

# Establishing risk factors and outcomes for congenital hypothyroidism with gland in situ using population-based data linkage methods: study protocol

Milagros Ruiz Nishiki <sup>1</sup>, Melissa Cabecinha <sup>2,3</sup>, Rachel Knowles,<sup>4</sup> Catherine Peters,<sup>5</sup> Helen Aitkenhead,<sup>6</sup> Adeboye Ifederu,<sup>6</sup> Nadia Schoenmakers,<sup>7</sup> Neil J Sebire,<sup>8</sup> Erin Walker,<sup>9</sup> Pia Hardelid<sup>1</sup>

**To cite:** Ruiz Nishiki M, Cabecinha M, Knowles R, *et al.* Establishing risk factors and outcomes for congenital hypothyroidism with gland in situ using population-based data linkage methods: study protocol. *BMJ Paediatrics Open* 2022;**6**:e001341. doi:10.1136/bmjpo-2021-001341

Received 5 November 2021  
Accepted 5 March 2022

## ABSTRACT

**Introduction** There has been an increase in the birth prevalence of congenital hypothyroidism (CH) since the introduction of newborn screening, both globally and in the UK. This increase can be accounted for by an increase in CH with gland in situ (CH-GIS). It is not known why CH-GIS is becoming more common, nor how it affects the health, development and learning of children over the long term. Our study will use linked administrative health, education and clinical data to determine risk factors for CH-GIS and describe long-term health and education outcomes for affected children.

**Methods and analysis** We will construct a birth cohort study based on linked, administrative data to determine what factors have contributed to the increase in the birth prevalence of CH-GIS in the UK. We will also set up a follow-up study of cases and controls to determine the health and education outcomes of children with and without CH-GIS. We will use logistic/multinomial regression models to establish risk factors for CH-GIS. Changes in the prevalence of risk factors over time will help to explain the increase in birth prevalence of CH-GIS. Multivariable generalised linear models or Cox proportional hazards regression models will be used to assess the association between type of CH and school performance or health outcomes.

**Ethics and dissemination** This study has been approved by the London Queen Square Research Ethics Committee and the Health Research Authority's Confidentiality Advisory Group CAG. Approvals are also being sought from each data provider. Obtaining approvals from CAG, data providers and information governance bodies have caused considerable delays to the project. Our methods and findings will be published in peer-reviewed journals and presented at academic conferences.

## INTRODUCTION

An increasing number of children are living with rare diseases as a result of improved drug treatments and neonatal care. Parents of affected children frequently report having to manage uncertainty about their child's current health status and prognosis,<sup>1 2</sup> and a lack of expertise about the condition among

## What is already known on this topic?

- ▶ Since the introduction of newborn screening programmes, there has been an increase in the birth prevalence of congenital hypothyroidism with gland in situ (CH-GIS).
- ▶ Rare conditions, such as CH-GIS, can be difficult to characterise in administrative health databases, due to a lack of specificity of clinical coding.
- ▶ Linkage between clinical, genetic and administrative health and education data can be an efficient way of monitoring long-term health and education outcomes for children with rare diseases like CH-GIS.

## What this study hopes to add?

- ▶ Linked clinical and administrative data will be used to determine what factors have contributed to the increase of the birth prevalence of CH-GIS in the UK.
- ▶ This study will establish risk factors for CH-GIS, and examine the association between CH-GIS, school performance and health outcomes.

health professionals.<sup>3</sup> The 2014 UK Strategy for Rare Diseases<sup>4</sup> stresses the lack of longitudinal, population-based data to assess the aetiology and health outcomes of rare conditions to support parents, clinicians and commissioners.

Existing data sources for research into the aetiology and long-term outcomes of rare diseases have major weaknesses. Disease registries contain detailed data on specific conditions, yet data collection requires substantial time input from clinicians and parents, leading to high costs and (non-random) loss to follow-up. Treating clinicians keep detailed records of phenotypes and treatments, however, these may not be stored in an easily analysable format. Further, these data sources only contain data on children with a particular



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Milagros Ruiz Nishiki; mili.ruiz@ucl.ac.uk

condition, and health or development outcomes cannot easily be compared with control groups. In contrast, administrative health and education data are universally and routinely collected for all children, including those with rare diseases. However, rare conditions are hard to characterise in administrative health databases due to lack of specificity of clinical coding. Further, these databases lack information on prescribed drugs, genotypes and biomarkers, which are important for understanding and eventually predicting the expected health and development trajectories of affected children. Instead, linkage between clinical and routinely collected health, vital statistics and education databases would allow long-term follow-up of children with rare conditions. We propose a follow-up study of cases and controls applying population-based data linkage methodologies to examine risk factors, and long-term health and education outcomes for children with congenital hypothyroidism with gland in situ (CH-GIS; a rare endocrine disorder).

### CH and CH-GIS

Newborn screening for CH was introduced in the late 1970s in the UK<sup>5</sup> and many other countries globally.<sup>6</sup> Where screening and early treatment for CH has been introduced, including in the UK, severe neurological impairment due to CH has been eliminated. Approximately 65% of children with CH are born with no, or an underdeveloped or malpositioned, thyroid gland.<sup>7–9</sup> This is known as thyroid dysgenesis, hypoplasia or ectopic thyroid, respectively. The remaining 35% of children with CH have normally situated and sized glands, but with malfunctions in thyroid hormone synthesis: this is known as CH-GIS. The birth prevalence of CH in the UK has increased from 29/10 000 births in 1982–1984 to 53/10 000 births in 2011–2012.<sup>10 11</sup> Many other countries report similar rises, which remain even after considering improvements in screening assays.<sup>9 12</sup> These increases are largely accounted for by a rise in birth prevalence of CH-GIS.<sup>13 14</sup> Several reasons may explain this observed increase, including improved survival of children born prematurely or with low birth weight,<sup>15</sup> or a change in the ethnic group composition of the population.<sup>16</sup> Variations in dietary iodine intake, and variations in the prevalence of autosomal recessive thyroid hormone defects, may contribute to these observed differences in CH by ethnic group.<sup>15</sup> To date, no study has examined multiple risk factors simultaneously.

