White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/235615/

Version: Published Version

**Article:**
Nath, S., Stone, T., Lam, J. et al. (2025) Educational outcomes for children and young people with cancer: study protocol for a population-based cohort study using linked education and hospital data from England. BMJ Open, 15 (11). e109410. ISSN: 2044-6055

https://doi.org/10.1136/bmjopen-2025-109410

**Reuse**
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

BMJ Open

# Educational outcomes for children and young people with cancer: study protocol for a population-based cohort study using linked education and hospital data from England

Selina Nath ,[1] Tony Stone,[1] Joseph Lam,[1] Richard G Feltbower ,[2] Darren Hargrave,[1,3] Martin G McCabe,[4] Harry Brown,[1] Lauren Cara Louise Hickinbottom,[1] Kella Jackson,[1] Lewis Paget,[1] Tilly Stanley,[1] Phoebe Skye Watts,[1] Katie Harron [1]

**To cite:** Nath S, Stone T, Lam J, *et al*. Educational outcomes for children and young people with cancer: study protocol for a population-based cohort study using linked education and hospital data from England. *BMJ Open* 2025;**15**:e109410. doi:10.1136/bmjopen-2025-109410

► Prepublication history for this paper is available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2025-109410).

Received 14 August 2025
Accepted 24 October 2025

Check for updates

© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY. Published by BMJ Group.

[1]UCL Great Ormond Street Institute of Child Health, University College London, London, UK
[2]University of Leeds, Leeds, UK
[3]Great Ormond Street Hospital for Children, London, UK
[4]Division of Cancer Sciences, University of Manchester, Manchester, UK

**Correspondence to**
Dr Selina Nath;
s.nath@ucl.ac.uk

## ABSTRACT

**Introduction** Childhood cancer survivors (CCSs) experience educational disruptions during and following treatment, yet robust, longitudinal evidence on educational performance remains limited. We will investigate differences in educational outcomes between CCSs and non-cancer peers during primary and secondary school. We will also explore how sociodemographic factors and age at diagnosis contribute to potential differences in General Certificate of Secondary Education (GCSE) examinations, a critical indicator of future academic and employment prospects.

**Methods and analysis** We will use the Education and Child Health Insights from Linked Data (ECHILD) to capture linked health and education data for children born in National Health Service (NHS)-funded hospitals in England. We will generate birth cohorts spanning September 1997 to August 2015 (estimated sample size: ~10 million), formed of pupils expected to have undertaken national curriculum assessments between academic years 2004/2005 and 2021/2022 including Key Stage (KS) 1, 2 and 4, corresponding to ages 7, 11 and 16 respectively. Cancer diagnosis will be identified from inpatient hospital records, using International Classification of Diseases, 10th Revision codes (ICD-10). We will investigate differences between CCS and their non-cancer peers in terms of their sociodemographic characteristics and describe trends in educational performances at all KSs, recorded Special Educational Needs and Disabilities (SEND) and school absences. Differences in KS4 (GCSE) performances between CCS and non-cancer peers will be quantified, according to and accounting for geographic region, sex, deprivation, ethnicity and birth characteristics. To assess whether cancer diagnosis disrupts academic trajectories, we will restrict analysis to those with KS2 attainment data and investigate KS4 performance. We will finally explore the influence of age at diagnosis on educational performance at KS4.

**Ethics and dissemination** Ethics approval was granted by NHS Health Research Authority Research Ethics Committee (20/EE/0180). Findings will be shared with academics, policymakers, children and families affected by childhood cancer, and published in journals. Code/metadata will be shared on ECHILD GitHub repository.

## STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ We use linked education and health data covering all births in National Health Service (NHS) hospitals in England.
⇒ The large sample size enables analyses stratified by cancer subgroups and sociodemographic factors, including up to 16 years of follow-up from birth.
⇒ Cancer diagnoses are identified from inpatient hospital records only (International Classification of Diseases, 10th Revision (ICD-10) codes), which may miss cases treated exclusively in outpatient settings.
⇒ Educational data are limited to state-funded schools and do not include children educated in private or alternative settings.
⇒ The Education and Child Health Insights from Linked Data (ECHILD) dataset lacks detailed clinical information on cancer treatment and primary care interactions.

## INTRODUCTION

### Background literature

Childhood cancer, although rare, remains the leading cause of mortality among children aged 1 to 9 years.[1 2] In England, approximately 1450 children (aged 0–14 years) and 668 young people (aged 15–19 years) are diagnosed annually,[3] with the most prevalent cancers during childhood being haematological malignancies (including leukaemia and lymphoma) and cancers of the central nervous system (CNS).[4] Advancements in treatment over the past decade have steadily increased survival rates, with over 80% of children living at least 5 years post-diagnosis.[5]

Despite these advancements, cancer and its prolonged treatments (often extending over months or years) may have short and long-term physical, psychological and social effects on survivors.[6–10] As the population of Childhood Cancer Survivors (CCSs) grows, there is a pressing need to identify key factors that shape their life trajectories and improve quality of life. The term 'childhood cancer survivor' is used in alignment with existing literature and insights shared from our patient and public involvement (PPI) experts by lived experience advisory groups for this study. However, we acknowledge this terminology may not resonate with all individuals who have experienced cancer during childhood.[11 12] Some may prefer to emphasise thriving beyond their diagnosis rather than the notion of survival, while others may identify differently based on their unique experiences and personal journeys. Additionally, although the term encompasses individuals diagnosed between birth and 18 years of age, we acknowledge that those in mid-to-late adolescence may identify as young people rather than children.

