FP cure model the MST estimates remained fairly stable over time, however both the mixture and non-mixture models suggest non-statistically significant decrease in MST over time. There were also substantial differences in the time when 90% of the uncured had died, again the mixture and non-mixture model estimates were similar and showed a decrease over time, while the FP model showed no difference in trend over time.

Figure 5.4 shows the predicted survival of the uncured from the three different models alongside the Kaplan-Meier observed survival for patients that died. For all age groups and time periods the predicted survival of the uncured is the same for the mixture and non-mixture model but there are differences between them and the FP cure model and the observed survival. The FP cure model more accurately models the observed survival particularly in the earliest and latest time periods for both children and TYA.

Table 5.4: Comparison of median survival time of the uncured and time when 90% uncured are dead from flexible parametric, mixture and non-mixture cure model

	Flexible Parametric		Mixture		Non-mixture	
	Median survival time of uncured (years) (95% CI)	**Time when 90% uncured dead (years) (95% CI)**	**Median survival time of uncured (years) (95% CI)**	**Time when 90% uncured dead (years) (95% CI)**	**Median survival time of uncured (years) (95% CI)**	**Time when 90% uncured dead (years) (95% CI)**
Age group						
0-14 years	1.3 (1.1, 1.4)	6.4 (5.8, 6.9)	1.3 (1.1, 1.5)	6.5 (5.4, 7.5)	1.3 (1.1, 1.5)	6.6 (5.5, 7.7)
15-29 years	1.4 (1.3, 1.6)	6.6 (6.1, 7.1)	1.6 (1.4, 1.8)	6.5 (5.7, 7.4)	1.6 (1.4, 1.7)	6.7 (5.8, 7.6)
Diagnosis period						
1990-1996	1.3 (1.1, 1.4)	6.3 (5.8, 6.8)	1.7 (1.4, 1.9)	8.9 (7.3, 10.5)	1.7 (1.4, 1.9)	9.1 (7.4, 10.8)
1997-2003	1.4 (1.2, 1.6)	6.6 (6.0, 7.1)	1.5 (1.3, 1.7)	6.4 (5.3, 7.5)	1.5 (1.3, 1.7)	6.5 (5.4, 7.6)
2004-2011	1.4 (1.2, 1.5)	6.6 (6.0, 7.1)	1.3 (1.1, 1.5)	4.9 (4.1, 5.6)	1.3 (1.1, 1.5)	4.9 (4.1, 5.7)

Figure 5.4: Predicted survival of the uncured from FP, mixture and non-mixture cure model and observed survival for those patients that died



5.3.4 Summary of cure model comparison

The percentage cured predicted for each model was similar. However, there were some differences in the predicted survival of the uncured particularly the time when 90% of uncured had died where the mixture and non-mixture models produced similar estimates which were slightly different to the FP model. The AIC suggested that the FP cure model was the best fitting model while the BIC suggested that the mixture or non-mixture model was best fitting. However, these model fit statistics should be interpreted cautiously with cure models (see Section 2.8). Plots of the predicted and observed survival of the uncured showed that the FP model was the best fitting. Therefore the FP model was selected to use in further analysis of cure models presented in this chapter.

5.4 Cure by diagnostic group

So far, cure models have been included for all cancers combined, however there is wide variation in survival by diagnostic group therefore separate models by cancer type are needed. Survival curves by age group, time period and diagnostic group were presented in Chapter 4 and showed that for all cancers the survival curves generally flattened out over time between 5-10 years from diagnosis. However the time at which this occurred varied by diagnostic group. For example, for CNS tumours it was longer at around 15-20 years.

The FP cure model was used to model cure by age group (children 0-14 years and TYA 15-29 years) and diagnosis period for the most common diagnostic groups in children and young people. Age groups and time periods were chosen to include at least 50 patients to ensure robust estimates as described in Chapters 3 and 4. From each model the percentage cured, MST of the uncured and time when 90% of the uncured had died were estimated.

5.4.1 Trends in cure by age group

Table 5.5 shows the cure model estimates by diagnostic group and age group. For children (0-14 years) diagnosed between 1990 and 2011, the percentage cured ranged from 55% for neuroblastoma (95% CI 46 to 62) and bone tumours (95% CI 43 to 65) to 84% for renal tumours (95% CI 76 to 89). The percentage cured was also high for leukaemia (79%, 95% CI 76 to 82) and lymphoma (82%, 95% CI 77 to 87)). There was little variation in the median survival time of

the uncured which ranged from 1.0 years for neuroblastoma (95% CI 0.7 to 1.2) to 1.9 years for bone tumours (95% CI 1.4 to 2.5). The time when 90% of the uncured had died ranged from 2.8 years for neuroblastoma (95% CI 2.5 to 3.0) to 9.4 years for CNS tumours (95% CI 8.4 to 10.4). For TYA (15-29 years) diagnosed between 1990 and 2011, the percentage cured ranged from 53% for leukaemia (95%CI 48 to 58) to 93% for germ cell tumours (95% CI 92 to 96). The median survival time of the uncured varied from 0.8 years for germ cell tumours (95% CI 0.4 to 1.2) to 1.9 years for lymphoma (95% CI 1.4 to 2.3) and CNS tumours (95% CI 1.4 to 2.4). The time when 90% of the uncured had died ranged from 3.9 years for germ cell tumours (95% CI 0.7 to 7.0) to 9.3 years for CNS tumours (95% CI 8.4 to 10.2).

For leukaemia and CNS tumours the percentage cured was higher for children compared to TYA but there was little difference in the survival of the uncured. The percentage cured for lymphoma, bone tumours and soft tissue sarcomas was similar for children and TYA (Table 5.5)

Table 5.5: Cure model results by age group and diagnostic group

Group	Percentage cured (95% CI)	Median survival time of uncured (years) (95% CI)	Time when 90% uncured dead (years) (95% CI)
Leukaemia			
0-14 years	79 (76, 82)	1.3 (0.9, 1.6)	5.7 (4.8, 6.7)
15-29 years	53 (48, 58)	1.2 (0.9, 1.5)	5.4 (4.5, 6.3)
Lymphoma			
0-14 years	82 (77, 87)	1.2 (0.7, 1.6)	7.5 (5.5, 9.4)
15-29 years	83 (81, 86)	1.9 (1.4, 2.3)	8.1 (6.6, 9.5)
CNS tumours			
0-14 years	63 (59, 68)	1.6 (1.1, 2.0)	9.4 (8.4, 10.4)
15-29 years	55 (50, 60)	1.9 (1.4, 2.4)	9.3 (8.4, 10.2)
Bone tumours			
0-14 years	55 (43, 65)	1.9 (1.4, 2.5)	5.9 (4.6, 7.3)
15-29 years	56 (47, 64)	1.7 (1.4, 2.1)	5.7 (4.3, 7.1)
Soft tissue sarcoma			
0-14 years	60 (52, 67)	1.5 (1.1, 1.8)	4.4 (3.3, 5.5)
15-29 years	58 (51, 65)	1.5 (1.2, 1.8)	4.5 (3.4, 5.6)
Germ cell tumours*			
15-29 years	94 (92, 96)	0.8 (0.4, 1.2)	3.9 (0.7, 7.0)
Neuroblastoma			
0-14 years	55 (46, 62)	1.0 (0.7, 1.2)	2.8 (2.5, 3.0)
Renal tumours			
0-14 years	84 (76, 89)	1.4 (0.7, 2.1)	4.3 (2.2, 6.3)
Carcinomas*			
15-29 years	71 (67, 75)	1.1 (0.9, 1.3)	4.7 (3.8, 5.7)

*based on Birch classification for TYA only

5.4.2 Trends in cure by period of diagnosis

Table 5.6 shows the cure model estimates by diagnostic group and period of diagnosis. For several diagnostic groups the percentage cured increased over time including leukaemia, lymphoma, CNS tumours, neuroblastoma (children only), and other epithelial tumours (TYA only) (Table 5.6). The percentage cured for germ cell tumours was high (>90%) across the whole time period and remained similar. The percentage cured for soft tissue sarcoma and renal tumours did not change over time. For all diagnostic groups there was little change in the survival of the uncured over time. For lymphoma patients the median survival of the uncured increased from 1.2 years (95% CI 0.9 to 1.6) for those diagnosed 1990-1996 to 2.0 years (95% CI 1.3 to 2.7) in 2004-2011 but this increase was not statistically significant.

Table 5.6: Cure model results by period of diagnosis and diagnostic group

Group	Percentage cured (95% CI)	Median survival time of uncured (years) (95% CI)	Time when 90% uncured dead (years) (95% CI)
Leukaemia			
1990-1996	58 (53, 63)	1.2 (0.9, 1.6)	5.5 (4.6, 6.3)
1997-2003	68 (63, 73)	1.3 (0.9, 1.7)	5.7 (4.7, 6.6)
2004-2011	81 (77, 85)	1.2 (0.7, 1.7)	5.7 (4.6, 6.8)
Lymphoma			
1990-1996	76 (72, 81)	1.2 (0.9, 1.6)	7.4 (5.6, 9.3)
1997-2003	83 (79, 87)	1.9 (1.3, 2.5)	8.1 (6.7, 9.6)
2004-2011	89 (86, 92)	2.0 (1.3, 2.7)	8.2 (6.8, 9.7)
CNS tumours			
1990-1996	52 (46, 58)	1.6 (1.1, 2.1)	9.1 (8.1, 10.1)
1997-2003	61 (55, 67)	1.6 (1.1, 2.2)	9.4 (8.4, 10.4)
2004-2011	65 (59, 71)	1.9 (1.3, 2.5)	9.5 (8.6, 10.5)
Soft tissue sarcoma			
1990-2000	59 (52, 66)	1.6 (1.1, 1.8)	4.6 (3.6, 5.6)
2001-2011	58 (51, 65)	1.4 (1.2, 1.8)	4.3 (3.1, 5.4)
Germ cell tumours* (15-29 years only)			
1990-1996	92 (88, 95)	N/A	N/A
1997-2003	96 (93, 98)	1.9 (0.8, 3.0)	5.6 (2.9, 8.4)
2004-2011	94 (91, 96)	1.0 (0.5, 1.6)	3.8 (1.1, 6.6)
Neuroblastoma (0-14 years only)			
1990-2000	40 (29, 51)	0.9 (0.6,1.2)	2.7 (2.4, 2.9)
2001-2011	70 (58, 79)	1.1 (0.7, 1.4)	2.9 (2.6, 3.1)
Renal tumours (0-14 years only)			
1990-2000	80 (68, 88)	1.3 (0.5, 2.1)	4.3 (2.2, 6.3)
2001-2011	89 (76, 95)	1.5 (0.5, 2.6)	4.6 (2.3, 6.8)
Carcinomas* (15-29 years only)			
1990-1996	61 (51, 69)	1.0 (0.7, 1.3)	4.5 (3.5, 5.5)
1997-2003	75 (67, 81)	1.3 (0.9, 1.7)	4.9 (4.0, 5.8)
2004-2011	74 (68, 80)	1.1 (0.8, 1.4)	4.7 (3.7, 5.7)

N/A model estimates not available for these groups
* Based on Birch Classification

5.5 Trends in cure for childhood acute lymphoblastic leukaemia

This section contains results from the analysis of cure models including children (aged 1-17 years) diagnosed with ALL (full methodology included in Section 3.5.2.3). Both overall survival and event-free survival (EFS) were included as outcomes in the cure models.Figure 5.5. shows the flowchart for the study population used in this analysis, including 7 patients who were excluded as they had missing data on white cell count.

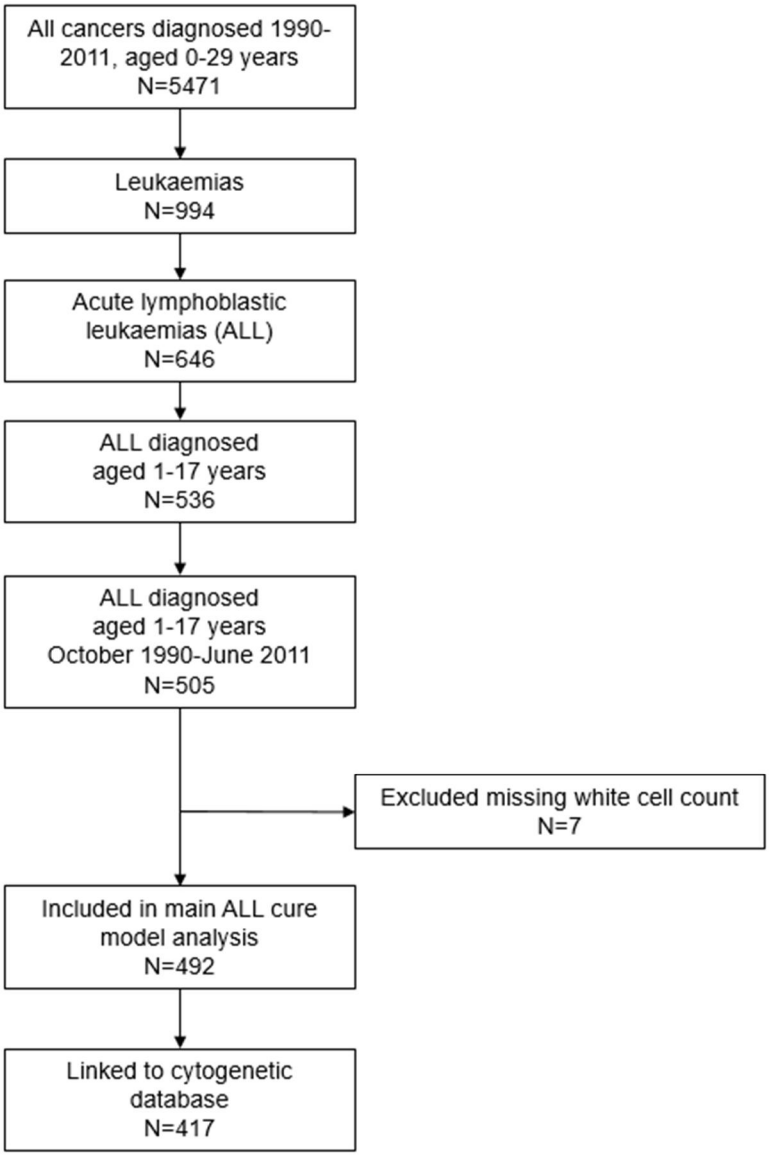


Figure 5.5: Flow chart of study sample included in ALL cure analysis

A total of 492 patients were included in the main analysis, of whom 81 (17%) died and 90 (18%) relapsed within the follow-up period (Table 5.7). The median

time to relapse was 2.5 years and among relapsing patients 53% (n=48) died during follow-up.

Table 5.7: Characteristics of ALL patients in Yorkshire aged 1-17 years

	N (% of total)	Deaths (% within each group)	Relapse (% within each group)
Full cohort	492	81 (16.5)	90 (18.3)
Diagnosis period			
Oct 1990-Mar 1997	138 (28.1)	32 (23)	53 (38)
Apr 1997-Sep 2003	163 (33.1)	27 (17)	24 (15)
Oct 2003-Jun 2011	191 (38.8)	22 (12)	13 (7)
Age group			
1-9 years	385 (78.2)	60 (16)	72 (19)
10-17 years	107 (21.8)	21 (20)	18 (17)
Sex			
Males	281 (57.1)	50 (18)	58 (21)
Females	211 (42.9)	31 (15)	32 (15)
White cell count			
$<50 \times 10^9$/L	391 (79.5)	54 (14)	67 (17)
$\geq 50 \times 10^9$/L	101 (20.5)	27 (27)	23 (23)

Figure 5.6 shows the overall survival and EFS trends over time. Five-year survival increased slightly from 86% (95% CI 79 to 91) in 1990-1997 to 89% (95% CI 84 to 93) in 2003-2011, while there was a significant increase in 5-year EFS over the same period from 62% (95% CI 53 to 69) to 86% (95% CI 81 to 91). The survival curves tended to flatten out around 8-10 year after diagnosis.

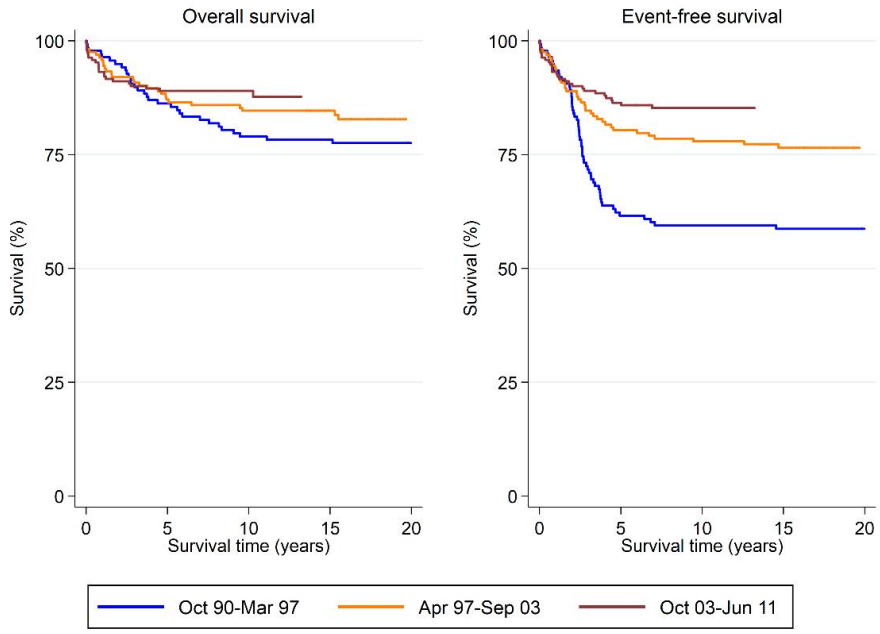


Figure 5.6: Overall survival and event-free survival by period of diagnosis, ALL patients aged 1-17 years

5.5.1 Overall survival

Results from the cure model based on overall survival are shown in Table 5.8. Results from the unadjusted and adjusted models were similar. The adjusted excess mortality rate ratio (EMRR) was 55% lower in 2003-2011 compared to 1990-1997 (Adjusted EMRR=0.45, 95% CI 0.26 to 0.80). The adjusted percentage cured increased from 77% (95% CI 70 to 84%) in 1990-1997 to 89% (95% CI 84 to 93%) in 2003-2011 and the median survival time of the uncured decreased from 3.2 years (95% CI 2.2 to 4.1) to 0.7 years (95% CI 0 to 1.5) over this time period. The adjusted EMRR was 2.4 times higher for those with a higher WCC compared to those with lower WCC (Adjusted EMRR=2.40, 95% CI 1.49 to 3.87). There were significant differences in the percentage cured by WCC, 87% (95% CI 84 to 90) for those with lower WCC and 72% (95% CI 63 to 81%) for those with higher WCC. The percentage cured was similar for those aged 1-9 years (84%, 95% CI 81, 88)) and those ages 10-17 years at diagnosis (81%, 95% CI 73 to 88) and males (83%, 95% CI 79 to 88) and females (85%, 95% CI 80 to 90). There were no differences in the MST of the uncured by age, sex or WCC.

Table 5.8: Unadjusted and adjusted cure model results for overall survival for ALL patients

	5- year survival (95% CI)	Excess mortality rate ratio (95% CI)		Cure percentage (95% CI)		Median survival time of the uncured (years) (95% CI)	
		Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
Diagnosis period							
Oct 1990-Mar 1997	86 (79, 91)	1.0 -	1.0 -	78 (70, 84)	77 (70, 84)	3.2 (2.2, 4.1)	3.2 (2.2, 4.1)
Apr 1997-Sep 2003	87 (81, 91)	0.70 (0.41, 1.19)	0.67 (0.39, 1.14)	84 (78, 89)	84 (78, 90)	2.0 (0.9, 3.2)	2.1 (0.8, 3.3)
Oct 2003-Jun 2011	89 (84, 93)	0.52 (0.30, 0.92)	0.45 (0.26, 0.80)	88 (82, 92)	89 (84, 93)	0.8 (0.1, 1.5)	0.7 (0, 1.5)
Age group							
1-9 years	89 (85, 92)	1.0 -	1.0 -	85 (80, 88)	84 (81, 88)	2.4 (1.4, 3.4)	1.7 (1.1, 2.5)
10-17 years	83 (75, 89)	1.29 (0.76, 2.17)	1.32 (0.78, 2.26)	81 (71, 87)	81 (73, 88)	2.2 (0.7, 3.6)	2.1 (0.9, 3.2)
Sex							
Males	88 (83, 91)	1.0 -	1.0 -	83 (78, 87)	83 (79, 88)	2.4 (1.4, 3.5)	1.9 (1.1, 2.7)
Females	88 (82, 92)	0.86 (0.54, 1.36)	0.90 (0.57, 1.43)	85 (79, 89)	85 (80, 90)	2.2 (0.9, 3.4)	1.8 (0.9, 2.7)
White cell count							
$<50 \times 10^9$/L	90 (87, 93)	1.0 -	1.0 -	87 (83, 90)	87 (84, 90)	1.7 (1.0, 2.4)	1.9 (1.1, 2.6)
$\geq 50 \times 10^9$/L	77 (68, 84)	2.29 (1.43, 3.68)	2.40 (1.49, 3.87)	72 (62, 80)	72 (63, 81)	1.5 (0.7, 2.2)	1.8 (1.0, 2.6)

Adjusted estimates are presented for each variable assuming that the distribution of the other variables is the same as the whole study population, therefore allowing direct comparison between groups

5.5.2 Event-free survival

Table 5.9 shows results of the EFS cure model. In these models, the percentage cured defines the group of patients free from relapse or who have not died and the uncured group contains a mixture of those who died and those who relapsed. The interpretation of the uncured group is not straightforward or clinically relevant therefore the focus of the results is on the percentage cured only.

The overall trends by risk factor are similar to the model for overall survival except that the estimates of the percentage cured are slightly lower in the EFS model. The adjusted percentage cured increased from 58% (95% CI 49 to 66) in 1990-97 to 86% (95% CI 81 to 91) in 2003-2011. There were significant differences in the percentage cured by WCC; 79% (95% CI 75 to 83) for those with lower WCC and 61% (95% CI 51 to 70%) for those with higher WCC. The percentage cured was similar for those aged 1-9 years and those aged 10-17 years and similar for males and females.

Table 5.9: Unadjusted and adjusted cure model results for event-free survival for ALL patients

	5-year event-free survival (95% CI)	**Cure percentage (95% CI)**	
		Unadjusted	**Adjusted**
Diagnosis period			
Oct 1990-Mar 1997	62 (53, 69)	59 (50, 67)	58 (49, 66)
Apr 1997-Sep 2003	80 (73, 86)	77 (70, 83)	77 (71, 83)
Oct 2003-Jun 2011	86 (81, 91)	85 (79, 90)	86 (81, 91)
Age group			
1-9 years	78 (73, 82)	76 (71, 80)	76 (72, 80)
10-17 years	77 (67, 84)	74 (65, 82)	73 (65, 81)
Sex			
Males	75 (70, 80)	73 (68, 78)	74 (69, 78)
Females	80 (74, 85)	78 (71, 83)	77 (72, 83)
White cell count			
$<50 \times 10^9$/L	81 (76, 84)	78 (74, 82)	79 (75, 83)
$\geq 50 \times 10^9$/L	65 (55, 74)	63 (53, 71)	61 (51, 70)

5.5.3 Cytogenetic risk groups

Linkage of cancer registrations to cytogenetic risk group was available for 417 ALL patients. There were differences by age for those that were matched to a record or not, however, there were no differences by diagnostic period, sex or WCC (Table 5.10). Some patients after linkage, were unable to be assigned a risk group and were categorised as “unknown”, while the not linked patients were kept as a separate group. However, estimates are not calculated for groups with fewer than 50 cases, therefore estimates are only provided for the good and intermediate risk group and the not linked cases (Table 5.11).

Table 5.10: Comparison of patient characteristics for those linked and not linked to cytogenetic risk group data

Variable	Linked N=417 n (%)	Not linked N=75 n (%)	Chi-squared p-value
Diagnosis period			
Oct 1990-Mar 1997	121 (29%)	17 (23%)	0.42
Apr 1997-Sep 2003	134 (32%)	29 (39%)	
Oct 2003-Jun 2011	162 (39%)	29 (39%)	
Age group			
1-9 years	334 (80%)	51 (68%)	0.02
10-17 years	83 (20%)	24 (32%)	
Sex			
Males	237 (57%)	44 (59%)	0.77
Females	180 (43%)	31 (41%)	
White cell count			
$<50 \times 10^9$/L	333 (80%)	58 (77%)	0.62
$\geq 50 \times 10^9$/L	84 (20%)	17 (23%)	

Based on cytogenetic risk group, the percentage cured was 90% for patients in the good risk group (95% CI 84 to 94), 75% for intermediate risk group (95% CI 66 to 82), and 90% for patients not linked (95% CI 81 to 95) (Table 5.11). There were no differences in the median survival time of the uncured between the risk groups.

Table 5.11: Cure model results for overall survival by cytogenetic risk group

Cytogenetic risk group	N	5-year survival (95% CI)	Cure percentage (95% CI)	Median survival time of the uncured (years) (95% CI)
Good	183	92 (87, 95)	90 (84, 94)	1.4 (0, 3.8)
Intermediate	124	84 (76, 89)	75 (66, 82)	2.9 (1.6, 4.2)
High	24	-	-	-
T-ALL	48	-	-	-
Unknown	38	-	-	-
Not linked	75	91 (81, 95)	90 (81, 95)	1.7 (0, 4.8)

5.5.4 Risk of relapse

In addition to estimating EFS, the cumulative incidence for the risk of relapse was estimated for each diagnostic period. There was a substantial reduction in the risk of relapse over time (Figure 5.7). 5-years post diagnosis the cumulative incidence of relapse reduced from 36% (95% CI 28 to 44) for those diagnosed 1990-1997 to 6% (95% CI 3 to 10) for those diagnosed 2003-2011.

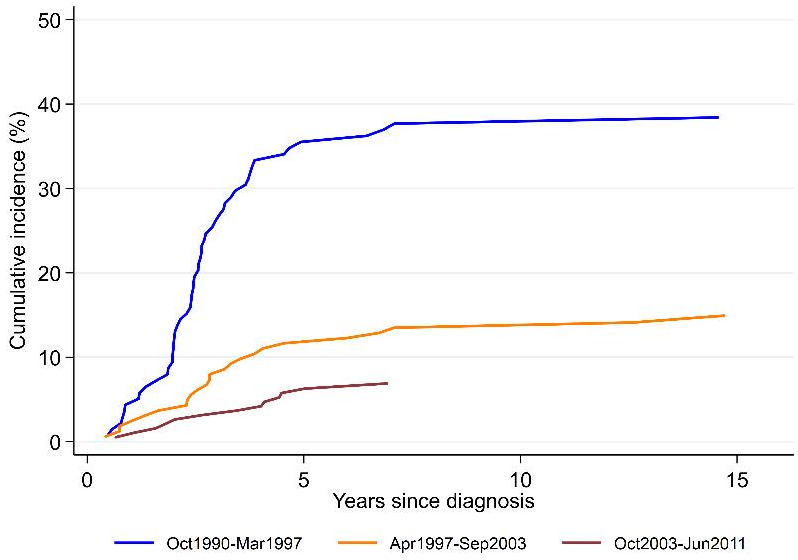


Figure 5.7: Cumulative incidence of relapse by diagnostic period

5.5.5 Checking model fit

A check of the model fit was carried out to ensure the validity of the model and the underlying cure model assumptions. Figure 5.8 shows the predicted survival curves, the proportion cured and yearly Ederer II life table estimates of survival for each diagnostic period. Model fit is adequate and the predicted cure proportion is very close to the empirical life table estimates of survival for each time period.

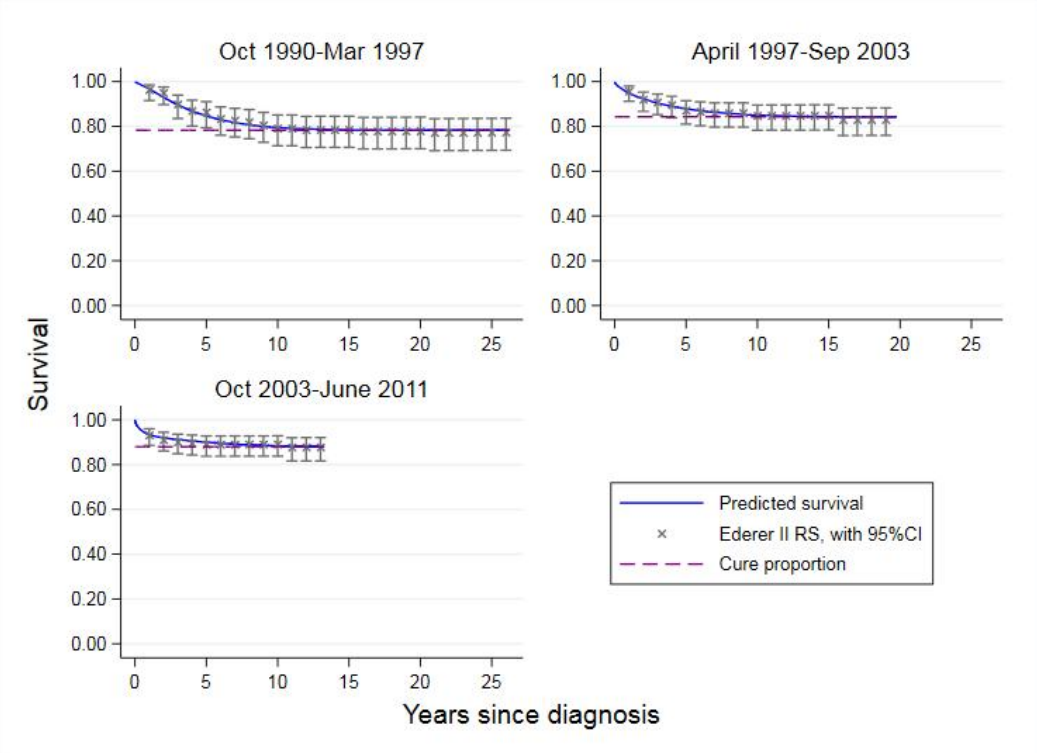


Figure 5.8: Plot of predicted relative survival from FP cure models for ALL patients and Ederer II lifetable estimates of relative survival

5.6 Discussion

5.6.1 Results in context

Results from this chapter show that it is feasible to model statistical cure for children and young people diagnosed with cancer. Comparison of results from different cure models, flexible parametric, mixture and non-mixture models, showed that the percentage cured was generally the same for all models, however, there were differences in the survival of the uncured. The two main assumptions of the cure model were satisfied: 1) that cure was a reasonable assumption and 2) that the survival of the uncured was modelled appropriately.

These assumptions were assessed graphically and through assessment of the model fit.

Model fit statistics such as the AIC and BIC, have limited use for cure models as they estimate the model fit over the whole time period and for cure models the interest is in the model fit towards the end of follow-up where data are more sparse, therefore model selection should not be based upon these criteria alone [28, 115, 116]. Based on the AIC the flexible parametric model was the best fitting, but based on the BIC the mixture or non-mixture model showed better model fit.

The survival of the uncured from the flexible parametric model more closely predicted the observed survival of this group therefore the flexible parametric model was chosen to use in further modelling in this chapter. This model is also recommended over the mixture and non-mixture models when survival is relatively high [28], as is the case for some diagnostic groups for children and young people with cancer.

Plots of survival from the cure models compared to lifetable estimates of survival also showed good agreement providing further evidence that the cure model assumptions were satisfied for this cohort.

The percentage cured is a useful measure of long-term survival and may be more informative for communicating prognosis to patients rather than focussing on the benchmark of 5-year survival. For all cancers combined, the percentage cured was 73% for children and 75% for TYA diagnosed between 1990 and 2011 and there was a significant increase over time in the percentage cured from 66% for those diagnosed 1990-1996 to 80% for those diagnosed 2004-2011, while over this time period the median survival time of the uncured group remained the same. There is great variation in survival by cancer type therefore cure models by diagnostic group are clinically informative. Analysis by age group and time period showed that the percentage cured ranged from 40% for children diagnosed with neuroblastoma in 1990-2000 to 95% for TYA with germ cell tumours diagnosed 1997-2003. There were some issues in calculating the median survival time of the uncured for germ cell tumours, which may be due to the high survival and limited number of deaths within this group.

Generally for most diagnostic groups there was in improvement in survival over time which was mainly driven by an increase in the percentage of patients cured rather than an increase in the survival of the uncured. There was no change in the percentage cured for germ cell tumours, which was very high in all time periods, or for soft tissue sarcomas which remained around 60%.

Due to the small sample size it was not possible to look in detail at further associations been clinical risk factors and cure for all diagnostic groups. However, a detailed analysis of children diagnosed with ALL was carried out. There are also clearly defined clinical risk stratification factors for childhood ALL and data were available on these for inclusion in statistical models.

An increase in the percentage of patients diagnosed with childhood ALL who have been cured with more contemporary therapeutic approaches was observed. However, there remained a relatively small group of patients where treatment was unsuccessful and whose survival was relatively short; the median survival time of the uncured diagnosed in most recent time period was around 1 year. The survival trends of patients who are not long-term survivors (the uncured) have not been described before, and the interpretation of trends in the survival of the uncured is difficult. Improvements in risk stratification and minimal residual disease monitoring [253] will have led to more patients moving to the cured group, leaving the most chemo-resistant patients in the uncured group. Due to the high proportion of patients cured these estimates are based on a relatively small sample size and should be interpreted with caution. Key prognostic post-relapse factors are duration of first remission, site of relapse and genetic subgroup [228]. This small group of uncured patients may contain a heterogeneous group in terms of molecular genetics and further investigation and examination of in this group is needed.

