Imperial College London

Title

neoWONDER: Neonatal Whole Population Data linkage to improving long-term health and wellbeing of preterm and sick babies

Version 1.0

24th March 2021

MAIN SPONSOR: Imperial College London

STUDY COORDINATION CENTRE: Imperial College London

IRAS Project ID: 293603

REC reference: xx

Protocol authorised by:

Name & Role **Date** **Signature**

Imperial College Peer Review Certificate no. XXX

Imperial College London

Study Management Group

Applicant and Chief Investigator:

Cheryl Battersby

NIHR Clinician Scientist and Clinical Senior Lecturer Imperial College London Honorary Consultant Neonatologist, Chelsea and Westminster Hospital

Chelsea and Westminster Hospital, 369 Fulham Road, SW10 9NH

Email: c.battersby@imperial.ac.uk

Sponsor

Imperial College London is the main research Sponsor for this study. For further information regarding the sponsorship conditions, please contact the Head of Regulatory Compliance at:

Joint Research Compliance Office Imperial College London and Imperial College Healthcare NHS Trust Room 215, Level 2, Medical School Building Norfolk Place London, W2 1PG

Tel: 0207 594 9459/ 0207 594 1862

Funder:

This project is funded by the National Institute for Health Research (NIHR) through the Advanced Fellowship awarded to Cheryl Battersby (reference: NIHR300617).

This protocol describes the neoWONDER study to link existing data to obtain long-term health and educational outcomes for babies born preterm or with surgical conditions. We have taken every care in its drafting, but corrections or amendments may be necessary. Please refer problems relating to this study to the Chief Investigator in the first instance. This study will adhere to the principles outlined in the UK Policy Frame Work for Health and Social Care Research. We will conduct the study in compliance with the protocol, the Data Protection Act and other regulatory requirements as appropriate.

STUDY SUMMARY

Aims: To obtain long-term health and education outcomes for babies born preterm and/or with surgical conditions by linking the National Neonatal Research Database (NNRD) to other existing health, educational, and environmental datasets. A de-identified linked dataset will be analysed:

1) To describe the long-term physical, mental health and educational outcomes

2) To determine the impact of post-discharge exposures (environmental and socio-economic factors) on modifying later health and educational outcomes

3) To evaluate the impact of neonatal care and interventions on the later health and educational outcomes

Population: Babies born and received care in neonatal units in England and Wales between 1st Jan 2007 and 31st December 2020; recorded gestational age of less than 32 weeks OR Any gestational age AND recorded to have received surgery and diagnosis of one of 6 conditions: necrotising enterocolitis, Hirschsprung's disease, gastroschisis, oesophageal atresia, congenital diaphragmatic hernia and posterior urethral valves.

Outcomes: Health (mortality, hospital admissions, diagnoses indicative of neurodisability or and/or chronic illness, health care utilisation), educational attainment (Early Years Foundation Stage Profile (EYFSP), key stage 1, key stage 2), school absence and special educational needs status.

Retrospective data linkage: The NNRD will be linked to the following datasets:

England: Hospital Episode for Statistics (HES), Death registrations from the Office for National Statistics (ONS), Mental Health Services Dataset (MHSDS), Personal Demographic Service (PDS), Paediatric Intensive Care Audit Network (PICANet), National Pupil Database (NPD), South London and Maudsley Clinical Record Interactive Search (CRIS); Wales: Secure Anonymised Information Linkage (SAIL) databank.

Data access: The de-identified linked dataset set containing education data will be accessed through the Office for National Statistics (ONS) Secure Research Service (SRS). A de-identified dataset without the education data will also be held on the Imperial College server.

Approvals: Research Ethics Committee and Confidential Advisory Group (CAG) approvals will be sought.

Abbreviations

CAG	Confidential Advisory Group
CI	Chief Investigator
CRIS	Clinical Record Interactive Search
HES	Hospital Episode Statistics
HQIP	Healthcare Quality Improvement Partnership
HRA	Health Research Authority
MHSDS	Mental Health Services Data Set
NDAU	Neonatal Data Analysis Unit
NHS	National Health Service
NNRD	National Neonatal Research Database
ONS SRS	Office for National Statistics Secure Research Service
PICANet	Paediatric Intensive Care Audit Network
REC	Research Ethics Committee
SAHSU	Small Area Health Statistics Unit
SEN	Special Educational Need
SLaM	South London and Maudsley NHS Foundation Trust

Table of contents

Contents

1 INTRODUCTION

1.1 Background

Early life medical and surgical care interventions in babies born preterm and/or with surgical conditions influence later life health and educational outcomes. Survivors are at risk of long-term neurological impairment and ongoing health, educational and social care needs. Improving long-term outcomes for this vulnerable patient group is important to patients, families and society. However, obtaining long-term outcomes post-discharge to evaluate the impact of interventions has been an ongoing challenge globally. Traditional consent-based surveillance studies are expensive and complex, The burden imposed on families results in progressive attrition over time. I will address this problem by obtaining long-term outcomes from linking together routine data sources. I will link an established source of routine data on all babies admitted to NHS neonatal units, the National Neonatal Research Database (NNRD), with other routine health, educational and environmental datasets in England and Wales. The NIHR has recognised the importance of this work, and has funded the neoWONDER (Neonatal Whole Population Data linkage to improve lifelong health and wellbeing of preterm babies) research programme through an Advanced Fellowship awarded to Cheryl Battersby (CI).

Importance of the long-term impact of preterm birth

Babies born very preterm (below 32 weeks gestation) represent 2-3% of all UK births (around 8,000 each year). Survival of the most premature babies born before 26 weeks has improved from 40% in 1995 to 56% in 2012 (1) but rates of disability remain unchanged (2). Very preterm babies are at risk of life-long complications that affect their physical and mental health and their need for health and social care (3). Cognitive impairment is the most prevalent disability among very preterm babies and contributes to poor educational attainment. Two-third require educational support (4); 23% have mental health problems such as autistic spectrum, attention deficit, hyperactivity and emotional disorders (5). There is a high risk of rehospitalisation and mortality in infancy (6); asthma and wheezing are highly prevalent (7).In later life, there are lower rates of employment, income, and self-esteem, higher risk of type 2 diabetes, cardiometabolic problems (8).

The societal cost in England to age 18 is estimated to be around £2.5 billion (5-9). Reducing complications from preterm birth will improve the life-long health and wellbeing of patients, their families and benefit the wider society by reducing demands on public services. These complications among very preterm survivors pose a concern for their health and well-being. However, no information on long-term outcomes have been available for very preterm babies born in the last decade in the UK to answer the question, *What neonatal interventions or factors post-discharge modify long-term outcomes?* Importantly, the impact of air pollution, other environmental and social factors remain under studied; findings will have important public health implications (9-14). As survival continues to improve, improving long-term outcomes is a national priority. To do this, long-term data are needed to evaluate the impact of neonatal interventions and inform strategies to improve outcomes through child and in adulthood (15).

Importance of the long-term impact of surgical conditions

There is increasing evidence suggesting that children with chronic health conditions have worse long-term outcomes including educational attainment (16). In particular, longer term outcomes for babies with surgical conditions are less well characterised. Major operations performed in the neonatal period or early infancy have the potential to have life-long consequences for children (17, 18). Many children who undergo surgery at this stage in life develop impaired cardiac, neurological, respiratory, gastrointestinal or bladder function, have a life-long requirement for engagement with healthcare, and report lower overall quality of life than their unaffected peers (19-21). Some of this impact is due to the underlying condition that required surgery, and some due to the interventions required. Given the overall impact of requiring surgery, children may also have lower educational attainment than children who do not require surgery.

This research study will investigate long-term outcomes of six common surgical conditions affecting gastrointestinal, respiratory and urological systems that require major surgery in the neonatal period. These include necrotising enterocolitis, oesophageal atresia, gastroschisis, posterior urethral valves, Hirschsprung's disease, and congenital diaphragmatic hernia; each year around 900 infants in England are affected (22, 23).

Why is this research important in terms of improving the health and/ or wellbeing of the public and/or to patients and health and care services?

This research will benefit children born preterm and/or with surgical conditions by establishing a sustainable approach to identify modifiable factors that influence long-term health and developmental outcomes. By demonstrating feasibility of this cost-efficient data linkage approach in the preterm and surgical population, more clinical questions can be addressed more rapidly to benefit more patients. Demonstrating proof of concept in this exemplar preterm and surgical population will support the continual linkage for future cohorts with complex conditions. Moreover, information on long-term outcomes will support counselling of families, decision-making, and inform future research and public policies to benefit patients and families.

This research will also improve the effectiveness and safety of health services by surveillance of long-term harm and benefits. Describing and quantifying long-term outcomes will improve care services by informing the planning of health, community and educational services to meet local needs.

Importance of long-term outcomes and linkage without consent

The cost and complexity of obtaining long-term outcome data mean we do not know at population level, the longer term outcomes for the 90,000 babies born very preterm in the UK over the last decade. Furthermore, Cochrane recognise the lack of long-term outcomes in randomised controlled trials as a knowledge gap (24). A search for ongoing *preterm* trials on the ISRCTN registry yielded 56 studies; none were powered on long-term outcomes (25). In a systematic review, over half of 76 neonatal trials did not report neurodevelopmental outcomes (26). Most reported outcomes are short-term (before discharge), which are poor predictors of longer-term functional outcomes (27). Lack of long-term data hinders the evaluation of meaningful benefit and inadvertent harm to patients. For example, antibiotics given to women in preterm labour increased the risk of cerebral palsy at 7 years (28); postnatal high dose dexamethasone increased the risk of cerebral palsy (23).

Obtaining long-term data using consent-based cohort studies is complex and expensive, with high attrition over time, limiting the validity of findings to a whole population. Another major drawback of opt-in consent based studies, is that seldom heard groups, including those whose English is not their first language, may not participate in opt-in studies.

The UK EPICure studies followed up babies born before 26 weeks in 1995 and 2006. 92% were assessed at 2.5 years and 71% at 11 years in EPICure 1 (283). Those lost to follow-up were more likely have a non-white ethnic origin, unemployed parents and cognitive impairment. 55% were followed up at 3 years in EPICure 2 (576). As survival improves and numbers rise, these studies are unfeasible and overburdensome for families. Recently, the US National Children's Study and the UK Early Life Study were both abandoned due to slow recruitment, resulting in a waste of US $1.2 billion and £9 million.