Educational outcomes are key indicators of future employment, income, progression and long-term quality of life.[10 13] Childhood cancer may adversely affect children's educational achievements due to treatment time and side effects, prolonged school absences, disrupted peer relationships and insufficient collaboration between schools, healthcare providers and families.[14 15] In the literature, educational outcomes of specific interest have been operationalised as attainment level and performance. Previous studies have predominantly focused on attainment level, defined as the highest level of educational level completed.[16–19] Fewer studies have looked at children's performance, defined as the marks or gradings obtained from an assessment.[17] Performance throughout both primary and secondary school lays an important foundation for learning,[20 21] and General Certificate of Secondary Education (GCSE) grades, especially in mathematics, are associated with higher future income and employment prospects.[22] Additional factors that may influence educational outcomes include educational support (special educational needs and disabilities; SEND) and school absenteeism.

Systematic reviews and meta-analyses on educational attainment literature have consistently reported survivors of CNS cancers to have lower educational attainment compared with non-cancer peers.[16–19] However, findings from studies that group all types of cancer survivors together show inconsistent and conflicting findings.[16 17] This is likely due to the inclusion of broad studies with varied designs, populations and outcomes.[23] Many studies were not population-based, had small sample sizes and lacked statistical power to detect differences, particularly given the rarity of childhood cancer and the diagnostic subgroups.[16] Meta-analyses also showed inconsistencies. Gummersall *et al*[18] reported that survivors were less likely to complete secondary school and university. Similarly, Saatci *et al* [19] found that CCSs were more likely to have only completed compulsory education (usually secondary school in most countries) and less likely to have completed university. There were no differences between CCSs and their peers in non-compulsory education before university, such as college-level education. Hernádfői *et al* [24] demonstrated a lack of evidence for differences in educational attainment across different levels; secondary education, college or bachelor's degree, and postgraduate studies. Variability in meta-analysis findings may reflect methodological heterogeneity. A meta-analysis[18] comparing pooled estimates from self-report and registry studies found that studies relying on self-reported data did not identify differences in educational attainment between CCSs and their peers. In contrast, registry-based studies showed CCSs were less likely to complete secondary education than their peers.

Registry studies conducted in Canada, Denmark, Finland, Italy, Norway and Sweden using national (or province) health and education records have consistently shown lower educational attainment among CCSs compared with their peers.[25–33] A recent study[34] combining data from Denmark, Finland and Sweden found that CCSs had a greater risk of not completing secondary school, compared with peers and siblings, with risks being greatest for CNS tumour survivors and weakest when survivors of all types of cancer were grouped together. However, these countries have different education systems to that found in England. Research in Scotland[35] has also shown that children with a history of cancer had higher odds of requiring special educational support, increased absenteeism and lower academic attainment, particularly those with haematological or CNS cancers.

A gap in the literature is the lack of longitudinal analyses of educational outcomes over time. The few studies that have explored pupil performances have often focused on single time-point assessments.[8 27 36–38] Overall, previous research has been limited by restricted access to multi-timepoint educational data and health records, preventing a full understanding of long-term academic trajectories. Another gap relates to inconsistencies in findings on how age at diagnosis affects educational outcomes. Some research indicates that being diagnosed before age 10 leads to poorer long-term academic outcomes, compared with peers without cancer,[18] due to treatment-related neurodevelopmental effects on the developing brain during early childhood.[39 40] Other studies suggest that diagnosis and treatment during adolescence (ages 10–19) may impact prefrontal cortex development, which plays a crucial role in executive functioning and learning.[41 42] Additionally, diagnosis near assessment time point may impact on performance due to school absences for treatment and recovery being in close proximity to examinations.[14 43]

This study will examine the impact of age and the timing of diagnosis in relation to educational performances. Findings could inform school-based interventions to better support children and young people (CYP) diagnosed with cancer and improve their educational performance during compulsory education, subsequently

impacting on their long-term outcomes beyond secondary school.

### Aim and objectives

The overarching aim of this study is to investigate educational performances of CCSs and evaluate how this differs according to sociodemographic characteristics and age at diagnosis, compared with non-cancer peers. We will also investigate how SEND and school absences vary between CCSs compared with non-cancer peers. Analyses will be stratified by cancer type (leukaemia, brain/CNS, lymphomas and any other cancers).

#### Objectives

##### *Objective 1*

To describe, for all cancers categorised together and stratified by cancer type, differences in educational outcomes between CCS and non-cancer peers for:

a. The proportion of children sitting examinations and proportion who did not complete examinations as expected at KS1 (age 7), KS2 (age 11) and KS4 (GCSEs, age 16).

   Of those who did sit examinations:

b. The proportion of children achieving the expected levels of the National Curriculum in line with Department for Education standards, and performances (standardised test scores), at KS1, KS2 and KS4.
c. The proportion of children with recorded SEND or an Education, Health and Care Plan (EHCP) by the end of each KS.
d. The total number of absences and proportion of children with persistent absences (>10% of missed sessions within a term) during the final year of each KS.

##### *Objective 2*

To explore factors contributing to differences in primary outcome KS4 GCSEs, between CCS and non-cancer peers by:

a. Quantifying the extent to which any differences in KS4 GCSE results between CCS and non-cancer peers are explained by sex, ethnicity, deprivation and region of England.
b. Evaluating the extent to which differences between CCS and non-cancer peers differ according to sex, ethnicity, deprivation and region of England (through interaction terms), to investigate potential inequalities.
c. Exploring whether cancer diagnosis disrupts educational performance trajectories, by comparing pre-morbid KS2 performance with post-diagnosis KS4 GCSE performance (restricting our cohort to those with complete data at KS2 and KS4 GCSE assessments, where diagnosis occurs between the two KSs). We will identify children whose performance declined, remained stable or improved between KS2 and KS4, to explore patterns of recovery and persistent disadvantage.