Despite the increasing proportion of children diagnosed with CH-GIS, the health and development outcomes for children with this condition are not well understood. While some studies have shown suboptimal school performance even in children with very mildly suppressed thyroid hormone levels in the neonatal period,<sup>17</sup> this needs to be weighed against potential risks of suboptimal education outcomes that have been associated with overtreatment.<sup>18 19</sup> Monitoring health and education outcomes for children with CH-GIS is recommended by the European Society for Paediatric

Endocrinology.<sup>20</sup> Many studies of outcomes for children with CH involve small cohorts who were born over 30 years ago.<sup>21–26</sup> Since screening and treatment thresholds have changed substantially during this period, up-to-date studies are urgently required.

### Aim

The overall aims of this study are to determine risk factors for CH-GIS and describe long-term health and education outcomes for affected children. We propose to use a clinical database containing clinical and treatment information for children with CH-GIS (to be developed as part of the project), linked to vital statistics and administrative health and education data to meet the research aims. This protocol describes the current study design, and the original design that had to be amended due to requests from data providers and information governance bodies.

### Objectives

The objectives of this research study are to:

1. Develop a coherent, structured and catalogued clinical database of children with CH, that can also be used for research.
2. Determine what factors have contributed to the increase in the birth prevalence of CH-GIS in the UK.
3. Determine health and education outcomes of children with CH-GIS compared with children with other forms of CH, and those who do not have CH.
4. Describe how a child's CH-GIS genotype affects treatment, health and educational outcomes.
5. Determine treatment trajectories for children with CH-GIS.

## METHODS AND ANALYSIS

### Study design

Objective 2 will be addressed via a population-based cohort study, and objectives 3–5 will be addressed using a follow-up study of cases and controls.

### Study population

For the cohort study (objective 2), the study population will include all children born in the North Thames region and screened at the North Thames Newborn screening laboratory between 1 January 2000 and 31 December 2020. This study period will allow us to examine changes in neonatal levels of thyroid stimulating hormone (TSH) concentrations over a 21-year period. We will propose to include all children, apart from where a child has opted out of their National Health Service (NHS) data being used for research.

For the follow-up study (objectives 3–5), the population consists of all children referred to the Great Ormond Street Hospital (GOSH) Endocrinology Clinic after a screen positive newborn bloodspot test result for CH between 2006 and 2020—this is also the population in the database to be established for objective 1. Fifteen controls will be selected per case of CH for the follow-up study.

## Data sources

### Cohort study data sources

#### *North Thames Newborn Screening database*

The population ‘spine’ for the cohort study will be the North Thames Newborn Screening (NBS) database, held at GOSH; North Thames covers the urban areas of West, North and East London and the suburban and rural areas of Essex, Hertfordshire and Bedfordshire. The database covers all babies in North Thames region screened at the GOSH Newborn Screening laboratory; over 99.9% of babies born in North Thames and approximately 20% of all births in England. The NBS database receives a daily feed of birth notifications from the NHS Personal Demographic Service dataset, a national electronic database of NHS patient details, as part of the Failsafe system, an IT solution for England that reduces the risk of babies missing or having delayed newborn blood spot screening.<sup>27</sup> The North Thames NBS database holds data on screening test results, babies’ and mothers’ NHS numbers, sex, dates of birth and sample collection, ethnicity, birth weight and gestational age.

Children in the cohort who have screened positive for CH and been referred to the GOSH endocrinology service will also be linked to diagnosis and follow-up data through a bespoke clinical management and research database created as part of this study (see below).

#### *GOSH CH database*

This clinical dataset contains information on 1800 children who have screened positive for CH through the North Thames NBS programme since 2006 and been referred to the Endocrinology Clinic at GOSH. Until 2018, all children with blood-spot TSH concentrations  $\geq 6$  mU/L, on first or repeat screening, were referred to this team. From 2018, the screening cut-off was set higher, at  $\geq 8$  mU/L.<sup>28</sup> Note that GOSH has historically had a lower screening threshold for a positive CH test compared with the national screening programme (which recommended  $\geq 10$  mU/L when GOSH used  $>6$  mU/L).<sup>7</sup>

The dataset holds information on final diagnosis and screening outcome (ie, true or false positive on screening), longitudinal data on blood tests at diagnosis and during follow-ups (at 2, 4 and 8 weeks and 6, 9 and 12 months after start of treatment, and as indicated afterwards), thyroid scans, and treatment dose, including starting dose, and changes in dosage during follow-up. The dataset also contains genetic data for ~50 children with CH from whole exome sequencing, targeted Sanger sequencing, and a customised next generation sequencing panel of CH-associated genes.<sup>29</sup> The CH clinical and genetic data were originally held in a variety of formats at GOSH, including Microsoft Word and Excel files, but we propose to consolidate the data to develop a clinical database that is also research ready (objective 1).

#### *Birth and death registration data*

The Office for National Statistics (ONS) collate data from registry offices on all births and deaths, which are then

shared with NHS Digital (NHSD), the NHS data provider for England. The birth registration database contains key sociodemographic information about registration type (sole or joint), and parental ages, occupations and countries of birth. Death registration data contains information about date of death and causes of death. All children from the NBS database will be linked to birth and death registration data.