Furthermore, specific to this study, contacting over 100,000 families who had preterm babies born in the last 14 years itself will require linking personal identifiers to obtain contact details. Importantly, contacting families or individuals with experience of preterm birth may cause unnecessary distress especially if the child has subsequently died or has complex needs.

A potential solution: data linkage without consent

This research study proposes a simpler and cost-efficient data linkage approach to obtain long-term data. The National Neonatal Research Database (NNRD) was established in 2007 and contains extracts from neonatal electronic patient records from admission to discharge from neonatal care. It is one of the world's largest neonatal databases with over 400 data fields and quality data (29, 30). It contains data for over 1 million babies from over 200 units in England, Scotland and Wales. The NNRD will be linked to routine records from health and education datasets that currently exist separately.

1.2 Lay summary

Aim

To improve the lifelong health and wellbeing of babies born preterm or sick by linking existing data to evaluate the long-term impact of neonatal interventions.

Background

In the United Kingdom each year, around 8,000 babies (2-3% of all births) are born before 32 weeks (more than 2 months early), and around 1000 born with conditions requiring surgery in the first few weeks of life. They require specialised care in neonatal units and are at risk of life-long disability and health problems. In England, the societal cost for caring for a preterm babies born in one year to 18 years old is estimated to be £2.5 billion.

Improving the life-long health and wellbeing of preterm and sick babies is a national priority. To do this, we need to monitor their progress over a long time and find out what neonatal

care and interventions (like feeding or breathing support) or factors after hospital discharge (like social or environmental) make a difference.

However, the major obstacle is obtaining this information, which is very complex and expensive. As a result, we have no information about the longer term progress of very preterm or sick babies born during the last decade.

We will make it possible to obtain long-term information by linking together data that exist separately. This is a simpler and more cost-efficient approach and can be continued in the future to track outcomes into adulthood. If shown to be possible, this approach can also benefit other patient groups with complex needs.

Design and methods

We will link an established database containing routine data from babies in NHS neonatal units, the National Neonatal Research Database, to other routine health and educational data. We will do this for around 100,000 babies born in England and Wales over a 14-year period (2007-2020).

The National Neonatal Research Database will be linked to health, education and environmental data in England and Wales to answer the following questions for babies born preterm or with surgical conditions:

What are the long-term health and educational outcomes?

Does air pollution or other environmental and socio-economic factors influence these outcomes?

What is the impact of certain neonatal interventions, for example mother's breast milk on later health and educational outcomes?

2 STUDY OBJECTIVES

I will link an established source of data on all babies admitted to NHS neonatal units, the National Neonatal Research Database, to other routine health, educational, and environmental datasets.

I will do this for over 100,000 babies born preterm (born before 32 weeks) or with surgical conditions in England and Wales over fourteen-year period (2007-2020).

Aims

1) Describe the long-term physical, mental health and educational outcomes
2) Determine the impact of post-discharge exposures (environmental and socio-economic factors) on modifying later health and educational outcomes
3) Evaluate the impact of an intervention, such as *mother's own breast milk,* on the later health and educational outcomes

3 STUDY DESIGN

3.1 Recruitment

There is no patient recruitment; this is a database retrospective analysis.

This study makes use of routinely available data and as such, no active study recruitment is required. To access data available in the NNRD, all neonatal units will be written to with information about the study and offered the opportunity to opt-out. This is an established process managed by the Neonatal Data Analysis Unit (NDAU) based at Imperial College. Caldicott Guardian Approval has been granted by neonatal units to submit data to the NNRD (31). The NDAU have an up-to-date contact database of all email addresses for lead neonatal clinicians of each neonatal unit across England and Wales. The NDAU manager will use this established database to send the letter.

3.2 Inclusion criteria

Eligible babies within the NNRD include those:

1. Born and admitted to a neonatal unit between 1st Jan 2007 and 31st December 2020
2. Received care in neonatal units in England and Wales
3. Have a recorded gestational age at birth <32 weeks
4. Gestational age at birth 32 weeks or more and received surgery with confirmed diagnosis of one of 6 conditions: necrotising enterocolitis, Hirschsprung's disease, gastroschisis, oesophageal atresia, congenital diaphragmatic hernia and posterior urethral valves.

Exclusion criteria

- Babies with missing data for principal background variables will be excluded. These are gestational age at birth, gender and place of birth.
- Babies cared for in neonatal units in Scotland which only started contributing data from 2015.

3.3 Data and sources

This study will only use routinely available data and will not actively recruit any babies or children. No identifiable data will be provided to the research study team. The final de-identified dataset will be formed by linking together the following:

1) National Neonatal Research Database (32): is managed by the Neonatal Data Analysis Unit (NDAU) and contains care data for all babies admitted to NHS neonatal units across UK since 2007. Its population coverage is internationally unique with 100% coverage of preterm babies born less than 32 weeks since 2012 and high representative coverage since 2008. Identifiers held on the NNRD include infant *NHS number*, *date of birth*, *gender, postcode.*

2) The Personal Demographic Service (PDS) (33) held and managed by NHS Digital contains demographic data for all NHS patients. *Identifiers include forename, surname, postcodes (including changes over time), gender, date of birth, NHS number.*

3) Death registrations from the Office for National Statistics (ONS) (34) held and managed by ONS collects data regarding date and cause of death in England and Wales. Contains information on all registered deaths. *Identifiers include NHS number, date of birth, gender*

4) Hospital Episode for Statistics (HES) (35) held and managed by NHS Digital contains data on hospital admissions to NHS hospitals, outpatient appointments, attendances at accident and emergency department/ Emergency Care Dataset (ECDS from 2019/20) across England. HES collects data related to patient demographics, diagnoses and clinical care. The NNRD has existing CAG approval to hold identifiers to link to HES. *Identifiers include NHS number, date of birth, gender*

5) SAIL (Secure Anonymised Information Linkage) databank Wales (36)

The SAILdatabank contains extended dataset for the population of Wales. This includes Patient Episode Database for Wales (PEDW), which is similar to HES for England but contains data within hospitals in Wales; Annual District Birth and Death Extracts, Congenital Anomaly Register, Education data for Wales, Critical Care Dataset, Patient Episode Database Wales, National Community Child Health Database, Welsh Demographic Service Dataset, Welsh Longitudinal General Practice dataset, Wales Results Reporting Service, Maternal Indicators Dataset. The SAILdatabank does not contain any identifiers. Third party linkage is carried out by NHS Wales Informatic Service (NWIS) which holds identifiers similarly to NHS Digital.

6) Paediatric Intensive Care Audit Network (PICANet) (37) is commissioned by the Healthcare Quality Improvement Partnership (HQIP) and is based at the Universities of Leeds and Leicester. It is a database recording details of the treatment of all critically ill children in paediatric intensive care units (PICU) across the UK. Data is collected by PICANet related to referrals, transports and admissions to PICU and includes: demographics of the child; clinical diagnoses; treatment received whilst in PICU and outcomes at the time of discharge. *Identifiers include NHS number, date of birth, gender.*

7) Mental Health Services Dataset (MHSDS) (38) is held and controlled by NHS Digital. The MHSDS also contains its predecessor, the Mental Health and Learning Disabilities Data Set and the Mental Health Minimum Data Set. This contains individual level data for all children accessing mental health care across the community, outpatient, and inpatient settings in England. *Identifiers include NHS number, date of birth, gender.*

8) The National Pupil Database (NPD) (39) is managed and controlled by the Department for Education and held in the ONS Safe haven. The NPD contains detailed information on the educational attainment, SEN, and attendance of children at state schools across England between the ages of 5-18 years. The NPD has previously been linked to health data and the ECHILD programme (40) has recently successfully linked NPD to HES data for individuals born since 1997 (41).

9) Clinical Record Interactive Search (CRIS) (42) is held and controlled by South London and Maudsley NHS Foundation (SLaM) King's College London. It was developed for use within the NIHR Maudsley Biomedical Research Centre (BRC). It provides authorised researchers with regulated, secure access to anonymised information extracted from South London and Maudsley NHS Foundation Trust electronic clinical records system. CRIS holds

Child and Adolescent Mental Health Service (CAMHS) datasets for South London and other regions and has been previously linked to Hospital Episode Statistics (HES) and National Pupil Database (NPD) (40).

10) Small Area Health Statistics Unit (SAHSU) (43) is based at Imperial College and is a nationally and internationally recognised institution for research into environment and health. Extensive environmental data are held on a stand-alone system the 'SAHSU Secure Research Systems' (SSRS), with controlled and restricted access. Confidential and sensitive data are held within SAHSU's secure encrypted database which can only be accessed by the database team. Researchers have access to only pseudonymised data. SAHSU has achieved ISO 27001:2013 certification in April 2019 and NHS Data Security and Protection Toolkit (DSPT) Assessment

3.4 Data items for extraction

1) NNRD: Baby's demographics (e.g. gestational age at birth, gender, birthweight, ethnicity); clinical care (e.g. days on respiratory support, medications received, length of stay); diagnoses (e.g. congenital anomalies, retinopathy of prematurity, bronchopulmonary dysplasia) and outcomes (e.g. survived neonatal care, discharged to a surgical unit).

2) ONS: Any death registration, including date of death and cause of death.

3) HES: Demographics (e.g. age, gender, ethnicity, deprivation); clinical care (e.g. where care was received, medications received, length of stay) and diagnoses (e.g. respiratory infection, appendicitis).

4) SAIL: From the PEDW, we will receive information about the child's demographics (e.g. age, gender, ethnicity, deprivation); clinical care (e.g. where care was received, medications received, length of stay) and diagnoses (e.g. respiratory infection, appendicitis). In addition, we will have education, birth and death data, critical care, primary and secondary care use, community child health resource use and maternity data.

5) PICANet: Child's demographics (e.g. age, gender, ethnicity, deprivation); clinical care (e.g. days on ventilation, medications, length of stay); diagnoses (e.g. respiratory infection, cardiac defect, cancer) and outcomes (e.g. survived PICU, died in PICU).