##### *Objective 3*

To investigate the sensitivity of KS4 GCSE outcomes to age at diagnosis by:

a. Evaluating differences in KS4 GCSE outcomes based on two categorical ages at diagnosis groupings.
   - Academically aligned categories, structured to reflect KSs: No diagnosis, diagnosis at age 0–7 years (ending KS1), 8–11 years (ending KS2) and 12–16 years (ending KS4)
   - Standard childhood cancer reporting categories, reflecting developmental stages and healthcare classifications: No diagnosis, diagnosis at age 0–4 (infancy and early childhood), 5–9 years (middle childhood), 10–14 years (early adolescence) and 15–16 years (late adolescence, overlapping with Teenagers and Young Adults classification).
   - We will additionally report on those diagnosed <1 years old, which represents a period of higher cancer incidence and unique developmental vulnerability.
b. Evaluating differences in KS4 GCSE outcomes according to a continuous variable for age at diagnosis (excluding those with no cancer diagnosis).

## METHODS AND ANALYSIS

### Study design

This is an observational population-based cohort study using linked administrative data from health and educational databases in England (Education and Child Health Insights from Linked Data (ECHILD)). We will develop birth cohorts of CYP, with longitudinal follow-up on education and health. The study commenced in July 2024 and is scheduled to conclude by July 2026.

### Data sources and linkage

ECHILD contains linked, longitudinal records from routinely collected health and education data.[44] Health-related data is derived from Hospital Episode Statistics (HES), a national database consisting of all information collected in the National Health Service (NHS) acute hospital care (including inpatient, outpatient, emergency services and diagnosis) and mortality data.[45] Using HES, children can be followed from their birth admission throughout all subsequent NHS hospital contacts.[46] In England, ~97% of children are born in NHS hospitals; however, HES does not include data from births in private hospitals or home settings.

Education data are available from the National Pupil Database (NPD), which captures comprehensive information on CYP attending state-funded schools in England.[47] This includes data on assessments conducted at ages 5, 7, 11, 16 and 18 years, SEND and school absences. In England, full-time education is compulsory from ages 5 to 16, with the participation extended to age 18 between years 2013–2015. Primary school education begins at age five and transitions to secondary school education at age 11. Pupils follow the National Curriculum, which is divided into phases called KS: assessments occur at age 7 (KS1), 11 (KS2) and 14 (KS3), with national examinations (GCSEs) administered at age 16 (KS4) and assessments

such as A-levels and other post-16 qualifications at age 18 (KS5). The database also records school-level characteristics, such as school type.

ECHILD does not capture data on CYP educated in the private sector (approximately 7% of pupils in England annually). Within ECHILD, ~99% of pupils in the NPD have been linked to their NHS ID. However, the NHS national data opt-out was applied to ECHILD, affecting approximately 5.8% of identified linkages. For further details on linkage methodology, see the ECHILD user guide.[48]

## Study population and follow-up

We will develop birth cohorts of children and young people using ECHILD. The study population will include all children born in NHS hospitals in England between first September 1997 and 31st August 2022 (HES birth cohort, linked to NPD (estimated sample size~10 million). Estimates are derived from previous ECHILD linkage reports published elsewhere.[48 49]

Educational outcomes will be obtained by linking the HES birth cohort to records in the NPD for children assessed at school during the academic years 2004/05 to 2021/22. We will extract relevant education and health data, including attainment and performance in KS assessments. Cancer diagnoses will be identified from HES Admitted Patient Care (APC) data, capturing the earliest diagnosis (if any) across four broad cancer categories: Leukaemia, Brain/CNS, Lymphomas and any other cancers. Death registration data from the Office for National Statistics (ONS), including date and cause of death, will be used to supplement diagnostic records in HES APC. To validate cancer diagnoses, we will cross-reference these with publicly accessible cancer incidence figures from NHS England's Cancer Data (National Cancer Registration and Analysis Service).[2]

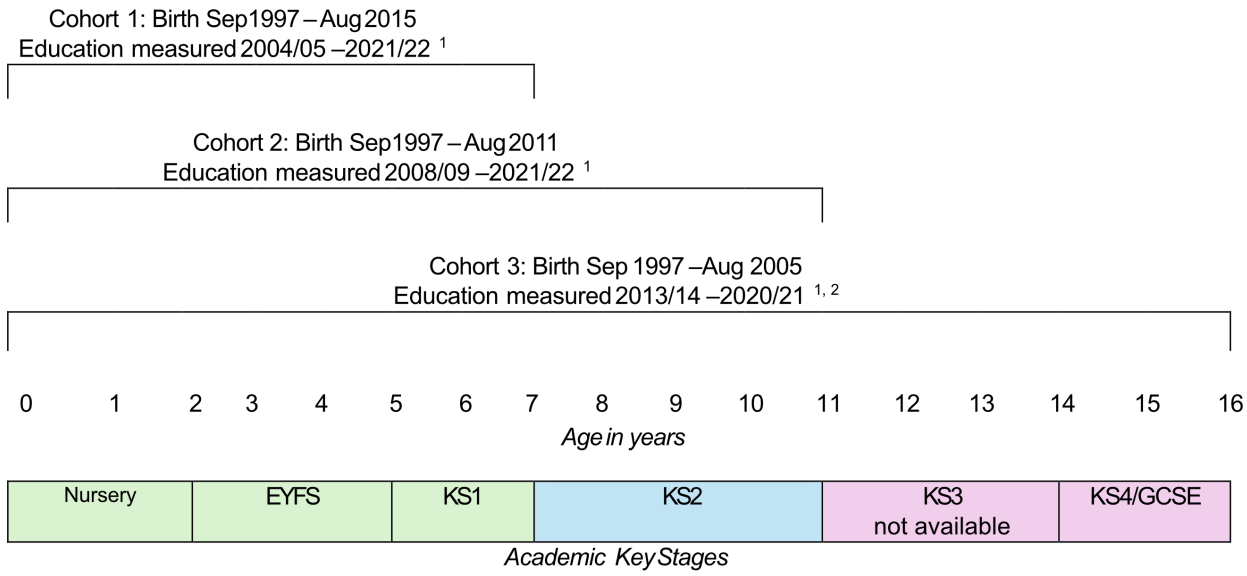
We will create three cohorts, each capturing a differing length of follow-up (See figures 1 and 2). Each sequential cohort will include a longer follow-up period but smaller population sample. Cohort 1 will include births between the academic years September 1997 to August 2015 and capture KS1 outcomes at age 7, excluding those born between September 2012 to August 2014 (due to their KS1 outcomes not being captured during the COVID pandemic). Cohort 2 will include births between September 1997 and August 2011 and capture KS2 outcomes at age 11, excluding those born between September 2008 and August 2010 (due to COVID).