### Follow-up study data sources

The follow-up study of cases and controls will include all children who have been referred to the GOSH Endocrinology Clinic after having tested positive through a blood spot test from the CH database. All children from the CH database will be linked to Hospital Episode Statistics (HES), the National Pupil Database (NPD) and the NHS Business Services Authority (NHSBSA) community dispensing data. HES and NPD data are already linked for the Education and Child Health Insights from Linked Data (ECHILD) database.<sup>30</sup> Controls will be selected from the HES portion of the ECHILD database; this minimises the risk of patient identification since controls will be selected using pseudonymised data.

#### *Hospital Episode Statistics*

HES contains data on all admissions to NHS hospitals,<sup>31</sup> including deliveries and births, as well as accident and emergency (A&E) attendances and outpatient bookings. HES datasets are supplied by NHSD.

#### *Department for Education National Pupil Database (NPD)*

The NPD holds data on all children in state primary and secondary schools; 93% of all school children in England.<sup>32</sup> The NPD contains pupil-level information on school test results (early years foundation stage (EYFS) profiles, key stages (KSs) and phonics) and special educational needs (SEN) provision.

#### *Education and Child Health Insights from Linked Data*

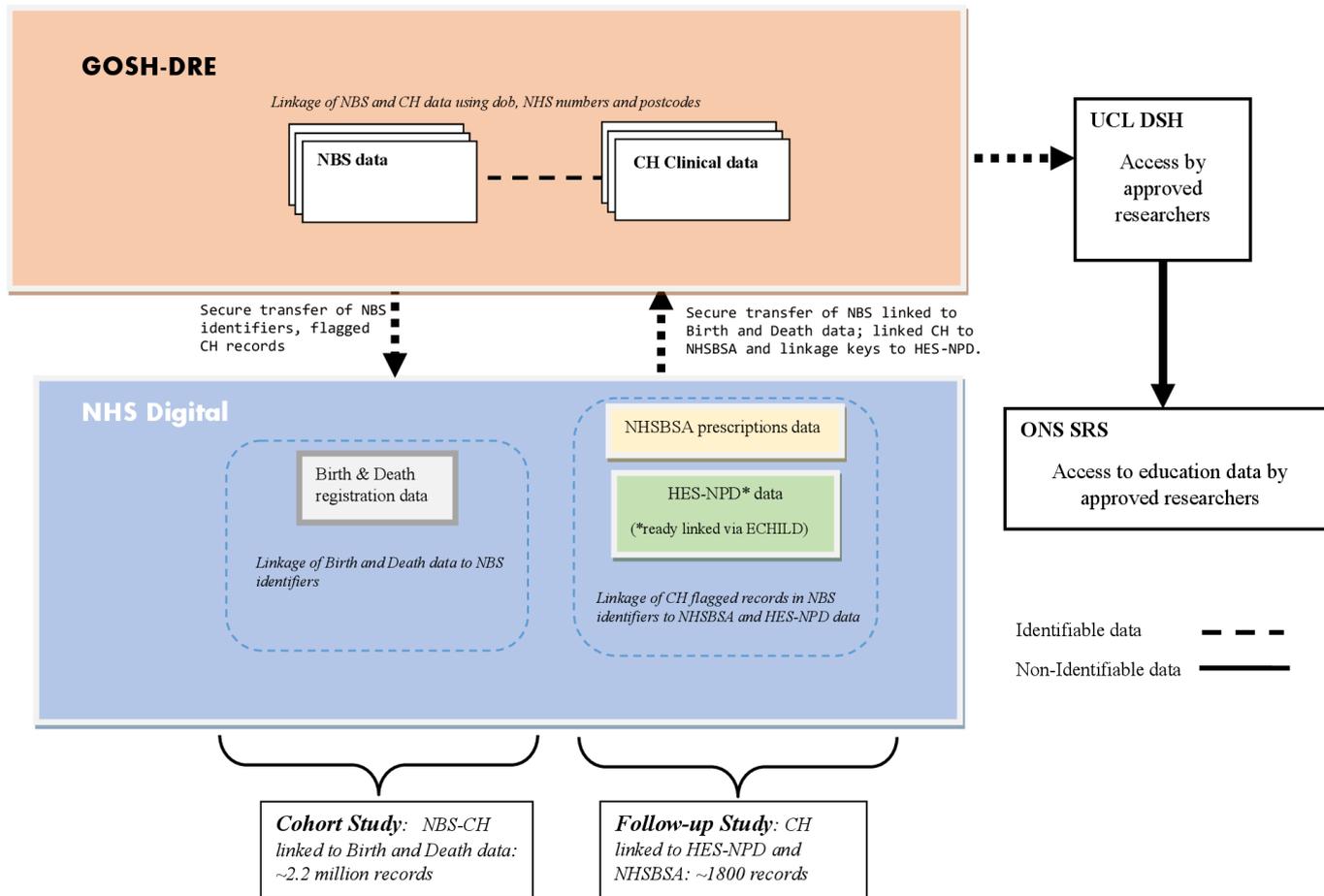
The ECHILD database<sup>30</sup> is a pseudonymised, linkable collection of HES, NPD, and social care data for a whole population-based cohort of children and young people in England, born between 1 September 1995 and 31 August 2020. ECHILD is held in the ONS Secure Research Service<sup>33</sup> (SRS). University College London (UCL) holds a copy of the HES portion of ECHILD in the UCL Data Safe Haven<sup>34</sup> (DSH).

#### *NHSBSA community dispensing data*

The NHSBSA dispensing database contains data relating to NHS prescriptions dispensed and submitted to NHS Prescription Services by community pharmacies in England. Note that we will only have NHSBSA community dispensing data for cases. NHSBSA dispensing data are supplied by NHSD.