Trends in EFS were also estimated and showed a similar pattern of trend, however, there was a greater increase in the percentage cured based on EFS which was mainly due to a significant reduction in the risk of relapse over time. Population-based estimates of EFS for ALL patients have not previously been reported, mainly due to lack of routinely collected data on relapse. The estimates of 5-year population-based EFS for ALL patients are similar to those reported in national clinical trials: between October 1990 and March 1997 estimated 5-year EFS was 62% compared with 63% reported in the UKALLXI study [254]; between April 1997 and September 2003 estimated 5-year EFS

was 80%, compared to 74% for ALL97 study and 80% for ALL97/99 study [229, 256]; and between October 2003 and June 2011 estimated 5-year EFS was 86% compared to 87% reported by the UKALL2003 study [253]. Similarly the UKALL2003 study found the 5-year cumulative incidence of relapse of 9% [253] compared to our findings of 6% during the same time period although those aged 18-24 years were not included in this analysis but they were included in UKALL2003. These findings provide evidence of the validity of these estimates and completeness of the ascertainment of relapse data for the population-based YSRCCYP and the potential to use routine cancer registry data to estimate long-term relapse incidence and event-free survival.

For patients diagnosed between 2003 and 2011, the 5-year survival estimate was very similar to the percentage of patients cured. The proportion cured for childhood ALL has been increasing since the 1970s reflecting major improvements in survival during this time [29, 30]. This increasing trend continued including patients diagnosed up to 2011, however the rate of increase may have slowed down; between 1997-2003 and 2003-2011 the percentage cured increased from 84% to 89%. This is consistent with population-based survival trends reported by clinical trial era [283, 284].

Cytogenetic information is important not only for predicting survival but also to identify patients at increased risk of relapse and those less likely to respond to treatment after relapse [228]. Cancer registries do not routinely collect this information, so this is a unique feature of this analysis and a major strength, although there may have been changes to cytogenetic information available over time. 5-year overall survival for those in the good risk group in patients in the ALL97/99 trial was 94%, and in this study the estimated percentage cured in this risk group to be 91% providing valuable information on the long-term survival for this group of patients.

5.6.2 Strengths and limitations

Previous studies have shown that flexible parametric cure models are generally robust to the number and position of knots [115]. For all cancers combined, the optimal flexible parametric cure model was chosen by comparing different models with different knot points (comparing models with between 3 and 7 knot points). Results showed that all models produced similar results in concordance with previous findings.

As well as reporting the percentage cured, results were presented on the survival of the uncured including the median survival time and the time when 90% of the uncured had died. Although the latter measure uses an arbitrary cut point it is a useful summary statistic when assessing cure models, but this measure did show the largest discrepancy when comparing flexible parametric, mixture and non-mixture cure models, as shown in Figure 5.4, where subtle differences in the survival curves of the uncured had larger implications on this outcome. Given the high survival rates observed these estimates are based on a small number of deaths and therefore replication of these findings in studies with larger samples is needed. A further limitation of this study is that when stratifying by diagnostic group, estimates by age group and time period may be based on small numbers and the survival of the uncured could not be estimated for TYAs with germ cell tumours, again mainly due to the high survival and small number of deaths within this group.

As indicated in Chapter 2, a key gap in the literature was a lack of population-based studies of cure in children and TYA cancers incorporating clinical risk factors, this was a key strength of the analysis presented here in particular for ALL patients where cure was estimated including age, sex and white cell count. Cytogenetic risk group was also included for those with available data. Cure was only estimated for cytogenetic risk groups with at least 50 cases, which meant that cure was only estimated for the good and intermediate risk groups. It was also estimated for the not linked group which included 15% of ALL patients. There were no differences in those cases without and without cytogenetic data apart from by age group with a higher rate of not linked cases in the older age group. Those not linked were most likely to be those not enrolled on clinical trials and difference in ALL trial recruitment rates by age have been reported; estimated between 85-99% for children [283] and 66-77% for those aged 15-17 years [285]. While recruitment rates for ALL trials are high they are not population-based. The estimated percentage cured for the not linked cases was similar to the good risk group, indicating generally good prognosis for this subgroup. Again a larger sample size may have enabled cure to be estimated for all cytogenetic risk groups.

All cases included in analysis had at least 5-years follow-up data but for some diagnostic groups a longer follow-up period may be needed to estimate cure robustly. The data showed that for ALL the survival curves tended to flatten out

after 8-10 years follow-up but there may remain some excess mortality after this, suggesting that, particularly for more recently diagnosed patients, a longer follow-up period may be needed.

5.6.3 Summary

This chapter used population-based data to estimate cure for children and TYA cancers incorporating clinical risk factors. Estimates of the proportion cured and the median survival time of the uncured by tumour type, over time and for children and TYAs were calculated. Generally for most diagnostic groups there was an improvement in survival over time which was mainly driven by an increase in the proportion of patients cured rather than an increase in the survival of the uncured. Statistical cure is measured at the population level and does not provide information on individual level cure. Overall survival and even event-free survival do not measure quality of survival or account for the late effects of earlier cancer treatment. In the next chapter long-term health outcomes and late effects of treatment are evaluated using linkage of cancer registrations to hospital admissions to investigate respiratory morbidity in long-term survivors.

Chapter 6 Respiratory late effects

6.1 Introduction

Respiratory conditions are one of the most common causes of late morbidity and mortality for long-term survivors of childhood cancer. Previous studies have described respiratory late effects based upon self-reported outcomes or hospital admissions for all respiratory conditions combined. However only one study (based in Scandinavia) has considered admissions for specific respiratory conditions [203]. This chapter provides a comprehensive analysis of respiratory morbidity in long-term survivors based on linked cancer registrations to hospital admissions to address Aim 2a of this thesis and addresses a key gap in the current knowledge identified in Chapter 2. Full details of the statistical methods are provided in Chapter 3 (Section 3.5.5). A description of admission patterns in the cancer survivor cohort is provided along with the cumulative incidence. Admission rates in cancer survivors were compared to rates in the general population using hospitalisation rate ratios. The association between treatment exposures and admission was investigated using competing risks regression models. Finally a description of trends in readmissions and subsequent mortality is provided.

6.2 Description of respiratory admissions

The analysis presented in this chapter is based upon 5-year survivors only. A total of 4235 individuals were included of whom 667 (15.7%) were admitted to hospital at least once for a respiratory condition.

Table 6.1 shows the patient characteristics by admission status. Younger patients, females, those diagnosed in the earlier time period (1990-1996) and those who received radiotherapy to the chest had higher rates of admission.

Table 6.1: Patient characteristics overall and by respiratory admission

	All (N=4235)		Respiratory admission (N=667)		No respiratory admission (N=3568)	
Characteristic	**n**	**%**	**n**	**%**	**n**	**%**
Age group (at diagnosis)						
0-4 years	756	17.8	144	21.6	612	17.2
5-9 years	426	10.1	73	10.9	353	9.9
10-14 years	426	10.1	77	11.5	349	9.8
15-19 years	623	14.7	97	14.5	526	14.7
20-24 years	838	19.8	110	16.5	728	20.4
25-29 years	1166	27.5	166	24.9	1000	28.0
Sex						
Males	2555	60.3	332	49.8	2223	62.3
Females	1680	39.7	335	50.2	1345	37.7
Period of diagnosis						
1990-1996	1185	28.0	253	37.9	932	26.1
1997-2003	1309	30.9	241	36.1	1068	29.9
2004-2011	1741	41.1	173	25.9	1568	44.0
Deprivation quintile						
1 (least deprived)	252	6.0	30	4.5	222	6.2
2	515	12.2	79	11.8	436	12.2
3	857	20.2	130	19.5	727	20.4
4	893	21.1	151	22.6	742	20.8
5 (most deprived)	1718	40.6	277	41.5	1441	40.4
Ethnicity						
White	3666	86.6	599	89.8	3067	87.9
South Asian	324	7.7	51	7.7	273	7.8
Other	165	3.9	17	2.6	148	4.2
Missing	80	1.9				
Pulmonary toxic chemotherapy	1342	31.7	213	31.9	1129	31.6
Radiotherapy to chest	139	4.0	40	6.0	129	3.6
Thoracic surgery	28	0.7	6	0.9	22	0.6

There were different admission patterns by diagnostic group (Figure 6.1). Overall 18% of children were admitted at least once and this ranged from 9% for neuroblastoma to 23% for renal tumours. For TYA, overall 14% were admitted which ranged from 9% for germ cell tumours to 23% for bone tumours.

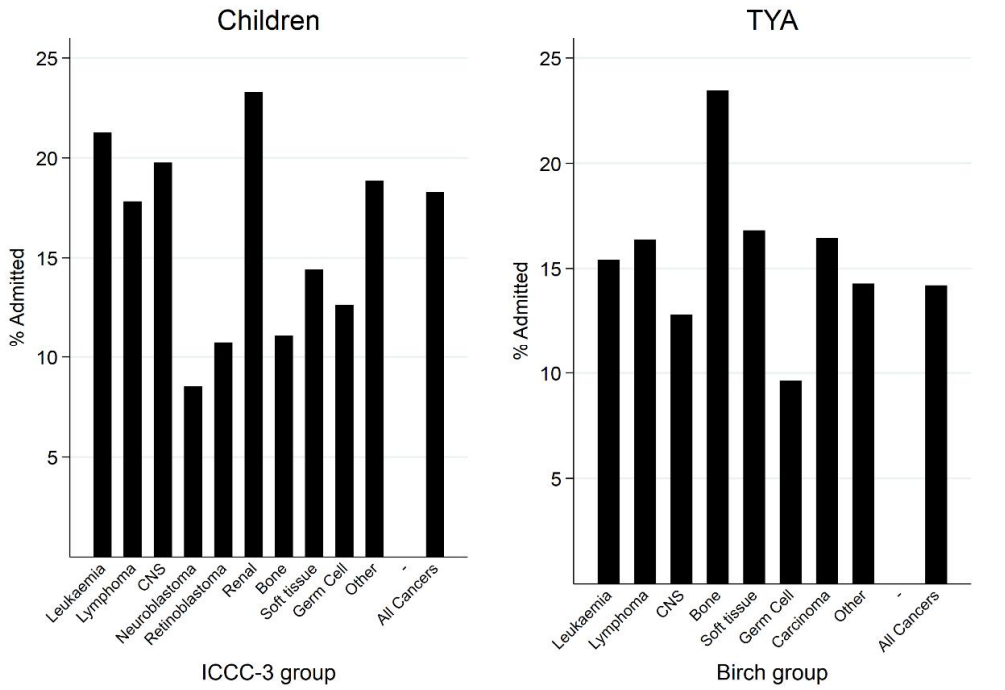


Figure 6.1: Percentage of survivors admitted for a respiratory condition by diagnostic group and age group

6.3 Comparison with general population

The observed and expected number of admissions for each respiratory condition and the corresponding hospitalisation rate ratio (HRR) are shown in Table 6.2 for the whole cohort and by age group at diagnosis. For all respiratory conditions the risk of hospitalisation was 1.86 (95% CI 1.73 to 2.01) times higher in cancer survivors compared to the general population. For each respiratory condition the excess risk was significantly higher and was highest for respiratory conditions due to external agents (HRR=162, 95% CI 73 to 360) and lung fibrosis (HRR=13, 95% CI 6.5 to 26). However, these two outcomes were based on fewer than 10 observed cases and wide variations in the estimated confidence intervals. The HRR was 3.9 (95% CI 3.27 to 4.59) for pneumonia, 3.6 (95% CI 2.70 to 4.78) for lower respiratory conditions and 1.5 (95% CI 1.34 to 1.69) for asthma.

The HRR was similar for children and TYA for any respiratory conditions (2.05, 95% CI 1.83 to 2.30 and 1.74, 95% CI 1.57 to 1.93, respectively) and asthma (1.54, 95% CI 1.28 to 1.85 for children and 1.49, 95% CI 1.29 to 1.73 for TYA). However, for pneumonia and chronic lower respiratory diseases the HRR was higher for children compared to TYA; 6.52 (95% CI 5.09 to 8.35) and 2.85 (95% CI 2.26 to 3.60), respectively for pneumonia and 9.40 (95% CI 5.85 to 15.1) and 2.66 (95% CI 1.86 to3.80) respectively for chronic lower respiratory diseases.

Table 6.2: Observed and expected respiratory admissions and hospitalisation rate ratio (HRR) by type of respiratory admission and age group

	Observed admissions	Expected admissions	HRR (95% CI)
All ages (0-29 years)			
Any respiratory admission	667	358	1.86 (1.73, 2.01)
Asthma	289	192	1.51 (1.34, 1.69)
Pneumonia	134	35	3.87 (3.27, 4.59)
Chronic lower respiratory disease	47	13	3.59 (2.70, 4.78)
Lung fibrosis	8	1	13.1 (6.5, 26.1)
Respiratory conditions due to other external agents	6	0.04	162 (73, 360)
Children (0-14 years)			
Any respiratory admission	294	143	2.05 (1.83, 2.30)
Asthma	115	75	1.54 (1.28, 1.85)
Pneumonia	63	10	6.52 (5.09, 8.35)
Chronic lower respiratory disease	17	2	9.40 (5.85, 15.1)
TYA (15-29 years)			
Any respiratory admission	373	214	1.74 (1.57, 1.93)
Asthma	174	117	1.49 (1.29, 1.73)
Pneumonia	71	25	2.85 (2.26, 3.60)
Chronic lower respiratory disease	30	11	2.66 (1.86, 3.80)

For children and TYA HRR not calculated for lung fibrosis or respiratory conditions due to other external agents due to small numbers

Stratifying by diagnostic group (Table 6.3) the excess risk for all respiratory conditions was higher in cancer survivors compared to the general population for all age groups for leukaemia, lymphoma and CNS tumours. For germ cell tumours for all ages and for TYA only there was no excess risk of admission compared to the general population.

Table 6.3: Observed and expected respiratory admissions and hospitalisation rate ratio (HRR) by diagnostic group

	Observed admissions	Expected admissions	HRR (95% CI)
All ages			
Leukaemia (n=725)	142	61	2.32 (1.97, 2.74)
Lymphoma (n=950)	158	82	1.92 (1.64, 2.24)
CNS tumours (n=619)	103	50	2.05 (1.69, 2.49)
Germ cell tumours (n=869)	88	73	1.21 (0.98, 1.49)
Children (0-14 years)			
Leukaemia (n=522)	111	46	2.40 (1.99, 2.89)
Lymphoma (n=202)	36	18	2.06 (1.48, 2.85)
CNS tumours (n=344)	68	31	2.22 (1.75, 2.81)
TYA (15-29 years)			
Leukaemia (n=201)	31	15	2.09 (1.47, 2.98)
Lymphoma (n=745)	122	65	1.88 (1.57, 2.25)
CNS tumours(n=273)	35	20	1.78 (1.28, 2.49)
Germ cell tumours (n=779)	75	65	1.16 (0.92, 1.45)
Carcinomas (n=383)	63	30	2.11 (1.65, 2.71)

For all ages and children diagnosis based on ICCC-3 classification, for TYA diagnosis based on Birch classification

Figure 6.2 compares the age at first admission for each respiratory condition in the cancer survivor cohort and the general population. For any respiratory admission the median age was similar in both groups: 27 years for cancer survivors and 25 years in the general population. The median age at admission was lower for cancer survivors compared to the general population for pneumonia (27 year and 31 year respectively) and chronic lower respiratory conditions (35 years and 43 years respectively).

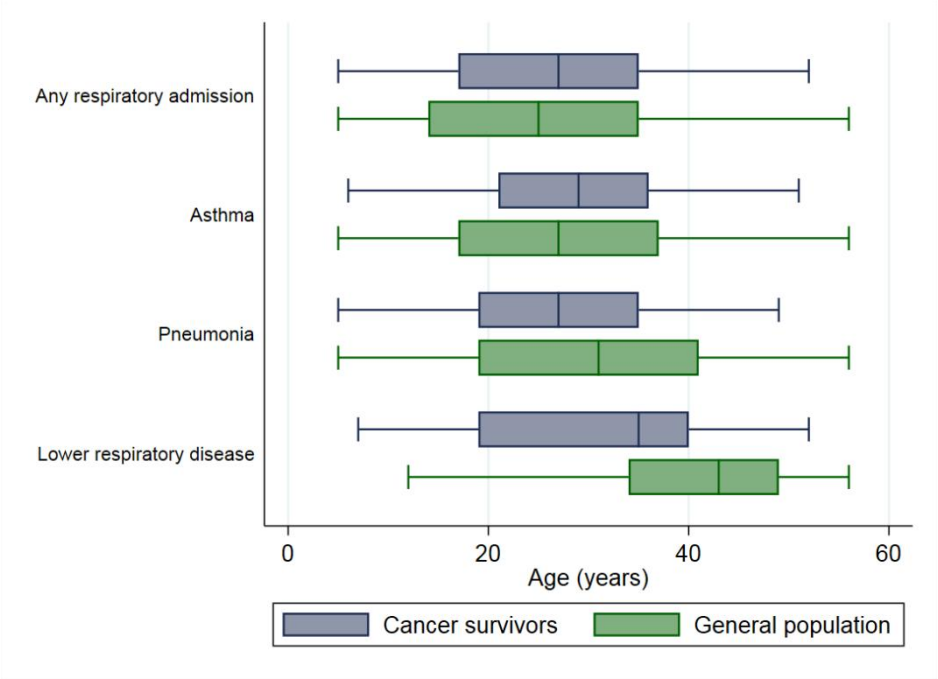


Figure 6.2: Box plot of age at first admission in cancer survivors compared to the general population

6.3.1 Comparison based on primary diagnosis only

A sensitivity analysis was carried out to estimate admission trends based on the primary diagnostic code each HES admission only with the results shown in Table 6.4. Based on primary admission only 352 admissions were observed, compared to 667 when all diagnostic codes within each admission were used. However when comparing to the admission rates in the general population the HRR was similar (1.82, 95% CI 1.64 to 2.02). The biggest difference in admissions were for asthma where based on primary diagnoses only 22 admissions were observed (compared to 289 admissions from analysis based on all diagnostic codes) giving a non-significant reduction in admissions compared to the general population (HRR=0.89, 95% CI 0.59 to 1.35). The number of observed admissions for pneumonia and chronic lower respiratory disease were lower when based on primary admission only, 96 for pneumonia compared to 134 based on all diagnostic codes and 13 for chronic lower respiratory conditions compared to 47 based on all diagnostic codes. However, the HRRs were similar to those based on all diagnostic codes: HRR= 3.82, 95% CI 3.12 to 4.66 for pneumonia and HRR=3.26, 95% CI 1.89 to 5.61. This trend was observed across both age groups.

Table 6.4: Observed and expected number of respiratory admissions and hospitalisation rate ratio (HRR) based on primary diagnosis for admissions only

	Observed admissions	Expected admissions	HRR (95% CI)
All ages (0-29 years)			
Any respiratory admission	352	194	1.82 (1.64, 2.02)
Asthma	22	25	0.89 (0.59, 1.35)
Pneumonia	96	25	3.82 (3.12, 4.66)
Chronic lower respiratory disease	13	4	3.26 (1.89, 5.61)
Children (0-14 years)			
Any respiratory admission	173	87	1.99 (1.71, 2.31)
Asthma	12	12	0.99 (0.56, 1.75)
Pneumonia	43	7	5.80 (4.30, 7.81)
Chronic lower respiratory disease	8	1	11.1 (5.54, 22.2)
TYA (15-29 years)			
Any respiratory admission	179	107	1.68 (1.45, 1.94)
Asthma	10	13	0.79 (0.42, 1.47)
Pneumonia	53	18	2.99 (2.28, 3.91)
Chronic lower respiratory disease	5	3	1.53 (0.64, 3.68)

HRR not calculated for lung fibrosis or respiratory conditions due to other external agents due to small numbers

6.4 Cumulative incidence

The cumulative incidence for admissions for any respiratory disease increased with increasing age without reaching at plateau (Figure 6.3). By age 40, the cumulative incidence for an admission for any type of respiratory condition was 49.3% (95% CI 44.6 to 53.7), asthma was 20.2% (95% CI 17.6 to 23.0), pneumonia was 13.2% (95% CI 8.2 to 19.5) and lower respiratory disease was 3.2% (95% CI 2.1 to 4.6). The cumulative incidence for lung fibrosis and conditions due to external agents at age 40 were less than 1%.

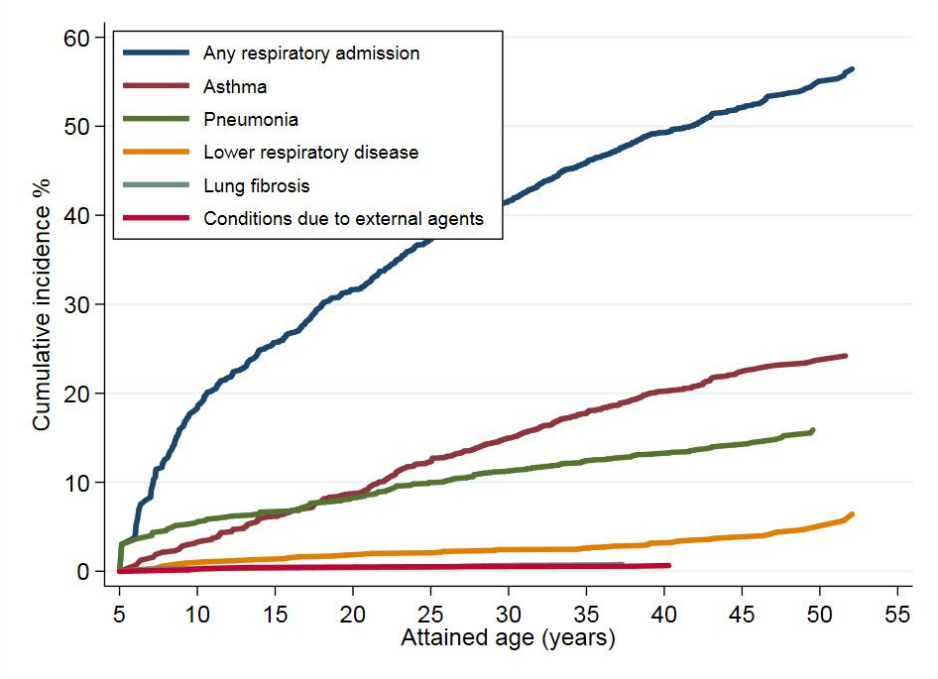


Figure 6.3: Cumulative incidence of respiratory admission by attained age

Sensitivity analysis was conducted to calculate the cumulative incidence based on primary admission only with the results shown in Figure 6.4. The cumulative incidence continued to increase without reaching a plateau. For all respiratory conditions the cumulative incidence by age 40 was 32.9% (95% CI 27.5 to 38.3), asthma was 2.4% (95% CI 1.2 to 4.4), pneumonia was 10.2% (95% CI 5.3 to 16.8) and chronic lower respiratory disease was 1.3% (95% CI 0.6 to 2.4). Compared to Figure 6.3, the biggest difference in cumulative incidence was observed for asthma.

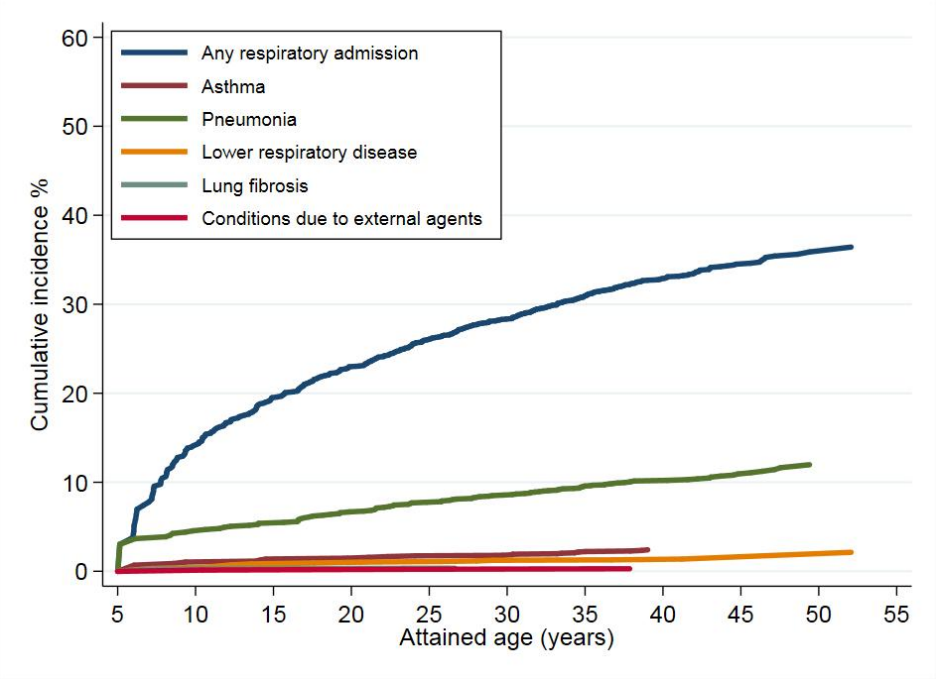


Figure 6.4: Cumulative incidence of respiratory admission by attained age based on primary diagnoses only

6.5 Association between treatment exposure and risk of admission

The association between each treatment exposure (pulmonary toxic chemotherapy, chest radiation and thoracic surgery) and risk of admission were investigated using competing risks regression models where death prior to admission was considered a competing risk. Results are presented for unadjusted and adjusted models in Table 6.5. For any respiratory admission after adjustment for potential confounders, those who received pulmonary toxic chemotherapy had an increased risk of admission (sHR=1.24, 95%C 1.02 to 1.51). There was no statistically significant association between receiving radiation to the chest (sHR=1.25, 95% CI 0.87 to 1.78) or thoracic surgery (sHR=1.38, 95% CI 0.59 to 3.24) and risk of admission.

Table 6.5: Association between treatment exposure and risk of admission for any respiratory disease, subdistribution Hazard Ratio (sHR) and 95% CI

Treatment exposure	Unadjusted sHR (95% CI)	Adjusted† sHR (95% CI)
Pulmonary toxic chemotherapy		
No	1.0 -	1.0 -
Yes	1.13 (0.96, 1.33)	1.24 (1.02, 1.51)
Chest radiotherapy		
No	1.0 -	1.0 -
Yes	1.23 (0.89, 1.71)	1.25 (0.87, 1.78)
Thoracic surgery		
No	1.0 -	1.0 -
Yes	1.17 (0.51, 2.68)	1.38 (0.59, 3.24)

† Adjusted for deprivation, diagnosis age, diagnosis year, diagnostic group, ethnicity and treatment exposures
Models exclude 80 individuals with missing ethnicity

Table 6.6 shows the associations between treatment exposures and admissions for asthma, pneumonia and chronic lower respiratory disease. For asthma there were no significant associations between any of the treatment exposures and admission, in both unadjusted and adjusted models. For pneumonia, exposure to pulmonary toxic chemotherapy was associated with an increased risk of admission (adjusted sHR=1.47, 95% CI 1.00 to 2.15), but there was no significantly increased risk for the other treatment exposures. For chronic lower respiratory disease the risk of admission increased for those who received thoracic surgery (adjusted sHR=8.74, 95% CI 2.61 to 29.3).

Table 6.6: Association between treatment exposure and risk of admission for asthma, pneumonia and chronic respiratory diseases, subdistribution Hazard Ratio (sHR) and 95% CI

	Asthma		Pneumonia		Chronic lower respiratory disease	
Treatment exposure	**Unadjusted sHR (95% CI)**	**Adjusted† sHR (95% CI)**	**Unadjusted sHR (95% CI)**	**Adjusted† sHR (95% CI)**	**Unadjusted sHR (95% CI)**	**Adjusted† sHR (95% CI)**
Pulmonary toxic chemotherapy						
No	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -
Yes	1.00 (0.77, 1.28)	0.84 (0.64, 1.12)	1.64 (1.16, 2.32)	1.47 (1.00, 2.15)	1.43 (0.80, 2.57)	0.98 (0.50, 1.92)
Chest radiotherapy						
No	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -
Yes	1.18 (0.72, 1.94)	1.22 (0.71, 2.10)	1.66 (0.90, 3.07)	1.24 (0.63, 2.41)	2.01 (0.79, 5.12)	1.31 (0.42, 4.08)
Thoracic surgery						
No	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -
Yes	0.86 (0.20, 3.58)	0.98 (0.23, 4.11)	1.86 (0.46, 7.51)	2.26 (0.57, 9.05)	7.94 (2.50, 25.3)	8.74 (2.61, 29.3)

† Adjusted for deprivation, diagnosis age, diagnosis year, diagnostic group, ethnicity and treatment exposures
Models exclude 80 individuals with missing ethnicity

Models were fitted including a two-way interaction between age group and each treatment exposure, the result are shown in Table 6.7. An interaction between thoracic surgery and age was not included due to small numbers of those receiving surgery. There was no significant interaction between pulmonary toxic chemotherapy and age for any of the outcomes. The association between chest radiation and age was statistically significant for any respiratory admission, asthma and chronic lower respiratory disease. For each of these outcomes the risk of admission was significantly higher for children who received radiation to the chest but not for TYA.

Table 6.7: Association between treatment exposure and risk of admission for respiratory disease including interaction with age group, adjusted subdistribution Hazard Ratio (sHR) and 95% CI

Outcome	Treatment exposure	Children sHR† (95% CI)	TYA sHR † (95% CI)	Interaction p-value
Any respiratory admission	**Pulmonary toxic chemotherapy**			
	No	1.0 -	1.0 -	0.43
	Yes	1.34 (1.01, 1.78)	1.17 (0.91, 1.49)	
	Chest radiotherapy			
	No	1.0 -	1.0 -	0.003
	Yes	3.32 (1.68, 6.57)	1.01 (0.68, 1.51)	
Asthma	**Pulmonary toxic chemotherapy**			
	No	1.0 -	1.0 -	0.51
	Yes	1.08 (0.69, 1.70)	0.90 (0.62, 1.30)	
	Chest radiotherapy			
	No	1.0 -	1.0 -	0.05
	Yes	3.20 (1.14, 8.99)	0.98 (0.53, 1.78)	
Pneumonia	**Pulmonary toxic chemotherapy**			
	No	1.0 -	1.0 -	0.76
	Yes	1.88 (1.08, 3.27)	1.67 (0.96, 2.93)	
	Chest radiotherapy			
	No	1.0 -	1.0 -	
	Yes	0.88 (0.12, 6.66)	1.25 (0.61, 2.57)	0.75
Chronic lower respiratory disease	**Pulmonary toxic chemotherapy**			
	No	1.0 -	1.0 -	
	Yes	2.03 (0.73, 5.60)	0.59 (0.26, 1.32)	0.06
	Chest radiotherapy			
	No	1.0 -	1.0 -	
	Yes	15.7 (3.6, 67.6)	0.49 (0.11, 2.26)	0.001

† sHRs from adjusted interaction model, adjusting for deprivation, diagnosis year, diagnostic group, ethnicity and treatment exposures

6.6 Checking proportionality assumption

The Fine and Gray competing risk regression models are based upon the proportional hazards assumption. This assumption was checked by fitting models including each variable in turn as a time varying coefficient, based on the model for any respiratory conditions. For categorical variables, this was testing by including dummy variables for each category. Table 6.8 shows the coefficients, 95% confidence intervals and p-values for each variable when included as a time varying coefficient in the model for any respiratory admission. There was no evidence that the proportional hazards assumption were violated for any of the variables included in the models.

Table 6.8: Estimated subdistribution hazard ratio (sHR) including each variable as time varying coefficient in model for any respiratory admission

Variable/category	sHR for time varying coefficient	(95% CI)	P
Pulmonary toxic chemotherapy	0.98	(0.95, 1.01)	0.27
Chest radiation	0.95	(0.90, 1.01)	0.11
Thoracic surgery	0.93	(0.77, 1.13)	0.46
Deprivation quintile 2	1.02	(0.94, 1.10)	0.61
Deprivation quintile 3	1.00	(0.93, 1.07)	0.96
Deprivation quintile 4	0.99	(0.92, 1.06)	0.74
Deprivation quintile 5	0.98	(0.91, 1.05)	0.58
Age at diagnosis	1.00	(1.00, 1.00)	0.17
Year of diagnosis	1.00	(1.00, 1.00)	0.88
Lymphoma	1.01	(0.96, 1.05)	0.77
CNS tumours	0.98	(0.93, 1.04)	0.51
Neuroblastoma	0.94	(0.84, 1.06)	0.32
Bone tumours	0.99	(0.91, 1.07)	0.74
Soft tissue sarcoma	1.02	(0.95, 1.09)	0.67
Germ cell tumours	1.04	(0.99, 1.09)	0.13
Other solid tumours	0.97	(0.92, 1.02)	0.26
South Asian ethnicity	1.02	(0.96, 1.09)	0.52
Other ethnicity	1.03	(0.93, 1.13)	0.62

6.7 Readmissions and subsequent mortality

Following the first admission for any respiratory disease trends in readmissions and subsequent mortality were described. This section of analysis is restricted to those with at least one admission for a respiratory condition (n=667), Overall 45% were readmitted for a respiratory condition at least once with 10% readmitted twice and 18% readmitted at least 3 times (Table 6.9). Focussing on those admitted for pneumonia only (n=134) 13% were readmitted for pneumonia once and 12% readmitted at least twice.