| Year of birth (1st September 1997 –31st August 2015) | Academic year (1st September 2002 –31st August 2022) | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2002 /03 | 2003 /04 | 2004 /05 | 2005 /06 | 2006 /07 | 2007 /08 | 2008 /09 | 2009 /10 | 2010 /11 | 2011 /12 | 2012 /13 | 2013 /14 | 2014 /15 | 2015 /16 | 2016 /17 | 2017 /18 | 2018 /19 | 2019 /20[1] | 2020 /21[1] | 2021 /22[2] |
| 1997 /98 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 | 14 | KS4 15 | KS4 16 | | | | | | | | |
| 1998 /99 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 | 14 | KS4 15 | KS4 16 | | | | | | | |
| 1999 /00 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 | 14 | KS4 15 | KS4 16 | | | | | | |
| 2000 /01 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 | 14 | KS4 15 | KS4 16 | | | | | |
| 2001 /02 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 | 14 | KS4 15 | KS4 16 | | | | |
| 2002 /03 | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 | 14 | KS4 15 | KS4 16 | | | |
| 2003 /04 | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 | 14 | KS4 15 | KS4 16 | | |
| 2004 /05 | | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 | 14 | KS4 15 | KS4 16 | |
| 2005 /06 | | | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 | 14 | KS4 15 | KS4 16 |
| 2006 /07 | | | | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 | 14 | KS4 15 |
| 2007 /08 | | | | | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 | 14 |
| 2008 /09 | | | | | | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 | 13 |
| 2009 /10 | | | | | | | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 | 12 |
| 2010 /11 | | | | | | | | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 | KS2 11 |
| 2011 /12 | | | | | | | | | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 | KS2 10 |
| 2012 /13 | | | | | | | | | | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 | KS2 9 |
| 2013 /14 | | | | | | | | | | | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 | KS2 8 |
| 2014 /15 | | | | | | | | | | | | | 0 | 1 | 2 | 3 | 4 | R 5 | KS1 6 | KS1 7 |

[1] Data not recorded during covid years for KS1 and KS2 (2019/20 and 2020/21). KS4 were teacher assessed.
[2] Data for KS4 in academic year 2021/22 not available within present version of the ECHILD research database.

KS1: Age 5 - 7 | KS2: Age 7 - 11 | KS4: Age 14 - 16 | Data unavailable [1,2]

**Figure 1** Expected age during primary and secondary school by year of birth and follow-up years in education. ECHILD, Education and Child Health Insights from Linked Data; KS, key stage; R, reception.




[1] Data not recorded during covid years for KS1 and KS2 (2019/20 and 2020/21). KS4 were teacher assessed.
[2] Data for KS4 in academic year 2021/22 not available within present version of the ECHILD research database.

**Figure 2** Cohort academic birth years, ages and education key stages. ECHILD, Education and Child Health Insights from Linked Data; EYFS, Early Years Foundation Stage; GCSE, General Certificate of Secondary Education; KS, key stage.

Cohort 3 will include births between September 1997 and August 2005 and capture KS4 outcomes at age 16.

## Key study measures

### Exposure: childhood cancer

Childhood cancer diagnoses will be identified using International Classification of Diseases, 10th Revision (ICD-10) codes as recorded in HES APC data.[50 51] Most children diagnosed with cancers will be admitted to hospital at some point during their diagnosis and treatment. Groups will include four broad diagnostic cancer types, three with greatest incidence in childhood, and all other cancers combined. The three broad diagnostic cancer groups with greatest incidence are Leukaemia, CNS, Lymphomas and any other cancers, accounting for, respectively, 32%, 24% and 12% of all childhood (0–14 years) cancers recorded in England in the period 1997 to 2016.[3] Children diagnosed with cancer will be compared with peers with no recorded diagnosis of cancer prior to the July following the relevant KS outcome assessment. Age at diagnosis will be derived by using HES birth date and diagnosis date and categorised using two age at diagnosis groupings. Academically aligned categories will be structured to reflect UK KSs: No diagnosis, diagnosis at age 0–7, age 8–11 and age 12–16, to correspond to ending at academic years of school for year 2 (6/7, KS1), year 6 (10/11, KS2) and year 11 (15/16, KS4). Standard childhood cancer reporting categories will also be used, to reflect developmental stages and healthcare classifications: No diagnosis, diagnosis at age 0–4 (infancy and early childhood), age 5–9 (middle childhood), age 10–14 years (early adolescence) and 15–16 years (late adolescence, overlapping with Teenagers and Young Adults classification). We will additionally report on those diagnosed <1 years old, which represents a period of higher cancer incidence and unique developmental vulnerability.