### Consolidating the CH database

The existing GOSH CH database will be consolidated to develop a clinical database that is also research ready



**Figure 1** Data flow and linkage diagram. CH, congenital hypothyroidism database; DRE, Digital Research Environment; ECHILD, Education and Child Health Insights from Linked Data; GOSH, Great Ormond Street Hospital; HES, Hospital Episode Statistics; NBS, Newborn Screening database; NHSBSA, NHS Business Services Authority; NPD, National Pupil Database; ONS, Office of National Statistics; SRS, Secure Research Service; UCL DSH, University College London Data Safe Haven.

(objective 1). Currently, the CH database contains data on patient's referral to GOSH, final CH diagnosis and treatments and is held in Microsoft Excel format. Longitudinal thyroid testing and thyroid treatment data are available from a separate Microsoft Word document that is updated at each patient appointment. Longitudinal testing and treatment data will be extracted from Microsoft Word documents and the Epic electronic health records software at GOSH, and linked to patients in the existing CH database. The resulting database will be imported to, and managed using, Research Electronic Data Capture (REDCap) electronic data capture tools hosted at GOSH.<sup>35 36</sup> REDCap is a secure, web-based software platform designed to support data capture for research studies. Moving forward, the database will be updated and maintained by the clinical team using REDCap.

### Data linkage

The suggested data flows for study data are outlined in figure 1.

For the cohort study, the GOSH NBS database will be linked to the CH clinical database deterministically, using NHS number, date of birth and postcode by the GOSH

Digital Research Environment (DRE) team. The GOSH DRE team will also link babies who were born to the same mother to create a 'family' ID using the mothers' NHS numbers, and link postcodes to lower super output areas to assign Index of Multiple Deprivation (IMD) deciles.<sup>37</sup> The DRE team will submit baby identifiers from the NBS database to NHSD, who will link the NBS data to births and deaths records via a well-established deterministic algorithm using NHS number, name, date of birth and postcode.

For the follow-up study of cases and controls, the cases will be the CH-NBS linked subsample of the NBS cohort. The GOSH DRE team will flag CH cases in the NBS dataset. NHSD will use these flagged identifiers to link CH records to (1) NHSBSA data dispensing data and (2) the ECHILD cohort. Encrypted ECHILD IDs will be returned to the UCL study team who will link the cases to the pseudonymised HES part of the ECHILD data stored on the UCL DSH via the existing encrypted IDs. Controls will be selected from the pseudonymised HES data, matched to cases by sex, month and year of birth and local authority at delivery. Since the NPD variables need to remain in the ONS SRS, variables from the CH

database and some variables from the HES database will be transferred to the ONS SRS for analysis of education outcomes.

### Data storage and access

The final linked cohort data will be stored, accessed, and analysed in UCL's DSH. Data for the follow-up study will also be stored, accessed and analysed via the DSH, with the exception of the linked NPD data, which will be kept in the ONS SRS, according to Department for Education (DfE) regulations. All researchers analysing data for this study will hold the Approved Researcher Status certified by the ONS.

### Outcomes of interest

The primary outcome of interest from the cohort study (objective 2) is type of CH (ie, no CH, agenesis, ectopic, hypoplastic or CH-GIS). The secondary outcome of interest is neonatal TSH levels from newborn screening. For the follow-up study, the outcomes for objectives 3 and 4 are:

- ▶ Age group-specific (<1 year, 1–4 years, 5+ years) A&E attendance and emergency hospital admission rates (from HES)
- ▶ EYFS phonics, KS1 and KS2 results (from NPD).
- ▶ SEN status (from NPD).
- ▶ Dispensing rates of drugs, including for severe ADHD (attention deficit hyperactivity disorder) and asthma (from NHSBSA dispensing data); as we will only have these data for cases, we will describe patterns of prescribing among cases only and this outcome will primarily be explored for objective 4.

For objective 5, the outcome is levothyroxine dose by age (from the CH clinical database or NHSBSA dispensing data).

### Other variables of interest

We will consider a number of variables as potential explanatory variables for the cohort study of temporal trends in TSH levels and CH-GIS, and as confounding variables for the follow-up study of cases and controls. These variables include sex, birth weight, parity, gestational age, maternal age at delivery, ethnicity, socioeconomic status (measured as parental occupation and IMD, the latter derived from postcode of residence at delivery). Note that there are no routinely collected, population level data on iodine status in pregnant women in England.

Data sources and corresponding variables or outcomes of interest are listed in [table 1](#).

### Statistical analysis

We will present the distribution of key baby and parental (as recorded on birth registration records) characteristics for the children in the cohort and case–control follow-up studies.

To address objective 2, the aetiology of CH for referred children using the CH database (eg, screening result, CH-GIS, other types of CH) will be established. The NBS database and birth registration data will be used to define known and potential risk factors for CH-GIS available within these datasets, including sex, birth weight, gestational age, multiple birth status, parental country of birth, ethnicity and socioeconomic position (using either parental occupation from the birth registration record

**Table 1** Data sources and variables or outcomes of interest

Study design	Data source	Variables of interest	Years requested
Cohort study	North Thames Newborn Screening database (NBS)	Sex, date of birth, date of sample collection, ethnicity, birth weight, gestational age, screening test results, IMD (derived from postcode of residence at birth)	2000–2020
	GOSH congenital hypothyroidism database (CH)	Final diagnosis, screening outcome, blood tests, thyroid scans, treatment doses, genotype	2006–2020
	Birth and Death Registration data	Date of death, cause of death, registration type, parents' age, parents' occupations, parent's birth countries	2000–2020
Follow-up study of cases and controls*	Hospital Episode Statistics (HES)	Hospital admissions, A&E attendances, outpatient visits (can be used to derive indicators of co-morbidity)	2006–2020
	National Pupil Database (NPD)	EYFS results KS results Phonics tests SEN status Free school meal eligibility Language spoken	2006–2020
	NHS Business Services Authority community dispensing data (NHSBSA)†	Prescriptions submitted; prescriptions dispensed	2006–2020

\*The cases for the follow-up study will consist of the population in the CH database. Controls will be selected from pseudonymised HES data which are part of the Education and Child Health Insights from Linked Data linked database.