Table 6.9: Readmissions following initial hospitalisation for respiratory condition

	N (%)
Readmission for any respiratory conditions	
(after admission for any respiratory disease N=667)	
No readmission	364 (54.6)
1 readmission	116 (17.4)
2 readmissions	66 (9.9)
3+ readmissions	121 (18.4)
Readmissions for pneumonia	
(after admission for pneumonia n=134)	
No readmission	101 (75)
1 readmission	17 (13)
2+ readmissions	16 (12)

Survival following first admission was calculated from date of first admission to date of death or censoring. Overall, 109 deaths were observed in those admitted for any respiratory conditions (Table 6.10). Figure 6.5 shows the Kaplan-Meier survival curve for those who first admission was for pneumonia compared to those admitted for other respiratory conditions. One year after first admission the survival rate was lower for those admitted for pneumonia (84%, 95% CI 74 to 90) compared to those admitted for other respiratory conditions (93%, 95% CI 91 to 95). After adjustment for potential confounders, the risk of death doubled for those whose first admission was for pneumonia compared to those whose first admission was for another respiratory disease (HR= 2.00, 95% CI 1.24 to 3.23).

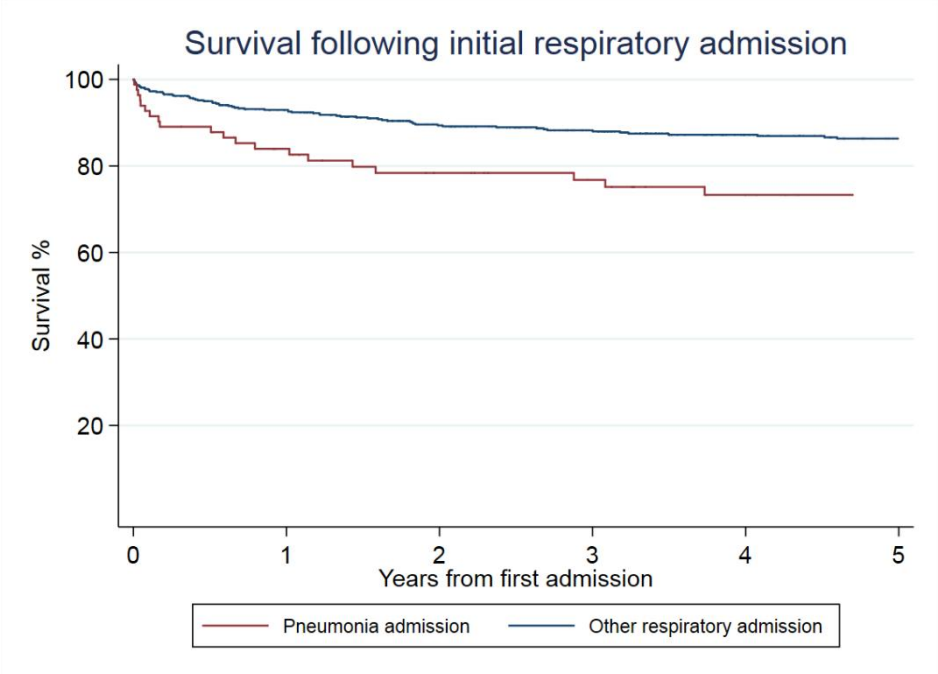


Figure 6.5: Kaplan-Meier survival following first respiratory admission type

Table 6.10: Mortality following first admission for respiratory disease for long-term survivors

First admission	N	Subsequent deaths (%)	1 year survival (%) † (95% CI)	Adjusted HR †† (95% CI)
All	667	109 (16%)	92 (89, 94)	-
Pneumonia	86	24 (28%)	84 (74, 90)	2.00 (1.24, 3.23)
Other respiratory conditions	581	85 (15%)	93 (91, 95)	1.0 -

† Survival measured from date of first admission 5-years post diagnosis
†† Adjusted for deprivation, diagnosis age, diagnosis year, diagnostic group, ethnicity, and treatment exposures

6.8 Discussion

6.8.1 Results in context

Long-term survivors for childhood and young adult cancers were hospitalised for respiratory conditions twice as often as general population controls. This excess risk varied by respiratory disease type and was greater for those diagnosed in childhood aged 0-14 compared to 15-29 year olds. The cumulative incidence of admissions continued to increase throughout life, reaching 50% by

age 40. Pulmonary toxic chemotherapy was associated with an increased risk of admission and in particular admissions for pneumonia, while radiation to the chest increased the risk of admission in children but not for TYA. Long-term survivors admitted for pneumonia had an increased risk of subsequent death following admission compared to those first admitted for other types of respiratory disease.

Linked hospital admissions were used as an objective measure of disease burden, compared to other studies of long-term survivors which rely upon self-reported outcomes. However, there are issues and complexities associated with this approach. Within each hospital admission the primary reason for admission is recorded along with (up to 19) secondary diagnostic codes mainly representing co-morbidities. The cumulative incidence of respiratory admission by age 40 based on all diagnostic codes was 50% compared to 33% when based on the primary admission only, the largest difference being in asthma admissions (20% falling to 2%), with no excess risk in hospitalisation compared to the general population. This implies that while asthma is a common co-morbidity in long-term cancer survivors it is not the main reason for subsequent hospitalisation. This is consistent with asthma more generally in the UK where the majority of asthma patients have mild disease mainly treated within primary care without experiencing exacerbation requiring hospitalisation [286].

Based on the primary admission rates only, the cumulative incidence results in this study (33% by age 40) are similar to other published studies. In the CCSS the cumulative incidence of any pulmonary condition by age 45 was 30% [210], in Switzerland the cumulative incidence of respiratory disease 35 years after cancer diagnosis was 21% [213], and the St Jude Lifetime Cohort Study found cumulative incidence by age 40 of 42% [40]. The cumulative incidence based on all diagnostic codes within each admission was 50% by age 40, which is slightly higher than those reported from other studies listed above. Direct comparison between studies is difficult and the differences in cumulative incidence may be due to different methods of event ascertainment (hospital admissions vs self-report vs clinical assessment), different time periods of recruitment or different age ranges included. Analyses presented in this chapter included those diagnosed up to 29 years whereas in other studies the upper age limit was 21 years. However, one notable difference was the cumulative incidence of lung fibrosis, the cumulative incidence of lung fibrosis was <1 % by age 40. This was substantially lower than that reported in other studies; it was

reported to be 5% in the CCSS at age 45 years [210] and 3% 35-years post diagnosis in Switzerland [213]. The lower rate observed in this cohort may be due to differences in treatments relating to better outcomes or due to coding issues associated with using routine health data. For example lung fibrosis may be not be coded within the HES record at all or may be coded as another respiratory condition. Furthermore, it is unclear from this analysis if the respiratory admissions are isolated late effects or due to complications of recurrent cancer. Relapse rates prior to first admission for respiratory conditions were slightly higher for those first admitted with pneumonia (29%) compared to those admitted for other respiratory conditions (19%).

Admissions for pneumonia and lower respiratory diseases were 3.5-4 times more likely in cancer survivors compared to the general population. In a large Scandinavian record linkage study, admissions for pneumonia were 2.8 times higher and admission for bronchitis and emphysema were double in cancer survivors compared to population controls [203]. One reason for the excess admissions in cancer survivors may be that they are more likely to be admitted for a respiratory condition, given that they have had cancer, than someone without cancer who would be managed for the same condition without a hospitalisation.

Pneumonia is the most common cause of respiratory deaths [83] and survivors with recurrent pneumonia are more likely to have limitations with daily living activities [210], therefore identifying those at greatest risk is important in order to identify preventative strategies, such as influenza and pneumococcal vaccinations. The analysis presented in this chapter identified a significant increased risk of pneumonia admissions for those treated with pulmonary toxic chemotherapy and an increased risk of subsequent mortality for those admitted with pneumonia compared to admissions for other respiratory conditions.

Higher excess risks of admission for those diagnosed in childhood compared to those diagnosed with cancer between 15-29 years for pneumonia and lower respiratory conditions were found which supports recent findings showing that children have a greater respiratory mortality (SMR of 6.8) than those aged 15 to 39 years at diagnosis (SMR 1.7) with differences in mortality from pneumonia evident (SMR 8.2 in children and 2.1 in AYA) [83]. The association between receiving pulmonary toxic chemotherapy and the risk of admission was similar in both age groups, however, there was a significant association between

radiation to the chest and hospital admissions for those diagnosed as children but not at older ages. This would appear to support previous studies that have identified those diagnosed at younger ages to be more likely to have abnormal respiratory function [207, 209, 217, 287].

6.8.2 Strengths and limitations

A key strength of this study is the inclusion of those diagnosed up to age 29 years, compared to previous studies of respiratory morbidity which only include those diagnosed up to age 21 [209-211, 213]. This also allowed the examination of differences in outcomes between children and TYA and two-way interactions between age group and treatment exposures to be included.

Another key strength of this analysis was the use of population-based data with general population controls to enable the calculation of the excess risk in long-term survivors. Treatment data were available, to enable specific treatment groups of interest to be identified, for example those who received specific chemotherapy drugs with known lung toxicity. However detailed treatment dose information was not available to allow further examination of dose-response on outcomes. Subgroup analysis with two-way interactions between age group and thoracic surgery and each outcome could not be considered due to small numbers.

Hospital admissions were used and an objective outcome to measure long-term morbidity due to respiratory conditions. This is a key strength compared to studies based on self-reported outcomes, which may suffer from recall bias, or studies based on clinical assessments, which may pick up asymptomatic respiratory conditions. However, the main limitation of this study is that hospital admission data may measure the severe end of the disease spectrum whilst many respiratory conditions, such as asthma, will be managed and treated within a primary care setting. Therefore these findings may be a potential underestimation of the true extent of respiratory disease burden.

HES data were available from 1997 onwards, and study patients who were diagnosed in 1990-1991 did not start follow-up 5-years from diagnosis but shortly afterwards (6/7 years post diagnosis) when admission data were available. Therefore for those diagnosed in the earlier time period there may be

an underestimation of admissions. However, for those diagnosed 1990-1996, 21% had at least 1 respiratory admission, compared to 10% of those diagnosed 2004-2011. Furthermore, the analysis considered time to first admission for each respiratory condition, incorporating the total person-time at risk, so this bias is likely to be small. For those admitted for at least one respiratory conditions 45% were readmitted at least once.

The limitations of linkage to HES data to quantify late effects within this cohort have previously been discussed in Section 4.8.2 and are considered further in Section 9.4.1.

6.8.3 Summary

Respiratory hospitalisations are significantly increased in CYA cancer survivors compared to the general population and contribute to a substantial burden of disease in long-term survivors. Respiratory morbidity was one of three health outcomes used to evaluate long-term outcomes in survivors in the next chapter; results from the investigation of subsequent malignant neoplasms are reported.

Chapter 7 Subsequent malignant neoplasms

7.1 Introduction

The leading causes of morbidity and mortality in long-term cancer survivors are subsequent malignant tumours, cardiovascular and respiratory diseases. In Chapter 7, respiratory morbidity was examined in terms of hospital admissions, and in Chapter 9 cardiovascular hospital admissions are investigated. The focus of this chapter is subsequent malignant neoplasms (SMN), addressing Aim 2b of this thesis. Long-term survivors of childhood and young adult cancer are at increased risk of developing a SMN in later life [33-35, 167, 168, 170, 171, 173-175, 288, 289]. This chapter describes the occurrence of subsequent tumours for patients in Yorkshire using the YSRCCYP data alongside national cancer registrations, with a particular focus on age at diagnosis and the time to develop a SMN. Many studies on SNM development include only 5-year survivors and examine the development of SMNs 5-years post diagnosis, excluding any early on-set SMNs, for example studies from the CCSS and BCCSS [33, 35, 167, 168, 175]. However, it has been estimated that up to 40% of SMNs occur in the first five years from diagnosis [170]. Studies on SMNs in TYA are more limited, however, recently it has been shown that TYA who develop an SMN within 1-5 years of their primary diagnosis have an increased risk of death compared to those with a longer latency period [180]. The analysis presented in this chapter also describes subsequent mortality in those who develop a SMN, to assess the impact of latency on risk of death. This area of research was identified as a gap in the current literature in Chapter 2. Full details of the statistical methods are provided in Chapter 3 (Section 3.5.6).

7.2 Description of SMNs

There were 5104 patients diagnosed with cancer between 1990 and 2010. Of these patients 140 (2.7%) developed a total of 158 SMNs over a total of 57,390 person years follow-up (range of follow-up 0 to 26.0 years). Table 7.1 shows the baseline characteristics of the cohort for those who did and did not develop a SMN. The following analysis is restricted to the first SMN diagnosed within each patient. SMNs were more commonly diagnosed in those with a primary tumour

diagnosed aged 25-29 years, those diagnosed in an earlier time period, a primary diagnosis of a CNS or germ cell tumour and those treated with radiotherapy for their first tumour.

Table 7.1: Patient characteristics of those with and without SMN diagnosed

		No SMN (n=4964)		SMN (n=152)	
Characteristic	**Group**	**n**	**%**	**n**	**%**
Age group	0-4 years	921	18.6	21	15.0
	5-9 years	503	10.1	17	12.1
	10-14 years	519	10.5	14	10.0
	15-19 years	730	14.7	16	11.4
	20-24 years	966	19.5	19	13.6
	25-29 years	1325	26.7	53	37.9
Sex	Males	2973	59.9	82	58.6
	Females	1991	40.1	58	41.4
Period of diagnosis	1990-1996	1550	31.2	69	49.3
	1997-2003	1623	32.7	37	26.4
	2004-2010	1791	36.1	34	24.3
Diagnostic group	Leukaemia	938	18.9	15	10.7
	Lymphoma	1002	20.2	21	15.0
	CNS tumours	801	16.1	48	34.3
	Neuroblastoma	153	3.1	0	0.7
	Retinoblastoma	60	1.2	2	1.4
	Renal tumours	145	2.9	5	3.6
	Hepatic tumours	37	0.8	0	-
	Bone tumours	221	4.5	3	2.1
	Soft tissue sarcoma	295	5.9	6	4.3
	Germ cell tumour	857	17.3	30	21.4
	Other	455	9.1	10	7.1
Ethnicity	White	4335	87.3	124	88.6
	South Asian	354	7.1	10	7.1
	Other	206	4.2	5	3.6
	Missing	69	1.4	1	0.7
Deprivation fifth	1 (least deprived)	313	6.3	9	6.4
	2	583	11.7	16	11.4
	3	1014	20.4	22	15.7
	4	1041	21.0	30	21.4
	5 (most deprived)	2013	40.6	63	45.0

Chemotherapy	Yes	3231	65.1	79	56.4
Radiotherapy	Yes	1375	27.7	54	38.6
Surgery	Yes	2441	49.3	83	59.3

The most common type of SMNs diagnosed were CNS tumours (n=48), and these were most frequently diagnosed following a primary CNS tumour (Table 7.2). SMNs were also more frequently diagnosed in those whose first tumour was a germ cell tumour (n=30) or a lymphoma (n=21).

Table 7.2: Diagnostic group of first and subsequent tumours

	Subsequent tumours										
First tumour	Leukaemia	Lymphoma	CNS	Digestive	Respiratory	Breast	Testicular	Thyroid	Soft tissue	Other	**Total**
Leukaemia	7		3					2	1	2	15
Lymphoma	5	5	1	1	1	1		4		3	21
CNS	1	2	33	2				1	4	5	48
Neuroblastoma											0
Retinoblastoma									1	1	2
Renal					1			2		2	5
Bone			2					1			3
Soft tissue	1			1	1			3			6
Germ cell	1	1	3	3	1		11		1	9	30
Other	1		2	1	1	2	1			2	10
Total	16	8	44	8	5	3	12	13	7	24	140

7.3 Cumulative incidence

The cumulative incidence increased with increasing follow-up and was similar for those diagnosed as children and TYAs (Figure 7.1). For both age groups combined, 10-years post diagnosis the cumulative incidence of SMN was 1.8% (95% CI 1.4 to 2.2) which increased to 3.7% (95% CI 3.1 to 4.5) 20-years from diagnosis.

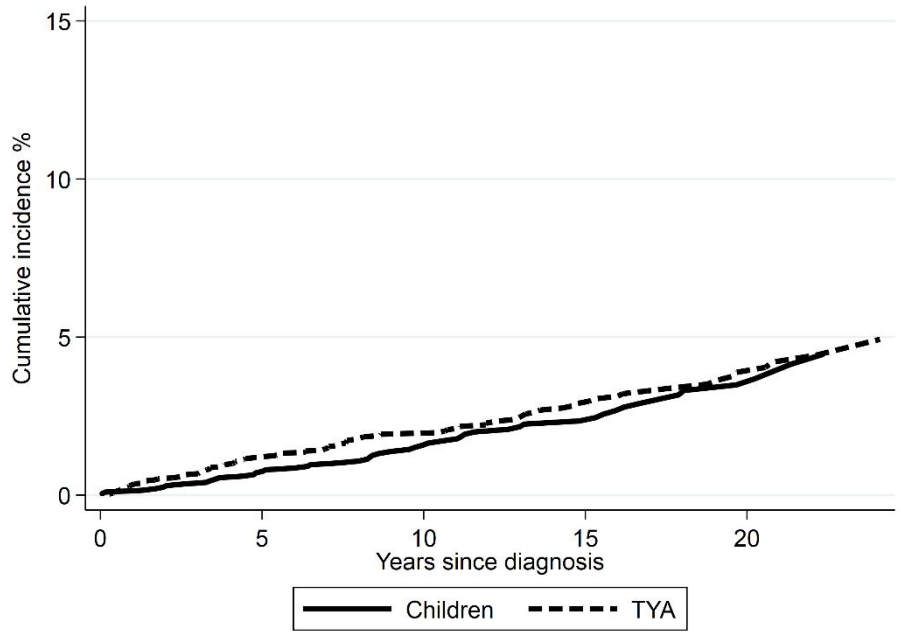


Figure 7.1: Cumulative incidence of SMN by age group

7.4 Comparison with general population

For all primary diagnostic groups combined, a total of 140 SMNs were observed, compared to 33 expected diagnoses giving an SIR of 4.3 (95% CI 3.6 to 5.0) (Table 7.3). SIRs and AERs were only estimated for primary diagnostic groups with at least 5 observed SMNs. A significant excess risk was observed for all primary diagnostic groups. The excess risk was highest for CNS tumours (SIR=13.3, 95% CI 10.0 to 17.7). Increased absolute excess risks were also found; the overall AER per 10,000 person years was 18.7 (95% CI 14.2 to 23.2). The AER was significantly higher for most primary diagnostic groups and was highest for survivors of CNS tumours (AER=53.6, 95% CI 36.6 to 70.5), followed by germ cell tumours (AER=18.5, 95% CI 8.3 to 28.8).

Table 7.3: Standardised incidence ratio (SIR) and absolute excess risk (AER) per 10,000 person years, for SMN by primary diagnostic group (all ages)

Primary diagnosis	Person-years	Observed	Expected	SIR (95% CI)	AER (95% CI)
All tumours	57390	140	32.9	4.26 (3.61, 5.03)	18.7 (14.2, 23.2)
Leukaemia	10342	15	3.4	4.37 (2.64, 7.26)	11.2 (3.1, 19.3)
Lymphoma	13000	21	9.5	2.22 (1.44, 3.41)	8.9 (0.6, 17.2)
CNS tumours	8291	48	3.6	13.3 (10.0, 17.7)	53.6 (36.6, 70.5)
Renal tumours	1896	5	0.7	7.41 (3.08, 17.8)	22.8 (-1.8, 47.5)
Soft tissue sarcoma	2867	6	1.5	4.04 (1.82, 8.99)	15.7 (-3.0, 34.4)
Germ cell tumours	11797	30	8.1	3.69 (2.58, 5.27)	18.5 (8.3, 28.8)
Other	4818	10	4.6	2.18 (1.17, 4.05)	11.2 (-4.3, 26.8)

SIR and AER not estimated for diagnostic groups with fewer than 5 observed cases

7.4.1 Comparison with SMN development 5-years post diagnosis only

For comparability with other studies the SIR and AER were also calculated based on 5-year survivors only (n=3932) and for SMNs that were diagnosed 5-years post diagnosis. Table 7.4 shows the results by primary diagnostic group with at least 5 observed SMNs. Overall, the excess risk was three times higher than expected (SIR=3.4, 95% CI 2.8 to 4.2) with significant excess risks observed for survivors of leukaemias, CNS tumours, soft tissue sarcomas and germ cell tumours. The AER for all cancers combined was 17.3 (95% CI 11.5 to 23.1). The AER was significantly higher for leukaemia, CNS tumours and germ cell tumours.

Table 7.4: Standardised incidence ratio (SIR) and absolute excess risk (AER) per 10,000 person years, by primary diagnostic group (all ages) for SMNs occurring 5-years post diagnosis

Primary diagnosis	Person-years	Observed	Expected	SIR (95% CI)	AER (95% CI)
All tumours	36229	89	26.3	3.38 (2.75, 4.17)	17.3 (11.5, 23.1)
Leukaemia	6512	11	2.7	4.15 (2.3, 7.5)	12.8 (1.7, 23.9)
Lymphoma	8435	13	7.8	1.66 (0.97, 2.87)	6.2 (-4.4, 16.8)
CNS tumours	5017	31	2.8	11.24 (7.9, 16.0)	56.3 (33.6, 79.0)
Soft tissue sarcoma	1796	5	1.2	4.2 (1.75, 10.1)	21.2 (-5.9, 48.4)
Germ cell tumours	7630	21	6.6	3.20 (2.08, 4.90)	18.8 (5.4, 32.3)

7.4.2 Comparison based on age at primary diagnosis

Figure 7.2 shows the SIR by main diagnostic group comparing age groups at primary diagnosis, while Figure 7.3 shows the comparisons in the AER by age and primary diagnostic group, to show estimates of the relative and absolute excess risks. These estimates are based on all SMNs occurring anytime since primary diagnosis. SIRs and AERs were not estimated for children with germ cell tumours due to small numbers but were for TYAs as this is one of the most common diagnostic groups within the age range. For all cancers combined the SIR was higher for those diagnosed as children (SIR=10.6, 95% CI 8.1 to 13.9) compared to those diagnosed as TYA (SIR=3.2, 95% CI 2.6 to 3.9). This was also observed for leukaemia, lymphoma and CNS tumours, with a significantly increased risk for lymphoma only; SIR for children 8.5 (95% CI 3.8 to 18.8) and SIR for TYA 1.7 (95% CI 1.03 to 2.9). For all diagnostic groups the absolute excess risk in the cancer survivors was higher than in the general population. Overall the absolute excess risk was similar for children and TYA survivors (AER= 20.6, 95% CI 14.1 to 27.0 and AER=17.4, 95% CI 11.3 to 23.5, per 10,000 person years, respectively). There was no significant increased absolute excess risk for TYA leukaemia and lymphoma survivors. The AER was highest for CNS survivors in both age groups (AER= 42.5, 95% CI 23.3 to 61.7 for children and AER=69.6, 38.6 to 100.6 for TYA)

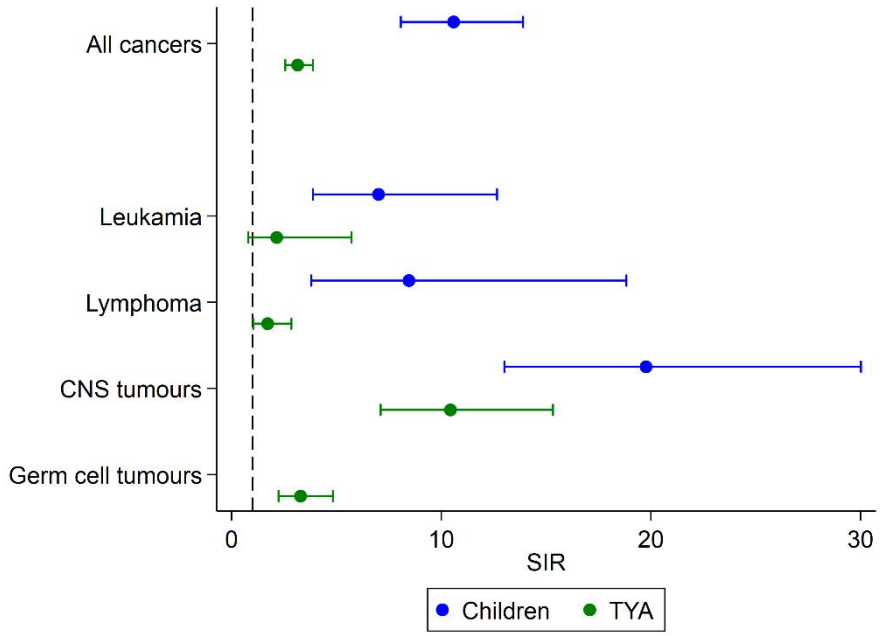


Figure 7.2: Standardised incidence ratio (SIR) of SMN by age and primary diagnostic group
The dashed line represents SIR=1 (excess risk equal in cancer survivors compared to general population)

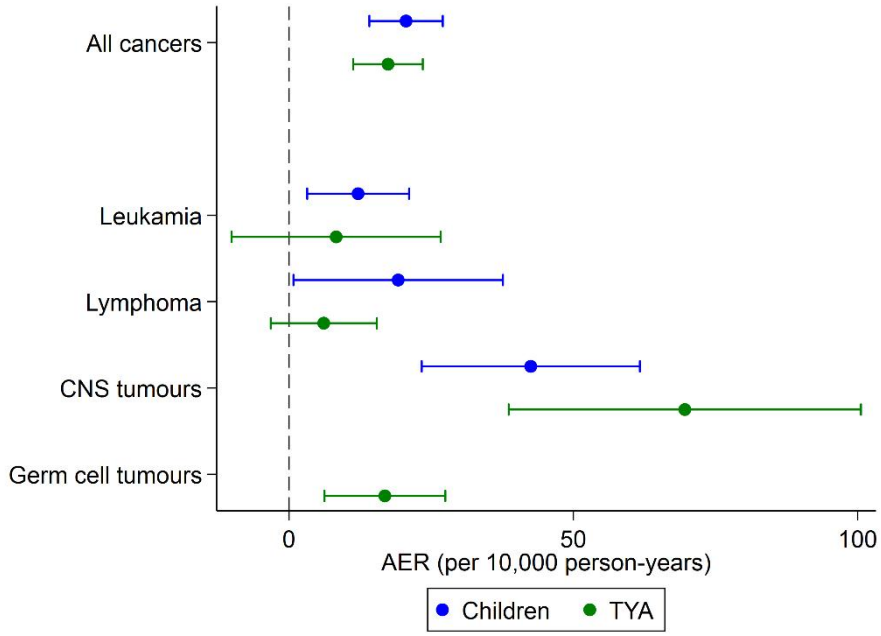


Figure 7.3: Absolute excess risk (AER) of SMN by age and primary diagnostic group
The dashed line represents AER=0 (excess risk equal in cancer survivors compared to general population)

7.5 Latency for SMNs

The median time to development of an SMN was 7.6 years (IQR 3.4 to 13.1), with 36% of all SMNs diagnosed within 5 years of the primary tumour, this percentage was higher for those whose first tumour was diagnosed as a TYA (42%) compared to those diagnosed as children (27%) (Table 7.5). The range in latency time was similar for those diagnosed as children and TYA, however the median time to SMN was slightly lower for TYA (6.9 years (IQR 3.1 to 12.8)) compared to children (9.3 years (IQR 4.8, 15.1)).

Table 7.5: Latency period by age group

Latency period	All ages N=140	Children N=52	TYA N=88
<5 years	51 (36%)	14 (27%)	37 (42%)
5 years-<10 years	35 (25%)	15 (29%)	20 (23%)
10 + years	54 (39%)	23 (44%)	31 (35%)

The type of SMN diagnosed varied by latency period (Table 7.6). Leukaemias and lymphomas accounted for a larger percentage of SMNs diagnosed within 5 years (22% and 12%, respectively) compared to SMNs diagnosed after 5 years (6% and 2% respectively). CNS tumours were the most frequently diagnosed SMN in both time periods; accounting for 25% of all SMNs within 5 years and 35% of all SMN after 5 years. Thyroid cancers were the next most common SMN type diagnosed after 5 years (13%).

Table 7.6: SMN types by latency period

SMN type	All N=140 n (%)	Latency <5 years N=51 n (%)	Latency ≥5 years N=89 n (%)
Leukaemia	16 (11%)	11 (22%)	5 (6%)
Lymphoma	8 (6%)	6 (12%)	2 (2%)
CNS	44 (31%)	13 (25%)	31 (35%)
Digestive	8 (6%)	2 (4%)	6 (7%)
Respiratory	5 (3%)	3 (6%)	2 (2%)
Breast	3 (2%)	1 (2%)	2 (2%)
Testicular	12 (9%)	6 (12%)	6 (7%)
Thyroid	13 (9%)	1 (2%)	12 (13%)
Soft tissue	7 (5%)	4 (8%)	3 (3%)
Other	24 (17%)	4 (8%)	20 (22%)

Comparing SMN types with at least 10 diagnoses, to ensure sufficient numbers per group, the median latency between primary tumour and SMN was shortest for leukaemias (4.1 years, IQR 1.9 to 8.8 years) and testicular cancers (4.9 years, IQR 2.6 to 12.1 years) and longest for thyroid cancers (15.1 years, IQR 9.9 to 20.4 years) (Figure 7.4).

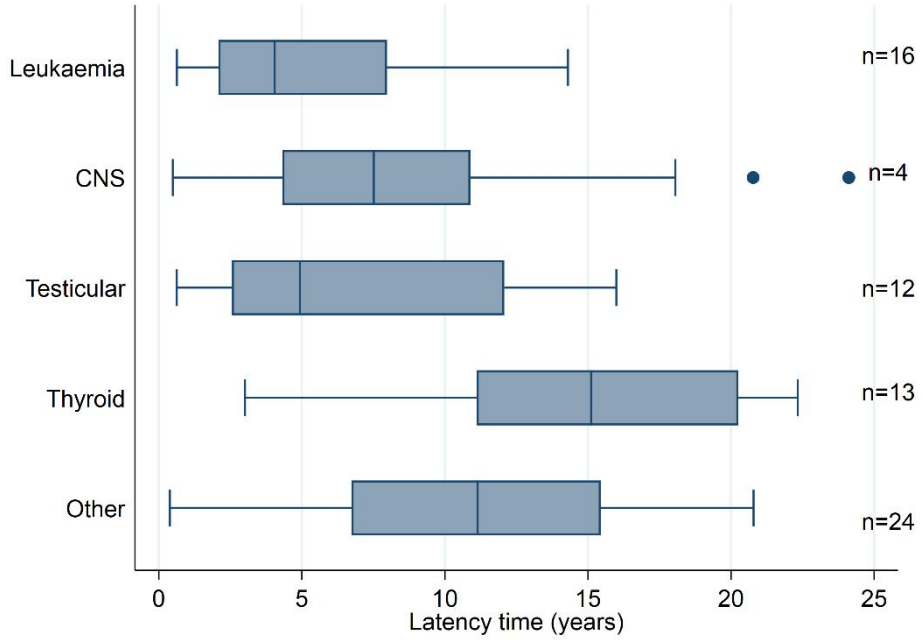


Figure 7.4: Box plot of latency time by SMN type

7.6 Survival following SMN

Patients who experienced an SMN had a higher percentage of deaths (41%) compared to those without an SMN (27%). This next section of analysis focusses on survival for those diagnosed with an SMN only (n=140). Follow-up time was calculated from diagnosis date of subsequent tumour to death or censoring date (31st December 2016) ensuring at least 1-year follow-up. Overall survival varied by latency period (Figure 7.5). Survival was similar for those with a latency period of <5 years and between 5-10 years from primary diagnosis, however, survival was higher for those with a latency period of 10 years or more. One-year survival was 96% (95% CI 86 to 99) for those with a latency period of 10+ years compared to 61% (95CI 46 to 73) for latency <5 years and 66% (95% CI 48 to 79) of those with latency of 5-10 years.

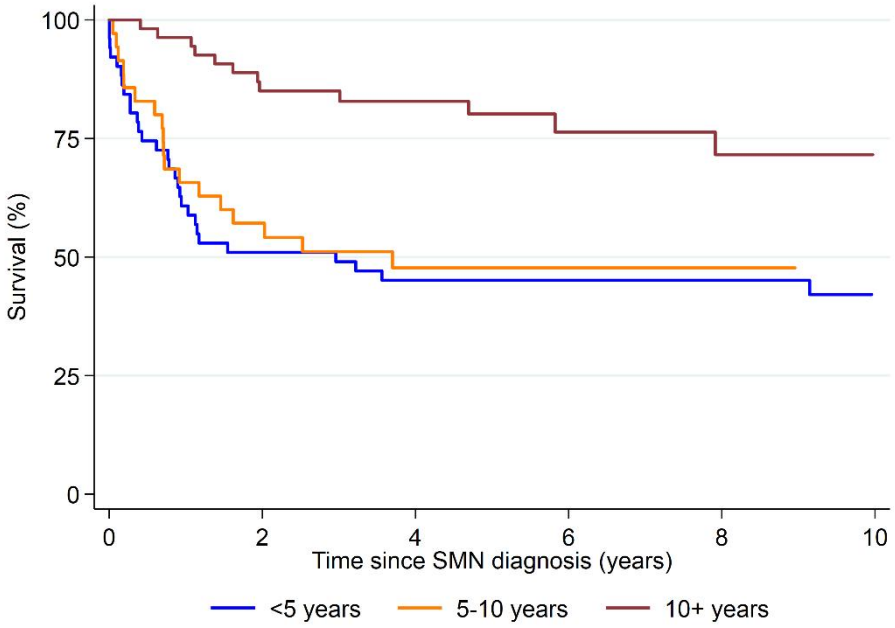


Figure 7.5: Kaplan-Meier survival from SMN by latency period

A flexible parametric survival model was used to examine the association between latency period and survival after adjustment for potential confounders with the results shown in Table 7.7. In both the unadjusted and adjusted models there was a significant association between latency period and survival. The risk of death was 87% lower for those with a latency period of 10 or more years

compared to those with a latency period of less than 5 years (HR=0.13; 95% CI 0.0 to 0.33).