### Outcomes: educational outcomes

We will evaluate educational performances at each KS. For KS1 and KS2, using a binary outcome (those who reached expected levels of the National Curriculum as defined with Department of Education standards), and a continuous standardised point score to measure performance.[52] Scores will be standardised within years, to make fair comparisons across years. Data on assessments at KS3 are not available in ECHILD. For KS4, we will assess the number and proportion of children achieving 5 A*- Cs (9 - 4 grades in reformed GCSEs), including English and Maths, and standardised point scores in English and Maths GCSE exams.[53] We will evaluate the number and proportion of children sitting examinations (ie,appearing in the school Census and sitting the exam) and those who did not complete the examinations as expected (ie, appearing in the Census but not having an assessment record) at KS1 (age 7), KS2 (age 11) and KS4 (GCSEs, age 16).

We will also evaluate the number and proportion with recorded SEND[16–19] or an EHCP. SEND provision will be grouped in the following descending hierarchy: 1. No SEND provision, 2. SEND support in mainstream school, 3. EHCP in mainstream school, 4. Enrolment in specialist provision. School absences will be defined as: 1. pupils who miss 10% or more of their possible sessions (persistent absence), 2. authorised/unauthorised absences for doctor/dentist appointments or illness and 3. Authorised/unauthorised absences for all other reasons than medical. Absence data is available from 2005/06 and reason for absence is available from

2006/07. Data were not recorded during academic years 2019/20 and 2020/21 for KS1 outcomes, KS2 outcomes and school absences due to the covid pandemic. KS4 outcomes data during these years were teacher assessed, rather than based on formal examination marks. KS4 data are not available for 2021/22 within the current version of ECHILD.

### Additional covariates of interest

We will derive a set of sociodemographic characteristics including year of birth, sex, ethnic group, region of England, and area-level deprivation (quintile of the Index of Multiple Deprivation; IMD). Infant birth characteristics will include gestational age at birth defined as very preterm (<32 weeks), moderate preterm (32–33 weeks), late preterm (34–36 weeks), and term (≥37 weeks), and congenital anomalies, identified using an International Classification of Diseases version 10 (ICD-10) code list.[51] We will also flag individuals with Down syndrome, neurofibromatosis or tuberous sclerosis (recorded in hospital records within the year from birth).[34]

## Statistical analysis plan

### Cohort description

We will use a CONSORT-style flowchart to illustrate the linkage process between HES and NPD records, and to show the inclusion and exclusion of children at each stage leading to the formation of the three cohorts for analysis. For each cohort, we will compare the characteristics of children who were linked with those who were not, as well as children with and without available data at each KS. We will report on follow-up completeness and attrition for children diagnosed with cancer and those without. Baseline sociodemographic and clinical characteristics will be described for children with and without a cancer diagnosis across the cohorts. Differences will be visualised graphically and evaluated using standardised mean differences to assess the size of any differences between groups. We will additionally report age at cancer diagnosis for affected children within the cohort.

#### *Objective 1*

We will quantify the proportion of children in each group with each binary outcome and describe mean standardised scores across groups for continuous outcomes. We will model the proportion of children achieving expected levels at each KS using generalised linear regression to generate relative risks and 95% CIs. We will use linear regression to estimate the difference in standardised scores at each KS between groups. Models will account for clustering of pupils within schools using robust standard errors.

#### *Objective 2*

We will adopt a causal inference framework to guide analysis. Building on our model for KS4 outcomes (Cohort 3) from Objective 1, we will evaluate the extent to which differences in outcomes between groups are explained by key covariates. Key confounders will be identified a priori using directed acyclic graphs (DAGs) based on a combination of inputs by experts with lived experience, clinical/epidemiological knowledge and existing literature. This approach supports transparency in assumptions and helps avoid inappropriate adjustment.[54] While we aim to strengthen causal interpretation, we recognise the inherent limitations of observational data and will interpret our findings with appropriate caution.

To determine whether the effect of having cancer differs according to sociodemographic group, we will include interaction terms within the models. This will enable us to identify any inequalities in outcomes which exist according to, for example, sex, ethnicity or deprivation.

To explore whether cancer diagnosis disrupts educational performance trajectories, we will restrict our analysis to the subset of children with complete data on assessments at KS2 and KS4, who had a cancer diagnosis between the two assessment time points. We will describe the number of children achieving expected levels at KS2 who went on to either achieve expected levels at KS4 or not. Additionally, we will explore patterns of recovery and persistent disadvantage by identifying children whose performance declined, remained stable or improved between KS2 and KS4.

#### *Objective 3*

Building on our model for KS4 outcomes (Cohort 3) from Objective 1 and 2, we will quantify the proportion of children achieving 5A*- Cs (9 - 4 grades in reformed GCSEs), including English and Maths and compare average standardised point scores at GCSE according to categorical age at diagnosis. Relative risks for each age group will be estimated from a generalised linear regression model. For those with cancer only, we will create a separate model including continuous age at diagnosis (in years) to evaluate the effect of each individual year on outcomes.

We will further explore the effect of timing of diagnosis in relation to outcomes by restricting our cohort to children aged 0–5 at diagnosis and comparing effect sizes for KS1, KS2 and KS4.

## Missing data

We will quantify missing data in covariates and assess whether this is associated with our primary outcome (GCSE grades at KS4). For ethnicity, missing data will be categorised as unknown. Decisions on whether to use complete case analysis or multiple imputation will be taken based on levels and patterns of missingness. We will conduct sensitivity analysis where required.

## Patient and public involvement and engagement

Previously published patient and public involvement and engagement (PPIE) with CYP and their carers identified educational outcomes as a key priority for data use,[55] and highlighted communication and involvement of CSS as a key component to public and patient trust on data use for research. Through several previous consultation

works conducted within our broader child health informatics team at Great Ormond Street Institute of Child Health, University College London,[56] acceptance was found among CYP on the use of administrative data for research, provided that they are included in the research process.