†NHSBSA will only be available for CH cases.

A&E, accident and emergency; EYFS, early years foundation stage; IMD, Index of Multiple Deprivation; KS, key stage; NHS, National Health Service; SEN, special educational needs.



or IMD). We will use linear, quantile, logistic or multinomial regression models to establish independent child, maternal and family risk factors for neonatal TSH levels and CH-GIS as appropriate, and determine how these have changed over time to determine which of these factors explain the increase in CH-GIS birth prevalence. We will explore the impact of changing screening thresholds and instrumentation (the North Thames Screening Programme changed from Autodelfia to a Genetic Screening Processor in 2019) during the study period in the regression models.

To determine the health and education outcomes of children with CH-GIS compared with children with other forms of CH, or those who do not have CH (objective 3), data from the follow-up study will be used. Appropriate statistical models, such as proportional hazard regression models, will be used to examine whether children with CH-GIS are more likely to experience the health outcomes of interest compared with other children, while adjusting for other risk factors as defined for objective 2, as well as risk factors that can be defined within HES (such as comorbidities). Death registration data will be used to calculate person-time at risk.

The association between thyroid CH-GIS and school performance will be examined using linked NPD data. Test scores will be standardised within each school year. Multivariable linear regression models will be used to assess the association between type of CH and school performance, adjusting for the risk factors defined for objective 2, and season of birth (eg, summer vs autumn born children). Multivariable logistic regression models will be used to determine if children with CH-GIS are more likely to have SEN status compared with other children. For both objectives 3 and 4, we will also explore the effect of different neonatal TSH levels within each CH diagnostic group.

The analyses for objective 4 will be exploratory at this stage, as the number of children with available genomic data is small (~50). The proportion of children with comorbidities and SEN status will be compared according to CH genotype using  $\chi^2$ -tests, t-tests, and linear and logistic regression models as appropriate. We will compare the proportion of children who are receiving treatment for asthma and severe ADHD according to genotype.

To address objective 5, two data-driven approaches will be used to predict levothyroxine treatment trajectories in children with CH-GIS: mixture cure models<sup>38</sup> and group-based trajectory models.<sup>39</sup>

### Sample size

The cohort analysis (objective 2) will include all children screened through the GOSH newborn blood spot screening programme between 1 January 2000 and 31 December 2020. There are approximately 2.2 million babies in the GOSH newborn screening database between 2000 and 2020; approximately 125 000 newborns are screened at GOSH every year.

The analysis of the follow-up study of cases and controls will include all cases from the CH database (approximately 1800 children), plus the control sample. To decide on the size of the control sample, we reviewed two similar studies: the Cleft Registry and Audit Network (CRANE) database study; a cohort study investigating differences in academic achievement between children with and without a cleft palate<sup>40</sup>; and a study of health and educational outcomes for children with Down Syndrome (Dr Maria Peppia, personal communication). The Down syndrome study used a control sample size of 9 controls per case. The CRANE study revealed that academic attainment of 5-year-old children with an oral cleft was between 0.2 and 0.4 SD lower than the national average. However, for our study, we expect to observe smaller differences in the outcomes of interest (ie, school attainment in children with CH). Assuming a significance level  $\alpha < 0.05$ , a minimum statistical power of 90%, and allowing for statistical adjustment of multiple confounders (ie, sociodemographic characteristics) some of which are less common (eg, premature birth), a more suitable control group would include 15 controls per case. Thus, the total sample for the follow-up study is 28 800 (1800 CH cases and 27 000 controls).

## ETHICS AND DISSEMINATION

### Ethics and information governance

This study has obtained, or is in the process of obtaining, ethical or information governance approvals from the following committees:

- ▶ London—Queen Square Research Ethics Committee (REC) (reference number: 20/LO/1216—Outcome: Favourable Opinion)
- ▶ The Antenatal and Newborn Screening Programme (ANNB) Research Advisory Committee (RAC) (full approval, reference ANNB\_NHSP\_046)
- ▶ Confidentiality Advisory Group (CAG) (conditional support, reference 21/CAG/0071)
- ▶ NHS Digital Data Access Request Service (DARS) (reference NIC-365469-G0P1Q)—pending
- ▶ Department for Education (DfE) (access to NPD via ONS SRS, reference DR211029.01)—pending.
- ▶ Internal review by The Research Data Access Group at GOSH DRIVE Unit<sup>41</sup> Information Governance Committee at GOSH (approved, reference: 19PE26)
- ▶ ONS Research Accreditation Panel (RAP)—pending.

### Changes to original study design

We originally proposed meeting objectives 3–4 using a cohort study design, creating the cohort using the NBS dataset as the cohort spine, with the entire cohort linked to the requested HES and NPD datasets. We also proposed linking the whole cohort to two further datasets held by Public Health England (PHE): the Newborn Hearing Screening Programme (NHSP) database and National Congenital Anomaly and Rare Disease Registration Service (NCARDRS). Using a cohort design for

objectives 3 and 4 would have allowed us to adjust analyses for comorbidities which are rare in the general population of children, but more common among children with CH and CH-GIS. Further, a cohort design where neonatal TSH levels for all children were linked to health and education outcomes would have allowed an exploration of the impact of mildly repressed thyroid hormone levels in the neonatal period on health and education outcomes.