Table 7.7: Association between latency period and subsequent mortality following SMN diagnosis, hazard ratio (HR), 95% CI

Latency period	Unadjusted HR (95% CI)	Adjusted HR† (95% CI)
<5 years	1.0 -	1.0 -
5-10 years	0.86 (0.48, 1.55)	0.40 (0.19, 0.84)
10+ years	0.29 (0.15, 0.57)	0.13 (0.05, 0.33)

† Adjusted for year of diagnosis of primary tumour, treatment received for primary tumour and SMN type

7.7 Discussion

7.7.1 Results in context

Subsequent malignant neoplasm developed in 3% of the study population. The cumulative incidence continued to increase throughout follow up, reaching 4% 20-years post diagnosis and was similar for children and TYA. In this analysis all those diagnosed under 30 years were included, whereas other studies have focussed on children only (with different definitions cut-offs at 15/18/21 years) [33-35, 167-171, 173, 175, 288-290] or TYA only (again with different age ranges being used, 15-24/15-39 years) [82, 174, 291] making direct comparison difficult. Other factors hindering direct comparability include different inclusions of primary tumour types, different definitions and inclusions of subsequent tumours, studies focussing on late SMNs only and different periods of follow-up and timescales used (for example cumulative incidence based on attained age and years from diagnosis). In this analysis 36% of SMNs were diagnosed as early onset SMNs therefore the estimated cumulative incidence may be higher than studies including late onset SMNs only. One of the more comparable studies to this was conducted in Canada on children (<15 years) diagnosed with primary cancer between 1985 and 2008 and found the cumulative incidence 15-years from diagnoses was 2.6% [170], compared to a cumulative incidence of 2.4% (95%C 1.7 to 3.2) 15-years post diagnosis in this study.

Compared to the general population the incidence of subsequent tumours was 4.3 times higher than expected, which was higher for children (SIR=10.62) than

for TYA (SIR=3.2). Again direct comparison with other studies is difficult and previously estimated SIRs range from 2 to 10 [33-35, 167-171, 173-175, 288, 290, 291]. A study based on SEER data [174] comparing SMNs in children and adolescents and young adults (15-39 years) found an SIR of 5.4 in children and 2.0 for AYA, so following a similar trend to that presented here of higher SIR in children compared to TYA. This was also observed across several primary diagnostic groups. The absolute excess risk, which measures the number of additional incident cases beyond those expected in the general population, was similar for children and TYAs, around 20 extra diagnoses per 10,000 person years. Both the SIR and AER were used as these provide complementary measures of both the relative and absolute excess risks.

A large study (including 200,000 survivors) based on 15-39 year old survivors in England and Wales examined subsequent neoplasms after 16 types of adolescent and young adult cancers and identified a small number of specific subsequent tumours that account for a substantial proportion of the excess number of neoplasms including many diagnoses of subsequent lung cancers [82]. This the largest study within this age group with extensive follow-up to identify specific types of subsequent tumours and highlights the unique features of the teenage and young adult population and differences compared to children. The results presented here included the 15-29 year age group, however, very few subsequent lung tumours were recorded, which may be due to the shorter follow-up period and relatively younger age of the cohort compared to the TYACSS study.

In the analysis presented in this chapter all tumours developing anytime following diagnosis were included, in contrast to the majority of studies in long-term childhood cancer survivors (i.e. those surviving at least 5-years) which only include tumours that develop 5-years post diagnosis [33, 35, 167, 168, 175, 288]. Sensitivity analysis was carried out restricting the estimation of standardised incidence ratios to those that developed 5-years post diagnosis only and found an SIR 3.4 times higher than expected, with significant excess risks by primary diagnostic groups. The number of primary diagnostic groups included were limited by sample size, however the highest SIR was observed for those diagnosed with primary CNS tumour with an estimated SIR of 11 (95% CI 8 to 16), this is higher than other previous estimates from the BCCSS (SIR= 2.7, 95% CI 2.4 to 3.2) [35] and the CCSS (SIR for medulloblastoma and PNET = 7.3, 95% CI 4.7 to 11.4, and SIR for other CNS = 6.9, 95% CI 4.0 to 11.7)

[167], but there are many factors previously mentioned why these findings may not be directly comparable. The estimated SIRs for those with primary leukaemias and soft tissue sarcomas were more similar; SIR for leukaemia in Yorkshire cohort was 4.2 (95% CI 2.3 to 7.5) compared with 4.3 (95% CI 3.6 to 5.2) in the BCCSS and 4.4 (95% CI 3.7 to 5.3) for ALL in the CCSS and SIR for soft tissue sarcoma was 4.2 (95% CI 1.8 to 10.1) in Yorkshire compared with 3.2 (95% CI 2.5, 4.1) in the BCCSS and 5.8 (95% CI 4.3 to 7.3) in the CCSS [35, 167]

The median time to SMN development was around 8 years. Just under half of all SMN in TYA were diagnosed within 5 years from the primary tumour, this figure was slightly lower for children at 27%. These findings are consistent with previous studies that identified a significant proportion of all SMNs were diagnosed within 5-years of diagnosis [170, 180]. The number of early onset SMNs was slightly higher in TYA compared to children, which may reflect the higher baseline incidence of tumours in the general population in this age group.

The types of SMNs diagnosed differed by latency period. Therapy related leukaemias are more likely to develop as early onset tumours while solid tumours generally had a longer latency period and may correspond to previously irradiated sites [181, 292].

CNS tumour were the most frequently diagnosed tumour type in both periods. Risk of CNS tumours are increased for those with previous CNS radiation and linked to genetic factors [292]. The analysis presented in this thesis included certain non-malignant tumours, such as those included in the ICCC-3 Classification [43], for example meningiomas and pilocytic astrocytomas. In fact, 34% of all diagnosed CNS SMNs were of non-malignant histology. A study in the Nordic countries reported that nearly a third of all SMNs (occurring any time since diagnosis) were CNS tumours and estimated an SIR of 10.4 (95% CI 6.7 to 16) for a CNS subsequent tumour following a primary CNS tumour, similar to our estimate of 13.3 (95% CI 10.1 to 17.7). Some of these excess brain tumours may be asymptomatic tumours picked up via routine imagining of CNS survivors [169].

Survival following SMN diagnosis is an important area of research [293] with limited studies examining the impact of SMNs on survival. The impact on latency on survival for AYAs with SMNs in the USA found higher death rates in those who developed SMN 1-5 years from primary tumour compared to SMNs that developed 6 or more years after primary tumour [180]. This study was a large population-based study including details on cause of death, however it only focussed on the 15-39 year age range. While in Canada those with early onset SMNs were 1.8 times more likely to die that those who developed an SMN after 5-years, this study was limited to children only [170]. Another US study of SMN developing before the age of 20 found those with a latency of less than 5-years had lower survival, but this study only included primary solid tumours [182]. A shorter latency period was associated with decreased survival for both children and TYAs combined in the results presented here. The mechanisms for this may be multi-factorial. Chemotherapy and radiotherapy are associated with increased risk of SMN development [181], but the impact of these factors on subsequent mortality is unknown. Treatment for the primary tumour may affect potential treatment options for SMNs due to the cumulative toxic effect which may then impact on prognosis. There are clear differences in the type of tumours diagnosed by latency and early onset SMNs may be more aggressive that those with longer latency. However, both SMN type and primary treatment exposure were adjusted for and the relationship between latency and survival was still evident. There may be other unmeasured and residual confounding factors that explain this relationship. Genetic predisposition and family history have a significant impact on SMN development [181], however it is not known how this impacts on survival. Genetic information was not available but this was included as a latent variable in the DAG (Section 3.5.6).

7.7.2 Strengths and limitations

Key strengths of this analysis are that it was population-based and was not based on patient recall or suffer from selection bias by only including those responding to questionnaires. Subsequent tumours were ascertained from both the YSRCCYP and national cancer registrations to identify tumours diagnosed outside of the Yorkshire region and those diagnosed after the age of 30. Both these registers have high ascertainment rates and potential losses to follow-up are minimal.

SMNs were coded according to the classification of multiple primary tumours defined by IACR/IARC where subsequent tumours were based on those with a

different histology to the primary tumour [165]. However, there is still a possibility that a relapse or recurrence of the primary tumour has been mistaken for a subsequent tumour. In additional certain non-malignant tumour types were included as these are commonly diagnosed and classified according to the ICCC-3 [43]. These were mainly benign CNS tumours that present with similar clinical symptoms and prognosis as malignant tumours.

The other main limitations are around the small sample size meaning subgroup analysis by, for example primary diagnosis, was not possible. Treatment information was extracted from the YSRCCYP database for all primary tumours, however treatment information for SMNs was not available. Follow-up was limited to SMN diagnosed under age 55, while the large cancer survivors studies (such as the BCCSS and CCSS) have follow-up up to age 60+, therefore different patterns of SMN may occur with longer follow-up as cancer rates in the general population start to increase steadily in these age groups. Only time to first SMN was considered in this analysis and some survivors in this study had more than one subsequent tumour diagnosed. Larger studies would be needed to examine these individuals in more detail. Further analysis incorporating all SMNs in addition to respiratory and cardiovascular admission in 5-year survivors is presented in Chapter 9.

The main limitation regarding the analysis by latency period is due to differences in survivorship and follow-up by latency period. Those with a longer latency period are more likely to be those with better initial prognosis and have been followed-up for longer by definition. In the latency survival analysis follow-up started from date of diagnosis of SMN and each individual was followed-up for at least one year. The median follow-up time for those with SMN diagnosed within 5 years was 3.2 years compared to 4.9 years for those with an SMN diagnosed after 5 years. From the Kaplan-Meier plot differences in survival by latency were evident after 1-year follow-up (63% in those with latency less than 5 years compared to 96% in those with latency of 10 or more years). The analysis of latency was based on 140 patients only, therefore replication of these findings in larger studies is needed. This would also allow examination of differences in subsequent mortality and latency by age group which was not possible in the analysis presented here due to small numbers.

7.7.3 Summary

Subsequent malignant neoplasms developed in 3% of the study population, with 36% of these occurring within 5-years of the primary diagnosis. The cumulative incidence continued to increase throughout follow up, reaching 4% twenty-years post diagnosis. A shorter latency period was associated with decreased survival. SMNs were the second health outcome to be investigated in this thesis. The next chapter evaluates the third long-term outcome, cardiovascular disease, and then goes on to explore the cumulative burden of multiple and recurrent morbidity outcomes.

Chapter 8 Cumulative burden of respiratory and cardiovascular morbidity and subsequent tumours

8.1 Introduction

The most common causes of late morbidity and mortality in long-term survivors of childhood and young adult cancers are subsequent neoplasms, cardiovascular and respiratory diseases. Chapters 7 and 8 provided a detailed description of respiratory morbidity and SMNs. The first section of this chapter focusses on hospital admissions for cardiovascular disease, including a comparison with rates in the general population (addressing Aim 2c of this thesis). Similar to the analysis for respiratory admissions, both admissions based on all diagnostic fields within each HES admission and the primary admission are compared. The next section in this chapter goes on to explore the cumulative burden of subsequent neoplasms, respiratory and cardiovascular diseases combined, to specifically address Aim 3 of this thesis. A lack of studies based on multiple and recurrent events was identified in the literature review in Chapter 2, therefore this chapter addresses this knowledge gap. The time to the first event for each outcome is described along with the cumulative burden estimated using the mean cumulative count (MCC). The association between treatment exposure and the cumulative burden is explored utilising multiple failure time models to account for all admissions and all subsequent tumours within each individual.

8.2 Description of cardiovascular admissions

The analysis presented in this section is based upon 5-year survivors diagnosed between 1990 and 2011, with follow-up to March 2017. A total of 4235 individuals were included of whom 427 (10.1%) were admitted to hospital at least once for a cardiovascular condition. As well as all cardiovascular conditions combined, seven specific conditions were included as shown in Table 8.1. Hypertension was the most common condition based on all diagnostic codes. Restricting admissions to cardiovascular conditions using only the primary diagnosis field within each HES record reduced the total number of

admissions to 201 (4.7%) of survivors, with fewer than 30 admissions for each specified condition.

Table 8.1: Number of survivors admitted for cardiovascular conditions

Cardiovascular disease	ICD10 codes	All diagnostic codes in HES	Primary admissions only
		n (%)	n (%)
All cardiovascular	I00-I99, G45	427 (10.1%)	201 (4.7%)
Hypertension	I10-I15	140 (3.3%)	9 (0.2%)
Coronary (Ischemic) heart disease	I20-I25	39 (0.9%)	20 (0.5%)
Cardiomyopathy and heart failure	I42, I43, I50, I51	64 (1.5%)	22 (0.5%)
Valvular heart disease	I34-I39	20 (0.5%)	3 (0.1%)
Pericardial disease	I30-I32	20 0.5%)	10 (0.2%)
Conduction disorders (Arrhythmias)	I44-I49	63 (1.5%)	26 (0.6%)
Cerebrovascular disease	I60-I69,G45	37 (0.9%)	23 (0.5%)

8.2.1 Cumulative incidence

The cumulative incidence for admissions for any cardiovascular disease increased with increasing age without reaching at plateau (Figure 8.1). By age 40, the cumulative incidence for an admission for any type of cardiovascular condition was 24.0% (95% CI 21.5 to 26.7), hypertension was 7.3% (95% CI 5.6 to 9.3), heart failure was 4.6% (95% CI 3.4 to 6.1), conduction disorders was 3.6% (95% CI 2.7 to 4.7), cerebrovascular disease was 2.5% (95% CI 1.6 to 3.7), pericardial disease was 1.4% (95% CI 0.9 to 2.2) and valvular heart disease was 1.3% (95% CI 0.6 to 2.4).

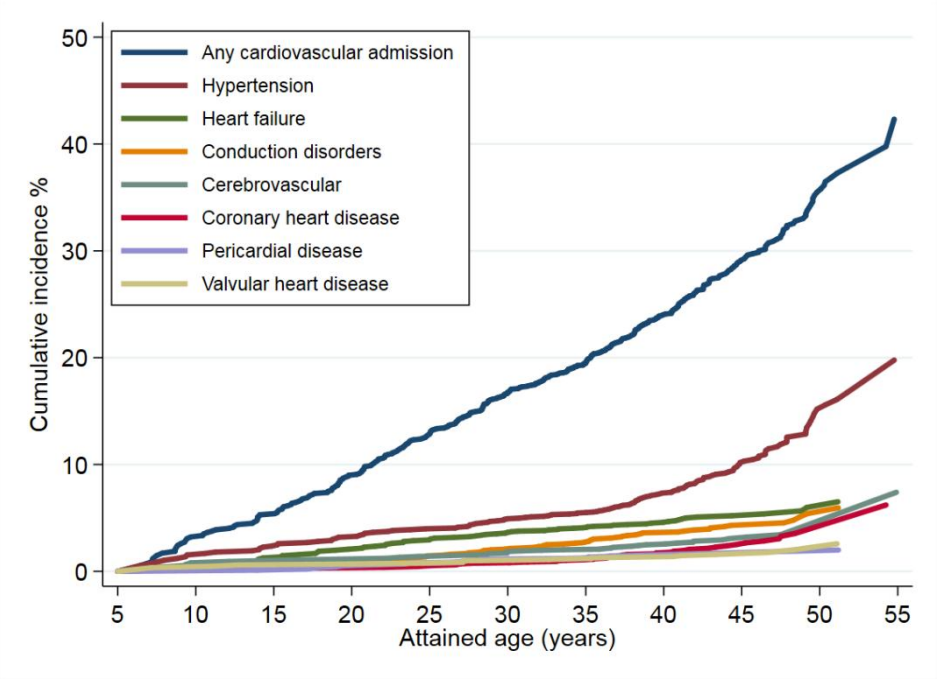


Figure 8.1: Cumulative incidence of cardiovascular admission by attained age

Sensitivity analysis was conducted to calculate the cumulative incidence based on primary admission only with the results shown in Figure 8.2. The cumulative incidence continued to increase with attained age. However, the incidence rates were substantially lower than when based on all diagnostic codes. By age 40, the cumulative incidence for any cardiovascular admission was 12.4% (95% CI 14.4 to 14.4). For all specified conditions by age 40 the cumulative incidence was between 1% and 2%.

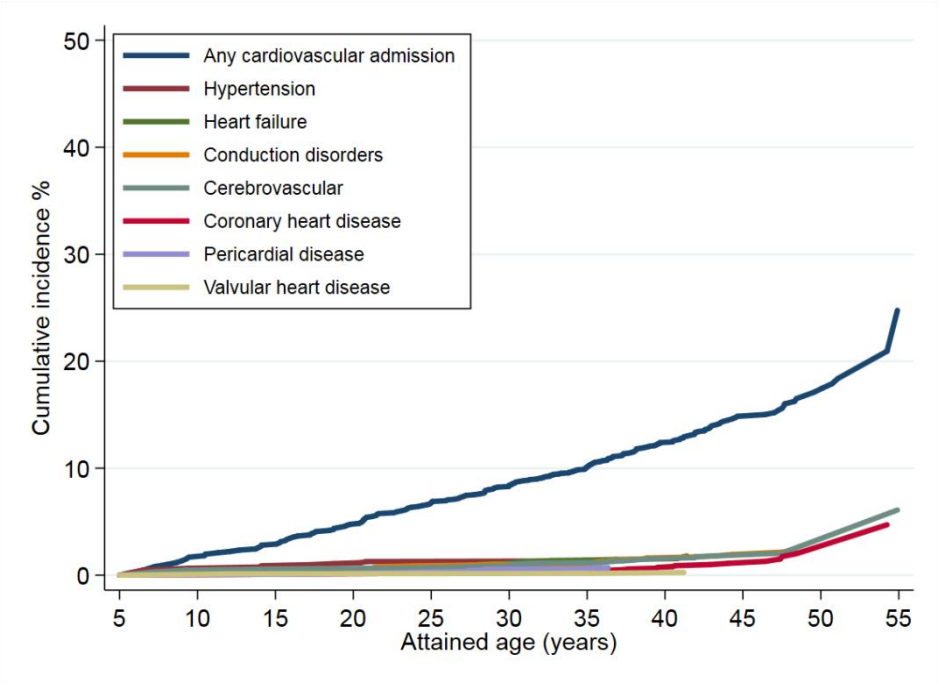


Figure 8.2: Cumulative incidence of cardiovascular admission by attained age based on primary diagnoses only

8.2.2 Comparison with general population

The observed and expected number of admissions for each cardiovascular condition and the corresponding hospitalisation rate ratio (HRR) are shown in Table 8.2. For all cardiovascular conditions the risk of hospitalisation was 1.67 times higher (95% CI 1.52 to 1.83) in cancer survivors compared to the general population. For each cardiovascular condition except coronary heart disease the excess risk was significantly higher in cancer survivors compared to the general population ranging from 1.56 (95% CI 1.32 to 1.84) for hypertension to 4.48 (95% CI 3.50 to 5.72) for cardiomyopathy and heart failure.

Table 8.2: Observed and expected cardiovascular admissions and hospitalisation rate ratio (HRR)

	Observed admissions	Expected admissions	HRR (95% CI)
All cardiovascular	427	255	1.67 (1.52, 1.83)
Hypertension	140	90	1.56 (1.32, 1.84)
Coronary heart disease	39	31	1.26 (0.92, 1.73)
Cardiomyopathy and heart failure	64	14	4.48 (3.50, 5.72)
Valvular heart disease	20	7	3.04 (1.96, 4.71)
Pericardial disease	20	6	3.63 (2.34, 5.62)
Conduction disorders	63	32	1.95 (1.52, 2.50)
Cerebrovascular disease	37	13	2.78 (2.01, 3.84)
Heart transplant	7	10	0.71 (033, 1.49)

Stratifying by age group at diagnosis (Table 8.3) the excess risk for all cardiovascular admissions was higher in children (HRR=3.11, 95% CI 2.64 to 3.67) compared to TYA (HRR=1.36, 95% CI 1.21 to 1.52). This trend of higher HRR in children compared to TYA was observed for all specific conditions with notable differences in the excess risk by age for hypertension (4.12, 95% CI 2.94 to 5.78 in children and 1.30, 95% CI 1.08 to 1.58 in TYA), coronary heart disease (3.50, 95% CI 1.67 to 7.34 in children and 1.11, 95% CI 0.78 to 1.56 in TYA), cardiomyopathy (14.4, 95% CI 10.3to 20.3 in children and 2.58, 95% CI 1.82 to 3.67 in TYA) and cerebrovascular disease (8.44, 95% CI 5.17 to 13.8 in children and 1.84, 95% CI 1.20 to 2.82 in TYA).

Table 8.3: Observed and expected cardiovascular admissions and hospitalisation rate ratio (HRR) by age at diagnosis

	Observed admissions	Expected admissions	HRR (95% CI)
Children			
All cardiovascular	141	45	3.11 (2.64, 3.67)
Hypertension	34	8	4.12 (2.94, 5.76)
Coronary heart disease	7	2	3.50 (1.67, 7.34)
Cardiomyopathy and heart failure	33	2	14.4 (10.3, 20.3)
Valvular heart disease	7	1.6	4.39 (2.09, 9.12)
Pericardial disease	11	1.5	7.18 (3.98, 12.97)
Conduction disorders	18	7	2.49 (1.57, 3.95)
Cerebrovascular disease	16	1.9	8.44 (5.17, 13.77)
TYA			
All cardiovascular	286	210	1.36 (1.21, 1.52)
Hypertension	106	81	1.30 (1.08, 1.58)
Coronary heart disease	32	29	1.11 (0.78, 1.56)
Cardiomyopathy and heart failure	31	12	2.58 (1.82, 3.67)
Valvular heart disease	13	5	2.61 (1.52, 4.50)
Pericardial disease	9	4	2.26 (1.18, 4.34)
Conduction disorders	45	25	1.79 (1.34, 2.40)
Cerebrovascular disease	21	11.4	1.84 (1.20, 2.82)

8.2.3 Comparison based on primary diagnosis only

Sensitivity analysis was carried out to estimate admission trends based on the primary diagnostic code only for each HES admission, with the results shown in Table 8.4. HRR were not calculated if there were fewer than 5 admission for a particular group. Based on primary admission only 201 admissions were observed, compared to 148 expected giving an HRR of 1.36 (95% CI 1.18 to 1.56). Significant excess risks were observed for all conditions except coronary heart disease (HRR=0.94, 95% CI 0.61 to 1.46). A significant excess risk of all cardiovascular admissions were observed for children (HRR=2.71, 95% CI 2.16 to 3.41) but not for TYA (HRR=1.05, 95% CI 0.88 to 1.25). Overall the excess risk was highest for cardiomyopathy and this was observed for both children (HRR= 17.2, 95% CI 9.5 to 31.0) and TYA (HRR=3.30, 95% CI 1.80 to 5.88), again with higher excess risk observed in children compared to TYA.

Table 8.4: Observed and expected number of admissions and hospitalisation rate ratio (HRR) based on primary diagnosis for admissions only

	Observed admissions	Expected admissions	HRR (95% CI)
All ages			
All cardiovascular	201	148	1.36 (1.18, 1.56)
Hypertension	9	4	2.05 (1.07, 3.94)
Coronary heart disease	20	22	0.94 (0.61, 1.46)
Cardiomyopathy and heart failure	22	4	5.48 (3.61, 8.32)
Valvular heart disease	<5	-	-
Pericardial disease	10	4	2.52 (1.35, 4.68)
Conduction disorders	26	16	1.58 (1.08, 2.32)
Cerebrovascular disease	23	10	2.25 (1.49, 3.38)
Children 0-14 years			
All cardiovascular	74	27	2.71 (2.16, 3.41)
Hypertension	8	0.7	10.9 (5.43, 21.7)
Coronary heart disease	<5	-	-
Cardiomyopathy and heart failure	11	0.6	17.2 (9.5, 31.0)
Valvular heart disease	<5	-	-
Pericardial disease	7	1	6.23 (2.97, 13.1)
Conduction disorders	7	4	1.83 (0.87, 3.83)
Cerebrovascular disease	10	1	7.83 (4.21, 14.5)
TYA 15-29 years			
All cardiovascular	127	120	1.05 (0.88, 1.25)
Hypertension	<5	-	-
Coronary heart disease	18	20	0.89 (0.56, 1.41)
Cardiomyopathy and heart failure	11	3.4	3.30 (1.80, 5.88)
Valvular heart disease	<5	-	-
Pericardial disease	<5	-	-
Conduction disorders	19	13	1.51 (0.96, 2.36)
Cerebrovascular disease	13	9	1.45 (0.84, 2.50)

HRR not calculated for groups where observed number of admission was fewer than 5

8.3 Cumulative burden subsequent neoplasms, cardiovascular and respiratory admissions

The analysis of the cumulative burden of all SMNs and all hospitalisations due to cardiovascular and respiratory conditions was restricted to 5-year survivors diagnosed between 1992 and 2010 to ensure consistent follow-up across all outcomes. All survivors were followed up till December 2015. A total of 3686 five-year survivors were included contributing to 28,335 person-years follow-up (starting 5-years post diagnosis). A total of 533 (15%) survivors were admitted to hospital at least once for a respiratory condition, 225 (6%) had 2 or more respiratory admissions, 316 (9%) were admitted at least once for a cardiovascular condition, 125 (3%) were admitted more than once for cardiovascular disease and 74 (2%) diagnosed with at least one SMN (Table 8.5).

Table 8.5: Summary of outcomes for cumulative burden analysis

Outcome	**n**	**%**
At least 1 respiratory admission	533	14.5
1 admission	*305*	*8.3*
2 admissions	*89*	*2.4*
3 admissions	*47*	*1.3*
4 admissions	*30*	*0.8*
5+ admissions	*59*	*1.6*
At least 1 cardiovascular admission	316	8.6
1 admission	*183*	*5.0*
2 admissions	*48*	*1.3*
3 admissions	*19*	*0.5*
4 admissions	*8*	*0.2*
5+ admissions	*50*	*1.4*
At least 1 SMN	74	2.0
1 SMN	*71*	*1.9*
2 SMNs	*3*	*0.1*
Combined outcomes		
Respiratory and cardiovascular admission	134	3.6
Respiratory and SMN	16	0.4
Cardiovascular and SMN	11	0.3
All 3 events	7	0.2

The median time from diagnosis to first event was 8.2 years (IQR 6.1 to 12.6) where the first event was a respiratory admission, 10.7 years (IQR 6.7 to 14.5) where the first event was a cardiovascular admission and 10.7 years (IQR 7.6 to 13.5) where the first event was a SMN (Figure 8.3). Subsequent events following first event are shown in Figure 8.3.

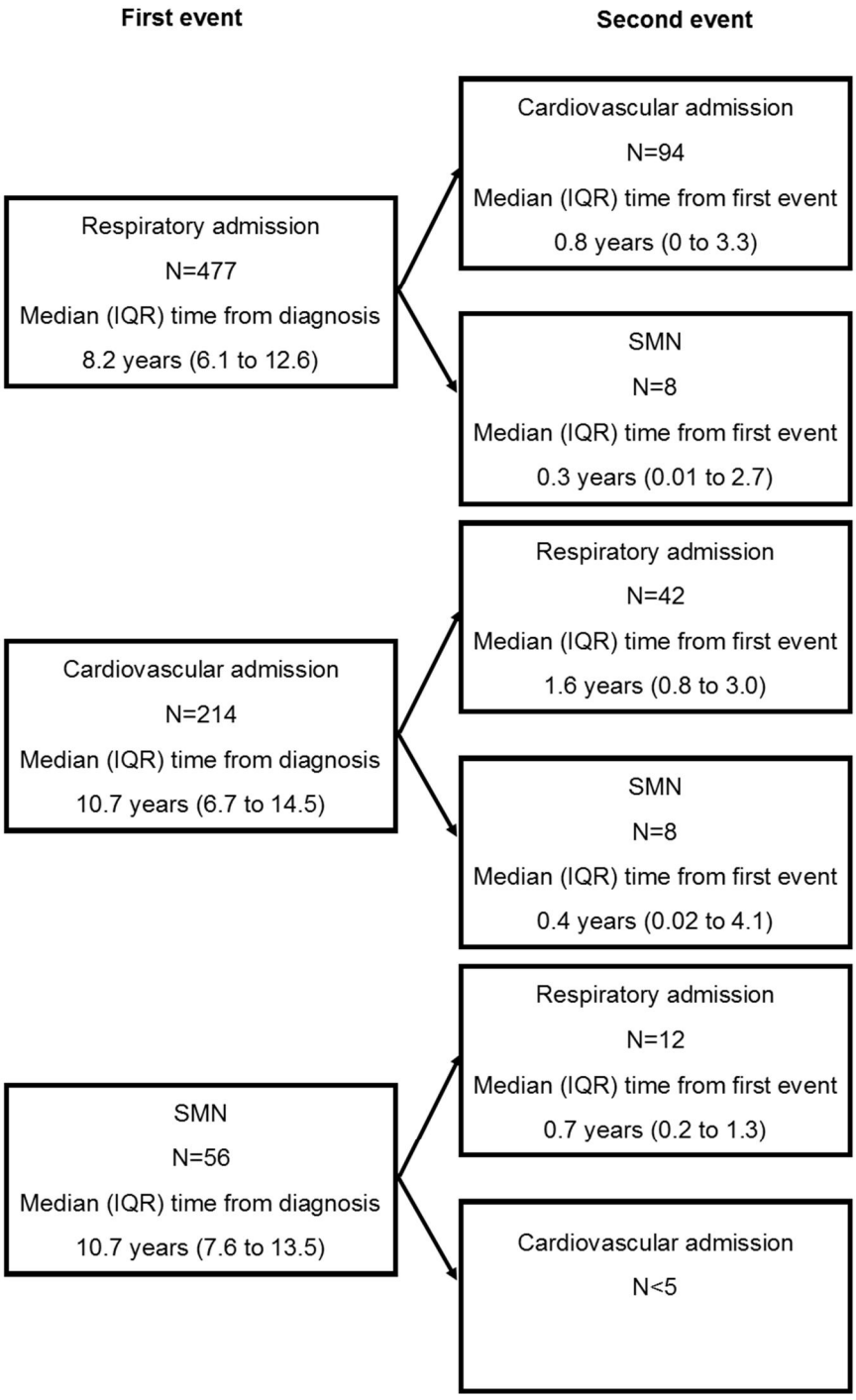


Figure 8.3: Flow chart showing order of events and time between events

Medians were only estimated for groups with more than 5 individuals. Event 3 not included due to small numbers (<5 in each group)

8.3.1 Overall burden

The cumulative incidence for admissions for respiratory disease, cardiovascular disease and SMN increased with attained age (Figure 8.4). By age 40 years, the cumulative incidence for an admission for respiratory disease was 51% (95% CI 46 to 55), cardiovascular disease was 25% (95% CI 22 to 28), and SMN was 6% (95% CI 5 to 8). By age 40 years, the cumulative incidence of being hospitalised for both a respiratory and cardiovascular condition was 12% (95% CI 10 to 14) for all other combinations of outcomes the cumulative incidence was less than 2%. The MCC showed that by age 40 years an average of 2.1 events per survivor were observed (95% CI 1.8 to 2.7).

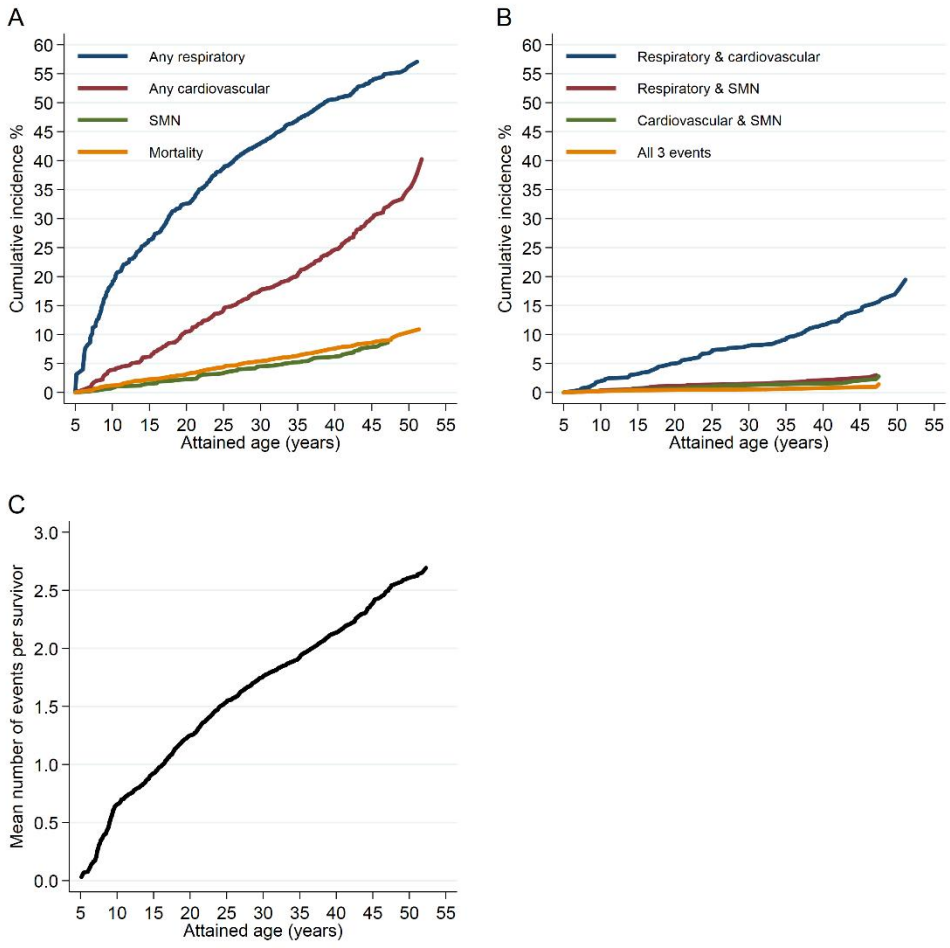


Figure 8.4: Cumulative incidence and cumulative burden of SMN, respiratory and cardiovascular admissions.