This study is conducted in consultation with a Young Person's experts by lived experience advisory group, including those who self-identify as experts by lived experience, survivors and warriors. Members were recruited via Young Lives vs Cancer. Their input has informed the development of research priorities and will guide interpretation of findings from the ECHILD linked health and education administrative data. Engagement is structured across four meetings aligned with key project milestones. Group members are invited to contribute as co-authors on research publications.

In the first meeting (April 2025), the advisory group helped identify research priorities. Educational outcomes and inadequate support, especially within the educational systems, emerged as central concerns. Members emphasised the lasting effects of cancer diagnosis and treatment, and the difficulty of reintegrating into education. They described a system that often overlooks emotional and social well-being, focusing narrowly on physical recovery and offering few accommodations. In the second meeting (July 2025), the group inputted into our analysis plan through discussions on covariates included in the analysis which cause inequalities (see Methods section), how age at diagnosis impacted on education and understanding of what is missing in our data. This knowledge will inform the construction of a DAG and provide contextual understanding.

In a separate consultation (June 2025), we engaged with a group of parents of young children diagnosed with cancer. Their feedback echoed concerns around inadequate educational support and highlighted the isolating experience of advocating for their children's rights to education. Parents described a prevailing expectation for children to simply 'slot back' into education following treatment, with limited awareness or provision for their emotional well-being. They also provided useful insights into potential sources of missing data in our cohort.

A key theme that emerged from our PPIE activities was the complexity of accessing support and navigating the education system. CYP were often expected to seamlessly resume their education post-treatment without adequate consideration of short- and long-term physical and emotional challenges. The rapidly changing and fluid nature of cancer treatment is not accommodated effectively by the current education system, such as Education and Healthcare plans. There was also reported inflexibility of exam boards and the wider education system with regards to formal examinations, which would be particularly pertinent to our research objective focusing on GCSE results. Although our dataset does not fully reflect these nuanced experiences, we are committed to integrating these insights into the interpretation of our findings to better understand the contextual realities. We also aim to embed lived experience voices in future mixed-methods research to more fully capture the narratives underlying the data.

We did not require ethics review to conduct this involvement activity as the young people and parents were advisors to the study team rather than research subjects.

## ETHICS AND DISSEMINATION

### Ethics

Approvals to access and to conduct statistical analyses of linked, de-identified data from HES and the NPD were obtained from NHS England (DARS-NIC-381972) and the Department for Education (DR200604.02B). Ethical approval for the ECHILD project was obtained from the NHS Health Research Authority Research Ethics Committee (REC references: 20/EE/0180 and 21/SW/0159), the National Research Ethics Service (17/LO/1494) and the UCL Great Ormond Street Institute of Child Health Joint Research and Development Office (20PE06). As the study uses de-identified, routinely collected administrative data, individual informed consent was not required and was waived by the approving ethics committees. All data were anonymised prior to access by the study team.

### Dissemination

We aim to ensure transparency, reproducibility and impact across academic, policy and public domains. Study methods and analysis code will be shared via the ECHILD GitHub repository and within the secure research environment of the Office for National Statistics (ONS). We will present preliminary findings to academic and policy audiences and engage our PPI advisory groups, young people and parents/carers, through structured consultations to share emerging results and gather input on interpretation of findings. Final outputs will include peer-reviewed articles, a funder report, and accessible materials such as infographics for non-academic audiences. The cohorts developed will support future research into long-term educational and health outcomes for CYP affected by cancer. Due to data-sharing agreements with NHS England and the Department for Education, the ECHILD database is not publicly available. Access is limited to accredited researchers via application to the ECHILD team and approval from ONS.

**Acknowledgements** We gratefully acknowledge all children, young people and families whose de-identified data will be used in this research. We would like to thank our young persons and parents PPIE lived experience advisory groups, who provided context and understanding of our data, and input into our study generously with their time and valuable insights. We would like to thank Young Lives vs Cancer, Children's Cancer and Leukaemia Group and Childhood Cancer Parents Alliance who assisted with recruitment for our PPIE groups. We thank Professor Ruth Gilber for input into the original grant proposal, and the ECHILD team members supporting this work—Ruth Blackburn, Vincent Grigori Nguyen, Farzan Ramzan and Milagros Ruiz Nishiki. We are grateful to the Office for National Statistics for providing the trusted research environment for the ECHILD database. The ECHILD database uses data from the Department for Education (DfE). The DfE does not accept responsibility for any inferences or conclusions derived by the authors. This

research used data provided by patients and collected by the NHS as part of their care and support. Source data can also be accessed by researchers by applying to NHS England. All research at Great Ormond Street Hospital NHS Foundation Trust and UCL Great Ormond Street Institute of Child Health is made possible by the NIHR Great Ormond Street Hospital Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

**Contributors** KH is the principal investigator of Cancer Research UK and Children with Cancer UK grant. TSto, RGF, DH and MGM are coinvestigators on the grant. SN wrote and produced the first draft of this manuscript. All authors reviewed, read and approved the final manuscript and contributed to and are responsible for the final design of the study. Guarantor: KH is the guarantor of this work.

**Funding** This work was supported and jointly funded by Cancer Research UK and Children with Cancer UK (CDEPIL-Jan24/100002). We gratefully acknowledge their support in enabling this study. The views expressed are those of the authors and not necessarily those of Cancer Research UK or Children with Cancer. The funders of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were involved in the design, or conduct, or reporting or dissemination plans of this research. Refer to the Methods section for further details.