The study received PHE RAC approval in November 2020, however the original study design was not approved by CAG. Feedback from CAG outlined concerns around the number of controls per case using a complete cohort design (~1200 controls/case). The feedback indicated that CAG recognised the public interest in conducting the study but suggested we apply for research database approval so that the NBS study cohort and linked NHSD data could be established as a database that could be accessed and analysed for other projects. The study was given a deferred outcome, with the protocol to be resubmitted once the concerns were addressed.

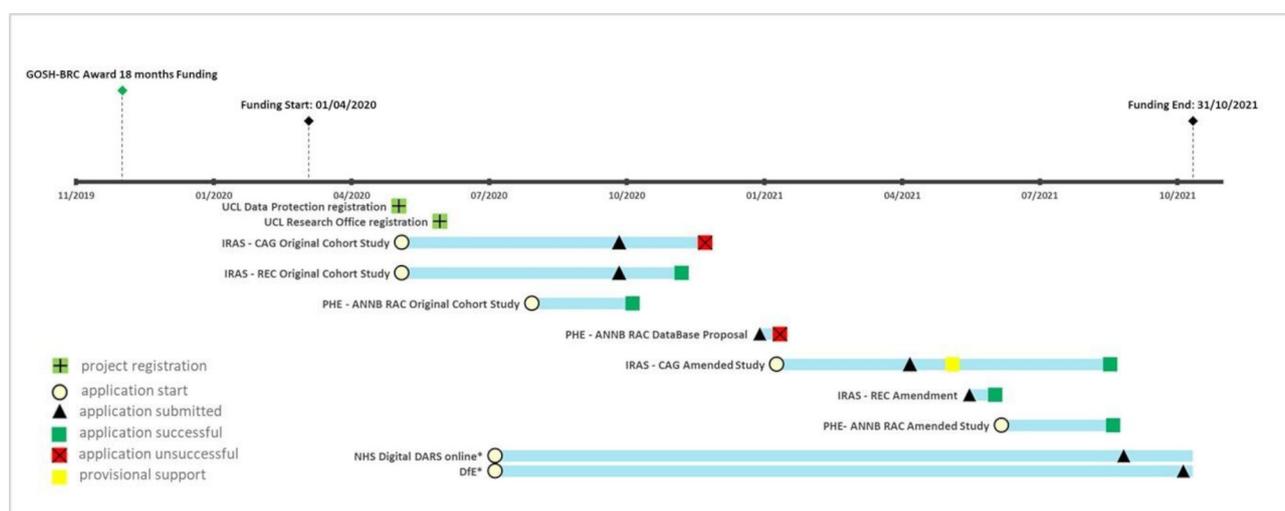
When these suggestions were subsequently reviewed by the RAC, we were informed that RAC were not prepared to approve establishing research database, as it would require senior NHS Screening Programme review. We were given two options—either wait for this review to take place (we were not given a time frame, but we deemed it highly unlikely to be within the time frame of the project) or amend the study design. This led to the current study design, using both a cohort and the follow-up study of cases and controls to address the research objectives. These issues caused considerable delays to the project (~6–7 months for an 18-month project). Further, due to the delays expected within PHE to provide linked data from NCARDRS and the NHSP, these datasets will no longer be used in this study.

Through our patient involvement activities (see below), we learnt that parents were interested in healthy growth in children with CH. We therefore explored linking the NBS and the CH database to the National Child Measurement Programme (NCMP) dataset (containing height and weight data for children starting primary and secondary school). We were also given approval by CAG to proceed with this linkage. However, due to the data sharing agreement for NCMP data, in which NHSD and PHE are joint controllers, the data cannot be linked at individual level to other datasets for research purposes.

Project approvals, data permissions and subsequent correspondence with governing bodies and data controllers have resulted in substantial delays (see figure 2). While we acknowledge that some of these steps are necessary for data protection and data governance issues, our experience clearly demonstrates that the process of linking multiple data sets is both complex and time consuming. The multiple applications to different ethics/governance committees and data providers required in England (unlike, for example, in Scotland<sup>42</sup> or Wales<sup>43</sup>), often providing very similar information about the study, cause significant delays to projects with tight funding timelines. Unfortunately, these experiences are not uncommon when applying for linkage to English NHS datasets for research.<sup>44</sup>

### Dissemination plan

Study outputs will be disseminated through Open Access publications and at academic conferences. Updates for children and families affected by CH will be distributed via the British Thyroid Foundation newsletter. Results will also be available to the wider public via the UCL Institute of Child Health Informatics Group website. Metadata and statistical code used in the analysis will be published on GitHub.



**Figure 2** Congenital hypothyroidism project timeline, from obtaining funding to present (September 2021). BRC, Biomedical Research Centre; CAG, Confidentiality Advisory Group; DSA, Data Sharing Agreement; DSPT, Data Security and Protection Toolkit; GOSH, Great Ormond Street Hospital; HRA, Health Research Authority; IRAS, Integrated Research Application System; NHS, National Health Service; REC, Research Ethics Committee; UCL, University College London. \*In progress.