A. Cumulative incidence for each event type (respiratory admission, cardiovascular admission, SMN). B. Cumulative incidence for combinations of events (date of last event used as follow-up end point). C. Mean cumulative count for all respiratory and cardiovascular admissions and all SMNs.

8.3.2 Treatment specific risks

The MCC for each treatment group are shown in Figure 8.5. There were no statistical differences in the MCC by treatment group. By age 40 years, the average number of events (per survivor) for those who received anthracyclines was 2.2 (95% CI 1.8 to 2.8) compared with 2.1 (95% CI 1.6 to 2.7) for those who did not; for those who received pulmonary toxic chemotherapy the MCC was 2.0 (95% CI 1.6 to 2.5) compared to 2.1 (95% CI 1.7 to 2.8) for those who did not; and the MCC for those who received radiation to the chest was 2.1 (95% CI 1.7 to 2.8) and 2.1 (95% CI 1.7 to 2.6) in those who did not.

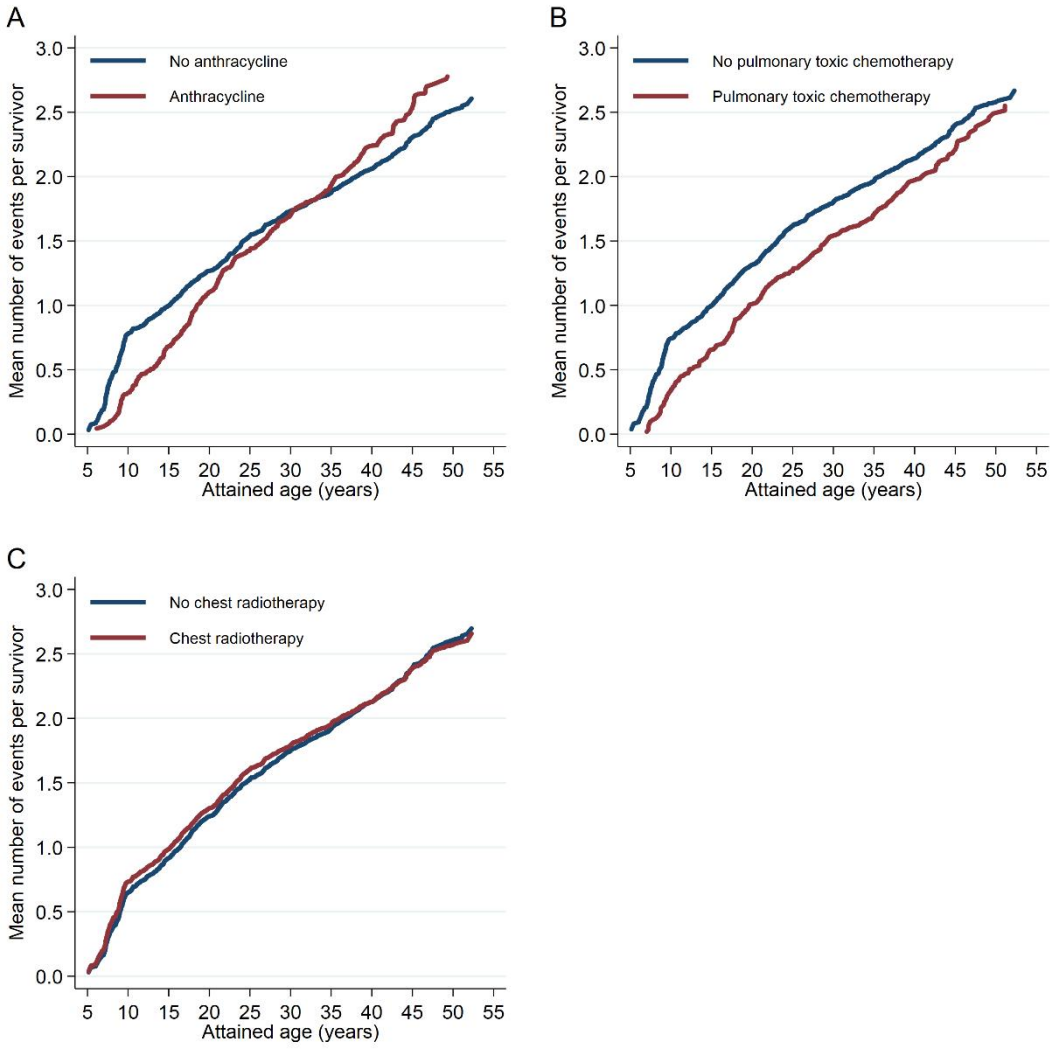


Figure 8.5: Mean cumulative count by treatment group

A. Anthracycline exposure, B. Pulmonary toxic chemotherapy exposure, C. Radiation to the chest exposure.

Table 8.6 shows the results of the unadjusted and adjusted models assessing the association between treatment exposures and risk of events (taking into account all and recurrent events). In unadjusted models there was no statistically significant association between any of the treatment exposures and risk of events. After adjustment for confounders the risk of an event was higher for those treated with pulmonary toxic chemotherapy (HR=1.22, 95% CI 1.04 to 1.44), there was no statistically significant association with anthracycline exposure (HR=0.94, 95% CI 0.79 to 1.11) or radiation to the chest (HR=1.11, 95% CI 0.86 to 1.43).

Table 8.6: Association between treatment exposures and risk of events, hazard ratios (HR) and 95% CI

Treatment	Unadjusted HR (95% CI)	Adjusted HR † (95% CI)
Anthracycline		
No	1.0 -	1.0 -
Yes	1.04 (0.93, 1.17)	0.94 (0.79, 1.11)
Pulmonary toxic chemotherapy		
No	1.0 -	1.0 -
Yes	1.07 (0.95, 1.20)	1.22 (1.04, 1.44)
Chest Radiotherapy		
No	1.0 -	1.0 -
Yes	1.12 (0.88, 1.41)	1.11 (0.86, 1.43)

†Adjusted for diagnostic group, age at cancer diagnosis, year or diagnosis, deprivation, ethnicity and the other treatment exposures
Models exclude 36 individuals with missing ethnic group

Sensitivity analysis was conducted to assess the association between treatment groups and each outcome individually using time to first event for each outcome within a competing risk framework (Table 8.7). In unadjusted analyses there was a significant association between anthracycline exposure and admission for respiratory conditions (sHR=1.28, 95% CI 1.08 to 1.53) and chest radiation and respiratory conditions (sHR=1.45, 95% CI 1.02 to 2.06). After adjustment for confounders, for those treated with pulmonary toxic chemotherapy there was an increased risk of respiratory admissions (sHR= 1.27, 95% CI 0.99 to 1.63) and SMN diagnosis (sHR=1.81, 95% CI 0.98 to 3.34).

Table 8.7: Association between treatment exposure and time to first event for each outcome, sub-distribution hazard ratios (sHR) and 95% CI

	Respiratory admission		Cardiovascular admission		SMN	
Treatment	**Unadjusted sHR (95% CI)**	**Adjusted sHR † (95% CI)**	**Unadjusted sHR (95% CI)**	**Adjusted sHR † (95% CI)**	**Unadjusted sHR (95% CI)**	**Adjusted sHR † (95% CI)**
Anthracycline						
No	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -
Yes	1.28 (1.08, 1.53)	0.94 (0.72, 1.23)	0.89 (0.70, 1.14)	0.84 (0.59, 1.20)	0.50 (0.28, 0.90)	0.50 (0.24, 1.05)
Pulmonary toxic chemotherapy						
No	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -
Yes	1.11 (0.93, 1.34)	1.27 (0.99, 1.63)	1.00 (0.79, 1.27)	1.09 (0.78, 1.51)	1.28 (0.79, 2.06)	1.81 (0.98, 3.34)
Chest Radiotherapy						
No	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -
Yes	1.45 (1.02, 2.06)	1.44 (0.97, 2.12)	1.16 (0.72, 1.85)	1.15 (0.69, 1.92)	0.83 (0.26, 2.65)	1.00 (0.31, 3.25)

†Adjusted for diagnostic group, age at cancer diagnosis, year of diagnosis and deprivation and the other treatment exposures

Models exclude 36 individuals with missing ethnic group

8.3.3 Differences by age

Differences by age group at diagnosis were evaluated by estimating the MCC separately for children and TYA, this was based upon time since diagnosis rather than attained age to enable direct comparison between age groups and follow-up periods. The MCC for each age group is shown in Figure 8.6. For children the mean number of event per survivor was 0.4 (95% CI 0.3 to 0.6) 10-years post diagnosis increasing to 1.2 (95% CI 0.9 to 1.4) 20-years after diagnosis. For TYA, 10-years post diagnosis the mean number of events was 0.3 (95% CI 0.2 to 0.4) increasing to 0.9 (95% CI 0.8 to 1.1) 20-years post diagnosis.

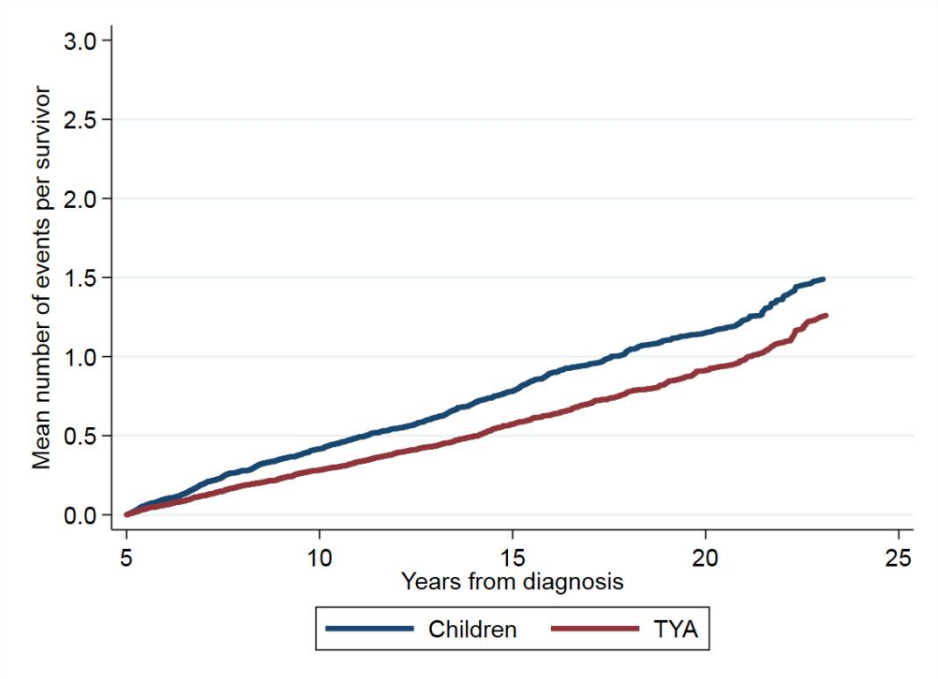


Figure 8.6: Mean cumulative count by age group

A two-way interaction between treatment exposure and age group was examined and separate models run for children and TYA to examine the association between treatment and risk of events. There were significant interactions between age group and chest radiotherapy (p=0.03) and some evidence of an interaction between pulmonary toxic chemotherapy and age group (p=0.08) (Table 8.8). The risk of experiencing an event was significantly increased for children treated with pulmonary toxic chemotherapy (HR=1.24, 95%C% 1.02 to 1.51) and radiation to the chest (HR=1.60, 95% CI 1.16 to 2.20) but no increased risk was observed for TYA (HR=1.07, 95% CI 0.89 to 1.28 for chemotherapy and HR=1.03, 95% CI 0.79 to 1.36 for radiotherapy).

Table 8.8: Association between treatment exposures and risk of events stratified by age at diagnosis, adjusted hazard ratios (HR) and 95% CI

Treatment	Children (N=1391) Adjusted HR † (95% CI)	TYA (N=2257) Adjusted HR † (95% CI)	Interaction p-value
Anthracycline			
No	1.0 -	1.0 -	
Yes	1.14 (0.91, 1.42)	0.97 (0.77, 1.23)	0.29
Pulmonary toxic chemotherapy			
No	1.0 -	1.0 -	
Yes	1.24 (1.02, 1.51)	1.07 (0.89, 1.28)	0.08
Chest Radiotherapy			
No	1.0 -	1.0 -	
Yes	1.60 (1.16, 2.20)	1.03 (0.79, 1.36)	0.03

†Adjusted for diagnostic group, year of diagnosis, deprivation and the other treatment exposures

Models excludes 36 with missing ethnicity (2 children and 34 TYA)

8.3.4 Differences by diagnostic group

Differences by diagnostic groups were estimated for leukaemia, lymphoma, CNS tumours, germ cell tumours and other solid tumour. For each of these diagnostic groups the MCC was calculated and models run to examine the association between treatment and risk of events within each diagnostic group. For CNS, germ cell tumours and other solid tumours models examining the relationship between anthracyclines and chest radiation were not included due to a small number of individuals treated with these within the diagnostic groups.

The MCC by diagnostic group is shown in Figure 8.7. The MCC was lowest for those diagnosed with germ cell tumours and highest for those with lymphomas and other solid tumours. By age 40 years, the mean number of events (per survivor) for those with germ cell tumours was 0.8 (95%C 0.6 to 1.0); 1.3 (95% CI 0.9 to 1.8) for CNS tumours; 2.2 (95% CI 1.6 to 2.8) for those with leukaemia, 2.4 (95% CI 1.5 to 3.6) for lymphoma, and 2.4 (95% CI 1.8 to 3.0) for those with other solid tumours.

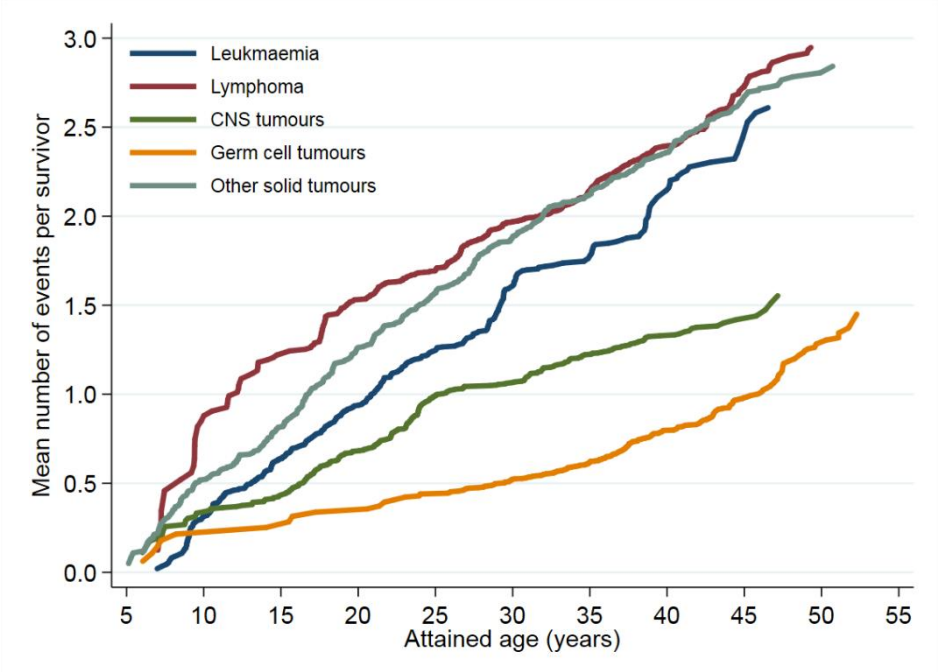


Figure 8.7: Mean cumulative count by diagnostic group

There were significant interactions between diagnostic group and pulmonary toxic chemotherapy (p<0.001) and chest radiation (p=0.004) (Table 8.9). The risk of experiencing an event was significantly increased for those diagnosed with a CNS tumour treated with pulmonary toxic chemotherapy (HR=3.21, 95%C% 2.36 to 4.35), and for those with leukaemia treated with radiation to the chest (HR=1.82, 95% CI 1.16 to 2.85).

Table 8.9: Association between treatment exposures and risk of events stratified by diagnostic group, adjusted hazard ratios (HR) and 95% CI

Treatment	Leukaemia (N=621) Adjusted HR † (95% CI)	Lymphoma (N=820) Adjusted HR † (95% CI)	CNS tumours (N=538) Adjusted HR † (95% CI)	Germ cell tumours (N=749) Adjusted HR † (95% CI)	Other solid tumours (N=922) Adjusted HR † (95% CI)	Interaction p-value
Anthracycline						
No	1.0 -	1.0 -	-	-	1.0 -	0.10
Yes	0.89 (0.67, 1.17)	0.92 (0.67, 1.27)	-	-	1.22 (0.96, 1.55)	
Pulmonary toxic chemotherapy						
No	1.0 -	1.0 -	1.0 -	1.0 -	1.0 -	<0.001
Yes	1.00 (0.72, 1.38)	1.17 (0.84, 1.62)	3.21 (2.36, 4.35)	1.18 (0.91, 1.53)	1.35 (0.96, 1.91)	
Chest Radiotherapy						
No	1.0 -	1.0 -	-	-	-	0.004
Yes	1.82 (1.16, 2.85)	0.85 (0.64, 1.13)	-	-	-	

†Adjusted for age at diagnosis, year of diagnosis, deprivation and the other treatment exposures

Not all models estimated due to small numbers

Models exclude 36 with missing ethnicity (2 lymphoma, 3 CNS tumours, 22 germ cell tumours and 9 other solid tumours)

8.3.5 General population rates

It is not possible to estimate the cumulative burden in relation to that in the general population as no reference data exist combining cancer incidence and cardiovascular and respiratory hospitalisations However, publically available national data are available by age group showing the number of hospital episodes for cardiovascular and respiratory conditions in England (Figure 8.8) and the number of new cancer cases diagnosed in the UK (Figure 8.9).

There are a higher number of admission for respiratory disease compared to cardiovascular disease up to age 40-44 years. From age 45 onwards admission for both disease types increase, but admissions for cardiovascular diseases increase at a faster rate. At age 60-64 years there are twice as many admissions for cardiovascular disease compared to respiratory disease (Figure 8.8). The number of new cancer cases increases with increasing age, with a steep increase from around age 45 onwards (Figure 8.9).

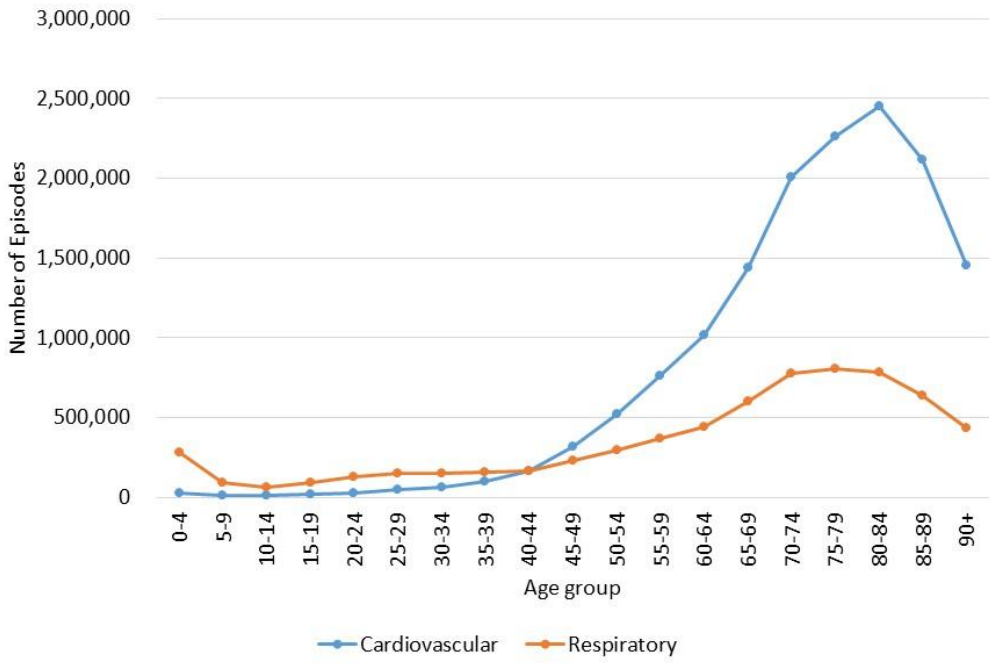


Figure 8.8: Number of HES episodes for cardiovascular and respiratory disease by age group, England, 2017-2018.

Source: [294]

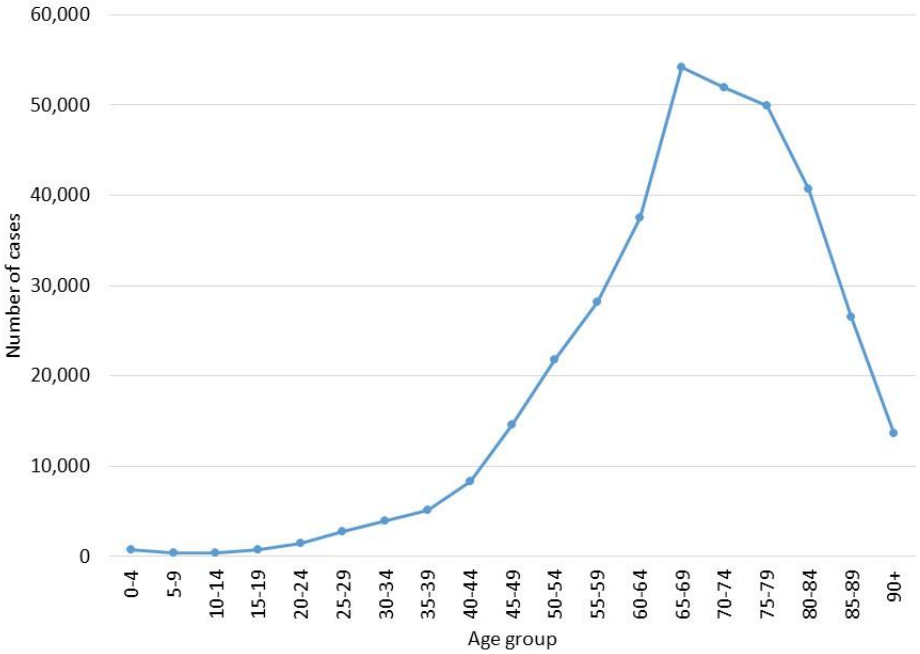


Figure 8.9: Number of new cancer cases diagnosed by age group, UK, 2014-2016.

Source: [2]

8.4 Discussion

8.4.1 Results in context

Cardiovascular conditions are one of the most common late effects in long-term survivors. Cardiovascular late effects in long-term survivors of cancer in children and young adults within Yorkshire have been described previously based on linked cancer registrations and hospital admissions [196, 281], therefore a thorough investigation of the relationship with previous treatment was not conducted here. However, the results presented in this chapter included a longer follow-up period than previously reported. The cumulative incidence and comparison of the incidence with the general population were reported for specific cardiovascular conditions in both children and TYA.

By age 40, the cumulative incidence for an admission for a cardiovascular condition was 24%. Previously reported cumulative incidences for the Yorkshire cohort were based on time from cancer diagnosis rather than attained age with a cumulative incidence of 8% for children and 14% for TYA 20-years post diagnosis [196]. The findings reported here are similar to other studies reporting the cumulative incidence by attainted age: from the AliCCS study, the

cumulative incidence for hospitalisations for cardiovascular conditions was 18% by age 40 [197], St Jude Life study report cumulative incidence of 17% for a severe cardiovascular event by age 40 [40]. Both these studies included those less than 18/20 years of age at diagnosis. Similarly to the results for respiratory admissions included in Chapter 6, restricting the results to the primary diagnostic code for each admissions reduces the cumulative incidence substantially, mainly driven by hypertension being recorded as a comorbidity.

Admissions for cardiovascular disease were almost twice as likely in cancer survivors compared to the general population, with a higher excess risk observed in children (HRR=3.1, 95% CI 2.6 to 3.7) compared to TYA (HRR=1.4, 95% CI 1.2 to 1.5). Previously reported results from Yorkshire (with a slightly different cancer diagnosis period and shorter follow-up period) reported admission rates were 2.6 (95% CI 1.9 to 3.6) times higher than the general population for children and for TYA admissions were 1.2 (95% CI 0.9 to 1.5) times higher but the excess risk in TYAs was not statistically significantly higher than in the general population [196]. Other studies based on hospital admissions for cardiovascular conditions also reported relative risks of admissions around 2 times higher than population controls [197, 203].

The second part of analysis presented in this chapter described the combined burden of hospitalisations for respiratory and cardiovascular diseases and SMNs in long-term survivors of childhood and young adult cancers. Long-term survivors experience a high burden of morbidity due to these conditions; by age 40-years each survivor experienced an average of 2 events. Respiratory admissions accounted for the largest proportion of these. After adjustment for potential confounders, there was a significant increased cumulative burden for those treated with pulmonary toxic chemotherapy but not for anthracyclines and radiation to the chest.

Overall there was no difference in the mean number of events by treatment group indicating that the burden of late effects is similar in both groups. There may be several explanations for this. The analysis was based on all cancers combined which may mask any differences by treatment within diagnostic groups. For example, further stratification by diagnostic group showed that the MCC was lowest for germ cell tumours, and treatment varies by diagnostic group. However, further stratification by diagnostic group and treatment was not possible due to small sample size. The SJLIFE study, which included 10-year

survivors and was based on clinical assessment and medical record validation, reported a cumulative burden by age 40 years of 2.55 for cardiovascular conditions, 0.75 for respiratory conditions and 0.45 for SMNs giving a total cumulative burden for these three conditions combined of 3.75 per individual [40]. This compares with our findings of the cumulative burden of 2.14 events per individual based upon hospital admissions including both the primary admission diagnosis as well as any recorded co-morbidities. As discussed in Chapter 6 in relation to respiratory admission hospital admission data capture the severe end of the disease spectrum and many respiratory and cardiovascular conditions will be managed and treated within a primary care setting. Therefore these findings may be a potential underestimation of the true extent of disease burden.

These findings provide an estimate of the health care usage due to these three conditions within long-term survivors which are useful to inform clinical guidelines and health service planning. The analysis was restricted to three disease areas that make significant contributions to the most common causes of late mortality within long-term childhood cancer survivors. The cumulative burden of long-term conditions has been assessed in other studies but based upon specific subgroups or outcomes. The MCC was used to quantify long-term conditions in survivors of Hodgkin lymphoma only [205, 295], to estimate the cumulative burden of SMNs [175], to examine differences in cumulative burden by ethnicity [296], haematopietic stem cell transplantation [297], and to estimate hospitalisations for neurologic disorders in CNS survivors [298]. In the calculations of the MCC every event was weighted equally for example, an admission where asthma was recorded as a comorbidity was given the same weight as the diagnosis of a SMN. From the cumulative incidence it can be seen that respiratory admissions had the largest contribution to the MCC. Other studies have shown that childhood and young adult cancer survivors have an increased risk of hospitalisations for cardiovascular and respiratory conditions compared to the general population, these studies have been based on time to first admission only and no estimates are given of the cumulative burden of all hospitalisations [199, 201-204, 291]. A Dutch study of hospitalisation in long-term childhood cancer survivors which included all admissions throughout follow-up found that those treated with surgery or radiotherapy had the highest rates of hospitalisation [200].

After adjustment for age at diagnosis, diagnostic group, deprivation, anthracycline exposure and radiation to the chest, it was found that those

treated with pulmonary toxic chemotherapy drugs had an increased cumulative burden of events. These findings concur with previous analysis based in Yorkshire on single late effects. No association was found between anthracyclines or chest radiation and admission for cardiovascular disease [196]. Results from Chapter 6 on respiratory admissions showed that respiratory admissions were associated with prior treatment with pulmonary toxic chemotherapy, while radiation to the chest was associated with an increased risk of hospitalisation in children but not TYA. In the UK, comprehensive risk stratified follow-up has been implemented for childhood and adolescent cancer survivors using a three tiered model based on cancer type and treatment intensity [299-301] and has been shown to predict increasing levels of moderate to severe late effects with increasing risk levels [302, 303]. Further work incorporating these risk stratification groups with the cumulative burden described here is needed.

8.4.2 Strengths and limitations

Key strengths of this analysis are that objective outcomes were used and the study was population-based. The MCC was used to estimate the cumulative burden, which is an important metric to quantify disease burden over time incorporating multiple and recurrent morbidities [304]. This measure accounts for the total person-time at risk and censoring within a competing risk framework, in this case death was considered a competing risk. An alternative would be to simply count the total number of events observed for each case, however this method would not account for differing follow-up periods or translate into an easily interpretable count per individual.

All diagnostic codes within each HES episode were used to identify respiratory and cardiovascular admissions, as shown in Chapter 6 for respiratory admissions and in this chapter for cardiovascular admissions, the cumulative incidence is reduced substantially when only the primary admission code was used. However, in order to capture all conditions and comorbidities all diagnostic codes were used in the cumulative burden analysis. A limitation of this is that all events, (SMN diagnoses and hospital admissions) are all weighted equally in the MCC, ideally different weightings could be used depending on the severity of the condition, and in the case of hospitalisations, if the diagnosis was the primary admission or a recorded comorbidity. This requires substantial further work to establish a criteria for weighting each condition and currently statistical methods accounting for different weightings in

the MCC calculation are not available. This is an important area for further research.

No reference data exist combining cancer incidence and cardiovascular and respiratory hospitalisations, therefore it was not possible to estimate the excess burden in cancer survivors compared to the general population. However, these conditions were selected based on prior knowledge of the increased risks compared to the general population [36, 37, 79, 81]. Furthermore, earlier results from this chapter for cardiovascular disease, Chapter 6 for respiratory disease and Chapter 7 for SMNs shows the excess risks for each specific outcome compared to the general population. Examination of national age-specific admission patterns and cancer incidence shows the risks increase steeply in the general population from mid 40s/50s onwards. With further follow-up of the Yorkshire cohort to older ages morbidity rates will increase, further highlighting the need for life-long follow-up care for this population of long-term survivors.

Another further limitation of this study is that in statistical models, death is considered as an event, again given equal weighting as hospitalisation and SMN diagnosis. An alternative would have been to consider deaths as a censoring event, but this would not be appropriate in this case. Right censoring should be non-informative meaning that it should not be related to the outcome of interest, however, if an individual dies they are unable to experience any further events. Therefore it was decided to classify deaths as an event and follow-up ended for each individual at this time. This approach has been recommended in analysis of hospital admission data [280]. An alternative approach would be to consider these models in a competing risk frame work using other more complex statistical methods such as multi-state models [268]. This issue is discussed further in Chapter 9 as an area for further research.

A final limitation of the modelling approach to acknowledge is that in the PWP multiple-failure time models only the first 10 admissions for both respiratory and cardiovascular conditions were included. The PWP models may fail to converge if the risk sets for recurrent events are too small [279]. Ten events was chosen as the threshold based on the included data, however less than 0.5% of the study population had more than 10 admissions for respiratory and cardiovascular admission so the loss of information for these individuals was minimal.

8.4.3 Summary

Cardiovascular admissions were almost twice as likely in cancer survivors compared to the general population, with a higher excess risk observed in children compared to TYA. The cumulative burden of all admissions for respiratory and cardiovascular conditions and all SMNs was estimated and by age 40, an individual experienced and average of 2 events. There was a significant increased cumulative burden for those treated with pulmonary toxic chemotherapy but not for anthracyclines or radiation to the chest. The next and final chapter of this thesis is the discussion chapter draws the final conclusions of this thesis together.

Chapter 9 Discussion

9.1 Introduction

This study describes survival trends and long-term outcomes for childhood and young adult cancer survivors using high-quality population-based cancer registry and linked hospital admission data, to provide vital intelligence to aid our understanding of the long-term health of young people diagnosed with cancer.

Chapters 5-8 of this thesis provide the evidence to the address the three aims set out in Chapter 1 and gaps in the current literature described in Chapter 2. A detailed discussion is provided at the end of each chapter summarising the key findings in the context of the current published literature and identifies the key strengths and limitations of each analysis. In this final discussion chapter the novel findings which have arisen from this thesis are summarised below with respect to the original aims, the clinical implications are discussed, strengths and limitations of the work identified and finally areas for further research are considered.

Aim 1: To assess the feasibility of applying cure models to CYA diagnosed with cancer using data from a regional population-based specialist cancer register

Cure models were successfully implemented to cancer registration data for children and young people using data from the YSRCCYP database (Chapter 5). Cure models provide additional metrics useful to identify and describe trends in survival. Additional measures include the percentage cured which is a summary of long-term survival and the median survival time of the uncured which gives information on those who are not long-term survivors. A flexible parametric cure model was used to obtain estimates of the proportion cured and the median survival time of the uncured by tumour type, over time and for children and TYAs. Generally for most diagnostic groups there was an improvement in survival over time which was mainly driven by an increase in the proportion of patients cured rather than an increase in the survival time of the uncured. Further detailed clinical risk factor data were incorporated into cure models for children with ALL. These models also revealed that the proportion

cured increased over time while there was little change in the median survival time of the uncured. By incorporating information on relapse, population-based estimates of event-free survival were also estimated for children diagnosed with ALL.