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

**ORCID iDs**
Selina Nath https://orcid.org/0000-0002-5976-1301
Richard G Feltbower https://orcid.org/0000-0002-1728-9408
Katie Harron https://orcid.org/0000-0002-3418-2856

## REFERENCES

1 Royal College of Paediatrics and Child Health. State of child health. London, Available: https://stateofchildhealth.rcpch.ac.uk/evidence/mortality/child-mortality
2 Public Health England: The National Cancer Registration and Analysis Service for England. Children, teenagers and young adults uk cancer statistics report 2021. 2021.
3 National Disease Registration Service. Children, teenagers and young adults uk cancer statistics report 2021. 2021. Available: https://digital.nhs.uk/ndrs/data/data-outputs/cancer-publications-and-tools/ctya-uk-cancer-statistics-report-2021/cancer-incidence
4 Office for National Statistics. Cancer registration statistics, england: 2017 - cancer diagnoses and age-standardised incidence rates for all types of cancer by age, sex and region including breast, prostate, lung and colorectal cancer. 2017. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2017
5 Office for National Statistics. Childhood cancer survival in england: children diagnosed from 1990 to 2014 and followed up to 2015 - long-term trends in the number of children (aged 0 to 14) surviving cancer 5 and 10 years after diagnosis. 2017. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/childhoodcancersurvivalinengland/childrendiagnosedfrom1990to2014andfollowedupto2015
6 Yeh JM, Ward ZJ, Stratton KL, *et al*. Accelerated Aging in Survivors of Childhood Cancer-Early Onset and Excess Risk of Chronic Conditions. *JAMA Oncol* 2025;11:535–43.
7 Stenmarker M, Mallios P, Hedayati E, *et al*. Morbidity and mortality among children, adolescents, and young adults with cancer over six decades: a Swedish population-based cohort study (the Rebuc study). *Lancet Reg Health Eur* 2024;42:100925.
8 Barrera M, Shaw AK, Speechley KN, *et al*. Educational and social late effects of childhood cancer and related clinical, personal, and familial characteristics. *Cancer* 2005;104:1751–60.
9 Brinkman TM, Li C, Vannatta K, *et al*. Behavioral, Social, and Emotional Symptom Comorbidities and Profiles in Adolescent Survivors of Childhood Cancer: A Report From the Childhood Cancer Survivor Study. *J Clin Oncol* 2016;34:3417–25.
10 Godono A, Felicetti F, Conti A, *et al*. Employment among Childhood Cancer Survivors: A Systematic Review and Meta-Analysis. *Cancers (Basel)* 2022;14:4586.
11 Cheung SY, Delfabbro P. Are you a cancer survivor? A review on cancer identity. *J Cancer Surviv* 2016;10:759–71.
12 Granek L, Nathan PC, Rosenberg-Yunger ZRS, *et al*. Psychological factors impacting transition from paediatric to adult care by childhood cancer survivors. *J Cancer Surviv* 2012;6:260–9.
13 Devine KA, Christen S, Mulder RL, *et al*. Recommendations for the surveillance of education and employment outcomes in survivors of childhood, adolescent, and young adult cancer: A report from the International Late Effects of Childhood Cancer Guideline Harmonization Group. *Cancer* 2022;128:2405–19.
14 Martinez-Santos A-E, Fernandez-De-La-Iglesia JDC, Sheaf G, *et al*. A systematic review of the educational experiences and needs of children with cancer returning to school. *J Adv Nurs* 2021;77:2971–94.
15 Wang M, Dzifa KL, Lei J, *et al*. The experiences of children and adolescents with cancer returning to school: A qualitative meta-synthesis. *J Pediatr Nurs* 2024;76:140–9.
16 Molcho M, D'Eath M, Alforque Thomas A, *et al*. Educational attainment of childhood cancer survivors: A systematic review. *Cancer Med* 2019;8:3182–95.
17 Frederiksen LE, Mader L, Feychting M, *et al*. Surviving childhood cancer: a systematic review of studies on risk and determinants of adverse socioeconomic outcomes. *Int J Cancer* 2019;144:1796–823.
18 Gummersall T, Skaczkowski G, Wilson C. Childhood cancer, age at diagnosis and educational attainment: A meta-analysis. *Crit Rev Oncol Hematol* 2020;145:102838.
19 Saatci D, Thomas A, Botting B, *et al*. Educational attainment in childhood cancer survivors: a meta-analysis. *Arch Dis Child* 2020;105:339–46.
20 CLIC Sargent for Children with Cancer. No child with cancer left out: the impact of cancer on children's primary school education. 2012.
21 CLIC Sargent for Children with Cancer. No teenager with cancer left out: the impact of cancer on young people's secondary school education. 2013.
22 Department for education. GCSE attainment and lifetime earnings: reseach report. 2021.
23 Gough D. Qualitative and mixed methods in systematic reviews. *Syst Rev* 2015;4:181.
24 Hernádfoi MV, Koch DK, Kói T, *et al*. Burden of Childhood Cancer and the Social and Economic Challenges in Adulthood: A Systematic Review and Meta-Analysis. *JAMA Pediatr* 2024;178:548–66.
25 Koch SV, Kejs AMT, Engholm G, *et al*. Educational attainment among survivors of childhood cancer: a population-based cohort study in Denmark. *Br J Cancer* 2004;91:923–8.
26 Ahomäki R, Harila-Saari A, Matomäki J, *et al*. Non-graduation after comprehensive school, and early retirement but not unemployment are prominent in childhood cancer survivors-a Finnish registry-based study. *J Cancer Surviv* 2017;11:284–94.
27 Harila-Saari AH, Lähteenmäki PM, Pukkala E, *et al*. Scholastic achievements of childhood leukemia patients: a nationwide, register-based study. *J Clin Oncol* 2007;25:3518–24.
28 Lönnerblad M, Suominen R, Harila-Saari A. Childhood lymphoma treatment impacts educational outcomes: a registry study from Sweden. *J Cancer Surviv* 2023;17:1072–81.
29 Lorenzi M, McMillan AJ, Siegel LS, *et al*. Educational outcomes among survivors of childhood cancer in British Columbia, Canada: report of the Childhood/Adolescent/Young Adult Cancer Survivors (CAYACS) Program. *Cancer* 2009;115:2234–45.
30 Boman KK, Lindblad F, Hjern A. Long-term outcomes of childhood cancer survivors in Sweden: a population-based study of education, employment, and income. *Cancer* 2010;116:1385–91.
31 Ghaderi S, Engeland A, Gunnes MW, *et al*. Educational attainment among long-term survivors of cancer in childhood and adolescence: a Norwegian population-based cohort study. *J Cancer Surviv* 2016;10:87–95.
32 Maule M, Zugna D, Migliore E, *et al*. Surviving a childhood cancer: impact on education and employment. *Eur J Cancer Prev* 2017;26:351–6.
33 Lönnerblad M, van't Hooft I, Blomgren K, *et al*. A nationwide, population-based study of school grades, delayed graduation, and qualification for school years 10-12, in children with brain tumors in Sweden. *Pediatr Blood Cancer* 2020;67:e28014.