## Patient and public involvement

Extensive work has been carried out to discuss the research priorities and methodologies with parents and young people. The following groups have been consulted about this project:

- ▶ The Young Person's Advisory Group (YPAG) and the Parents and Carers Advisory Group (PCAG) at GOSH. The YPAG has been consulted three times at meetings where PH presented the project and explained how the study would involve linking newborn screening data from the laboratory at GOSH, to clinical data held by the endocrinology team at GOSH, plus national hospital admissions and school's data.
- ▶ The British Thyroid Foundation (BTF): PH spoke to several parents whose children have CH as well as young people with CH about this study, after the BTF published the study summary on the BTF website and asked interested parents/young people to contact PH. PH also presented the research project at one of the BTF Children's Conference and will continue to update parents on the progress of the research via their newsletter and Facebook page.
- ▶ The Genetic Alliance: PH has presented the study at a meeting with representatives from rare disease organisations.

The researchers will continue to work with the BTF, YPAG, PCAG and The Genetic Alliance to provide updates on the research progress, interpret findings and support dissemination.

### Author affiliations

<sup>1</sup>UCL Great Ormond Street Institute of Child Health Population Policy and Practice, London, UK

<sup>2</sup>Institute of Child Health, UCL, London, UK

<sup>3</sup>Research Department of Primary Care and Population Health, UCL, London, UK

<sup>4</sup>Life Course Epidemiology and Biostatistics, University College London, London, UK

<sup>5</sup>Endocrinology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

<sup>6</sup>Department of Chemical Pathology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

<sup>7</sup>University of Cambridge Metabolic Research Laboratories, Wellcome Trust-Medical Research Council Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

<sup>8</sup>Paediatric Pathology, Great Ormond Street Hospital for Children, London, UK

<sup>9</sup>UCL Partners, London, UK

**Acknowledgements** We are grateful for the support of the British Thyroid Foundation and Genetic Alliance, and for all the input from parents, children and young people via PPI activities.

**Contributors** PH, RK and CP conceived the study, developed the study protocol, and developed the methodology. PH, MR and MC drafted the manuscript, and all authors contributed to and reviewed the protocol manuscript.

**Funding** This work is funded by the NIHR GOSH BRC. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. NS is funded by the Wellcome Trust (219496/Z/19/Z)

**Competing interests** RK holds an honorary clinical consultant contract with the ANNB screening programme, now hosted by NHS England and Improvement (formally with Public Health England). RK is also the chair of the ANNB Research Advisory Committee. MRN holds an honorary contract with NHS Digital. No other authors have any competing interests to declare.

**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.

**Ethics approval** Not applicable.

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

**Data availability statement** No data are available. Not applicable.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

### ORCID iDs

Milagros Ruiz Nishiki <http://orcid.org/0000-0003-2265-2889>

Melissa Cabecinha <http://orcid.org/0000-0001-6869-4692>

## REFERENCES

- 1 Smith J, Cheater F, Bekker H. Parents' experiences of living with a child with a long-term condition: a rapid structured review of the literature. *Health Expect* 2015;18:452–74.
- 2 Khangura SD, Tingley K, Chakraborty P, *et al*. Child and family experiences with inborn errors of metabolism: a qualitative interview study with representatives of patient groups. *J Inherit Metab Dis* 2016;39:139–47.
- 3 Pelentsov LJ, Fielder AL, Esterman AJ. The supportive care needs of parents with a child with a rare disease: a qualitative descriptive study. *J Pediatr Nurs* 2016;31:e207–18.
- 4 Department of Health and Social Care Publication. The UK strategy for rare diseases. DEP heal soc care Publ, 2013. Available: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/260562/UK\\_Strategy\\_for\\_Rare\\_Diseases.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/260562/UK_Strategy_for_Rare_Diseases.pdf) [Accessed 1 Sep 2021].
- 5 Downing M, Pollitt R. Newborn bloodspot screening in the UK--past, present and future. *Ann Clin Biochem* 2008;45:11–17.
- 6 Fisher DA. Screening for congenital hypothyroidism. *Trends Endocrinol Metab* 1991;2:129–33.
- 7 Langham S, Hindmarsh P, Krywawych S, *et al*. Screening for congenital hypothyroidism: comparison of borderline screening cut-off points and the effect on the number of children treated with levothyroxine. *Eur Thyroid J* 2013;2:180.
- 8 Deladoëy J, Ruel J, Giguère Y, *et al*. Is the incidence of congenital hypothyroidism really increasing? A 20-year retrospective population-based study in Québec. *J Clin Endocrinol Metab* 2011;96:2422–9.
- 9 Albert BB, Cutfield WS, Webster D, *et al*. Etiology of increasing incidence of congenital hypothyroidism in New Zealand from 1993–2010. *J Clin Endocrinol Metab* 2012;97:3155–60.
- 10 Pollitt RJ. Evidence or enthusiasm? why yields from UK newborn screening programmes for congenital hypothyroidism are increasing. *Arch Dis Child* 2016;101:120–3.
- 11 Knowles RL, Oerton J, Cheetham T, *et al*. Newborn screening for primary congenital hypothyroidism: estimating test performance at different TSH thresholds. *J Clin Endocrinol Metab* 2018;103:3720–8.
- 12 McGrath N, Hawkes CP, McDonnell CM, *et al*. Incidence of congenital hypothyroidism over 37 years in Ireland. *Pediatrics* 2018;142.
- 13 Barry Y, Bonaldi C, Goulet V, *et al*. Increased incidence of congenital hypothyroidism in France from 1982 to 2012: a nationwide multicenter analysis. *Ann Epidemiol* 2016;26:100–5.
- 14 Hertzberg V, Mei J, Therrell BL. Effect of laboratory practices on the incidence rate of congenital hypothyroidism. *Pediatrics* 2010;125(Suppl 2):S48–53.
- 15 Hinton CF, Harris KB, Borgfeld L, *et al*. Trends in incidence rates of congenital hypothyroidism related to select demographic factors: data from the United States, California, Massachusetts, New York, and Texas. *Pediatrics* 2010;125(Suppl 2):S37–47.
- 16 Peters C, Brooke I, Heales S, *et al*. Defining the newborn blood spot screening reference interval for TSH: impact of ethnicity. *J Clin Endocrinol Metab* 2016;101:3445–9.
- 17 Lain SJ, Bentley JP, Wiley V, *et al*. Association between borderline neonatal thyroid-stimulating hormone concentrations and educational and developmental outcomes: a population-based record-linkage study. *Lancet Diabetes Endocrinol* 2016;4:756–65.