6) MHSDS and mental datasets: Constant supervision and care required due to disability indicator, Looked after child indicator, Child protection plan indication code, care professional

service or team type association (mental health), disability code (behaviour/emotional/hearing/manual dexterity/learning disability/mobility and gross motor, self-care/sight/speech/ no disability)

7) NPD: The School Census (including alternative provision census and pupil referral unit census, Children in Need, Children Looked After, School Absences, School Exclusions); Early Years Foundation Stage Profile (EYFSP) data at age 5, Phonics data, Attainment, Good for Development (reaching expected attainment), Key Stage 1 data, Key Stage 2 data with attainment scores, Education, health and Care plan (EHC), Special Educational Needs include: primary and secondary SEN type Specific learning difficulty (SPLD), Moderate learning difficulty (MLD), Severe learning difficulty (SLD), Profound & multiple learning difficulty (PMLD), Social, emotional and mental health (SEMH), Speech, language and communication needs (SLCN), Hearing impairment (HI), Visual impairment (VI), Multi-sensory impairment (MSI), Physical disability (PD), Autistic spectrum disorder (ASD), Other difficulty / disorder (OTH), Disability Access Funding (For 3-4 year olds), Socio-economic status (free school meals).

8) CRIS: Mental health assessment, diagnoses, attendance to CAMHS, medication, physical health review, admissions to secondary care

9) **SAHSU:** environmental datasets (air pollution, access to green space, walkability index)

Detailed data dictionaries and information about the extraction will be developed before data is provided to the research team. Researchers will only have access to de-identified data.

3.5 Data linkage using identifiers

Identifiers are necessary to conduct the linkage between the NNRD and other health and education databases. NHS number, date of birth, gender will be used to link the NNRD to HES, ONS, MHSDS, PICANet, CRIS.

Forename, surname, date of birth, postcode, gender are required to link the NNRD to education because education records do not hold NHS numbers (they hold Unique Pupil numbers instead). However, the NNRD does not contain forename, surname or recent postcodes (only the postcode at the time of discharge from neonatal care) . Therefore, the NHS number from NNRD needs to be linked to the Personal Demographic Service (PDS) in NHS Digital prior to linkage to the NPD.

Due to the need to access personal identifiable data for the linkage, we will apply for Confidentiality Advisory Group (CAG) approval to use identifiers without consent. This will enable the linkage to be legally permissible under the Health Research Authority's support under Section 251 of the NHS Act 2006. The proposed data flow for this section can be found in the Appendix 13.1 and 13.2 for England and Wales, respectively.

Identifiers will be temporarily used, in a secure environment, for accurate linkage. Once third party linkage has been carried out, all personal information (NHS number, forename, surname, date of birth, and postcodes) are removed and the linked records will only retain the anonymised unique ID and securely transferred to the Neonatal Data Analysis Unit (NDAU) and to the ONS SRS for linkage to the NPD. No researchers will have access to these identifiers and individual children cannot be identified in the data set.

3.5.1 Split file and third party data linkage

A "Split-file" process will be used to separate personal identifiers from the clinical dataset so that only identifiers (without clinical data) are shared with the independent third party for linkage. NHS Digital for England, and NHS Wales Informatics Service (NWIS) for Wales will the third parties to conduct the linkage. The data flows are designed such that no organisation will hold data they do not already hold, and researchers will only analyse de-identified data without any identifiers. No clinical data will be transferred from the NNRD/PICANet/CRIS to the third parties. NHS Digital (for England) or NHS Wales Informatics Service (NWIS) carries out the linkage and then personal identifiers are removed with only the unique ID retained. These de-identified data are then securely transferred to the NDAU to be linked back to clinical data using the unique ID.

Split file process of the clinical data: NNRD, PICANet, CRIS

Please see Appendix 13.3 illustrating split file process linking NNRD to HES.

Step 1: Split NNRD, PICANet, CRIS into 2 files: File 1 (identifiers) and File 2 clinical data (no identifiers). Both hold a unique ID. File 1 is sent to the NHS Digital which acts as a Trusted Third Party (TTP) and File 2 is retained in the NDAU (without identifiers).

Step 2: Send File 1 (identifiers, unique ID) to NHS digital to create NNRD-PICANet/CRIS linkage files.

Step 3: Identifiers are then removed with unique ID retained, and transferred back to the NDAU.

Step 4: File 2 (clinical data, unique ID, no identifiers) is then linked back to File 1 using the unique ID without identifiers.

3.6 NNRD linkage to HES/ONS/MHSDS

NNRD (File 1 identifiers) will be sent to NHS Digital to be linked to HES, ONS mortality, MHSDS data using the NHS number and date of birth. Identifiers are then moved with retained unique ID and transferred back to the NDAU Imperial College. The unique ID will be used to link File 2 to File 2 at the NDAU.

3.7 NNRD linkage to NPD

Linkage between the NNRD and NPD requires additional identifiers such as forename, surname, recent postcodes (which is not held on the NNRD). This is because there is no common identifier on both NHS and NPD records. NPD is educational data and holds Unique Pupil Numbers rather than NHS number. As the forename and postcode on the NNRD is likely to have changed by the time the child starts school (4-5 years following discharge from neonatal care), linkage to the Personal Demographic Service (PDS) is necessary to obtain most up to date identifiers to accurately link to the NPD. Hence, File 1 (identifiers) will be transferred to NHS Digital Personal Demographic Service first to link to subsequent postcode addresses following neonatal unit discharge (See Appendix 13.3 for split-file process).

Importantly, the Department for Education holds the National Pupil Database (NPD) in the Office for National Statistics Secure Research Service (ONS SRS) and data from the NPD cannot leave the ONS SRS. Therefore, identifiers will need to be transferred from PDS to the ONS SRS.

The ONS SRS is set up for accredited researchers to access data from the National Pupil Database (NPD). Data cannot be downloaded from the SRS, and ONS procedures ensure it operates within a legal framework without disclosure of sensitive information. Researchers working on this project will obtain the necessary ONS accreditation

Step 1: File 1 (identifiers only +unique ID) will be transferred to the NHS Digital PDS. The PDS will use these identifiers to identify additional identifiers such as subsequent postcodes or change of names. Forename, surname, date of birth, gender and postcodes will be

securely transferred to the ONS SRS to be used to link to educational data within the NPD. A logic model, designed to maximize the chance of a reliable postcode match (given the variation over time), will be used. This is an established model developed to improve the linkage of health and NPD data. The quality of linkage will be evaluated with the data controllers and un-linked records will be reviewed and the sensitivity of the probabilistic matching algorithms maximised.

Step 2: File 2- NDAU send clinical data and unique ID (without identifiers) to ONS SRS securely and linked to NPD using unique ID.

3.8 NNRD to SAIL databank (Wales)

Step 1: NDAU send File 1 (personal identifiers only) to The NHS Wales Informatics Service (NWIS) third party for linkage to SAIL databank

Step 2: NDAU send File 2 (NNRD clinical data with unique ID and no identifiers) by secure transfer to SAIL databank to link

Safehaven for final linked dataset: SAIL databank, which does not hold any identifiers.

4 Data access

4.1 NDAU and NNRD

The Neonatal Data Analysis Unit (the NDAU) is a research unit in the Faculty of Medicine, Imperial College London and is based at Chelsea and Westminster Hospital. The NDAU was established to improve the quality of operational clinical data captured at the point of clinical care and promote their best use to support neonatal services and facilitate research. The NDAU is led by a multi-professional Steering Board that includes parent and patient representatives. NDAU complies with the National Statistics Practice and Protocol on Data Access and Confidentiality.

The Neonatal Data Analysis Unit (NDAU) at Imperial College is the data controller of the NNRD. NDAU have a robust information governance framework for information management and have deployed a range of privacy enhancing technologies, physical security measures and audit procedures to safeguard the data according to rigorous standards.

The NNRD is formed from clinical data about mother and baby that are extracted from routinely recorded NHS care records. Data recorded are securely transferred to servers based within the Chelsea and Westminster NHS Foundation trust. Following receipt of these data, any personal information (for example NHS numbers) are removed and held separately on the Chelsea and Westminster server. The de-identified data (without personal information) is transferred to Imperial College server.

All researchers using the data must be affiliated to the Neonatal Data Analysis Unit at Imperial College and a data sharing agreement as appropriate.

4.2 Data security

All documentation relevant to the study will be stored on the Imperial College server system known as the N Drive. The relevant study folder is only accessible to the CI and other nominated research team staff. Access to this area is via a request to IT Services by the CI (Cheryl Battersby) in order to control access.

No documents can be moved or copied to any other area of the N Drive, or any other location, and no data can be taken off the physical University premises. When vacant, office are locked with a key, and the main entrance has a code protected padlock. Furthermore, computers have screen lock applied if individuals are not present at the desk. No personal data will be stored on paper. Researchers working away from normal office location will access these data via secure Virtual Private Network (VPN) or remote desk topping.

Data sharing agreements will be drawn up between data controllers (Table 1). All data will be transferred securely via approved secure file transfer systems. All files will be password protected and passwords will be provided verbally over the phone.

4.3 Opt-out for the NNRD

There is an established mechanism opt-out if parents do not want their child's information to be used. There is information on the NDAU website of how to do this; they should tell the neonatal unit staff where their baby is being cared for, who will contact the NDAU on the parents' behalf to ensure their child's information is not sent to the Neonatal Data Analysis Unit and not included in the NNRD. This will not affect a baby's care in any way.

The NDAU have created resources about the NNRD for parents and staff: Information for parents leaflet, the Neonatal Data Analysis Unit poster "Why your baby's data are important

(pdf)" and The National Neonatal Research Database further information. These provide more information about the NNRD and the services and projects it supports (44).