34 Mogensen H, Tettamanti G, Frederiksen LE, *et al*. Educational attainment in survivors of childhood cancer in Denmark, Finland, and Sweden. *Br J Cancer* 2024;130:260–8.
35 Baughan N, Pell JP, Mackay DF, *et al*. Educational outcomes in childhood cancer survivors: A Scotland-wide record-linkage study of 766,217 schoolchildren. *PLoS ONE* 2023;18:e0286840.
36 Andersen KK, Duun-Henriksen AK, Frederiksen MH, *et al*. Ninth grade school performance in Danish childhood cancer survivors. *Br J Cancer* 2017;116:398–404.
37 Lähteenmäki PM, Harila-Saari A, Pukkala EI, *et al*. Scholastic achievements of children with brain tumors at the end of comprehensive education: a nationwide, register-based study. *Neurology (ECronicon)* 2007;69:296–305.
38 Lähteenmäki PM, Sankila R, Pukkala E, *et al*. Scholastic achievement of children with lymphoma or Wilms tumor at the end of comprehensive education--a nationwide, register-based study. *Int J Cancer* 2008;123:2401–5.
39 Jacola LM, Edelstein K, Liu W, *et al*. Cognitive, behaviour, and academic functioning in adolescent and young adult survivors of childhood acute lymphoblastic leukaemia: a report from the Childhood Cancer Survivor Study. *Lancet Psychiatry* 2016;3:965–72.
40 Jacola LM, Krull KR, Pui C-H, *et al*. Longitudinal Assessment of Neurocognitive Outcomes in Survivors of Childhood Acute Lymphoblastic Leukemia Treated on a Contemporary Chemotherapy Protocol. *J Clin Oncol* 2016;34:1239–47.
41 Prasad PK, Hardy KK, Zhang N, *et al*. Psychosocial and Neurocognitive Outcomes in Adult Survivors of Adolescent and Early Young Adult Cancer: A Report From the Childhood Cancer Survivor Study. *J Clin Oncol* 2015;33:2545–52.
42 John TD, Sender LS, Bota DA. Cognitive Impairment in Survivors of Adolescent and Early Young Adult Onset Non-CNS Cancers: Does Chemotherapy Play a Role? *J Adolesc Young Adult Oncol* 2016;5:226–31.
43 Vance YH, Eiser C. The school experience of the child with cancer. *Child Care Health Dev* 2002;28:5–19.
44 Mc Grath-Lone L, Libuy N, Harron K, *et al*. Data Resource Profile: The Education and Child Health Insights from Linked Data (ECHILD) Database. *Int J Epidemiol* 2022;51:17–17f.
45 Herbert A, Wijlaars L, Zylbersztejn A, *et al*. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017;46:1093–1093i.
46 Zylbersztejn A, Lewis K, Nguyen V, *et al*. Evaluation of variation in special educational needs provision and its impact on health and education using administrative records for England: umbrella protocol for a mixed-methods research programme. *BMJ Open* 2023;13:e072531.
47 Jay MA, McGrath-Lone L, Gilbert R. Data Resource: the National Pupil Database (NPD). *Int J Popul Data Sci* 2019;4:1101.
48 Ramzan F, Mc Grath-Lone L, Blackburn R, *et al*. Education and child health insights from linked data (echild): an introductory guide for researchers. 2023. Available: https://docs.echild.ac.uk
49 Feng Q, Ireland G, Gilbert R, *et al*. Data Resource Profile: A national linked mother-baby cohort of health, education and social care data in England (ECHILD-MB). *Int J Epidemiol* 2024;53:dyae065.
50 NHS England. Hospital episode statistics data dictionary. 2025. Available: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary#download-hes-data-dictionaries
51 World Health Organization. International Statistical Classification of Diseases and Related Health Problems (10th Ed.).2016. Available: https://icd.who.int/browse10/2016/en
52 Department for Education. Statistics: key stage 2. 2023. Available: https://www.gov.uk/government/collections/statistics-key-stage-2
53 Department for Education. Revised gcse and equivalent results in england, 2015 to 2016. 2017. Available: https://assets.publishing.service.gov.uk/media/5a815d9140f0b62302696ea2/SFR03_2017.pdf
54 Tennant PWG, Murray EJ, Arnold KF, *et al*. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol* 2021;50:620–32.
55 Hughes NF, Fern LA, Polanco A, *et al*. Patient and public involvement to inform priorities and practice for research using existing healthcare data for children's and young people's cancers. *Res Involv Engagem* 2023;9:71.
56 ECHILD: Education and Child Health Insights from Linked Data. Major events 2025. n.d. Available: https://www.echild.ac.uk/engagements