Aim 2: To evaluate the long-term health outcomes for CYA cancer survivors

Linkage of cancer registrations and hospital admissions data were used to address Aim 2 with a particular focus on three outcomes that contribute most to late mortality and morbidity in long-term child and young adult cancer survivors.

a. Respiratory morbidity

This was addressed using hospitalisations for respiratory conditions occurring at least 5-years post diagnosis (Chapter 6). The cumulative incidence of hospitalisation for respiratory conditions increased with attained age reaching 50% by age 40. Long-term survivors were hospitalised for respiratory conditions twice as often as general population comparisons; this excess risk varied by respiratory disease type and was greater for those diagnosed in childhood aged 0-14 compared to 15-29 year olds. Pulmonary toxic chemotherapy was associated with an increased risk of admissions for all respiratory disease especially pneumonia, while radiation to the chest increased the risk of admissions in children but not for TYA. Long-term survivors admitted for pneumonia had an increased risk of subsequent death following admission compared to those admitted for other types of respiratory disease.

b. Subsequent malignant neoplasm

This analysis was based on cancer registrations ascertained from the YSRCCYP and NCRAS presented in Chapter 7. Subsequent malignant neoplasms developed in 3% of the study population, with 36% of these occurring within 5-years of the primary diagnosis. The cumulative incidence continued to increase throughout follow up, reaching 4% twenty-years post diagnosis and patterns were similar for children and TYA. Compared to the general population the incidence of subsequent tumours was 5 times higher than expected, with a higher excess risk for children than for TYA. A shorter latency period was associated with decreased survival.

c. Cardiovascular morbidity

Linked cancer registration and hospital admission data were used to estimate the cumulative incidence of cardiovascular morbidity resulting in hospitalisation and comparison with the general population (Chapter 8). By age 40, the cumulative incidence for an admission for a cardiovascular condition was 24%. Admissions for cardiovascular disease were almost twice as likely in cancer survivors compared to the general population, with a higher excess risk observed in children compared to TYA.

Aim 3: To assess the cumulative burden of subsequent tumours, cardiovascular and respiratory morbidity for CYA cancer survivors

Novel methodology incorporating multiple and recurrent events was used to estimate the cumulative burden of combined admissions for respiratory and cardiovascular conditions and SMN diagnoses (Chapter 8). The mean cumulative count was used to estimate the total burden of all hospitalisations for respiratory and cardiovascular disease and all SMNs within a competing risks framework while also accounting for the total person-time at risk. By age 40, an individual experienced and average of 2 events, mainly driven by respiratory admissions. After adjustment for potential confounders, there was a significant increased cumulative burden for those treated with pulmonary toxic chemotherapy but not for anthracyclines or radiation to the chest.

9.2 Originality of study findings

As identified in the literature review in Chapter 2, there are limited studies based on cure models for children and TYA diagnosed with cancer. Cure models were systematically included for all diagnostic groups within children and young people. Furthermore, for children with leukaemia clinical prognostic risk factors, including cytogenetic risk groups were incorporated into a cure model to provide estimates of long-term survival and cure. This work was published in the British Journal of Haematology in 2018 [305].

This is the first study to describe the relationships between treatment exposures and hospital admissions for specific respiratory condition in survivors of CYA cancers. This work was published in the International Journal of Cancer in 2019 [306]. SMNs have not been examined within the Yorkshire register previously. The focus on this analysis was on early as well as late onset tumours and the impact of latency on survival.

Many studies of the late effects in childhood cancer survivors focus on single disease areas and time to first event. However, long-term survivors are at increased risk of multiple morbidities and recurrent disease. This is the first study to examine the combined burden of all hospital admissions for respiratory and cardiovascular disease and SMNs in long-term survivors using statistical methods that account for multiple and recurrent events. This work has been submitted to the International Journal of Cancer (September 2019) and is currently under review.

9.3 Clinical implications of study findings

Five-year survival estimates are commonly used as a measure of comparison for survival, often across different groups of patients, for example, by age, sex or geographical area. These five-year survival rates are often used a measure of cure [307], however for some cancer sites this period of follow-up may be too short to fully assess cure particularly for childhood cancers with high survival rates [308, 309]. Indeed, national survival figures reported by the ONS for childhood cancer now also include 10-year survival estimates [12]. Cure models provide additional metrics to those obtained from standard survival analysis by estimating the proportion of patients cured and whose life expectancy is the same as the general population. These metrics are of interest to many groups including clinicians, epidemiologists, public health professionals and cancer patients themselves. The work presented in this thesis provides estimates for long-term prognosis. Variation in the percentage cured ranged from 40% for children diagnosed with neuroblastoma to 95% for TYA with germ cell tumours.

Despite high survival and cure rates, survivors of CYA cancers are at increased risk of late effects of treatment and in this sense may never be defined as being cured of their cancer [309, 310]. Two separate meeting of specialists, including oncologists, psychologists, general practitioners, epidemiologists and cancer survivors, have been convened to establish a definition of cure: one for long-term survivors of childhood cancer in 2006 [311] and one for long-term survivors of adult cancer in 2014 [312]. From both these meetings a consensus statement was published to define what is meant by cure. For childhood cancer survivors cure refers to cure from the original cancer regardless of any potential for, or presence of, remaining disabilities or side effects of treatment. Children treated for cancer can be considered cured when the chance that they die from their cancer is no greater than that of age matched peers in the general population

dying from any cause [311]. For long-term survivors of adult cancers a patient can be defined as cured if their life expectancy is the same as that of the age and sex matched general population [312]. Both these definitions are based on long-term survival but do not take into account the quality of life of the surviving cancer patient or the late side effects of treatment and are in essence the same definition of cure that is measured at the population level from a cure model. However, for cancer survivors cure may mean the return to the potential for a normal life [313]. Therefore, in addition to estimating survival and cure for CYA cancer survivors it is essential to examine long-term quantitative and qualitative health outcomes of this group of survivors.

Life-long clinical follow-up of childhood and adolescent cancer survivors is recommended [72]. In the UK, a risk stratified three-level aftercare model has been implemented based upon cancer type and treatment intensity [299-301]. The model has been shown to predict increasing levels of morbidity with increasing risk levels with levels 1 and 2 having substantially lower levels of follow-up care at hospital compared to individuals in level 3 [303]. In the current climate of limited resources within the NHS in England, follow-up care service need to be planned appropriately to ensure the best outcomes possible for individuals. The work presented in this thesis adds to the evidence base, particularly around health service usage, identifying those at greatest risk of presenting at hospital with cardiovascular and respiratory conditions, which will aid strategies for prevention and early identification and treatment for long-term health conditions. There is a growing need for increased self-management of conditions with support and access to specialist services when required. For example, for respiratory health it is vitally important long-term survivors do not smoke, are a healthy weight and regularly exercise. However if presenting with symptoms of cough they may need appropriate follow-up and access to specialist services.

Teenagers and young adults with cancer are a unique group with growing recognition that the late effects of treatment experienced in this group should not be based on extrapolation of findings of studies in long-term childhood cancer survivors [67, 73]. Where possible in this thesis outcomes were considered separately for children (aged 0-14 years) and TYA (aged 15-29 years). For respiratory admissions, subsequent tumours and cardiovascular admissions the excess risks compared to the general population (measured by hospitalisation rate ratios or standardised incidence rate) were greater in

children than in TYA. Further analysis investigation the association between treatment exposure and later outcomes, observed that radiation to the chest was associated with an increased risk of respiratory admission and total cumulative burden in children but not in TYA.

Many studies of the late effects in childhood cancer survivors focus on single disease areas, however long-term survivors are at increased risk of multiple morbidities and recurrent disease. Assessing the impact of multimorbidity is a growing area of research and an emerging priority for health care services [314, 315]. Common epidemiological measures such as the incidence or prevalence of single conditions do not adequately capture the wider burden of late effects in long-term cancer survivors and studies which capture complex disease trajectories are needed [304]. The results presented in Chapter 8 identified that long-term survivors experienced a high disease burden due to respiratory and cardiovascular disease and SMNs. Long-term follow up care is needed that accounts for the complexity of health needs for this high risk population.

9.4 Strengths and limitations

9.4.1 Data quality

This study used high-quality population-based data. The YSRCCYP has high ascertainment and minimal loss to follow-up. Approximately 5550 children and young people with cancer were included in the full cohort available for inclusion in analysis which included 4200 long-term survivors. This included those diagnosed over a 21 year period from 1990 to 2011 with a minimum follow-up period of 5-years.

Key strengths of this work were the availability and inclusion of clinical data such as treatment and relapse information and additionally for ALL patients cytogenetic risk groups. Population-based estimates for event-free survival (based on time to relapse) for ALL patients with the YSRCCYP cohort were similar to those reported from national ALL clinical trials [229, 253, 254, 256] providing evidence of the validity of the estimates and completeness of the ascertainment of relapse data within the study sample.

While detailed treatment data were available for inclusion in statistical analysis (including identification of specific chemotherapy drugs) there was a lack of dose information meaning that more detailed treatment exposure data could not be included.

The study sample included all those diagnosed with cancer under the age of 30, providing vital evidence for the TYA age group in addition to children. There are many studies of the late effects of childhood cancer survivors such as that from the BCCSS and CCSS which include those diagnosed under 15 year and 21 years respectively. However, studies examining late effects including teenagers and young adults are more limited. For example, previous studies of respiratory morbidity have only included those aged up to age 21 [209-211, 213]. The TYACSS study has shown that SMNs diagnosed in TYAs aged 15-39 years follow a different pattern to those developed in childhood survivors indicating that further follow-up and investigation of this unique group is needed. There is a growing international consensus that the TYA age range should include those aged 15-39 years [68]. However, for the analysis presented in this thesis data were only available for TYAs aged 15-29 years at diagnosis. A limitation of the analysis is that in most analyses only two broad age groups were considered (children vs TYA) rather than look at variation in outcomes within these age ranges. These age groups were chosen to ensure consistency in the analyses presented and sufficient numbers within each age group to obtain robust estimates. The only exception is the analysis of cure for ALL patients which included those aged 1-17 years and two age groups within this, 1-9 years and 10-17 years, which were selected based on risk stratification criteria [229, 253].

Morbidity outcomes were based on hospital admissions as an objective outcome measure, compared to other studies based upon self-reported outcomes which may suffer from recall and selection bias as they are reliant upon completion of questionnaires from responders. Or studies based on clinical assessment which are mainly single-centred and may pick up asymptomatic outcomes. However, there are several limitations to this type of analysis where outcomes are based upon HES data. Linkage to HES admission data were available for 90% of the study cohort. There were some differences in individuals linked and not linked to HES and results based on hospital admissions need to be interpreted with this in mind. Certain groups may be underrepresented in the analysis presented here including those diagnosed with germ cell tumours, who also have high survival rates, and those diagnosed at

older ages. This may results in an underestimation of the burden of late effects within these groups. Reassuringly there were no differences in linkage rates by deprivation.

Linkage of the YSRCCYP to HES data was conducted by NHS Digital using standard procedures including four unique identifiers: NHS number, date of birth, sex and postcode. Despite this it is still possible that there is a mismatch within the linkage process and the cancer registration record was incorrectly matched to the wrong HES data. These errors are likely to be small. Those not linked to any HES records include a mix of those diagnosed with cancer who never engage with health services, those whose cancers do not require an inpatient hospital stay, those admitted to private hospitals and those not linked due to linkage errors. The linkage rate for the cohort included in this thesis is comparable to other English cancer registry linked HES admission studies [240, 281, 282].

HES data were available from 1997 onwards, study patients who were diagnosed in 1990-1991 did not start follow-up exactly 5-years from diagnosis but shortly after when admission data were available (potentially 6-7 years post diagnosis). This may bias the findings slightly, however, these individuals were included in analysis and the statistical methods used accounted for the person-time at risk. The linkage rate to HES was also slightly lower for these individuals as no records of admission were available around the time of their diagnosis and treatment. Therefore for those diagnosed in the earlier time period the identification of late effects based on hospital admissions may be an underestimation of morbidity.

The analysis of hospital admission was based on 5-year survivors and included admissions 5-years post diagnosis only. An obvious limitation of this is that any admissions in the time period from the end of treatment to 5-years post diagnosis are not captured. The cumulative incidence plots, particularly for respiratory admissions shows a large jump immediately 5-year post diagnosis which suggests there are many admissions prior to this which were not included, again perhaps suggesting an underestimation of the true risk of late effects in this cohort.

HES data are not collected primarily for research purposes but are the basis for “Payment by Results” for hospitals to be reimbursed for the care they provide [316]. This has implications for the accuracy and coding of admissions which has improved over time with more diagnostic codes being used and improvements in coding accuracy [225]. Diagnostic coding practices may also vary between hospitals particularly with the recoding of comorbidities. For both respiratory and cardiovascular outcomes sensitivity analysis was conducted comparing admissions based on the primary diagnosis and including all diagnostic codes with the biggest differences in cumulative incidence found for asthma and hypertension which were more commonly recorded as a comorbidities rather than the primary reason for admission.

A further limitation of using hospital admissions as an outcome measure is that hospital admissions are likely to measure the severe end of the disease spectrum whilst many respiratory and cardiovascular conditions will be managed and treated within a primary care setting. However, conditions such as asthma and hypertension were commonly recorded as comorbidities in HES as observed in the analysis of respiratory and cardiovascular admissions. An alternative approach would have been to use HES Outpatient records. On initial inspection of these data over 99% of records had the diagnosis field coded as “Other ill-defined and unspecified causes of morbidity”, meaning limited analyses could be undertaken on these data. Furthermore, outpatient data were only available from 2003 onwards.

A further advantage of using HES data for outcomes was that data on general population controls were available. Admissions for the same disease types were matched by age, sex and admission year for the Yorkshire region so that the excess risk in the cancer survivor cohort could be estimated. This is advantageous over studies reliant upon sibling controls as siblings will have been exposed to the same genetic and similar environmental risks as the cancer survivor therefore the true excess risk may be underestimated.

Follow-up was limited to examine late effects in those aged up to 56 years only. Further follow-up of this cohort to older ages is needed as the risk for cardiovascular and respiratory disease and cancers increase in the general population beyond this point in life (as shown in Figure 8.8 and Figure 8.9). Analysis presented here included long-term survivors diagnosed since the 1990s. This include those more recently diagnosed compared to other studies:

the BCCSS includes those diagnosed since 1940 [80], the CCSS since 1970 [78] and TYACSS since 1971 [317]. Treatment for childhood cancers are evolving to reduce late-effects while maintaining high survival rates therefore the estimates of the late effects obtained from these studies and indeed the results presented in this thesis may not be applicable to children and young people diagnosed with cancer today.

Finally, the results generated from this thesis are based upon one region only. However, the Yorkshire region population comprises 2 million under 30 years and is comparable in size to Scandinavian countries such as Denmark and Norway and is representative of the England and Wales in terms of childhood cancer incidence [318] therefore these results should be generalizable to the wider population. However, some analyses were limited by small sample size when examining subgroups, for example some temporal trends by diagnostic group, some two-way interactions between age group and treatment exposures and the analysis by cytogenetic risk group for ALL patients.

9.4.2 Modelling issues

The main strengths of the analysis presented in this thesis are the use of appropriate statistical methodology to address study aims as described in detail in Chapter 3. However, the limitations of each of these methods need to be acknowledged.

A comparison of the mixture, non-mixture and flexible parametric cure models showed consistency in the predicted percentage cured but differences in the survival time of the uncured group. The flexible parametric cure model was selected to conduct further analysis as the survival of the uncured using this model more closely predicted the survival of those that died than the other methods and this model is recommended over the other models when survival is high [28] as in the case when considering CYA cancer survival. However, the survival estimates of the uncured should be interpreted with caution given that it is based on a relatively small sample size. The small sample size for some diagnostic groups when stratified by age group and period of diagnosis limited further analyses.

Time to cure was not predicted for the cure models as this is not recommended for use with the flexible parametric cure model due to placement of the last knot at the last observed death time to allow estimation of the percentage cured [123]. The time to cure varied by diagnostic group as shown in the Kaplan-Meier plots of survival included in Section 4.4. In this study each individual had at least 5-years of follow-up data but a longer follow-up period may be needed for some diagnostic groups, for example CNS tumours in TYAs, to obtain more robust estimated of the proportion cured.

The selection of potential confounders was based on causal inference methodology using DAGs to make clear the underlying assumptions. The DAGs were developed based on published evidence and clinical expertise but incorrect specification of the DAG may result in incorrect adjustment for potential confounders and statistical inference. However, this method was chosen over other methods to select confounders for adjustment such as methods based on p-values or stepwise regression procedures which are not recommended as they do not incorporate the causal relationships of interest or adequately control for confounding [266]. By using DAGs the assumptions underlying each model are explicit, which aids transparency and reproducibility of the research [264, 266].

Data on lifestyle factors, such as smoking, obesity and exercise, and genetic rick factors, including family history and genetic predisposition, were not available to include as potential confounders in adjusted models. However, the choice of confounder adjustment sets were based upon DAGs, where these confounders were included in the DAGs as unmeasured latent variables. This ensured that the selected adjustment set did not include these variables, although they were accounted for in the underlying causal structure of model assumptions. The major limitation of not having information on lifestyle or genetic risk factors is that it is not possible to assess the impact of these on late-effects. However, that was not the focus on the analysis presented in this thesis and the interest was mainly in assessing the association between treatment exposures and late effects.

For most variables included in the statistical models data were complete, however, for ethnicity and stage missing data was an issue. Ethnic group was based on a combination of HES data and the naming algorithm software Onomap. In theory by using this approach an ethnic group should have been

assigned to all individuals in the study sample, but the licence for the Onomap software expired in July 2016 at which point not all individuals in the cohort had been assigned an ethnic group based on the results from this software. Therefore a small percentage (<2%) of individuals did not have an ethnic group assigned. These individuals were excluded from models which adjusted for ethnicity. Sensitivity analysis was conducted running models without adjusting for ethnic group and results were similar. A new version of the Onomap software should be available soon (https://www.onomap.org/) at which time and ethnic group could be assigned to all individuals.

The levels of missing data by stage varied by diagnostic group from 20% for leukaemia to 43% for both lymphoma and germ cell tumours. For some diagnostic groups no stage data were available at all. Although stage was included in all DAGs as a potential cofounder minimum adjustment sets that included variables other than stage were selected. An alternative to this would be to use missing data techniques to impute stage for those with missing values, however this was out with the scope of this project and is discussed further in recommendations for further research (Section 9.5).

The work presented on cumulative burden used the mean cumulative count to quantify the cumulative disease burden within a competing risk framework accounting for death, and multiple failure-time survival models to examine the association with previous cancer treatment. The advantage of using these methods are that multiple and recurrent events can be examined and analyses are not limited to the first occurrence of an event, while also incorporating the total person-time at risk for each individual. The main limitation of these analyses are that all events are weighted equally which may not be appropriate. For example if an individual is admitted to hospital for pneumonia this is weighted the same as an admission for another condition where hypertension is recorded as a co-morbidity. In the Prentice, Williams and Petersen models death was also considered as an event, again given an equal weighting to all other events. Methods that assign weights to different outcomes are not readily available for implementation in standard statistical software packages and is a potential area for further research. Alternative methods, such as multi-state modelling, may also be applicable and are discussed further in Section 9.5. Sensitivity analysis was conducted looking at time to first admission for each outcome using standard competing risk regression (where death was treated as

a competing risk) and a similar association between treatment exposures and outcomes was found providing validity and robustness to the study findings.

9.5 Recommendations for future research

High quality data are essential to fully explore the complex disease trajectories of long-term CYA cancer survivors. Data linkage of routinely collected clinical datasets offers a rich data source to explore these pathways. Advantages of using routinely available data are that it is population-based with objective outcomes, and places no additional burden on individuals compared to traditional patient-completed questionnaires. Further work should explore the feasibility of linkage to other datasets such as primary care data. Research evaluating the consistency and triangulation between incidence of late effects based on hospital admissions, primary-care and self-reported data are needed. Further research utilising prescribing data could also be undertaken as a measure of long-term morbidity. Such studies have been undertaken in Finland for example, on long-term CYA cancer survivors in relation to medication for the metabolic syndrome and cardiovascular conditions [319, 320].

The analysis conducted in this thesis was undertaken within a causal inference framework to assess relationships between treatment exposures and hospital admissions. Another area of future research is to develop a risk prediction model to identify which survivors are more likely to be admitted to hospital for respiratory and cardiovascular conditions.

The outcomes selected for inclusion in this study were chosen as these are the most common causes of late mortality and morbidity in CYA cancer survivors. However, other disease areas could be considered, both individually as outcomes and to be incorporated into the cumulative burden analysis.

This study included those age 0-29 years, the full AYA age range includes those 15-39 years and further work examining these outcome in the full 15-39 age range is needed.

In this study cytogenetic risk data were available for some leukaemia patients. Further work incorporating genetic, biological and molecular data with cancer

registration data are needed. This is particularly important in the context of personalised medicine where is it recognised that individualised and targeted therapies are needed to obtain the best outcomes for individuals [321].

Methods to deal with missing data are commonly used in epidemiological studies [231, 232]. However, methods to deal with missing data within a competing risk framework have so far focussed on cause-specific models [322, 323] and methods to deal with missing data applied to subdistribution hazards models are needed.

Finally the models used to explore the relationship between treatment exposures and cumulative burden were not conducted within a competing risk framework, with death treated as an event and given the same weightings as the other events of interest. An alternative approach for further consideration are multi-state models where an illness-death model would allow a distinction between multiple events (such as hospitalisations) and deaths to be incorporated [268, 279].

9.6 Conclusions

This thesis aimed to investigate the long-term survival and adverse health outcomes for children and young people diagnosed with cancer in Yorkshire. Generally for most diagnostic groups there was an improvement in survival over time which was mainly driven by an increase in the proportion of patients cured rather than an increase in the survival of the uncured. Despite the high survival rates long-term survivors have significant burden of adverse health outcomes in later life. The results presented provide an evidence base to aid risk-stratification for the long-term follow-up care for this high risk population.

References

1. Bray, F., Ferlay, J., Soerjomataram, I., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J. Clin., 2018. **68**(6): 394-424.

2. Cancer Research UK. *Cancer incidence statistics* http://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence. [Accessed: 13th September 2019]

3. Cancer Research UK. *Cancer Mortality Statistics.* http://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality. [Accessed: 13th September 2019]

4. Murphy, M.F., Bithell, J.F., Stiller, C.A., et al., *Childhood and adult cancers: contrasts and commonalities.* Maturitas, 2013. **76**(1): 95-8.

5. Kaatsch, P., *Epidemiology of childhood cancer.* Cancer Treat. Rev., 2010. **36**(4): 277-85.

6. Pritchard-Jones, K., Pieters, R., Reaman, G.H., et al., *Sustaining innovation and improvement in the treatment of childhood cancer: lessons from high-income countries.* Lancet Oncol., 2013. **14**(3): e95-e103.

7. Barr, R.D., Ferrari, A., Ries, L., et al., *Cancer in Adolescents and Young Adults: A Narrative Review of the Current Status and a View of the Future.* JAMA pediatrics, 2016. **170**(5): 495-501.

8. Birch, J.M., Alston, R.D., Kelsey, A.M., et al., *Classification and incidence of cancers in adolescents and young adults in England 1979-1997.* Br. J. Cancer, 2002. **87**(11): 1267-74.

9. Sender, L. and Zabokrtsky, K.B., *Adolescent and young adult patients with cancer: a milieu of unique features.* Nat. Rev. Clin. Oncol., 2015. **12**(8): 465-80.

10. Stiller, C., *Childhood Cancer in Britain*. 2007, Oxford: Oxford University Press.

11. Cancer Research UK. *Children's cancer statistics* http://www.cancerresearchuk.org/health-professional/cancer-statistics/childrens-cancers. [Accessed: 13th September 2019]

12. Office for National Statistics. *Cancer survival in England: adult, stage at diagnosis and childhood - patients followed up to 2018.* https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalinengland/stageatdiagnosisandchildhoodpatientsfollowedupto2018#childhood-cancer-survival-continues-to-improve. [Accessed: 13th September 2019]

13. O'Leary, M., Krailo, M., Anderson, J.R., et al., *Progress in childhood cancer: 50 years of research collaboration, a report from the Children's Oncology Group.* Semin. Oncol., 2008. **35**(5): 484-93.

14. Birch, J.M., Pang, D., Alston, R.D., et al., *Survival from cancer in teenagers and young adults in England, 1979-2003.* Br. J. Cancer, 2008. **99**(5): 830-5.

15. Cancer Research UK. *Young people's cancers statistics* *https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/young-peoples-cancers*. [Accessed: 13th September 2019]

16. O'Hara, C., Moran, A., Whelan, J.S., et al., *Trends in survival for teenagers and young adults with cancer in the UK 1992-2006.* Eur. J. Cancer, 2015. **51**(14): 2039-48.

17. Johnston, W.T., Lightfoot, T.J., Simpson, J., and Roman, E., *Childhood cancer survival: a report from the United Kingdom Childhood Cancer Study.* Cancer Epidemiol., 2010. **34**(6): 659-66.

18. Stark, D., Bowen, D., Dunwoodie, E., et al., *Survival patterns in teenagers and young adults with cancer in the United Kingdom: Comparisons with younger and older age groups.* Eur. J. Cancer, 2015. **51**(17): 2643-54.

19. Gatta, G., Zigon, G., Capocaccia, R., et al., *Survival of European children and young adults with cancer diagnosed 1995-2002.* Eur. J. Cancer, 2009. **45**(6): 992-1005.

20. Steliarova-Foucher, E., Stiller, C., Kaatsch, P., et al., *Geographical patterns and time trends of cancer incidence and survival among children and adolescents in Europe since the 1970s (the ACCIS project): an epidemiological study.* Lancet, 2004. **364**(9451): 2097-105.

21. Trama, A., Botta, L., Foschi, R., et al., *Survival of European adolescents and young adults diagnosed with cancer in 2000-07: population-based data from EUROCARE-5.* Lancet Oncol., 2016. **17**(7): 896-906.

22. Dickman, P.W. and Adami, H.O., *Interpreting trends in cancer patient survival.* J. Intern. Med., 2006. **260**(2): 103-17.

23. Ellis, L., Woods, L.M., Esteve, J., et al., *Cancer incidence, survival and mortality: explaining the concepts.* Int. J. Cancer, 2014. **135**(8): 1774-82.

24. De Angelis, R., Capocaccia, R., Hakulinen, T., et al., *Mixture models for cancer survival analysis: Application to population-based data with covariates.* Stat. Med., 1999. **18**(4): 441-454.

25. Sposto, R., *Cure model analysis in cancer: an application to data from the Children's Cancer Group.* Stat. Med., 2002. **21**(2): 293-312.

26. Lambert, P.C., Thompson, J.R., Weston, C.L., and Dickman, P.W., *Estimating and modeling the cure fraction in population-based cancer survival analysis.* Biostatistics, 2007. **8**(3): 576-94.

27. Othus, M., Barlogie, B., Leblanc, M.L., and Crowley, J.J., *Cure models as a useful statistical tool for analyzing survival.* Clin. Cancer Res., 2012. **18**(14): 3731-6.

28. Yu, X.Q., De Angelis, R., Andersson, T.M., et al., *Estimating the proportion cured of cancer: some practical advice for users.* Cancer Epidemiol., 2013. **37**(6): 836-42.

29. Shah, A., Stiller, C.A., Kenward, M.G., et al., *Childhood leukaemia: long-term excess mortality and the proportion 'cured'.* Br. J. Cancer, 2008. **99**(1): 219-223.

30. Gatta, G., Rossi, S., Foschi, R., et al., *Survival and cure trends for European children, adolescents and young adults diagnosed with acute lymphoblastic leukemia from 1982 to 2002.* Haematologica, 2013. **98**(5): 744-752.

31. Mertens, A.C., Liu, Q., Neglia, J.P., et al., *Cause-specific late mortality among 5-year survivors of childhood cancer: the Childhood Cancer Survivor Study.* J. Natl. Cancer Inst., 2008. **100**(19): 1368-79.

32. Reulen, R.C., Winter, D.L., Frobisher, C., et al., *Long-term cause-specific mortality among survivors of childhood cancer.* JAMA, 2010. **304**(2): 172-9.

33. Meadows, A.T., Friedman, D.L., Neglia, J.P., et al., *Second neoplasms in survivors of childhood cancer: findings from the Childhood Cancer Survivor Study cohort.* J. Clin. Oncol., 2009. **27**(14): 2356-62.

34. Olsen, J.H., Moller, T., Anderson, H., et al., *Lifelong cancer incidence in 47,697 patients treated for childhood cancer in the Nordic countries.* J. Natl. Cancer Inst., 2009. **101**(11): 806-13.

35. Reulen, R.C., Frobisher, C., Winter, D.L., et al., *Long-term risks of subsequent primary neoplasms among survivors of childhood cancer.* JAMA, 2011. **305**(22): 2311-9.

36. Geenen, M.M., Cardous-Ubbink, M.C., Kremer, L.C., et al., *Medical assessment of adverse health outcomes in long-term survivors of childhood cancer.* JAMA, 2007. **297**(24): 2705-15.

37. Oeffinger, K.C., Mertens, A.C., Sklar, C.A., et al., *Chronic health conditions in adult survivors of childhood cancer.* N. Engl. J. Med., 2006. **355**(15): 1572-82.

38. Phillips, S.M., Padgett, L.S., Leisenring, W.M., et al., *Survivors of childhood cancer in the United States: prevalence and burden of morbidity.* Cancer Epidemiol. Biomarkers Prev., 2015. **24**(4): 653-63.

39. Oeffinger, K.C. and Robison, L.L., *Childhood cancer survivors, late effects, and a new model for understanding survivorship.* JAMA, 2007. **297**(24): 2762-4.

40. Bhakta, N., Liu, Q., Ness, K.K., et al., *The cumulative burden of surviving childhood cancer: an initial report from the St Jude Lifetime Cohort Study (SJLIFE).* Lancet, 2017. **390**(10112): 2569-2582.

41. Woodward, E., Jessop, M., Glaser, A., and Stark, D., *Late effects in survivors of teenage and young adult cancer: does age matter?* Ann. Oncol., 2011. **22**(12): 2561-8.

42. Fritz, A., Percy, C., Jack, A., et al., *International Classification of Diseases for Oncology. Third edition*. 2000, Geneva: World Health Organisation.

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

44. Bleyer, W.A., *Cancer in older adolescents and young adults: Epidemiology, diagnosis, treatment, survival, and importance of clinical trials.* Med. Pediatr. Oncol., 2002. **38**(1): 1-10.

45. Barr, R.D., Holowaty, E.J., and Birch, J.M., *Classification schemes for tumors diagnosed in adolescents and young adults.* Cancer, 2006. **106**(7): 1425-1430.

46. Jensen, O.M., Parkin, D.M., MacLennan, R., et al., *Cancer Registration: Principles and Methods. IARC Scientific Publications No. 95*, International Agency for Research on Cancer. 1991: Lyon, France.

47. Parkin, D.M., *The evolution of the population-based cancer registry.* Nat. Rev. Cancer, 2006. **6**(8): 603-12.

48. Public Health England. *National Cancer Registration and Analysis Service (NCRAS)* https://www.gov.uk/guidance/national-cancer-registration-and-analysis-service-ncras. [Accessed: 13th September 2019]

49. National Cancer Registration and Analysis Service. *Cancer Outcomes and Services Dataset (COSD)* http://www.ncin.org.uk/collecting_and_using_data/data_collection/cosd. [Accessed: 13th September 2019]

50. Siesling, S., Louwman, W.J., Kwast, A., et al., *Uses of cancer registries for public health and clinical research in Europe: Results of the European Network of Cancer Registries survey among 161 population-based cancer registries during 2010-2012.* Eur. J. Cancer, 2015. **51**(9): 1039-49.

51. Steliarova-Foucher, E., Stiller, C., Colombet, M., et al., *Registration of childhood cancer: Moving towards pan-European coverage?* Eur. J. Cancer, 2015. **51**(9): 1064-79.

52. Birch, J.M., Marsden, H.B., and Swindell, R., *Incidence of malignant disease in childhood: a 24-year review of the Manchester Children's Tumour Registry data.* Br. J. Cancer, 1980. **42**(2): 215-23.

53. Craft, A.W., Amineddine, H.A., Scott, J.E., and Wagget, J., *The Northern region Children's malignant disease registry 1968-82: incidence and survival.* Br. J. Cancer, 1987. **56**(6): 853-8.

54. Muir, K.R., Parkes, S.E., Mann, J.R., et al., *Childhood cancer in the West Midlands: incidence and survival, 1980-1984, in a multi-ethnic population.* Clin. Oncol. (R. Coll. Radiol.), 1992. **4**(3): 177-82.

55. McKinney, P.A., Parslow, R.C., Lane, S.A., et al., *Epidemiology of childhood brain tumours in Yorkshire, UK, 1974-95: geographical distribution and changing patterns of occurrence.* Br. J. Cancer, 1998. **78**(7): 974-9.

56. Parkin, D.M., Kramarova, E., Draper, G.J., et al., *International Incidence of Childhood Cancer, Vol. II. IARC Scientific Publications No. 144*, International Agency for Research on Cancer. 1998: Lyon.

57. Stiller, C.A., *International patterns of cancer incidence in adolescents.* Cancer Treat. Rev., 2007. **33**(7): 631-45.

58. Barr, R.D., *Adolescents, young adults, and cancer--the international challenge.* Cancer, 2011. **117**(10 Suppl): 2245-9.