4.4 Data controllers and approvals

The data controllers of the other databases that will be linked to the NNRD, such as PICANet, CRIS, NHS Digital datasets have their respective robust data governance structure and processes in place, summarised in Table 1 (37, 42, 45). For example, access to the CRIS South London mental dataset is through applications to the CRIS. The analyses carried out using CRIS are closely reviewed, monitored and audited by a CRIS Oversight Committee, which carries representation from the Maudsley Caldicott Guardian and is chaired by a service user. The CRIS Oversight Committee is responsible for ensuring all research applications comply with ethical and legal guidelines. The data is used in an entirely anonymised and data-secure format and all patients have the choice to opt-out of their anonymised data being used. Other data security policies are described in detail elsewhere by the data controllers (32-38, 42, 43).

Table 1. Data source and data controllers

Database	**Data controller**	**Approval committee for data requests**
NNRD	Neonatal Data Analysis Unit (NDAU) Imperial College London	NNRD steering board
PICANet	Healthcare Quality Improvement Partnership (HQIP)	HQIP
CRIS	South London and Maudsley (SLAM)	CRIS oversight committee
ONS, HES, MSDS, PDS	NHS Digital	CAG and REC approvals as appropriate
SAIL	SAIL databank – does not hold identifiers (these are linked by third party NWIS)	SAIL Information Governance Review Panel (IGRP)
SAHSU	Imperial College SAHSU	SAHSU

4.5 Study outputs

Under the General Data Protection Regulation, the ‘data controller’ is responsible for what happens to data which is collected. Neonatal Data Analysis Unit (NDAU) at Imperial College London will act as the data controller for this study.

Imperial College London will keep primary research data for 10 years after the study has completed in line with Medical Research Council guidance for clinical studies; these data will not contain any identifying information.

The final de-identified NNRD-HES/ONS/MSDS-PICANet-CRIS-NPD linked dataset (without identifiers) will be accessed via researchers in the ONS SRS safehaven. Researchers will have undertaken the relevant courses to achieve ONS researcher accreditation.

5 COHORTS FOR DATA LINKAGE

Four cohorts from the NNRD will be linked to other health and education databases. Whilst linkage between health datasets will include up to those born in 2020, the linkage to the NPD and CRIS will be limited to children born up to the end of 2016 as the youngest cohort will be at least school-age by 2020.

Cohort 1-3 will include only those born in England only; cohort 4 only those in Wales (Appendix 13.4 for diagram illustration). Population: around 8,000 babies are born <32 weeks in England and Wales each year The analysis is planned for 2021 (therefore data up till end of 2020).

Cohort 1 Born 2007-2020 in England: link to health data (HES,ONS, PICANet, MHSDS)

-Preterm babies born less than 32 weeks and surgical babies (all gestations) with one of six surgical diagnoses: necrotising enterocolitis, Hirschsprung's disease, gastroschisis, oesophageal atresia, congenital diaphragmatic hernia and posterior urethral valves.

Cohort 2 Born 2007-2016 in England: link to school age outcomes (NPD and CRIS)

-Preterm babies born less than 32 weeks gestation

Cohort 3 Born 2012-2016 in England: link to school-age outcomes (NPD and CRIS)

-Surgical babies (all gestations) with one of six surgical diagnoses: necrotising enterocolitis, Hirschsprung's disease, gastroschisis, oesophageal atresia, congenital diaphragmatic hernia and posterior urethral valves.

Cohort 4 Born 2012-2020 in Wales: link to SAIL databank (contains health, education, social data)

Preterm babies born less than 32 weeks

6 WORKSTREAMS AND ANALYSIS PLAN

6.1 Workstream 1 (Cohort 1): Describe long-term physical and mental health outcomes and post-discharge influencing factors for very preterm babies born in England (cohort 1)

Inclusion criteria: preterm babies born <32 weeks born in neonatal units in England 2007-2020 (14 years)

Maximum follow up would be 14 years and minimum follow up of 1 year. The study would therefore include a total of: 840,000 person follow-up years.

Data sources: Link NNRD to Hospital Episode Statistics (HES), Paediatric Intensive Care (PICAnet), MHSDS, death data (Office for National Statistics (ONS)), environmental datasets (air pollution, access to green space, walkability index via the Small Area Health Statistics Unit SAHSU Imperial College)

Estimated sample size: 102,000

6.1.1 Health outcomes

Source: HES, ONS, MHSDS, PICANet, CRIS (for South London network)

- All-cause mortality
- Cause-specific mortality
- Mental health and behavioural conditions (first diagnosis of any of the conditions (identified with predefined ICD 10/SNOMED codes including conduct disorder, emotional disorders such as depression and anxiety, hyperactivity disorders e.g. ADHD, social and communication disorders e.g. Autism Spectrum Disorders
- Chronic conditions (first diagnosis of any of the conditions below (using ICD 10/SNOMED codes including cancer and blood conditions, cardiovascular, respiratory conditions, musculoskeletal/ dermatological conditions, neurological conditions, metabolic/ endocrine/ digestive conditions, renal/ genitourinary conditions)
- Health resource use – frequency of visits, use of secondary care and Paediatric intensive care, length of stay

- Number and proportion Incidence of CAMHS referrals and Prevalence of mental health problems(who accessed CAMHS services, with mental health diagnoses, accessing secondary health services)
- Health economic evaluation including health and educational resource needs

Analyses

- Assess data linkage rates: Apply deterministic (using actual identifier) and probabilistic (if not full complement of identifiers available) linkage
- Determine mortality rates and causes of death following discharge
- Determine frequency of access to primary health and secondary health services
- Describe diagnoses post discharge from neonatal care
- Determine the potential modifying impact of post-discharge influences (air pollution and socioeconomic factors) on long-term physical health outcomes.
- Determine trends over time and geographical variation

6.1.2 Statistical analysis

A formal statistical analysis plan will be developed before receiving data. The proposed analysis of the linked dataset will focus on initial descriptive analysis. Important consideration will be given to data quality and completeness, which are always potential issues within routine data analysis. We will describe the characteristics of admission neonatal care and subsequent readmissions including length of stay, interventions or procedures, healthcare resource use. Summary statistics will be produced for all four cohorts.

We will investigate the risk of admission to paediatric care for subgroups of children born preterm and/or with surgical conditions. To explore the timing of readmission and determine if there is a correlation with diagnoses or demographics, survival analysis approaches such as Cox regression or flexible parametric modelling will be used; outcome would be presentation or admission to secondary care. Model assumptions will be carefully considered, for example the assumption of proportional hazards when using the Cox model

Time to death for each child during the follow-up period (from birth to December 2020).
A survival analysis with cox proportional hazards modelling will be conducted. Follow-up

time will be censored. The absolute mortality rates and hazard ratios with 95% confidence. A logistic regression will be undertaken to determine the odds ratio (with 95% confidence intervals) of these outcomes (amongst the included population) at age 2 years, 5 years and 12 years, during the study follow-up period.

Subgroup analyses: Geographic variation, temporal trends, gender, only South London network with CRIS matches

6.2 Workstream 2: Determine educational outcomes for very preterm babies in England (cohort 2)

Inclusion criteria: preterm babies born <32 weeks born in neonatal units in England 2007-2016 (10 years). With data up to end of 2020, maximum follow up is 13 years and minimum follow up of 4 years.

Data sources: NNRD to NPD

Estimated sample size: 80,000

Outcomes

School attendance rates, Special school attendances rates, statement of educational needs, Educational attainment: Early Years Foundation Stage (age 4), at Key Stage 1 (Age 7), and at Key stage 2 (Age 11). The mean academic attainment scores of the included population at the early years foundation key stage, key stage 1 and key stage 2, during the study follow-up period

Analyses: Determine the impact of air pollution and socioeconomic factors on modifying long-term educational health outcomes.

To explore trends over time and geographical differences

Subgroup analyses: geographic variation (area of birth); temporal trends (year of birth); eligibility for free school meals; looked after child; first language spoken at home; school type (mainstream or special school); by school ages.

Covariates: maternal age, ethnicity, month of birth

6.3 Workstream 3: Evaluate the impact on the later health and educational outcomes of very preterm babies of an exemplar intervention, *mother's own breast milk (cohort 1 and 2)*

Background

This workstream will test the utility of the linked dataset generated from workstreams 1 and 2, to address one of the top priorities for preterm babies, "What is the optimum milk feeding strategy for the best long-term outcomes of premature babies". Given that mother's milk is an affordable and modifiable intervention, it should be exploited for population health benefits. However, the long-term benefit of maternal breast milk whether own mother's milk or donated mother's milk) has been difficult to demonstrate due to the potential wider social and environmental confounders.

Study design: Quasi-experimental methodology to enable the comparison of long-term outcomes between similar populations with the same propensity for either intervention. Causal inference methods; analysis plan will be developed with a senior statistician to explore propensity matching, instrumental variables and advanced statistical methods.

Data sources: NNRD, health, education datasets

Estimated sample size: 70,000

Population: Very preterm babies born <32 weeks and discharged from neonatal units in England 2007-2016

Intervention and comparison groups: Three groups: No breast milk, Own mothers' milk, Donated mother's milk. I will determine the distribution of the length of exposure of each intervention during neonatal care and make the appropriate cut points for comparison

Outcomes: Long-term physical health and educational outcomes

6.4 Workstream 4: Describe the health and educational progress of very preterm babies in Wales (cohort 4)

Inclusion criteria: Preterm babies born <32 weeks in neonatal units in Wales 2012-2019 (11 neonatal units in the Wales Neonatal Network)

With data up to end of 2020, maximum follow up would be 8 years and minimum follow up of 1 year.

Data sources

NNRD will be linked to the SAIL extended dataset: Annual District Birth and Death Extracts, Congenital Anomaly Register, Education data for Wales, Critical Care Dataset, Patient Episode Database Wales, National Community Child Health Database, Welsh Demographic Service Dataset, Welsh Longitudinal General Practice dataset, Wales Results Reporting Service, Maternal Indicators Dataset (46).

Estimated Sample size

3500

Data Analysis: Descriptive analytics; Summary statistics; Multivariable logistic regression

1) Mortality rates and causes of death following neonatal discharge

2) Health outcomes following discharge including diagnoses, treatment, medication, and access to primary and secondary health care and procedures

3) Educational attainement- Key stage 1 for babies born in 2012, Statement of special educational needs, special school and school attendance

4) Compare outcomes to that of term counterparts born in the same year

6.5 Workstream 5: Health and education outcomes of babies with surgical conditions (cohorts 1 and 3)

To investigate the long-term outcomes and academic attainment of children who have undergone early childhood surgery and identify factors associated with attainment of a Good Level of Development (GLD) on the Early Years Foundation Stage Profile (EYFSP).