- 18 Bongers-Schokking JJ, Resing WCM, de Rijke YB, *et al.* Cognitive development in congenital hypothyroidism: is overtreatment a greater threat than undertreatment? *J Clin Endocrinol Metab* 2013;98:4499–506.
- 19 Cherella CE, Wassner AJ. Update on congenital hypothyroidism. *Curr Opin Endocrinol Diabetes Obes* 2020;27:63–9.
- 20 Léger J, Olivieri A, Donaldson M, *et al.* European Society for paediatric endocrinology consensus guidelines on screening, diagnosis, and management of congenital hypothyroidism. *Horm Res Paediatr* 2014;81:80–103.
- 21 Salerno M, Micillo M, Di Maio S, *et al.* Longitudinal growth, sexual maturation and final height in patients with congenital hypothyroidism detected by neonatal screening. *Eur J Endocrinol* 2001;145:377–83.
- 22 Rovet JF. Children with congenital hypothyroidism and their siblings: do they really differ? *Pediatrics* 2005;115:e52–7.
- 23 Rovet JF, Ehrlich R. Psychoeducational outcome in children with early-treated congenital hypothyroidism. *Pediatrics* 2000;105:515–22.
- 24 Simons WF, Fuggle PW, Grant DB, *et al.* Educational progress, behaviour, and motor skills at 10 years in early treated congenital hypothyroidism. *Arch Dis Child* 1997;77:219–22.
- 25 Tillotson SL, Fuggle PW, Smith I, *et al.* Relation between biochemical severity and intelligence in early treated congenital hypothyroidism: a threshold effect. *BMJ* 1994;309:440–4.
- 26 Dimitropoulos A, Molinari L, Etter K, *et al.* Children with congenital hypothyroidism: long-term intellectual outcome after early high-dose treatment. *Pediatr Res* 2009;65:242–8.
- 27 NBSFS user guide - GOV.UK. Available: <https://www.gov.uk/government/publications/newborn-blood-spot-screening-failsafe-solution-user-guide/newborn-blood-spot-screening-failsafe-solution-user-guide> [Accessed 7 Sep 2021].
- 28 Laboratory guide to screening for CHT in the UK - GOV.UK. Available: <https://www.gov.uk/government/publications/congenital-hypothyroidism-screening-laboratory-handbook/a-laboratory-guide-to-newborn-blood-spot-screening-in-the-uk-for-congenital-hypothyroidism> [Accessed 1 Nov 2021].
- 29 Peters C, Nicholas AK, Schoenmakers E, *et al.* *DUOX2/DUOX2* Mutations Frequently Cause Congenital Hypothyroidism that Evades Detection on Newborn Screening in the United Kingdom. *Thyroid* 2019;29:790–801.
- 30 Welcome to the ECHILD project website | UCL Great Ormond Street Institute of Child Health - UCL – University College London. Available: <https://www.ucl.ac.uk/child-health/research/population-policy-and-practice-research-and-teaching-department/cenb-clinical-20> [Accessed 1 Sep 2021].
- 31 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.
- 32 Jay MA, McGrath-Lone L, Gilbert R. Data resource: the National pupil database (NPD). *Int J Popul Data Sci* 2019;4:1101.
- 33 Office for National Statistics. ONS research and data access policy. Available: <https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/onsresearchanddataaccesspolicy> [Accessed 7 Sep 2021].
- 34 Data Safe Haven (DSH) | Information Services Division - UCL – University College London. Available: <https://www.ucl.ac.uk/isd/services/file-storage-sharing/data-safe-haven-dsh> [Accessed 7 Sep 2021].
- 35 Harris PA, Taylor R, Thielke R, *et al.* Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
- 36 Harris PA, Taylor R, Minor BL, *et al.* The REDCap Consortium: building an international community of software platform partners. *J Biomed Inform* 2019;95:103208.
- 37 English indices of deprivation 2019 - GOV.UK. Available: <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019> [Accessed 7 Sep 2021].
- 38 Amico M, Van Keilegom I. Cure models in survival analysis. *Annu Rev Stat Appl* 2018;5:311–42.
- 39 Frankfurt S, Frazier P, Syed M, *et al.* Using group-based trajectory and growth mixture modeling to identify classes of change trajectories. *Couns Psychol* 2016;44:622–60.
- 40 Fitzsimons KJ, Deacon SA, Copley LP, *et al.* School absence and achievement in children with isolated orofacial clefts. *Arch Dis Child* 2021;106:154–9.
- 41 Home | GOSH drive. Available: <https://www.goshdrive.com/> [Accessed 25 Oct 2021].
- 42 ISD Services. Electronic data research and innovation service (eDRIS) | Isd Scotland. Available: <https://www.isdscotland.org/Products-and-services/Edris/> [Accessed 25 Oct 2021].
- 43 SAIL Databank - The Secure Anonymised Information Linkage Databank. Available: <https://saildatabank.com/> [Accessed 25 Oct 2021].
- 44 Taylor JA, Crowe S, Espuny Pujol F, *et al.* The road to hell is paved with good intentions: the experience of applying for national data for linkage and suggestions for improvement. *BMJ Open* 2021;11:e047575.