59. Ries, L.A.G., Smith, M.A., Gurney, J.G., et al., *Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975-1995*, National Cancer Institute SEER Program. 1999.

60. Public Health England, *Trends in five-year survival for teenagers and young adults with cancer in the UK*, NCIN data briefing. 2014.

61. Feltbower, R.G., Siller, C., Woodward, E., et al., *Treatment and survival patterns for germ cell tumors among 13- to 24-year olds in Yorkshire, UK.* Pediatr. Blood Cancer, 2011. **56**(2): 282-8.

62. Alston, R.D., Rowan, S., Eden, T.O., et al., *Cancer incidence patterns by region and socioeconomic deprivation in teenagers and young adults in England.* Br. J. Cancer, 2007. **96**(11): 1760-6.

63. Geraci, M., Eden, T.O., Alston, R.D., et al., *Geographical and temporal distribution of cancer survival in teenagers and young adults in England.* Br. J. Cancer, 2009. **101**(11): 1939-45.

64. Bleyer, A., O'Leary, M., Barr, R., and Ries, L.A.G., *Cancer Epidemilogy in Older Adolescents and Young Adults 15 to 29 Years of Age, Including SEER Incidence and Survival: 1975-2000*, National Cancer Institute. 2006.

65. van Laar, M., McKinney, P.A., Parslow, R.C., et al., *Cancer incidence among the south Asian and non-south Asian population under 30 years of age in Yorkshire, UK.* Br. J. Cancer, 2010. **103**(9): 1448-52.

66. Bleyer, A., Budd, T., and Montello, M., *Adolescents and young adults with cancer: the scope of the problem and criticality of clinical trials.* Cancer, 2006. **107**(7 Suppl): 1645-55.

67. Fidler, M.M., Frobisher, C., Hawkins, M.M., and Nathan, P.C., *Challenges and opportunities in the care of survivors of adolescent and young adult cancers.* Pediatr. Blood Cancer, 2019. **66**(6): e27668.

68. Trama, A., Botta, L., and Steliarova-Foucher, E., *Cancer Burden in Adolescents and Young Adults: A Review of Epidemiological Evidence.* Cancer J., 2018. **24**(6): 256-266.

69. Public Health England, *Childhood Cancer Statistics, England Annual report 2018. http://www.ncin.org.uk/view?rid=3715*. 2018.

70. Abrams, A.N., Hazen, E.P., and Penson, R.T., *Psychosocial issues in adolescents with cancer.* Cancer Treat. Rev., 2007. **33**(7): 622-30.

71. Stark, D., Bielack, S., Brugieres, L., et al., *Teenagers and young adults with cancer in Europe: from national programmes to a European integrated coordinated project.* Eur. J. Cancer Care (Engl.), 2016. **25**(3): 419-27.

72. National Institute for Health and Clinical Excellence, *Improving outcomes in children and young people with cancer*. 2005.

73. Gupta, S., *Adolescents and young adults with cancer and the risk of subsequent primary neoplasms: not just big children.* Lancet Oncol., 2019. **20**(4): 466-467.

74. Steliarova-Foucher, E., Fidler, M.M., Colombet, M., et al., *Changing geographical patterns and trends in cancer incidence in children and adolescents in Europe, 1991-2010 (Automated Childhood Cancer Information System): a population-based study.* Lancet Oncol., 2018. **19**(9): 1159-1169.

75. Rossi, S., Baili, P., Capocaccia, R., et al., *The EUROCARE-5 study on cancer survival in Europe 1999-2007: database, quality checks and statistical analysis methods.* Eur. J. Cancer, 2015. **51**(10): 2104-2119.

76. Gatta, G., Botta, L., Rossi, S., et al., *Childhood cancer survival in Europe 1999-2007: results of EUROCARE-5--a population-based study.* Lancet Oncol., 2014. **15**(1): 35-47.

77. Surveillence Epidemilogy and End Results Program. *http://seer.cancer.gov/*. [Accessed: 13th September 2019]

78. Robison, L.L., Armstrong, G.T., Boice, J.D., et al., *The Childhood Cancer Survivor Study: a National Cancer Institute-supported resource for outcome and intervention research.* J. Clin. Oncol., 2009. **27**(14): 2308-18.

79. Armstrong, G.T., Chen, Y., Yasui, Y., et al., *Reduction in Late Mortality among 5-Year Survivors of Childhood Cancer.* N. Engl. J. Med., 2016. **374**(9): 833-42.

80. Hawkins, M.M., Lancashire, E.R., Winter, D.L., et al., *The British Childhood Cancer Survivor Study: Objectives, methods, population structure, response rates and initial descriptive information.* Pediatr. Blood Cancer, 2008. **50**(5): 1018-25.

81. Fidler, M.M., Reulen, R.C., Winter, D.L., 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**: i4351.

82. Bright, C.J., Reulen, R.C., Winter, D.L., et al., *Risk of subsequent primary neoplasms in survivors of adolescent and young adult cancer (Teenage and Young Adult Cancer Survivor Study): a population-based, cohort study.* Lancet Oncol., 2019. **20**(4): 531-545.

83. Fidler, M.M., Reulen, R.C., Bright, C.J., et al., *Respiratory mortality of childhood, adolescent and young adult cancer survivors.* Thorax, 2018. **73**(10): 959-968.

84. Winther, J.F., Kenborg, L., Byrne, J., et al., *Childhood cancer survivor cohorts in Europe.* Acta Oncol., 2015. **54**(5): 655-68.

85. Tulla, M., Berthold, F., Graf, N., et al., *Incidence, Trends, and Survival of Children With Embryonal Tumors.* Pediatrics, 2015. **136**(3): e623-32.

86. Kaatsch, P., Steliarova-Foucher, E., Crocetti, E., et al., *Time trends of cancer incidence in European children (1978-1997): report from the Automated Childhood Cancer Information System project.* Eur. J. Cancer, 2006. **42**(13): 1961-71.

87. National Cancer Registration and Analysis Service and Teenage Cancer Trust, *13-24 year olds with cancer in England. Incidence, mortality and survival*, Public Health England. 2018: London.

88. Stiller, C.A., Desandes, E., Danon, S.E., et al., *Cancer incidence and survival in European adolescents (1978-1997). Report from the Automated Childhood Cancer Information System project.* Eur. J. Cancer, 2006. **42**(13): 2006-18.

89. Clark, T.G., Bradburn, M.J., Love, S.B., and Altman, D.G., *Survival analysis part I: basic concepts and first analyses.* Br. J. Cancer, 2003. **89**(2): 232-8.

90. Collett, D., *Modelling Survival Data in Medical Research*. 1994, London: Chapman & Hall.

91. Kaplan, E.L. and Meier, P., *Nonparametric-Estimation from Incomplete Observations.* J Am Stat Assoc, 1958. **53**(282): 457-481.

92. Machin, D., Cheung, Y.B., and Parmar, M., *Survival analysis: a practical approach*. 2006: John Wiley & Sons.

93. Royston, P. and Lambert, P.C., *Flexible parametric survival analysis using Stata: beyond the Cox model*. 2011, Texas: Stata Press.

94. Cox, D.R., *Regression models and life tables (with discussion).* Journal of the Royal Statistical Society, 1972. **34**: 187-220.

95. Cleves, M., *An introduction to survival analysis using Stata*. 2008, Texas: Stata Press.

96. Bradburn, M.J., Clark, T.G., Love, S.B., and Altman, D.G., *Survival analysis part II: multivariate data analysis--an introduction to concepts and methods.* Br. J. Cancer, 2003. **89**(3): 431-6.

97. Esteve, J., Benhamou, E., and L, R., *Statistical Methods in Cancer Research, volume IV. Descriptive Epidemiology*, in *IARC Scientific Publications No. 128*, IARC. 1994, International Agency for Research on Cancer: Lyon.

98. Sarfati, D., Blakely, T., and Pearce, N., *Measuring cancer survival in populations: relative survival vs cancer-specific survival.* Int. J. Epidemiol., 2010. **39**(2): 598-610.

99. Ederer, F., Axtell, L.M., and Cutler, S.J., *The relative survival rate: a statistical methodology.* Natl. Cancer Inst. Monogr., 1961. **6**: 101-21.

100. Cho, H., Howlader, N., Mariotto, A.B., and Cronin, K.A., *Estimating relative survival for cancer patients from the SEER Program using expected rates based on Ederer I versus Ederer II method. https://surveillance.cancer.gov/reports/tech2011.01.pdf*, in *Surveillance Research Program NCI Technical Report 2011-01*. 2011.

101. Rutherford, M.J., Dickman, P.W., and Lambert, P.C., *Comparison of methods for calculating relative survival in population-based studies.* Cancer Epidemiol., 2012. **36**(1): 16-21.

102. Hakulinen, T., Seppa, K., and Lambert, P.C., *Choosing the relative survival method for cancer survival estimation.* Eur. J. Cancer, 2011. **47**(14): 2202-10.

103. Dickman, P.W., Lambert, P.C., Coviello, E., and Rutherford, M.J., *Estimating net survival in population-based cancer studies.* Int. J. Cancer, 2013. **133**(2): 519-21.

104. Office for National Statistics. *National Life Tables: England, https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesenglandreferencetables*. [Accessed: 13th September 2019]

105. Perme, M.P., Stare, J., and Estève, J., *On estimation in relative survival.* Biometrics, 2012. **68**(1): 113-120.

106. Lambert, P.C., Dickman, P.W., and Rutherford, M.J., *Comparison of different approaches to estimating age standardized net survival.* BMC Med. Res. Methodol., 2015. **15**: 64.

107. Poole, J., Bannon, F., McPhail, S., et al., *Standard Operating Procedure: Guideline on Population Based Cancer Survival* 2016, United Kingdom and Ireland Association of Cancer Registries.

108. Dickman, P.W., Sloggett, A., Hills, M., and Hakulinen, T., *Regression models for relative survival.* Stat. Med., 2004. **23**(1): 51-64.

109. Royston, P. and Parmar, M.K., *Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects.* Stat. Med., 2002. **21**(15): 2175-2197.

110. Lambert, P.C. and Royston, P., *Further development of flexible parametric models for survival analysis.* Stata Journal, 2009. **9**(2): 265.

111. Peng, Y. and Taylor, J.M.G., *Cure Models*, in *Handbook of Survival Analysis*, J.P. Klein, et al., Editors. 2014. p. 113-134.

112. Boag, J.W., *Maximum likelihood estimates of the proportion of patients cured by cancer therapy.* Journal of the Royal Statistical Society. Series B (Methodological), 1949. **11**(1): 15-53.

113. Berkson, J. and Gage, R.P., *Survival curve for cancer patients following treatment.* J Am Stat Assoc, 1952. **47**(259): 501-515.

114. Lambert, P.C., *Modeling of the cure fraction in survival studies.* Stata Journal, 2007. **7**(3): 351-375.

115. Andersson, T.M., Dickman, P.W., Eloranta, S., and Lambert, P.C., *Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models.* BMC Med. Res. Methodol., 2011. **11**: 96.

116. Woods, L.M., Morris, M., and Rachet, B., *No 'cure' within 12 years of diagnosis among breast cancer patients who are diagnosed via mammographic screening: women diagnosed in the West Midlands region of England 1989-2011.* Ann. Oncol., 2016. **27**(11): 2025-2031.

117. Woods, L.M., Rachet, B., Lambert, P.C., and Coleman, M.P., *'Cure' from breast cancer among two populations of women followed for 23 years after diagnosis.* Ann. Oncol., 2009. **20**(8): 1331-1336.

118. Verdecchia, A., De Angelis, R., Capocaccia, R., et al., *The cure for colon cancer: Results from the EUROCARE study.* Int. J. Cancer, 1998. **77**(3): 322-329.

119. Smoll, N.R., Schaller, K., and Gautschi, O.P., *The cure fraction of glioblastoma multiforme.* Neuroepidemiology, 2012. **39**(1): 63-9.

120. Airtum Working Group, *Italian cancer figures, report 2014: Prevalence and cure of cancer in Italy.* Epidemiol. Prev., 2014. **38**(6 Suppl 1): 1-122.

121. Dal Maso, L., Guzzinati, S., Buzzoni, C., et al., *Long-term survival, prevalence, and cure of cancer: a population-based estimation for 818 902 Italian patients and 26 cancer types.* Ann. Oncol., 2014. **25**(11): 2251-60.

122. Tai, P., Yu, E., Cserni, G., et al., *Minimum follow-up time required for the estimation of statistical cure of cancer patients: verification using data from 42 cancer sites in the SEER database.* BMC Cancer, 2005. **5**: 48.

123. Andersson, T.M.L. and Lambert, P.C., *Fitting and modeling cure in population-based cancer studies within the framework of flexible parametric survival models.* Stata Journal, 2012. **12**(4): 623-638.

124. Cvancarova, M., Aagnes, B., Fossa, S.D., et al., *Proportion cured models applied to 23 cancer sites in Norway.* Int. J. Cancer, 2013. **132**(7): 1700-10.

125. Edlinger, M., Ulmer, H., Cvancarova, M., and Oberaigner, W., *Site-specific proportion cured models applied to cancer registry data.* Cancer Causes Control, 2014. **25**(3): 365-73.

126. Bejan-Angoulvant, T., Bouvier, A.M., Bossard, N., et al., *Hazard regression model and cure rate model in colon cancer relative survival trends: are they telling the same story?* Eur. J. Epidemiol., 2008. **23**(4): 251-9.

127. Eloranta, S., Lambert, P.C., Andersson, T.M., et al., *The application of cure models in the presence of competing risks: a tool for improved risk communication in population-based cancer patient survival.* Epidemiology, 2014. **25**(5): 742-8.

128. Eloranta, S., Lambert, P.C., Cavalli-Bjorkman, N., et al., *Does socioeconomic status influence the prospect of cure from colon cancer - A population-based study in Sweden 1965-2000.* Eur. J. Cancer, 2010. **46**(16): 2965-2972.

129. Ito, Y., Nakayama, T., Miyashiro, I., et al., *Trends in 'cure' fraction from colorectal cancer by age and tumour stage between 1975 and 2000, using population-based data, Osaka, Japan.* Jpn. J. Clin. Oncol., 2012. **42**(10): 974-83.

130. Shack, L.G., Shah, A., Lambert, P.C., and Rachet, B., *Cure by age and stage at diagnosis for colorectal cancer patients in North West England, 1997-2004: a population-based study.* Cancer Epidemiol., 2012. **36**(6): 548-53.

131. Francisci, S., Capocaccia, R., Grande, E., et al., *The cure of cancer: a European perspective.* Eur. J. Cancer, 2009. **45**(6): 1067-79.

132. Silversmit, G., Jegou, D., Vaes, E., et al., *Cure of cancer for seven cancer sites in the Flemish Region.* Int. J. Cancer, 2017. **140**(5): 1102-1110.

133. Gieser, P.W., Chang, M.N., Rao, P.V., et al., *Modelling cure rates using the Gompertz model with covariate information.* Stat. Med., 1998. **17**(8): 831-9.

134. Lee, J.W., *Two-sample rank tests for acceleration in cure models.* Stat. Med., 1995. **14**(19): 2111-8.

135. Andersson, T.M., Lambert, P.C., Derolf, A.R., et al., *Temporal trends in the proportion cured among adults diagnosed with acute myeloid*

leukaemia in Sweden 1973-2001, a population-based study. Br. J. Haematol., 2010. **148**(6): 918-24.

136. Nesbit Jr, M.E., Buckley, J.D., Feig, S.A., et al., *Chemotherapy for induction of remission of childhood acute myeloid leukemia followed by marrow transplantation or multiagent chemotherapy: A report from the Childrens Cancer Group.* J. Clin. Oncol., 1994. **12**(1): 127-135.

137. Shah, A., Andersson, T.M., Rachet, B., et al., *Survival and cure of acute myeloid leukaemia in England, 1971-2006: a population-based study.* Br. J. Haematol., 2013. **162**(4): 509-16.

138. Bouliotis, G. and Bessell, E.M., *Hodgkin disease (1973-2002): long-term survival and cure fractions.* Leuk. Lymphoma, 2015. **56**(5): 1278-1285.

139. Glimelius, I., Ekberg, S., Jerkeman, M., et al., *Long-term survival in young and middle-aged Hodgkin lymphoma patients in Sweden 1992-2009-trends in cure proportions by clinical characteristics.* Am. J. Hematol., 2015. **90**(12): 1128-34.

140. Sposto, R., London, W.B., and Alonzo, T.A., *Criteria for optimizing prognostic risk groups in pediatric cancer: Analysis of data from the Children's Oncology Group.* J. Clin. Oncol., 2007. **25**(15): 2070-2077.

141. Pfirrmann, M., Saussele, S., Hochhaus, A., et al., *Explaining survival differences between two consecutive studies with allogeneic stem cell transplantation in patients with chronic myeloid leukemia.* J. Cancer Res. Clin. Oncol., 2014. **140**(8): 1367-81.

142. Gatta, G., van der Zwan, J.M., Casali, P.G., et al., *Rare cancers are not so rare: the rare cancer burden in Europe.* Eur. J. Cancer, 2011. **47**(17): 2493-511.

143. Crocetti, E., Trama, A., Stiller, C., et al., *Epidemiology of glial and non-glial brain tumours in Europe.* Eur. J. Cancer, 2012. **48**(10): 1532-1542.

144. Gatta, G., Ferrari, A., Stiller, C.A., et al., *Embryonal cancers in Europe.* Eur. J. Cancer, 2012. **48**(10): 1425-1433.

145. Trama, A., Mallone, S., Nicolai, N., et al., *Burden of testicular, paratesticular and extragonadal germ cell tumours in Europe.* Eur. J. Cancer, 2012. **48**(2): 159-169.

146. Hunsberger, S., Albert, P.S., and London, W.B., *A finite mixture survival model to characterize risk groups of neuroblastoma.* Stat. Med., 2009. **28**(8): 1301-1314.

147. Ventura, L. and Mezzetti, M., *Estimating cancer incidence using a Bayesian back-calculation approach.* Stat. Med., 2014. **33**(25): 4453-68.

148. Wang, L., Du, P., and Liang, H., *Two-component mixture cure rate model with spline estimated nonparametric components.* Biometrics, 2012. **68**(3): 726-35.

149. Weston, C.L. and Thompson, J.R., *The definition of start time in cancer treatment studies analysed by non-mixture cure models.* Journal of Applied Statistics, 2009. **36**(1): 39-52.

150. Weston, C.L. and Thompson, J.R., *Modeling survival in childhood cancer studies using two-stage non-mixture cure models.* Journal of Applied Statistics, 2010. **37**(9): 1523-1535.

151. Rowland, J.H., Kent, E.E., Forsythe, L.P., et al., *Cancer survivorship research in Europe and the United States: where have we been, where are we going, and what can we learn from each other?* Cancer, 2013. **119 Suppl 11**: 2094-108.

152. Vassal, G., Schrappe, M., Pritchard-Jones, K., et al., *The SIOPE strategic plan: A European cancer plan for children and adolescents.* J Cancer Policy, 2016. **8**: 17-32.

153. NHS England, *Cancer Reform Strategy* 2007.

154. Richards, M., Corner, J., and Maher, J., *The National Cancer Survivorship Initiative: new and emerging evidence on the ongoing needs of cancer survivors.* Br. J. Cancer, 2011. **105 Suppl 1**: S1-4.

155. Maddams, J., Brewster, D., Gavin, A., et al., *Cancer prevalence in the United Kingdom: estimates for 2008.* Br. J. Cancer, 2009. **101**(3): 541-7.

156. Department of Health, MacMillan Cancer Support, and NHS Improvement, *Living with and Beyond Cancer: Taking Action to Improve Outcomes (an update to the 2010 The National Cancer Survivorship Initiative Vision)*. 2013.

157. Mertens, A.C., Yasui, Y., Neglia, J.P., et al., *Late mortality experience in five-year survivors of childhood and adolescent cancer: the Childhood Cancer Survivor Study.* J. Clin. Oncol., 2001. **19**(13): 3163-72.

158. Moller, T.R., Garwicz, S., Barlow, L., et al., *Decreasing late mortality among five-year survivors of cancer in childhood and adolescence: a population-based study in the Nordic countries.* J. Clin. Oncol., 2001. **19**(13): 3173-81.

159. Armstrong, G.T., Liu, Q., Yasui, Y., et al., *Late mortality among 5-year survivors of childhood cancer: a summary from the Childhood Cancer Survivor Study.* J. Clin. Oncol., 2009. **27**(14): 2328-38.

160. Garwicz, S., Anderson, H., Olsen, J.H., et al., *Late and very late mortality in 5-year survivors of childhood cancer: changing pattern over four decades--experience from the Nordic countries.* Int. J. Cancer, 2012. **131**(7): 1659-66.

161. Brewster, D.H., Clark, D., Hopkins, L., et al., *Subsequent mortality experience in five-year survivors of childhood, adolescent and young adult cancer in Scotland: a population based, retrospective cohort study.* Eur. J. Cancer, 2013. **49**(15): 3274-83.

162. Bhatia, S., Armenian, S.H., Armstrong, G.T., et al., *Collaborative Research in Childhood Cancer Survivorship: The Current Landscape.* J. Clin. Oncol., 2015. **33**(27): 3055-64.

163. Hudson, M.M., Ness, K.K., Gurney, J.G., et al., *Clinical ascertainment of health outcomes among adults treated for childhood cancer.* JAMA, 2013. **309**(22): 2371-81.

164. Gibson, T.M., Mostoufi-Moab, S., Stratton, K.L., et al., *Temporal patterns in the risk of chronic health conditions in survivors of childhood cancer diagnosed 1970-99: a report from the Childhood Cancer Survivor Study cohort.* Lancet Oncol., 2018. **19**(12): 1590-1601.

165. International Agency for Research on Cancer, World Health Organisation, International Association of Cancer Registries, and European Network of Cancer Registries, *International Rules for Multiple Primary Cancers (ICD-O Third Edition)*, IARC. 2004: Lyon.

166. National Cancer Institute. *NCI Dictionary of Cancer Terms, http://www.cancer.gov/publications/dictionaries/cancer-terms*. [Accessed: 13th September 2019]

167. Friedman, D.L., Whitton, J., Leisenring, W., et al., *Subsequent neoplasms in 5-year survivors of childhood cancer: the Childhood Cancer Survivor Study.* J. Natl. Cancer Inst., 2010. **102**(14): 1083-95.

168. Turcotte, L.M., Whitton, J.A., Friedman, D.L., et al., *Risk of Subsequent Neoplasms During the Fifth and Sixth Decades of Life in the Childhood Cancer Survivor Study Cohort.* J. Clin. Oncol., 2015. **33**(31): 3568-75.

169. Maule, M., Scelo, G., Pastore, G., et al., *Second malignancies after childhood noncentral nervous system solid cancer: Results from 13 cancer registries.* Int. J. Cancer, 2011. **129**(8): 1940-52.

170. Pole, J.D., Gu, L.Y., Kirsh, V., et al., *Subsequent Malignant Neoplasms in a Population-Based Cohort of Pediatric Cancer Patients: A Focus on the First 5 Years.* Cancer Epidemiol. Biomarkers Prev., 2015. **24**(10): 1585-92.

171. Inskip, P.D. and Curtis, R.E., *New malignancies following childhood cancer in the United States, 1973-2002.* Int. J. Cancer, 2007. **121**(10): 2233-40.

172. Kaatsch, P., Reinisch, I., Spix, C., et al., *Case-control study on the therapy of childhood cancer and the occurrence of second malignant neoplasms in Germany.* Cancer Causes Control, 2009. **20**(6): 965-80.

173. Jenkinson, H.C., Hawkins, M.M., Stiller, C.A., et al., *Long-term population-based risks of second malignant neoplasms after childhood cancer in Britain.* Br. J. Cancer, 2004. **91**(11): 1905-10.

174. Lee, J.S., DuBois, S.G., Coccia, P.F., et al., *Increased risk of second malignant neoplasms in adolescents and young adults with cancer.* Cancer, 2016. **122**(1): 116-23.

175. Turcotte, L.M., Liu, Q., Yasui, Y., et al., *Temporal Trends in Treatment and Subsequent Neoplasm Risk Among 5-Year Survivors of Childhood Cancer, 1970-2015.* JAMA, 2017. **317**(8): 814-824.

176. Allodji, R.S., Hawkins, M.M., Bright, C.J., et al., *Risk of subsequent primary leukaemias among 69,460 five-year survivors of childhood cancer diagnosed from 1940 to 2008 in Europe: A cohort study within PanCareSurFup.* Eur. J. Cancer, 2019. **117**: 71-83.

177. Bright, C.J., Hawkins, M.M., Winter, D.L., et al., *Risk of Soft-Tissue Sarcoma Among 69 460 Five-Year Survivors of Childhood Cancer in Europe.* J. Natl. Cancer Inst., 2018. **110**(6): 649-660.

178. Fidler, M.M., Reulen, R.C., Winter, D.L., et al., *Risk of Subsequent Bone Cancers Among 69 460 Five-Year Survivors of Childhood and Adolescent Cancer in Europe.* J. Natl. Cancer Inst., 2018. **110**(2).

179. Chao, C., Bhatia, S., Xu, L., et al., *Incidence, Risk Factors, and Mortality Associated With Second Malignant Neoplasms Among Survivors of Adolescent and Young Adult Cancer.* JAMA Netw Open, 2019. **2**(6): e195536.

180. Goldfarb, M., Rosenberg, A.S., Li, Q., and Keegan, T.H.M., *Impact of latency time on survival for adolescents and young adults with a second primary malignancy.* Cancer, 2018. **124**(6): 1260-1268.

181. Choi, D.K., Helenowski, I., and Hijiya, N., *Secondary malignancies in pediatric cancer survivors: perspectives and review of the literature.* Int. J. Cancer, 2014. **135**(8): 1764-73.

182. Vasudevan, V., Cheung, M.C., Yang, R., et al., *Pediatric solid tumors and second malignancies: characteristics and survival outcomes.* J. Surg. Res., 2010. **160**(2): 184-9.

183. Keegan, T.H.M., Bleyer, A., Rosenberg, A.S., et al., *Second Primary Malignant Neoplasms and Survival in Adolescent and Young Adult Cancer Survivors.* JAMA oncology, 2017. **3**(11): 1554-1557.

184. Brown, A.L., Arroyo, V.M., Agrusa, J.E., et al., *Survival disparities for second primary malignancies diagnosed among childhood cancer survivors: A population-based assessment.* Cancer, 2019. **125**(20): 3623-3630.

185. Teepen, J.C., Kremer, L.C., van der Heiden-van der Loo, M., et al., *Clinical characteristics and survival patterns of subsequent sarcoma, breast cancer, and melanoma after childhood cancer in the DCOG-LATER cohort.* Cancer Causes Control, 2019. **30**(9): 909-922.

186. Lipshultz, S.E., Cochran, T.R., Franco, V.I., and Miller, T.L., *Treatment-related cardiotoxicity in survivors of childhood cancer.* Nat. Rev. Clin. Oncol., 2013. **10**(12): 697-710.

187. Steiner, R.K., Franco, V.I., and Lipshultz, S.E., *How do we improve the long-term consequences of cardiotoxicity in survivors of childhood cancer?* Prog. Pediatr. Cardiol., 2014. **36**(1–2): 27-30.

188. Armstrong, G.T. and Ross, J.D., *Late Cardiotoxicity in Aging Adult Survivors of Childhood Cancer.* Prog. Pediatr. Cardiol., 2014. **36**(1-2): 19-26.

189. Tukenova, M., Guibout, C., Oberlin, O., et al., *Role of cancer treatment in long-term overall and cardiovascular mortality after childhood cancer.* J. Clin. Oncol., 2010. **28**(8): 1308-15.

190. Mulrooney, D.A., Yeazel, M.W., Kawashima, T., et al., *Cardiac outcomes in a cohort of adult survivors of childhood and adolescent cancer: retrospective analysis of the Childhood Cancer Survivor Study cohort.* BMJ, 2009. **339**: b4606.

191. Fidler, M.M., Reulen, R.C., Henson, K., et al., *Population-Based Long-Term Cardiac-Specific Mortality Among 34 489 Five-Year Survivors of Childhood Cancer in Great Britain.* Circulation, 2017. **135**(10): 951-963.

192. Henson, K.E., Reulen, R.C., Winter, D.L., et al., *Cardiac Mortality Among 200 000 Five-Year Survivors of Cancer Diagnosed at 15 to 39 Years of Age: The Teenage and Young Adult Cancer Survivor Study.* Circulation, 2016. **134**(20): 1519-1531.

193. Scholz-Kreisel, P., Spix, C., Blettner, M., et al., *Prevalence of cardiovascular late sequelae in long-term survivors of childhood cancer: A systematic review and meta-analysis.* Pediatr. Blood Cancer, 2017. **64**(7): e26428.

194. Armstrong, G.T., Oeffinger, K.C., Chen, Y., et al., *Modifiable risk factors and major cardiac events among adult survivors of childhood cancer.* J. Clin. Oncol., 2013. **31**(29): 3673-80.

195. Chow, E.J., Chen, Y., Hudson, M.M., et al., *Prediction of Ischemic Heart Disease and Stroke in Survivors of Childhood Cancer.* J. Clin. Oncol., 2018. **36**(1): 44-52.

196. van Laar, M., Feltbower, R.G., Gale, C.P., et al., *Cardiovascular sequelae in long-term survivors of young peoples' cancer: a linked cohort study.* Br. J. Cancer, 2014. **110**(5): 1338-41.

197. Gudmundsdottir, T., Winther, J.F., de Fine Licht, S., et al., *Cardiovascular disease in Adult Life after Childhood Cancer in Scandinavia: A population-based cohort study of 32,308 one-year survivors.* Int. J. Cancer, 2015. **137**(5): 1176-86.

198. Chao, C., Xu, L., Bhatia, S., et al., *Cardiovascular Disease Risk Profiles in Survivors of Adolescent and Young Adult (AYA) Cancer: The Kaiser Permanente AYA Cancer Survivors Study.* J. Clin. Oncol., 2016. **34**(14): 1626-33.

199. Brewster, D.H., Clark, D., Hopkins, L., et al., *Subsequent hospitalisation experience of 5-year survivors of childhood, adolescent, and young adult cancer in Scotland: a population based, retrospective cohort study.* Br. J. Cancer, 2014. **110**(5): 1342-50.

200. Sieswerda, E., Font-Gonzalez, A., Reitsma, J.B., et al., *High Hospitalization Rates in Survivors of Childhood Cancer: A Longitudinal Follow-Up Study Using Medical Record Linkage.* PLoS One, 2016. **11**(7): e0159518.

201. Lorenzi, M.F., Xie, L., Rogers, P.C., et al., *Hospital-related morbidity among childhood cancer survivors in British Columbia, Canada: report of the childhood, adolescent, young adult cancer survivors (CAYACS) program.* Int. J. Cancer, 2011. **128**(7): 1624-31.

202. Zhang, Y., Lorenzi, M.F., Goddard, K., et al., *Late morbidity leading to hospitalization among 5-year survivors of young adult cancer: a report of the childhood, adolescent and young adult cancer survivors research program.* Int. J. Cancer, 2014. **134**(5): 1174-82.

203. de Fine Licht, S., Rugbjerg, K., Gudmundsdottir, T., et al., *Long-term inpatient disease burden in the Adult Life after Childhood Cancer in Scandinavia (ALiCCS) study: A cohort study of 21,297 childhood cancer survivors.* PLoS Med., 2017. **14**(5): e1002296.

204. Kurt, B.A., Nolan, V.G., Ness, K.K., et al., *Hospitalization rates among survivors of childhood cancer in the Childhood Cancer Survivor Study cohort.* Pediatr. Blood Cancer, 2012. **59**(1): 126-32.

205. Bhakta, N., Liu, Q., Yeo, F., et al., *Cumulative burden of cardiovascular morbidity in paediatric, adolescent, and young adult survivors of Hodgkin's lymphoma: an analysis from the St Jude Lifetime Cohort Study.* Lancet Oncol., 2016. **17**(9): 1325-34.

206. Feijen, E.A., Font-Gonzalez, A., van Dalen, E.C., et al., *Late Cardiac Events after Childhood Cancer: Methodological Aspects of the Pan-European Study PanCareSurFup.* PLoS One, 2016. **11**(9): e0162778.

207. Huang, T.T., Hudson, M.M., Stokes, D.C., et al., *Pulmonary Outcomes in Survivors of Childhood Cancer A Systematic Review.* Chest, 2011. **140**(4): 881-901.

208. Versluys, A.B. and Bresters, D., *Pulmonary Complications of Childhood Cancer Treatment.* Paediatr. Respir. Rev., 2016. **17**: 63-70.

209. Mertens, A.C., Yasui, Y., Liu, Y., et al., *Pulmonary complications in survivors of childhood and adolescent cancer - A report from the childhood cancer survivor study.* Cancer, 2002. **95**(11): 2431-2441.

210. Dietz, A.C., Chen, Y., Yasui, Y., et al., *Risk and impact of pulmonary complications in survivors of childhood cancer: A report from the Childhood Cancer Survivor Study.* Cancer, 2016. **122**(23): 3687-3696.

211. Huang, T.T., Chen, Y., Dietz, A.C., et al., *Pulmonary Outcomes in Survivors of Childhood Central Nervous System Malignancies: A Report From the Childhood Cancer Survivor Study.* Pediatr. Blood Cancer, 2014. **61**(2): 319-325.

212. Green, D.M., Zhu, L., Wang, M., et al., *Pulmonary Function after Treatment for Childhood Cancer. A Report from the St. Jude Lifetime Cohort Study (SJLIFE).* Ann Am Thorac Soc, 2016. **13**(9): 1575-85.