Cohort population: All infants requiring surgery for one of six predetermined conditions (necrotising enterocolitis, Hirschsprung's disease, gastroschisis, oesophageal atresia, congenital diaphragmatic hernia and posterior urethral valves) born in the UK during the study period (01 January 2012 – 31 August 2016) will be eligible and will be identified through the National Neonatal Research Database. The surgical codes that will be extracted are included in the Appendix 13.5.

Data source: NNRD linked to ONS, HES, NPD

Outcomes: The primary outcome is attainment of a Good Level of Development (GLD) on the Early Years Foundation Stage Profile (EYFSP). Secondary outcomes include mortality, hospital admissions, school absence and special educational needs status.

Statistical analysis All statistical analyses will be conducted using Stata Version 16 (Statacorp LLC) software. We will perform multiple linear regression for covariates with continuous outcomes and logistic regression for dichotomous outcomes, in order to analyse the contribution of various determinants on the development of outcomes.

Estimated sample size : 8000 (Since 2007 for health outcomes; 4500 from 2016 for education outcomes).

6.6 Potential missing data

Data quality and completeness of the NNRD, PICAnet, CRIS data is high (30). Logic and range checks are applied upon entry of data into the neonatal electronic patient record. At the NDAU, the data are then cleaned and missing data are distinguished from non-missing data. However, data checks on all data will still be completed to investigate for missing or implausible data values. If levels of missing data are high (>10%) for key variables, reasons for this will be investigated and methods to account for this may be used. If key variables are missing, a sensitivity analysis will be undertaken to impute poor scores and rerun the initial analyses to assess the impact on the results or run repeated analyses with and without the babies with missing data. Conversely, if we believe data are missing at random we will investigate whether multiple imputation may be appropriate. At this stage we are unaware of the level of missing data which will be encountered. Robust standard operating procedures will be developed for handling each variable if appropriate based on level of missingness.

7 PATIENT AND PUBLIC INVOLVEMENT AND ENGAGEMENT

7.1 Patient and parent focus groups for linking together existing datasets for preterm babies

The neoWONDER research project was developed with adults born preterm and parents of preterm children. Four parents and ex-patients helped co-develop this proposal. They are in full support of the project and wrote these accounts (website www.neowonder.org.uk).

GM: “I was born prematurely at 26 weeks in the late 70’s. My birth history was complex and was the recipient of several pioneering treatments, meaning I spent the first 6 months of my life in NICU. This has inspired me to have a continuing interest in neonatal research, particularly the long term impact of premature birth. I wanted to get involved in this research

because I could relate to it and have lived experience of many of the issues children born prematurely face."

JK: "…I am the mother of George, who was born premature at 27 weeks in March 2019 at home. As a mother of an ex-premature baby and after having spent over 3 months in a neonatal unit, I always have questions in mind about the development, education, mental health of my son and every other premature child. This is one of the reasons I decided to get involved in neoWONDER to get answers to all these questions for myself, other parents, and the medical society."

TB: "My son was born at 24 weeks 5 years ago. Being both a mother of a child born extremely premature and a paediatric doctor, I can relate to the struggles parents may face. He received very good care on the neonatal unit but since discharge from hospital, we have found it difficult to access the support and services he needs. I want to participate in neoWONDER to raise awareness of these challenges and also make my contribution to improve long-term outcomes of extreme premature babies."

TH: "…My eldest two, now almost 11 year old twins, were born at 27 weeks gestation, spending almost 3 months in a neonatal intensive care unit before coming home. Significant amount of time spent in and out of hospitals for ongoing concerns, monitoring and intervention greatly influenced my choice of career path working in the early years foundation stage. I fully support the neoWonder's aim to ascertain the long-term outcomes of preterm birth and inform neonatal interventions. I feel privileged to be involved and welcome the opportunity to contribute to its future success."

Bliss, the national charity for preterm babies has helped recruit a large study group of almost 400 parents and ex-patients to our neoWONDER parent patient group through their social media presence and online profile.

The patient and public involvement workstream "Parent and patient perspectives on linkage between existing data to evaluate long-term health and wellbeing of preterm babies" commenced in November 2020 and is ongoing. This workstream received ethics approval from Yorkshire and The Humber-Leeds East REC 20/YH/0330 (IRAS project ID 291612). This workstream includes 3 focus groups, a national survey seeking the views of parents and patient on data linkage without consent, and 10 in-depth interviews. The focus groups and national survey have been completed and interviews are taking place in March and April 2021.

- Three parent and patient focus groups were conducted in October 2020 to design a national survey seeking views from parents and patients on obtaining long-term outcomes using data linkage without explicit consent
- The focus group participants were all supportive of the need for linking data to obtain long-term outcomes and some were surprised that this was not already happening. In particular, all believed that health and educational outcomes for children are inextricably linked and very important. Quotes from the sessions:
- *"I assumed our data are already being used and am surprised that it is not. How a child develops is so important, just as important as health…"*
- *"Very happy for my baby's data to be used and linked to find out outcomes- I think it is important that we stress these data already exist; all we are asking is to link them together- makes complete sense"*
- Following revisions led by the parents and patients, the online survey was launched in December for a 6 week period. Respondents were recruited via dissemination through charities (BLISS, Smallest Things, Twins Trust), 20 neonatal units via posters, Social media pages of parents and health professionals (Twitter, Facebook, Instagram), and through the neoWONDER patient/parent group.

The respondents

There were 508 respondents; over half had experience of very preterm birth. There was good representation of views from across the United Kingdom, with the highest representation from the Southwest (24%), Northeast (14%) and London (13%). Of the respondents: around 42% and 44% had experience of caring for a child with ongoing health needs or educational needs, respectively.

Importance of long-term outcomes in preterm babies

- 97% believed it was important to have better information on how preterm children develop as they grow up
- 100% believed information about longer term health is important
- 98% rated educational progress important or very important.
- 93% believed health and learning/ education are closely related
- 92% were supportive of linking existing records to find out what happens to preterm children as they grow up.

Data linkage and anonymised data

- The survey provided an explanation that linkage between health and education records requires the temporary use of personal identifiers, which are then removed and not seen by the researchers for analysis.
- 76% were fine about the temporary use of identifiers e.g. name, date of birth, post code for linkage; 24% were unsure.
- Common reasons given for the uncertainty were related to the need for more information about the use of data. Some quotes:
- *"What my address and her date of birth would be used for and where this information may be visible or accessible"*
- *"It depends on how the information would be used, by who, and how available to the public it would be"*
- *"Who will be able to see this information?"*
- *"I would worry a little about the security of the data."*
- "*Reassurance that data were not used for non-research related purposes e.g. being sold to insurance firms"*
- *"I would prefer that a unique identifier is used to link the data. Indeed, think that all this data should be linked for all citizens anyway"*
- 91% were happy for researchers to use "sensitive" information, such as whether a child has special educational needs, or Education, health and care plan (EHC) or free school meals if it was anonymous.
- When asked till what age should data be linked to for preterm children, All but 1% said were in favour for linking to obtain long-term data. 12% wanted lifelong linkage; 61% agreed with concept of linkage but wanted their children to be consulted for their views when older; 26% were in favour for linking till adulthood (either aged 18 or completed full-time education)

Use of anonymised data without consent, notification and opt-out/dissent

Within the survey, there was an explanation for the need to seek permission from the Confidential Advisory Group (CAG) to link these data without explicit consent as it was i)

impractical to seek permission from 90,000 families retrospectively ii) the need to include data from the whole population for meaningful research.

- 70% were happy for researchers to use existing medical and school records without consent if they were anonymous. 14% were unsure as they wanted more information and time to understand. 16% said no, they were not happy. The reasons given in free text comments were related to the need for more information. Some quotes:
- *"I'd like to discuss the process more so that I understand"*
- *"I would like to know who would have access to this personal information"*
- *"I worry how long this information would be stored for and how it would be used.*
- *If it is linked there may be the opportunity to use it to effect things such as availability of travel insurance - I.e health issue"*
- *"Assurance that these not follow a child into adulthood, where they have the right to make decisions about their own info"*
- When asked views on the best methods or channels for notification/opt-out, respondents were supportive of dissemintating information through Parent and patient networks, Social media, Charities, Schools, SENCOs (Special Educational Needs Co-Ordinator), GP surgeries, Health visitors, Community Paediatricians, Hospital follow-up appointments, towards the end of their neonatal stay.

Making data linkage routine in the future

- Regarding whether data linkage should be routine going forwards for preterm babies, 76% yes, 22% maybe. Only 2% said no
- When asked about their views on using a common anonymised identifier on both health and education records to avoid the use of identifiers, 84% said yes; 16% not sure.

The findings of the survey indicated that the majority of respondents were supportive of the concept of linking together existing routine data without consent. The reason for uncertainty was due to the need for more information and clarity on who will have access to the data, and reassurance that identifiers and data will not be sold.

A working group including parents, patients and health professionals are co-developing a leaflet and video to explain the temporary use of identifiers in the linkage process.

Ten survey respondents who have expressed mistrust or uncertainty about data linkage will be invited for an in-depth interview to explore their views on the video and leaflet. The feedback will contribute to the final version of the resources that will disseminate information to parents, patients and the public.

7.2 Parent and public involvement and engagement for surgical conditions and long-term outcomes

7.2.1 Parent Advisory Group feedback

A focus group meeting took place on the 25th February. Parents/relatives represented: Children aged 2.5 years to 19 with Hirschsprung's disease, exomphalos, NEC, short bowel, two support charities (NEC, short bowel), one Facebook support group.

Parents were strongly in favour of linking health and education data to provide information on their children's long term outcomes. They were happy that this could be done without requiring consent as long as information was made available (eg via a website, charities, support groups) that the research was taking place, and that the results were made available to them through similar routes in a timely manner. Parents were particularly strongly in favour of research to show the impact of their child's condition on the child's educational outcomes.