213. Kasteler, R., Weiss, A., Schindler, M., et al., *Long-term pulmonary disease among Swiss childhood cancer survivors.* Pediatr. Blood Cancer, 2017. **65**(1): e26749.

214. Armenian, S.H., Landier, W., Francisco, L., et al., *Long-Term Pulmonary Function in Survivors of Childhood Cancer.* J. Clin. Oncol., 2015. **33**(14): 1592-1600.

215. Mulder, R.L., Thonissen, N.M., van der Pal, H.J.H., et al., *Pulmonary function impairment measured by pulmonary function tests in long-term survivors of childhood cancer.* Thorax, 2011. **66**(12): 1065-1071.

216. Motosue, M.S., Zhu, L., Srivastava, K., et al., *Pulmonary function after whole lung irradiation in pediatric patients with solid malignancies.* Cancer, 2012. **118**(5): 1450-1456.

217. Record, E., Williamson, R., Wasilewski-Masker, K., et al., *Analysis of Risk Factors for Abnormal Pulmonary Function in Pediatric Cancer Survivors.* Pediatr. Blood Cancer, 2016. **63**(7): 1264-71.

218. Office for National Statistics, *2011 Census: Aggregate data (England and Wales) [computer file]. UK Data Service Census Support. Downloaded from http://infuse.mimas.ac.uk. This information is licensed under the terms of the Open Government Licence*

[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2].

219. Office for National Statistics, *Census gives insights into characteristics of Yorkshire and the Humber's population*. 2012: London.

220. Kay, I., *Portrait of Yorkshire and the Humber.* Regional Trends, 2009. **41**(1): 52-64.

221. Office for National Statistics. *Open Geography portal http://geoportal.statistics.gov.uk/*. [Accessed: 13th September 2019]

222. NHS Digital. *Hospital Episode Statistics https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics*. [Accessed: 13th September 2019]

223. NHS Digital, *HES Data Dictionary Admitted Patient Care*. 2018, NHS Digital

224. Health and Social Care Information Centre, *Methodology for creation of the HES Patient ID (HESID).* http://content.digital.nhs.uk/media/1370/HES-Hospital-Episode-Statistics-Replacement-of-the-HES-patient-ID/pdf/HESID_Methodology.pdf 2014.

225. Herbert, A., Wijlaars, L., Zylbersztejn, A., et al., *Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC).* Int. J. Epidemiol., 2017. **46**(4): 1093-1093i.

226. Health and Social Care Information Centre, *Methodology to create provider and CIP spells from HES APC data.* http://content.digital.nhs.uk/media/11859/Provider-Spells-Methodology/pdf/Spells_Methodology.pdf. 2014.

227. Office for National Statistics. *Population Estimates.* https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates. [Accessed: 10th September 2019]

228. Irving, J.A., Enshaei, A., Parker, C.A., et al., *Integration of genetic and clinical risk factors improves prognostication in relapsed childhood B-cell precursor acute lymphoblastic leukemia.* Blood, 2016. **128**(7): 911-22.

229. Moorman, A.V., Ensor, H.M., Richards, S.M., et al., *Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial.* Lancet Oncol., 2010. **11**(5): 429-38.

230. van Laar, M., McKinney, P.A., Stark, D.P., et al., *Survival trends of cancer amongst the south Asian and non-south Asian population under 30 years of age in Yorkshire, UK.* Cancer Epidemiol., 2012. **36**(1): e13-8.

231. Sterne, J.A., White, I.R., Carlin, J.B., et al., *Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls.* BMJ, 2009. **338**: b2393.

232. White, I.R., Royston, P., and Wood, A.M., *Multiple imputation using chained equations: Issues and guidance for practice.* Stat. Med., 2011. **30**(4): 377-399.

233. UK Data Service. *Census Support. Deprivation Data* https://census.ukdataservice.ac.uk/get-data/related/deprivation. [Accessed: 13th September 2019]

234. Townsend, P., Phillimore, P., and Beattie, A., *Health and deprivation: inequality and the North*. 1988, London: Croom Helm.

235. Pearce, N., *The ecological fallacy strikes back.* J. Epidemiol. Community Health, 2000. **54**(5): 326-7.

236. Smith, L., Norman, P., Kapetanstrataki, M., et al., *Comparison of ethnic group classification using naming analysis and routinely collected data:*

application to cancer incidence trends in children and young people. BMJ Open, 2017. **7**(9): e016332.

237. NHS Executive, *Collection of ethnic group data for admitted patients (EL/94/77)*, NHS Executive. 1994: Leeds.

238. Mathur, R., Bhaskaran, K., Chaturvedi, N., et al., *Completeness and usability of ethnicity data in UK-based primary care and hospital databases.* Journal of public health, 2014. **36**(4): 684-92.

239. Downing, A., West, R.M., Gilthorpe, M.S., et al., *Using routinely collected health data to investigate the association between ethnicity and breast cancer incidence and survival: what is the impact of missing data and multiple ethnicities?* Ethn. Health, 2011. **16**(3): 201-12.

240. National Cancer Intelligence Network, *Cancer Incidence and Survival By Major Ethnic Group, England, 2002-2006*. 2009.

241. OnoMAP. *http://www.onomap.org/*. [Accessed: 13th September 2019]

242. Mateos, P., Longley, P.A., and O'Sullivan, D., *Ethnicity and Population Structure in Personal Naming Networks.* PLoS One, 2011. **6**(9): e22943.

243. Cummins, C., Winter, H., Cheng, K.K., et al., *An assessment of the Nam Pehchan computer program for the identification of names of south Asian ethnic origin.* J. Public Health Med., 1999. **21**(4): 401-6.

244. Nanchahal, K., Mangtani, P., Alston, M., and dos Santos Silva, I., *Development and validation of a computerized South Asian Names and Group Recognition Algorithm (SANGRA) for use in British health-related studies.* J. Public Health Med., 2001. **23**(4): 278-85.

245. Lakha, F., Gorman, D.R., and Mateos, P., *Name analysis to classify populations by ethnicity in public health: validation of Onomap in Scotland.* Public Health, 2011. **125**(10): 688-96.

246. Feltbower, R.G., Kinsey, S.E., Richards, M., et al., *Survival following relapse in childhood haematological malignancies diagnosed in 1974-2003 in Yorkshire, UK.* Br. J. Cancer, 2007. **96**(7): 1147-52.

247. National Cancer Intelligence Network, *Bladder Cancer Incidence, Mortality and Survival Rates in the United Kingdom*. 2013.

248. National Cancer Intelligence Network, *Non-melanoma skin cancer in England, Scotland, Northern Ireland, and Ireland. NCIN Data Briefing*. 2013.

249. Bray, F. and Ferley, J., *Cancer Incidence in Five Continents Vol. X. Chapter 7: Age Standardisation*, D. Forman, et al., Editors. 2014, International Agency for Research on Cancer: Lyon, France.

250. Public Health England, *Childhood cancer registration in England: 2015 to 2016*, Report on behalf of the Children Teenagers and Young Adults Site

Specific Clinical Reference Group National Cancer Registration and Analysis Service. 2016.

251. Andersson, T.M., Eriksson, H., Hansson, J., et al., *Estimating the cure proportion of malignant melanoma, an alternative approach to assess long term survival: a population-based study.* Cancer Epidemiol., 2014. **38**(1): 93-9.

252. Eriksson, H., Lyth, J., and Andersson, T.M., *The proportion cured of patients diagnosed with Stage III-IV cutaneous malignant melanoma in Sweden 1990-2007: A population-based study.* Int. J. Cancer, 2016. **138**(12): 2829-36.

253. Vora, A., Goulden, N., Wade, R., et al., *Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): a randomised controlled trial.* Lancet Oncol., 2013. **14**(3): 199-209.

254. Hann, I., Vora, A., Harrison, G., et al., *Determinants of outcome after intensified therapy of childhood lymphoblastic leukaemia: results from Medical Research Council United Kingdom acute lymphoblastic leukaemia XI protocol.* Br. J. Haematol., 2001. **113**(1): 103-14.

255. Mitchell, C.D., Richards, S.M., Kinsey, S.E., et al., *Benefit of dexamethasone compared with prednisolone for childhood acute lymphoblastic leukaemia: results of the UK Medical Research Council ALL97 randomized trial.* Br. J. Haematol., 2005. **129**(6): 734-45.

256. Mitchell, C., Payne, J., Wade, R., et al., *The impact of risk stratification by early bone-marrow response in childhood lymphoblastic leukaemia: results from the United Kingdom Medical Research Council trial ALL97 and ALL97/99.* Br. J. Haematol., 2009. **146**(4): 424-36.

257. Vora, A., Mitchell, C.D., Lennard, L., et al., *Toxicity and efficacy of 6-thioguanine versus 6-mercaptopurine in childhood lymphoblastic leukaemia: a randomised trial.* Lancet, 2006. **368**(9544): 1339-48.

258. Coviello, V. and Boggess, M., *Cumulative incidence estimation in the presence of competing risks.* The Stata Journal, 2004. **4**(2): 103-112.

259. Greenland, S., Pearl, J., and Robins, J.M., *Causal diagrams for epidemiologic research.* Epidemiology, 1999. **10**(1): 37-48.

260. Greenland, S., *Quantifying biases in causal models: classical confounding vs collider-stratification bias.* Epidemiology, 2003. **14**(3): 300-6.

261. Hernan, M.A., Hernandez-Diaz, S., and Robins, J.M., *A structural approach to selection bias.* Epidemiology, 2004. **15**(5): 615-25.

262. Schisterman, E.F., Cole, S.R., and Platt, R.W., *Overadjustment bias and unnecessary adjustment in epidemiologic studies.* Epidemiology, 2009. **20**(4): 488-95.

263. Williamson, E.J., Aitken, Z., Lawrie, J., et al., *Introduction to causal diagrams for confounder selection.* Respirology, 2014. **19**(3): 303-11.

264. Pearce, N. and Lawlor, D.A., *Causal inference-so much more than statistics.* Int. J. Epidemiol., 2016. **45**(6): 1895-1903.

265. Williams, T.C., Bach, C.C., Matthiesen, N.B., et al., *Directed acyclic graphs: a tool for causal studies in paediatrics.* Pediatr. Res., 2018. **84**(4): 487-493.

266. Lederer, D.J., Bell, S.C., Branson, R.D., et al., *Control of Confounding and Reporting of Results in Causal Inference Studies: Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals.* Ann Am Thorac Soc, 2019. **16**(1): 22-28.

267. Textor, J., van der Zander, B., Gilthorpe, M.S., et al., *Robust causal inference using directed acyclic graphs: the R package 'dagitty'.* Int. J. Epidemiol., 2016. **45**(6): 1887-1894.

268. Putter, H., Fiocco, M., and Geskus, R.B., *Tutorial in biostatistics: Competing risks and multi-state models.* Stat. Med., 2007. **26**(11): 2389-2430.

269. Lau, B., Cole, S.R., and Gange, S.J., *Competing risk regression models for epidemiologic data.* Am. J. Epidemiol., 2009. **170**(2): 244-56.

270. Andersen, P.K., Geskus, R.B., de Witte, T., and Putter, H., *Competing risks in epidemiology: possibilities and pitfalls.* Int. J. Epidemiol., 2012. **41**(3): 861-70.

271. Austin, P.C., Lee, D.S., and Fine, J.P., *Introduction to the Analysis of Survival Data in the Presence of Competing Risks.* Circulation, 2016. **133**(6): 601-609.

272. Wolbers, M., Koller, M.T., Stel, V.S., et al., *Competing risks analyses: objectives and approaches.* Eur. Heart J., 2014. **35**(42): 2936-41.

273. Austin, P.C. and Fine, J.P., *Practical recommendations for reporting Fine-Gray model analyses for competing risk data.* Stat. Med., 2017. **36**(27): 4391-4400.

274. Fine, J.P. and Gray, R.J., *A proportional hazards model for the subdistribution of a competing risk.* J Am Stat Assoc, 1999. **94**(446): 496-509.

275. Breslow, N.E. and Day, N., *Statistical methods in cancer research. Vol. 2. The design and analysis of cohort studies. Lyon, France: International Agency for Research on Cancer, 1987*. 1987, IARC scientific publications.

276. Office for National Statistics. *Cancer Registration Statistics England* *http://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatisticscancerregistrationstatisticsengland*. [Accessed: 13th September 2019]

277. Dong, H., Robison, L.L., Leisenring, W.M., et al., *Estimating the burden of recurrent events in the presence of competing risks: the method of mean cumulative count.* Am. J. Epidemiol., 2015. **181**(7): 532-40.

278. Prentice, R.L., Williams, B.J., and Peterson, A.V., *On the Regression-Analysis of Multivariate Failure Time Data.* Biometrika, 1981. **68**(2): 373-379.

279. Amorim, L.D. and Cai, J., *Modelling recurrent events: a tutorial for analysis in epidemiology.* Int. J. Epidemiol., 2015. **44**(1): 324-33.

280. Westbury, L.D., Syddall, H.E., Simmonds, S.J., et al., *Identification of risk factors for hospital admission using multiple-failure survival models: a toolkit for researchers.* BMC Med. Res. Methodol., 2016. **16**: 46.

281. van Laar, M., *Variation in Cancer Outcomes Amongst Children and Young Adults in Yorkshire*. 2015, University of Leeds.

282. Shah, A., Diggens, N., Stiller, C., and Murphy, M.F., *Investigation of potential indicators of co-morbidity on outcomes for children with leukaemia using linked Hospital Episode Statistics data. http://www.ncin.org.uk/view?rid=2241*. 2013.

283. Stiller, C.A., Kroll, M.E., and Pritchard-Jones, K., *Population survival from childhood cancer in Britain during 1978-2005 by eras of entry to clinical trials.* Ann. Oncol., 2012. **23**(9): 2464-9.

284. van Laar, M., Kinsey, S.E., and Feltbower, R.G., *Survival of childhood acute lymphoid leukaemia in Yorkshire by clinical trial era, 1990-2011.* Br. J. Haematol., 2015. **171**(1): 116-9.

285. Hough, R., Sandhu, S., Khan, M., et al., *Are survival and mortality rates associated with recruitment to clinical trials in teenage and young adult patients with acute lymphoblastic leukaemia? A retrospective observational analysis in England.* BMJ Open, 2017. **7**(10): e017052.

286. Bloom, C.I., Nissen, F., Douglas, I.J., et al., *Exacerbation risk and characterisation of the UK's asthma population from infants to old age.* Thorax, 2018. **73**(4): 313-320.

287. De, A., Guryev, I., LaRiviere, A., et al., *Pulmonary function abnormalities in childhood cancer survivors treated with bleomycin.* Pediatr. Blood Cancer, 2014. **61**(9): 1679-84.

288. Teepen, J.C., van Leeuwen, F.E., Tissing, W.J., et al., *Long-Term Risk of Subsequent Malignant Neoplasms After Treatment of Childhood Cancer in the DCOG LATER Study Cohort: Role of Chemotherapy.* J. Clin. Oncol., 2017. **35**(20): 2288-2298.

289. Tukenova, M., Diallo, I., Hawkins, M., et al., *Long-term mortality from second malignant neoplasms in 5-year survivors of solid childhood tumors: temporal pattern of risk according to type of treatment.* Cancer Epidemiol. Biomarkers Prev., 2010. **19**(3): 707-15.

290. Zong, X., Pole, J.D., Grundy, P.E., et al., *Second malignant neoplasms after childhood non-central nervous system embryonal tumours in North America: A population-based study.* Eur. J. Cancer, 2017. **84**: 173-183.

291. Bhuller, K.S., Zhang, Y., Li, D., et al., *Late mortality, secondary malignancy and hospitalisation in teenage and young adult survivors of Hodgkin lymphoma: report of the Childhood/Adolescent/Young Adult Cancer Survivors Research Program and the BC Cancer Agency Centre for Lymphoid Cancer.* Br. J. Haematol., 2016. **172**(5): 757-68.

292. Moppett, J., Oakhill, A., and Duncan, A.W., *Second malignancies in children: the usual suspects?* Eur. J. Radiol., 2001. **38**(3): 235-48.

293. Morton, L.M., Swerdlow, A.J., Schaapveld, M., et al., *Current knowledge and future research directions in treatment-related second primary malignancies.* EJC Suppl, 2014. **12**(1): 5-17.

294. NHS Digital. *Hospital Admitted Patient Care Activity, 2017-18. https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2017-18*. [Accessed: 13th September 2019]

295. de Vries, S., Schaapveld, M., van Nimwegen, F.A., et al., *High burden of subsequent malignant neoplasms and cardiovascular disease in long-term Hodgkin lymphoma survivors.* Br. J. Cancer, 2018. **118**(6): 887-895.

296. Liu, Q., Leisenring, W.M., Ness, K.K., et al., *Racial/Ethnic Differences in Adverse Outcomes Among Childhood Cancer Survivors: The Childhood Cancer Survivor Study.* J. Clin. Oncol., 2016. **34**(14): 1634-43.

297. Eissa, H.M., Lu, L., Baassiri, M., et al., *Chronic disease burden and frailty in survivors of childhood HSCT: a report from the St. Jude Lifetime Cohort Study.* Blood Adv, 2017. **1**(24): 2243-2246.

298. Kenborg, L., Winther, J.F., Linnet, K.M., et al., *Neurologic disorders in 4858 survivors of central nervous system tumors in childhood-an Adult Life after Childhood Cancer in Scandinavia (ALiCCS) study.* Neuro Oncol., 2019. **21**(1): 125-136.

299. Wallace, W.H., Blacklay, A., Eiser, C., et al., *Developing strategies for long term follow up of survivors of childhood cancer.* BMJ, 2001. **323**(7307): 271-4.

300. Jefford, M., Rowland, J., Grunfeld, E., et al., *Implementing improved post-treatment care for cancer survivors in England, with reflections from Australia, Canada and the USA.* Br. J. Cancer, 2013. **108**(1): 14-20.

301. Glaser, A., Levitt, G., Morris, P., et al., *Enhanced quality and productivity of long-term aftercare of cancer in young people.* Arch. Dis. Child., 2013. **98**(10): 818-24.

302. Edgar, A.B., Duffin, K., Borthwick, S., et al., *Can intensity of long-term follow-up for survivors of childhood and teenage cancer be determined by therapy-based risk stratification?* BMJ Open, 2013. **3**(8).

303. Frobisher, C., Glaser, A., Levitt, G.A., et al., *Risk stratification of childhood cancer survivors necessary for evidence-based clinical long-term follow-up.* Br. J. Cancer, 2017. **117**(11): 1723-1731.

304. Landier, W., Skinner, R., Wallace, W.H., et al., *Surveillance for Late Effects in Childhood Cancer Survivors.* J. Clin. Oncol., 2018. **36**(21): 2216-2222.

305. Smith, L., Glaser, A.W., Kinsey, S.E., et al., *Long-term survival after childhood acute lymphoblastic leukaemia: population-based trends in cure and relapse by clinical characteristics.* Br. J. Haematol., 2018. **182**(6): 851-858.

306. Smith, L., Glaser, A.W., Peckham, D., et al., *Respiratory morbidity in young people surviving cancer: Population-based study of hospital admissions, treatment-related risk factors and subsequent mortality.* Int. J. Cancer, 2018. **145**(1): 20-28.

307. Johnson, P., Greiner, W., Al-Dakkak, I., and Wagner, S., *Which Metrics Are Appropriate to Describe the Value of New Cancer Therapies?* BioMed research international, 2015. **2015**: 865101.

308. Parkin, D.M. and Hakulinen, T., *Chapter 12. Analysis of Survival* in *Cancer Registration: Principles and Methods*, O.M. Jensen, et al., Editors. 1991: Lyon

309. Schwartz, C.L., *Health status of childhood cancer survivors: cure is more than the eradication of cancer.* JAMA, 2003. **290**(12): 1641-3.

310. Zwaan, C.M. and Sposto, R., *Cure trends in acute lymphoblastic leukemia: is it time for a revised concept of cure?* Haematologica, 2013. **98**(5): 655-6.

311. Haupt, R., Spinetta, J.J., Ban, I., et al., *Long term survivors of childhood cancer: cure and care. The Erice statement.* Eur. J. Cancer, 2007. **43**(12): 1778-80.

312. Tralongo, P., Dal Maso, L., Surbone, A., et al., *Use of the word "cured" for cancer patients—implications for patients and physicians: the Siracusa charter.* Current Oncology, 2015. **22**(1): e38.

313. Barr, R.D. and Sala, A., *Quality-adjusted survival: a rigorous assessment of cure after cancer during childhood and adolescence.* Pediatr. Blood Cancer, 2005. **44**(3): 201-4.

314. The Lancet, *Making more of multimorbidity: an emerging priority.* Lancet, 2018. **391**(10131): 1637.

315. Stafford, M., Steventon, A., Thorlby, R., et al., *Briefing: Understanding the health care needs of people with multiple health conditions*. 2018, The Health Foundation: London.

316. Department of Health, *A simple guide to Payment by Results. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213150/PbR-Simple-Guide-FINAL.pdf*. 2012.

317. Bright, C.J., Hawkins, M.M., Guha, J., et al., *Risk of Cerebrovascular Events in 178 962 Five-Year Survivors of Cancer Diagnosed at 15 to 39 Years of Age: The TYACSS (Teenage and Young Adult Cancer Survivor Study).* Circulation, 2017. **135**(13): 1194-1210.

318. Feltbower, R.G., Lewis, I.J., Picton, S., et al., *Diagnosing childhood cancer in primary care--a realistic expectation?* Br. J. Cancer, 2004. **90**(10): 1882-4.

319. Kero, A.E., Madanat-Harjuoja, L.M., Jarvela, L.S., et al., *Health conditions associated with metabolic syndrome after cancer at a young age: A nationwide register-based study.* Cancer Epidemiol., 2016. **41**: 42-9.

320. Kero, A.E., Madanat-Harjuoja, L.M., Jarvela, L.S., et al., *Cardiovascular medication after cancer at a young age in Finland: A nationwide registry linkage study.* Int. J. Cancer, 2016. **139**(3): 683-90.

321. NHS England, *Improving Outcomes through Personalised Medicine. https://www.england.nhs.uk/wp-content/uploads/2016/09/improving-outcomes-personalised-medicine.pdf*. 2016.

322. Bartlett, J.W. and Taylor, J.M., *Missing covariates in competing risks analysis.* Biostatistics, 2016. **17**(4): 751-63.

323. Lau, B. and Lesko, C., *Missingness in the Setting of Competing Risks: from missing values to missing potential outcomes.* Curr Epidemiol Rep, 2018. **5**(2): 153-159.

324. Andrae, B., Andersson, T.M., Lambert, P.C., et al., *Screening and cervical cancer cure: population based cohort study.* BMJ, 2012. **344**: e900.

325. Ater, J.L., Zhou, T., Holmes, E., et al., *Randomized study of two chemotherapy regimens for treatment of low-grade glioma in young children: a report from the Children's Oncology Group.* J. Clin. Oncol., 2012. **30**(21): 2641-7.

326. Frazier, A.L., Hale, J.P., Rodriguez-Galindo, C., et al., *Revised risk classification for pediatric extracranial germ cell tumors based on 25 years of clinical trial data from the United Kingdom and United States.* J. Clin. Oncol., 2015. **33**(2): 195-201.

327. Gamel, J.W., George, S.L., Edwards, M.J., and Seigler, H.F., *The long-term clinical course of patients with cutaneous melanoma.* Cancer, 2002. **95**(6): 1286-93.

328. Gatta, G., Capocaccia, R., Hakulinen, T., et al., *Variations in survival for invasive cervical cancer among European women, 1978-89.* Cancer Causes Control, 1999. **10**(6): 575-581.

329. Ito, Y., Nakayama, T., Tsukuma, H., et al., *Role of age and tumour stage in the temporal pattern of 'cure' from stomach cancer: a population-based study in Osaka, Japan.* Cancer Epidemiol., 2012. **36**(2): 128-32.

330. Rahimzadeh, M., Baghestani, A.R., Gohari, M.R., and Pourhoseingholi, M.A., *Estimation of the cure rate in Iranian breast cancer patients.* Asian Pacific Journal of Cancer Prevention: Apjcp, 2014. **15**(12): 4839-42.

331. Rutqvist, L.E. and Wallgren, A., *Long-term survival of 458 young breast cancer patients.* Cancer, 1985. **55**(3): 658-65.

332. Weston, C.L., Douglas, C., Craft, A.W., et al., *Establishing long-term survival and cure in young patients with Ewing's sarcoma.* Br. J. Cancer, 2004. **91**(2): 225-32.

Appendix A Summary of studies utilising cure modes in children and young adults with cancer

Table A.9.1: Summary of studies included in literature review of applications of cure models in children and young adults with cancer

Author, year	Study type	Study population	Cancer type	Type of cure model
AIRTUM working group, 2014 [120]	Cancer registry	Italy All ages (0-44 years separate age group) Diagnosed 1976-2009 N=1,624,533	All cancers combined and 50 separate diagnostic groups	Mixture
Andersson et al, 2009 [135]	Cancer registry	Sweden 19-80 years (19-40 years separate age group) Diagnosed 1973-2001 N=6439	AML	Mixture
Andersson et al, 2014 [251]	Cancer registry	Sweden All ages (0-30 years separate age group) Diagnosed 1996-2005 N=5850	Cutaneous malignant melanoma	Flexible parametric
Andrae et al, 2012 [324]	Cancer registry	Sweden 23-65 years Diagnosed 1999-2001 N= 1230	Cervical cancer (including screening history)	Mixture
Ater et al, 2012 [325]	Clinical trial	USA Children under 10 Recruited 1997-2005 n=274	Low grade glioma	Non-mixture
Bejan-Angoulvant et al, 2008 [126]	Cancer registry	France 15 years + (15-44 years separate age group) Diagnosed 1980-1996 N=9998	Colon cancer	Mixture
Bouliotis et al, 2015 [138]	Cancer registry	Nottingham, UK All ages (0-26 years separate age group) Diagnosed 1973-2002 N=768	Hodgkin lymphoma	Non-mixture model
Crocetti et al, 2012 [143]	Cancer registry	Europe All ages Diagnosed 1995-2002 n=44,947	CNS tumours; glial and non-glial tumours	Mixture

Author, year	Study type	Study population	Cancer type	Type of cure model
Cvancarova et al, 2012 [124]	Cancer registry	Norway All ages Diagnosed 1963-2007 N=627,346	All cancers combined and 23 most common sites	Mixture
Dal Maso et al, 2014 [121]	Cancer registry	Italy 15-74 years (15-44 years separate age group) Diagnosed 1985-2005 N=818,902	All cancers combined and 26 most common sites	Mixture
De Angelis et al, 1999 [24]	Cancer registry	Finland All ages (0-44 years separate age group) Diagnosed 1953-1992 N=22,617	Colon cancer	Mixture
Edlinger et al, 2014 [125]	Cancer registry	Tyrol, Austria All ages Diagnosed 2005-2009 N=16,144	All cancers combined and 25 most common sites	Mixture
Eloranta et al, 2010 [128]	Cancer registry	Sweden All ages (0-50 years separate age group) Diagnosed 1965-2000 N=58,873	Colon cancer	Mixture
Eloranta et al, 2014 [127]	Cancer registry	Sweden 19-80 years (19-50 year separate age group) Diagnosed 1973-2007 N=121,886	Melanoma, colon cancer and AML	Flexible parametric
Eriksson et al, 2016 [252]	Cancer registry	Sweden All ages (0-50 years separate age group) Diagnosed 1990-2007 N=856	Melanoma	Flexible parametric
Francisci et al, 2009 [131]	Cancer registry	Europe 15-99 years (15-44 years separate age group) Diagnosed 1988-1999 N=5,967,548	All cancers combined and five common sites: breast (women only), lung, prostate, colorectal and stomach	Mixture
Frazier et al, 2015 [326]	Clinical trial	UK and USA Paediatric age range (Does not specify upper age limit)	Extracranial germ cell tumours	Non-mixture

Author, year	Study type	Study population	Cancer type	Type of cure model
		Diagnosed 1985-2009 N=519		
Gamel et al, 2002 [327]	Hospital data	USA 11-92 years Diagnosed 1978-1996 N=5342	Melanoma	Mixture
Gatta et al, 1999 [328]	Cancer registry	Europe 15-99 years Diagnosed 1979-1989 N=40,906	Cervical cancer	Mixture
Gatta et al, 2012 [144]	Cancer registry	Europe All ages Diagnosed 1995-2002 N=3322	Embryonal tumours:	Mixture
Gatta et al, 2013 [30]	Cancer registry	Europe Children 0-14 years TYA, 15-24 years Diagnosed 1982-2002 N=22,886	ALL	Mixture
Gieser et al, 1998 [133]	Clinical trial	USA Age range not specified Period of recruitment not specified N=763	ALL	Mixture cure model
Glimelius et al, 2015 [139]	Cancer registry	Sweden 18-59 years (18-29 years separate age group) Diagnosed 1992-2009 N=1947	Hodgkin lymphoma	Flexible parametric
Hunsberger et al, 2009 [146]	Clinical trial	USA Age range not specified Recruited 1986-2001 N=2558	Neuroblastoma	Mixture
Ito et al, 2012 [129]	Cancer registry	Osaka, Japan All ages (<50 years separate age group) Diagnosed 1975-2000 N=33,885	Colorectal	Mixture
Ito et al, 2012 [329]	Cancer registry	Osaka, Japan 15-99 years (15-39 years separate age group) Diagnosed 1975-2000 N=66,032	Stomach cancer	Mixture
Lee, 1995 [134]	Clinical trial	302 boys (no further details given)	ALL	Mixture

Author, year	Study type	Study population	Cancer type	Type of cure model
Nesbit et al, 1994 [136]	Clinical trial	USA 0-21 years Diagnosed 1979-1983 N=381	AML (study of bone marrow transplantation)	Not specified
Pfirrmann et al, 2014 [141]	Clinical trial	Germany 11-58 years Diagnosed 1995-2004 N=256 patients receiving allogeneic hematopoietic stem cell transplantation	Chronic myeloid leukaemia	Non-mixture
Rahimzadeh et al, 2014 [330]	Hospital data	Tehran, Iran Age range 23 to 79 years Treated 2006-2008 N=305	Breast cancer	Non-mixture
Rutqvist and Wallgren, 1985 [331]	Hospital data	Sweden <40 years Women undergoing surgery between 1921 and 1959 N=458	Breast cancer	Mixture
Shack et al, 2012 [130]	Cancer registry	North West England 15-99 years (15-44 years separate age group) Diagnosed 1997-2004 N=25,563	Colorectal cancer	Mixture
Shah et al, 2008 [29]	Cancer registry	Great Britain 0-14 years Diagnosed 1971-2000 N=13,069	Leukaemia	Mixture
Shah et al, 2013 [137]	Cancer registry	England 15-99 years (15-24 years and 25-39 years separate age groups) Diagnosed 1971-2006 N=48,380	AML	Mixture
Silversmit et al, 2017 [132]	Cancer registry	Belgium 15-99 years (15-34/44/49/54/59 as separate age group depending on cancer) Diagnosed 1999-20100 N=94,891	Cervix, colon, corpus uteri, melanoma, oesophagus, pancreas, stomach	Mixture
Smoll et al, 2012 [119]	Cancer registry	USA, SEER All ages (16-39 years separate age group)	Glioblastoma multiforme	Non-mixture

Author, year	Study type	Study population	Cancer type	Type of cure model
		Diagnosed 2001-2006 n=11,189		
Sposto, 2002 [25]	Clinical trial	USA Age range not specified Diagnosed 1978-1990 NHL n=345 Hodgkins lymphoma and NHL n=97 ALL (two studies n=636 and n=942)	NHL, Hodgkin Lymphoma and ALL	Compariso n of Cox and mixture and non-mixture cure models
Sposto et al, 2007 [140]	Clinical trial	USA Age range not specified Hodgkin lymphoma n=780 Neuroblastoma n=317	Hodgkin lymphoma and neuroblastoma	Non-mixture cure model
Trama et al, 2012 [145]	Cancer registry	Europe All ages Diagnosed 1995-2002 n=25,769	Testicular, paratesticular and extragonadal germ cell tumours	Mixture
Ventura et al, 2014 [147]	Cancer registry	Tuscany, Italy 15-94 years (15-44 years separate age group) Diagnosed 1987-2005 N=not stated	Stomach cancer in women	Bayesian mixture model
Verdecchia et al, 1998 [118]	Cancer registry	Europe 15-99 years (15-44 years separate age group Diagnosed 1978-1985 N=74,475	Colon cancer	Mixture
Wang et al, 2012 [148]	Cancer registry	USA SEER 5-101 years (<40 years as separate age group) Dates of diagnosis not stated n=637	Melanoma	Mixture
Weston et al, 2004 [332]	Clinical trial	UK Age range 1-34 years Recruited 1987-1993 Ewings Tumour Study ET-1 n=142 ET-2 n=243	Ewings sarcoma	Non-mixture

Author, year	Study type	Study population	Cancer type	Type of cure model
Weston et al, 2009 [149]	Clinical trial	UK, Germany, Netherlands and Austria Age range not stated Recruited 1987-1999 Ewings Tumour Study ET-2 n=242, Eicess-92 n=647	Ewings sarcoma	Non-mixture
Weston et al, 2010 [150]	Clinical trial	UK, Germany, Netherlands and Austria Age range not stated Recruited 1987-1999 Ewings Tumour Study ET-2 n=242, Eicess-92 n=647	Ewings sarcoma	Non-mixture
Woods et al, 2009 [117]	Cancer registry	West Midlands, England and New South Wales, Australia 15-99 years (15-39 years separate age group) Diagnosed 1980-1995 N=80,313	Breast	Mixture and non-mixture