Parent of 19 year old, just about to have 9th major operation:

"His whole education has been affected. Because he has problems with repeated vomiting every time he felt sick he would be sent home for 48 hours. He missed so much school they tried to send us to court. He did his GCSEs two weeks after a major operation and although he did them at home no one gave him any extra help because he just had recent surgery and so he failed and had to resit them all. He has been registered with special educational needs and it was suggested that he should be sent to a special school but we just wanted him supported in a mainstream school."

Grandparent of a 13 year old with short bowel and long term total parenteral nutrition (charity head):

"Understanding school expectations is crucial. We were told [our granddaughter] may never go to school and she's done brilliantly. Because many of the parents who contact our charity have very variable school experiences I've written a booklet explaining the problems that these children may have in schools to enable parents to set up a dialogue [with their

children's teachers]. I believe education is just one part of a childs quality of life and this information will be really important to enhance schools understanding otherwise children will not get what they need."

Parent of a 2 1/2 year old with exomphalos:

"I am already really anxious about what will happen when my daughter goes to school and would really like to know what her outcomes of the education system might be. I feel a lot of attention has been on preterm children who spend a lot of time in a neonatal unit when our children [those with surgical problems] spend just as much time in hospital so we really need to know how it affects them longer term."

Mother of 10 year old with Hirschsprung's disease:

"The link between health and school outcomes are crucial for a child's quality of life. The surgical team need a hand in hand approach with the schools to offer more holistic care. We need the information to be able to best support a child in their school environment."

8 ANTICIPATED IMPACT

8.1 Preterm and surgical babies and their parents

- Demonstrating feasibility of NNRD multi-domain linkage supports the case to continue tracking outcomes into adulthood for this and future cohorts
- Information will help professionals to counsel parents and optimise educational and support for children and families
- Knowledge of influencing factors may influence health, educational, environmental and social policies
- Statistical methodology developed to large datasets can be applied to resolve other clinical uncertainties more quickly
- **Other patient groups**
- Linkage methods can benefit other groups of patients e.g. term babies with brain injury at birth and children on chronic disease registers e.g. cystic fibrosis
- **Public**
- Ensuring routine data that already exists is linked to maximise benefit, and provides a cost-efficient approach to address research questions
- **Academics**

- Clinical trialists, health economists (long-term neonatal outcomes) and organisations interested in exemplar linkage studies (e.g. Health Data Research UK)

8.2 Dissemination

Animation videos and leaflets*, 'What is data linkage and how can it benefit you, your baby and other families?*' and written reports summarising the research findings will be produced. These will be disseminated to **families** through UK neonatal units and the charity BLISS using their established communication channels (newsletters, social media, and volunteers).

Findings will be shared with **health professionals and academics** through peer-reviewed scientific publications, conference presentations, social media, and a study website.

A written report will summarise the findings for **public service providers and policy makers.**

9 REGULATORY ISSUES

9.1 ETHICS APPROVAL

No patient identifiable information will be used in this study and only existing anonymised data held in the NNRD will be used. The Neonatal Data Analysis Unit (NDAU) holds UK Research Ethics Committee approval, 16/LO/1093, and Confidential Advisory Group (CAG) approval, ECC 8-05(f/2010), to form the NNRD.

The Study Coordination Centre has obtained approval from the xxx Research Ethics Committee (REC) and Health Regulator Authority (HRA).

Each of the databases included in this proposal have received appropriate Caldicott Guardian approvals and necessary information governance checks; the individual databases also have their own CAG approvals to hold identifiable data.

9.2 CONSENT

Not applicable

9.3 SPONSOR

The sponsor is Imperial College London; insurance policies are held that apply to this study.

9.4 INDEMNITY

Imperial College London holds negligent harm and non-negligent harm insurance policies which apply to this study.

9.5 FUNDING

This project is funded by the National Institute for Health Research (NIHR) via the Advanced Fellowship programme (reference: NIHR300617).

9.6 AUDITS

The study may be subject to inspection and audit by Imperial College London under their remit as sponsor and other regulatory bodies to ensure adherence to GCP and the UK Policy Frame Work for Health and Social Care Research.

10 STUDY MANAGEMENT

This project will be based at the Neonatal Data Analysis unit, Chelsea and Westminster Hospital, Imperial College Campus. The NDAU manages and hosts the National Neonatal Research Database. All necessary governance structures are in place. The UK Neonatal Collaborative (all units in the UK) submit data with Caldicott Guardian approvals.

An independent steering committee comprising experts in data linkage, preterm birth and members of the public/ex-patients will meet annually to oversee the project. A project advisory group will meet more regularly to oversee the day-to-day running of the project and will comprise researchers, collaborators and parents/ex-patients. They will be updated at least quarterly on the progress of the project. Formal meeting minutes will be documented, reviewed and approved at each meeting. These will be posted on the neoWONDER website.

The research will also be monitored within the strict policies of Imperial College Neonatal Data Analysis Unit, NHS Digital, the Department for Education and the ONS. Annual progress reports will also be submitted annually to the funder, NIHR.

11 PUBLICATION POLICY

The results will be disseminated to:

Health professionals and academics through peer-reviewed scientific journals, conference presentations, website. The UK Neonatal Collaborative will be named collaborators and will be acknowledged in all academic publications.

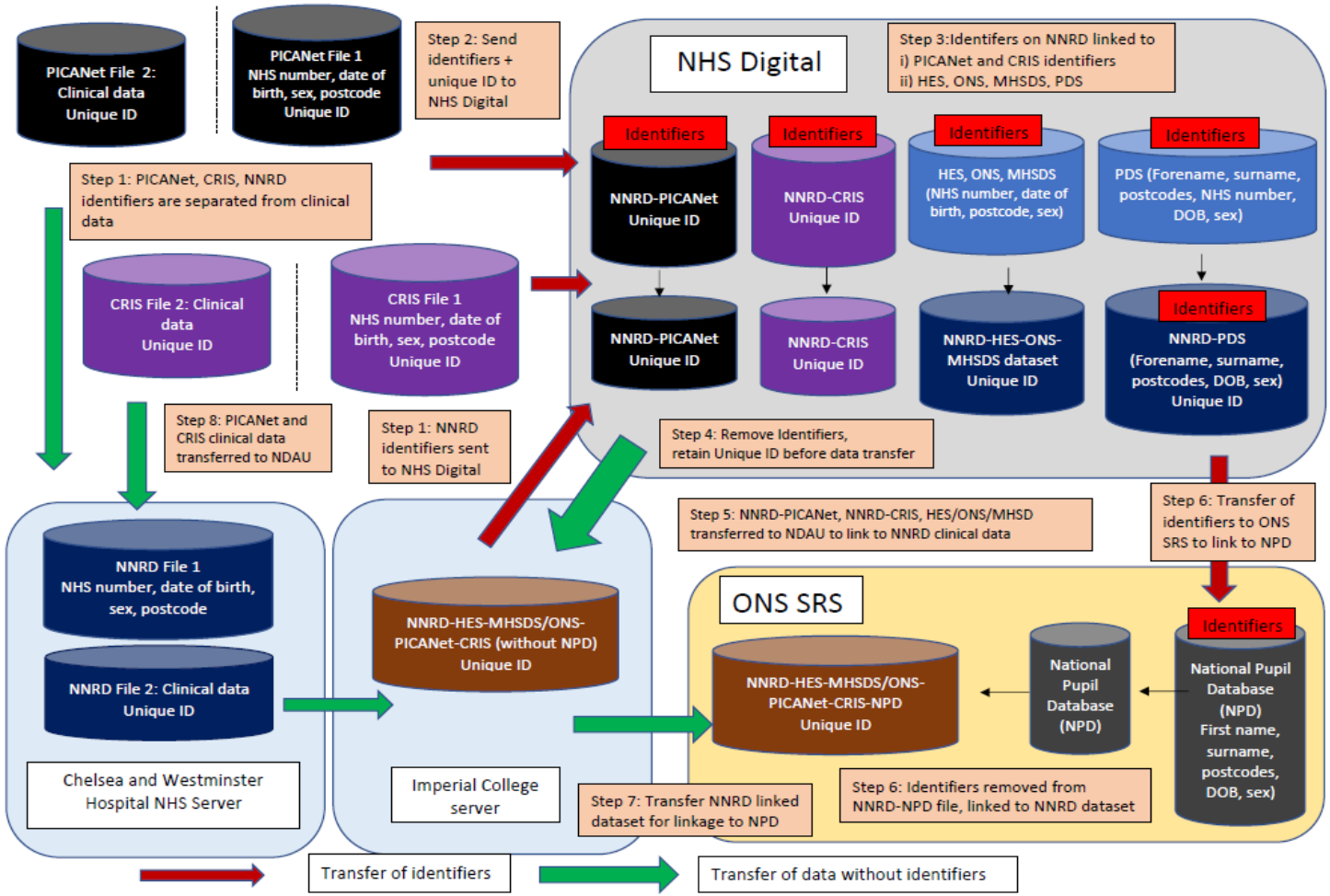
Patients and public through co-production of information leaflets and animation videos will be disseminated through UK neonatal units, BLISS charity established communication channels, social media

12 CONFIDENTIALITY

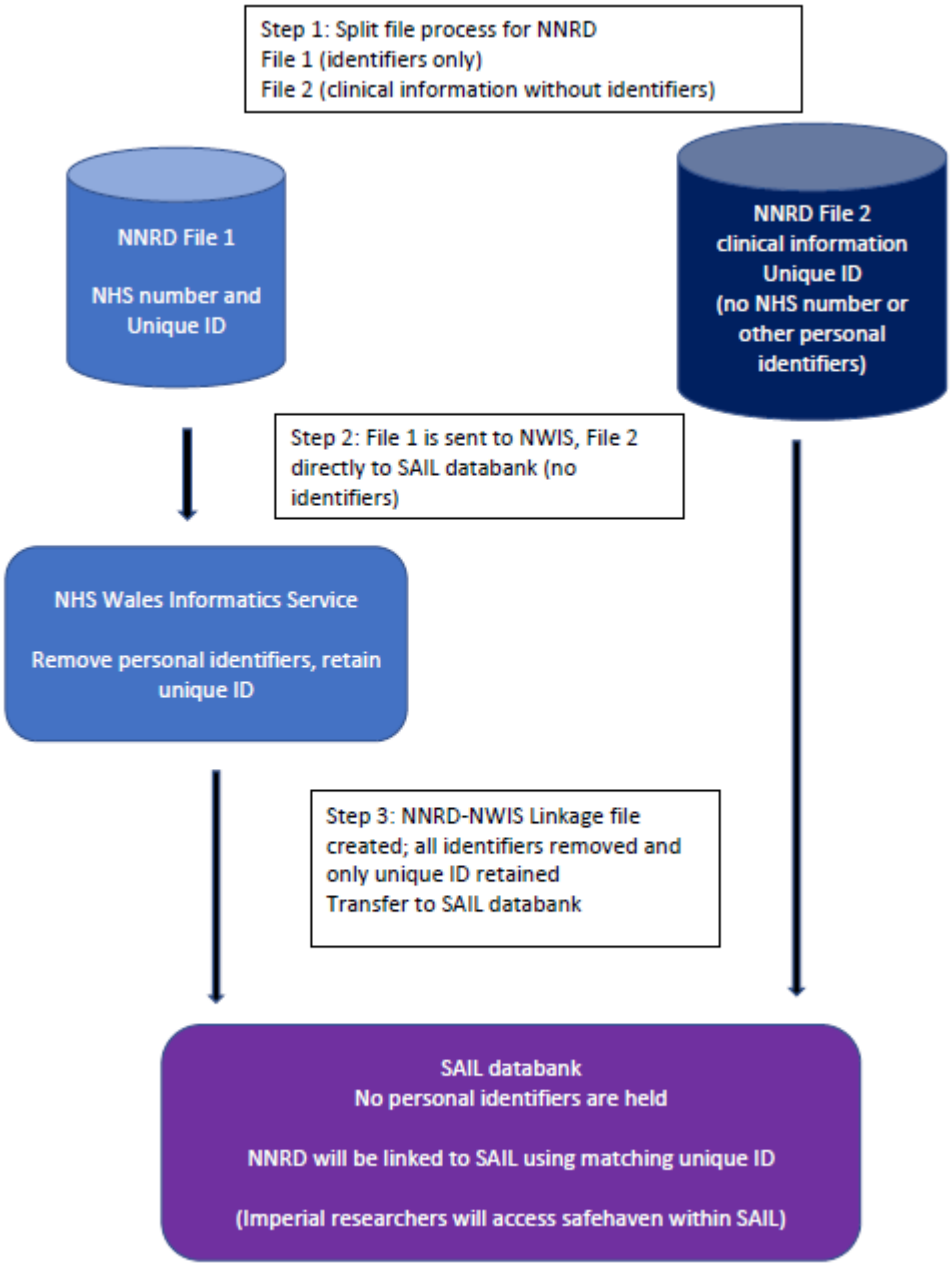
The Chief Investigator will preserve the confidentiality of participants taking part in the study and is registered under the Data Protection Act.

13 APPENDICES

13.1 Data flow: England



13.2 Data flow Wales



13.3 Split-file linkage process

Appendix: Split-file linkage process illustrated by NNRD-HES linkage

Step 1
Split NNRD into File 1 (identifiers) and File 2 (no identifiers). Both have unique Study ID

File 1 NNRD (identifiers only) + Unique ID

Unique ID	NHS number	Date of birth	Sex	Postcode
2123	12345678	19/01/2009	M	BR1 8ZZ

File 2 NNRD clinical data (without identifiers)+ Unique ID

Unique ID	NNRD Clinical 1	NNRD Clinical 2
2123	XX	XX

Step 2 Send File (identifiers) to NHS Digital to create NHS-HES linkage file
Link File 1 NNRD to HES by matching NHS number, date of birth, sex, postcode

File 1 NNRD (identifiers only)

Unique ID	NHS number	Date of birth	Sex	Postcode
2123	12345678	19/01/2009	M	BR1 8ZZ

+ HES data

NHS number	Date of birth	Sex	Postcode	HES Clinical 1	HES Clinical 2
12345678	19/01/2009	M	BR1 8ZZ	xx	xx

NNRD-HES file

Unique ID	NHS number	Date of birth	Sex	Postcode	HES Clinical 1	HES Clinical 2
2123	12345678	19/01/2009	M	BR1 8ZZ	XX	XX

Step 3
Remove identifiers, retain unique ID and transfer NNRD-HES file to NDAU

Unique ID	HES Clinical 1	HES Clinical 2
2123	XX	XX

Step 4: Merge NNRD-HES linkage file to NNRD File 2 clinical data using unique ID

NNRD-HES linkage file : File 2 NNRD clinical data (without identifiers)+ Study unique ID

NNRD

Unique ID	NNRD Clinical 1	NNRD Clinical 2
2123	XX	XX

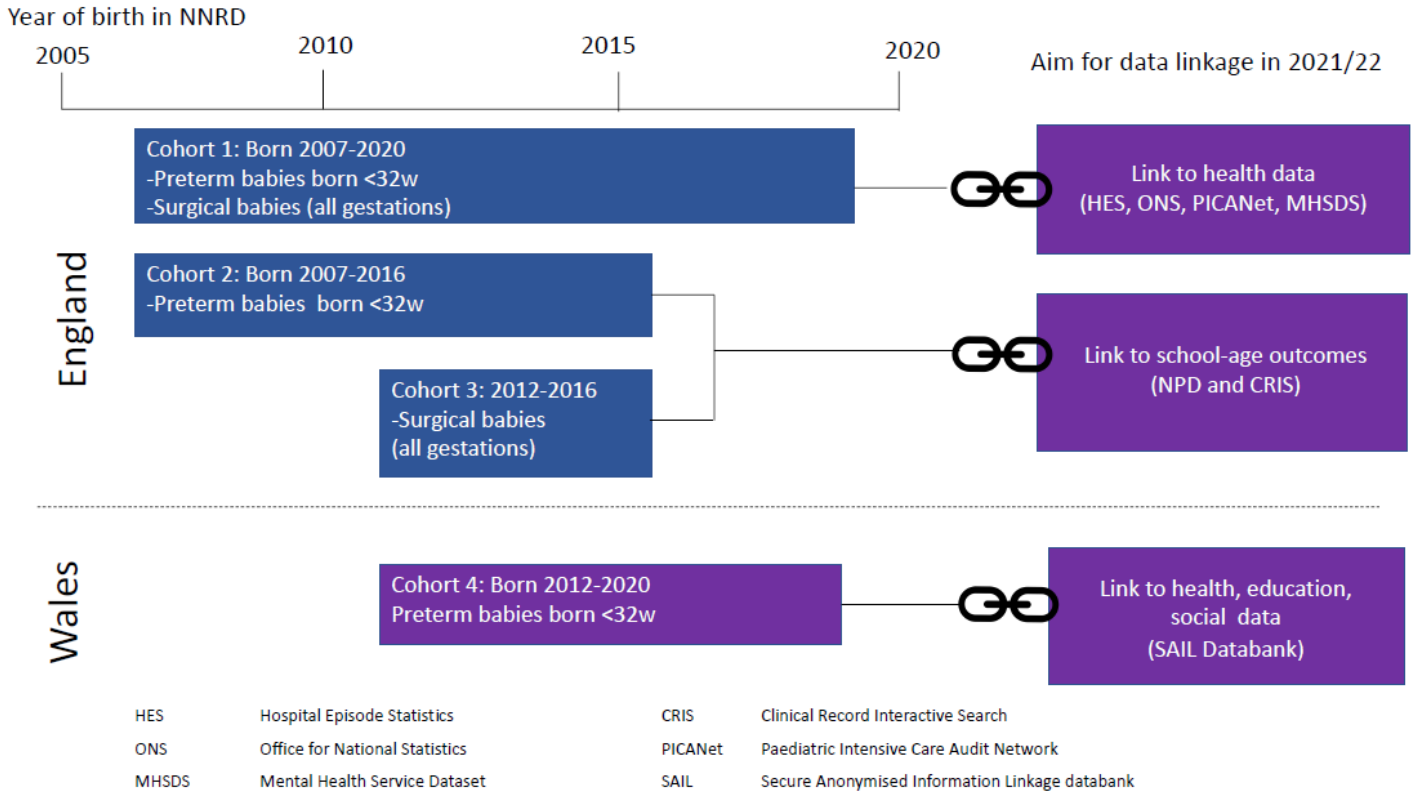
+ HES

Unique ID	HES Clinical 1	HES Clinical 2
2123	XX	XX

Linked NNRD-HES dataset without identifiers

Study Unique ID	HES Clinical 1	HES Clinical 2	NNRD Clinical 1	NNRD Clinical 2
2123	XX	XX	XX	XX

13.4 Cohort linkage



HES	Hospital Episode Statistics	CRIS	Clinical Record Interactive Search
ONS	Office for National Statistics	PICANet	Paediatric Intensive Care Audit Network
MHSDS	Mental Health Service Dataset	SAIL	Secure Anonymised Information Linkage databank

13.5 Surgical codes

Global Codes for NNRD/Clevermed:

- 10708: Necrotising enterocolitis – perforated [diag]
- 15809: Necrotizing enterocolitis [diag]
- 10501: Hirschsprung's disease – long-segment [diag]
- 10502: Hirschsprung's disease – short-segment [diag]
- 10503: Hirschsprung's enterocolitis [diag]
- 100039: Pull through for Hirschsprungs disease (any technique) [vonsurg]
- 11270: Definitive Hirschsprung's surgery – Duhamel [op]
- 11271: Definitive Hirschsprung's surgery – Lester Martin [op]
- 11272: Definitive Hirschsprung's surgery – Transanal pull through [op]
- 11273: Definitive Hirschsprung's surgery – Rehbein [op]
- 11274: Definitive Hirschsprung's surgery – Soave-Boley [op]
- 11275: Definitive Hirschsprung's surgery – Swenson [op]
- 11499: Laparoscopic assisted definitive Hirschsprung's surgery [op]
- 16226: Hirschsprungs Disease [diag]
- 16227: Hirschsprung's disease (long segment) [diag]
- 16228: Hirschsprung's disease (short segment) [diag]
- 11058: Total intestinal aganglionosis [diag]
- 100039: Pull through for Hirschsprungs disease (any technique) [op]
- 11277: Delayed closure gastroschisis [op]
- 11654: Primary repair of gastroschisis [op]
- 11708: Repair gastroschisis using prosthesis (specify type) [op]
- 11784: Silo insertion for reduction of gastroschisis [op]
- 1006730: Closure of gastroschisis includes closure of exomphalos [opcs]
- 16499: Gastroschisis [diag]
- 100035: Gastroschisis repair (primary or staged) [vonsurg]
- 200310: Gastroschisis [vonbirthde]
- 10464: Gastroschisis [diag]
- 200303: Esophageal atresia [vonbirthde]
- 10740: Oesophageal atresia with distal trachea-oesophageal fistula [diag]
- 10741: Oesophageal atresia without distal fistula [diag]
- 11658: Primary repair of oesophageal atresia [op]
- 1010245: Oesophageal atresia without tracheoesophageal fistula [diag]
- 1010246: Oesophageal atresia with tracheoesophageal fistula [diag]
- 16195: Atresia of oesophagus without fistula [diag]
- 16196: Atresia of oesophagus with trachea-oesophageal fistula (TOF) [diag]
- 100010: Tracheoesophageal atresia and/or/fistula repair [vonsurg]
- 200303: Esophageal atresia [vonbirthde]
- 10246: Congenital diaphragmatic hernia [diag]
- 10694: Morgagni diaphragmatic hernia [diag]
- 10905: Recurrent congenital diaphragmatic hernia [diag]

- 11597: Other repair of diaphragmatic hernia (specify) [op]
- 11657: Primary repair of congenital diaphragmatic hernia [op]
- 11660: Prosthetic repair of congenital diaphragmatic hernia (specify) [op]
- 1006132: Fetoscopic insertion of tracheal plug for congenital diaphragmatic hernia [opcs]
- 1006148: Percutaneous insertion of tracheal plug for congenital diaphragmatic hernia
- 1006671: Repair of congenital diaphragmatic hernia [opcs]
- 100044: Repair of diaphragmatic hernia [vonsurg]
- 200602: Congenital diaphragmatic hernia [vonbirthde]
- 1001921: Repair of diaphragmatic hernia using thoracic approach NEC [opcs]
- 1001923: Repair of diaphragmatic hernia using abdominal approach NEC [opcs]
- 1001924: Other specific repair of diaphragmatic hernia [opcs]
- 1001925: Unspecified repair of diaphragmatic hernia [opcs]
- 10854: Posterior urethral valves (PUV) [diag]
- 1004687: Endoscopic destruction of urethral valves [opcs]
- 11598: Other urinary diversion (specify) [op]
- 11444: Hook ablation of posterior urethral valve [op]
- 16357: Congenital posterior urethral valves (PUV) [diag]
- 100058: Resection of posterior urethral valves [vonsurg]
- 11338: Endoscopic resection of posterior urethral valve [op]

14 REFERENCES

1. Santhakumaran S, Statnikov Y, Gray D, Battersby C, Ashby D, Modi N. Survival of very preterm infants admitted to neonatal care in England 2008–2014: time trends and regional variation. Archives of Disease in Childhood - Fetal and Neonatal Edition. 2017.

2. Moore T, Hennessy EM, Myles J, Johnson SJ, Draper ES, Costeloe KL, et al. Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: the EPICure studies. Bmj. 2012;345:e7961.

3. Johnson S, Hennessy E, Smith R, Trikic R, Wolke D, Marlow N. Academic attainment and special educational needs in extremely preterm children at 11 years of age: the EPICure study. Arch Dis Child Fetal Neonatal Ed. 2009;94(4):F283-9.

4. O'Brien F, Roth S, Stewart A, Rifkin L, Rushe T, Wyatt J. The neurodevelopmental progress of infants less than 33 weeks into adolescence. Arch Dis Child. 2004;89(3):207-11.

5. Johnson S, Hollis C, Kochhar P, Hennessy E, Wolke D, Marlow N. Psychiatric disorders in extremely preterm children: longitudinal finding at age 11 years in the EPICure study. Journal of the American Academy of Child and Adolescent Psychiatry. 2010;49(5):453-63.e1.

6. Houweling LMA, Bezemer ID, Penning-Van Beest FJA, Meijer WM, Van Lingen RA, Herings RMC. First year of life medication use and hospital admission rates: Premature compared with term infants. Journal of Pediatrics. 2013;163(1):61-6.e1.

7. Boyle EM, Poulsen G, Field DJ, Kurinczuk JJ, Wolke D, Alfirevic Z, et al. Effects of gestational age at birth on health outcomes at 3 and 5 years of age: population based cohort study. Bmj. 2012;344:e896.

8. Raju TNK, Buist AS, Blaisdell CJ, Moxey-Mims M, Saigal S. Adults born preterm: a review of general health and system-specific outcomes. Acta Paediatrica. 2017;106(9):1409-37.

9. ElHassan NO, Bai S, Gibson N, Holland G, Robbins JM, Kaiser JR. The impact of prematurity and maternal socioeconomic status and education level on achievement-test scores up to 8th grade. PLoS One. 2018;13(5):e0198083.

10. Benavente-Fernandez I, Synnes A, Grunau RE, Chau V, Ramraj C, Glass T, et al. Association of Socioeconomic Status and Brain Injury With Neurodevelopmental Outcomes of Very Preterm Children. JAMA Netw Open. 2019;2(5):e192914.

11. Bangma JT, Kwiatkowski E, Psioda M, Santos HP, Jr., Hooper SR, Douglass L, et al. Early life antecedents of positive child health among 10-year-old children born extremely preterm. Pediatr Res. 2019.

12. Santos J, Pearce SE, Stroustrup A. Impact of hospital-based environmental exposures on neurodevelopmental outcomes of preterm infants. Curr Opin Pediatr. 2015;27(2):254-60.

13. Seeni I, Ha S, Nobles C, Liu D, Sherman S, Mendola P. Air pollution exposure during pregnancy: maternal asthma and neonatal respiratory outcomes. Annals of epidemiology. 2018;28(9):612-8.e4.

14. Lavigne E, Belair MA, Rodriguez Duque D, Do MT, Stieb DM, Hystad P, et al. Effect modification of perinatal exposure to air pollution and childhood asthma incidence. The European respiratory journal. 2018.

15. Bell EF, Rysavy MA. What Parents Want to Know after Preterm Birth. The Journal of Pediatrics. 2018;200:10-1.

16. Taras H, Potts-Datema W. Chronic Health Conditions and Student Performance at School. J Sch Health. 2005;75(7):255-66.

17. Stam H, Hartman EE, Deurloo JA, Groothoff J, Grootenhuis MA. Young adult patients with a history of pediatric disease: impact on course of life and transition into adulthood. J Adolesc Health. 2006;39(1):4-13.

18. Svoboda E, Fruithof J, Widenmann-Grolig A, Slater G, Armand F, Warner B, et al. A patient led, international study of long term outcomes of esophageal atresia: EAT 1. J Pediatr Surg. 2018;53(4):610-5.

19. Peetsold M, Heij H, Deurloo J, Gemke R. Health-related quality of life and its determinants in children and adolescents born with oesophageal atresia. Acta Paediatr. 2010;99(3):411-7.

20. Bojanic K, Grizelj R, Vukovic J, Omerza L, Grubic M, Caleta T, et al. Health-related quality of life in children and adolescents with congenital diaphragmatic hernia: A cross-sectional study. Health and Quality of Life Outcomes. 2018;16.

21. Onishi S, Nakame K, Kaji T, Kawano M, Moriguchi T, Sugita K, et al. The bowel function and quality of life of Hirschsprung disease patients who have reached 18 years of age or older - the long-term outcomes after undergoing the transabdominal soave procedure. J Pediatr Surg. 2017;52(12):2001-5.

22. Allin B, Long A-M, Gupta A, Knight M, Lakhoo K, Kazmierski M, et al. A UK wide cohort study describing management and outcomes for infants with surgical Necrotising Enterocolitis. Sci Rep. 2017;7(1):41149.

23. Bradnock TJ, Knight M, Kenny S, Nair M, Walker GM. Hirschsprung's disease in the UK and Ireland: incidence and anomalies. Arch Dis Child. 2017;102(8):722-7.

24. Cochrane Neonatal Reviews. Gaps in research 2018 [Available from: https://neonatal.cochrane.org/gaps-research.

25. ISRCTN registry. BMC [Available from: https://www.isrctn.com/.

26. Webbe JWH, Ali S, Sakonidou S, Webbe T, Duffy JMN, Brunton G, et al. Inconsistent outcome reporting in large neonatal trials: a systematic review. Archives of Disease in Childhood - Fetal and Neonatal Edition. 2019:fetalneonatal-2019-316823.

27. Marlow N. Measuring neurodevelopmental outcome in neonatal trials: a continuing and increasing challenge. Archives of Disease in Childhood - Fetal and Neonatal Edition. 2013;98(6):F554-F8.

28. Kenyon S, Pike K, Jones DR, Brocklehurst P, Marlow N, Salt A, et al. Childhood outcomes after prescription of antibiotics to pregnant women with spontaneous preterm labour: 7-year follow-up of the ORACLE II trial. Lancet. 2008;372(9646):1319-27.

29. Statnikov Y, Ibrahim B, Modi N. A systematic review of administrative and clinical databases of infants admitted to neonatal units. Archives of Disease in Childhood - Fetal and Neonatal Edition. 2017.

30. Battersby C, Statnikov Y, Santhakumaran S, Gray D, Modi N, Costeloe K, et al. The United Kingdom National Neonatal Research Database: A validation study. PLoS One. 2018;13(8):e0201815.

31. Neonatal Data Analysis Unit. National Neonatal Research Database (NNRD) [Available from: www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data-analysis-unit.

32. Neonatal Data Analysis Unit. National Neonatal Research Database London [Available from: www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data-analysis-unit.

33. NHS Digital. Personal Demographic Service (PDS) 2021 [Available from: www.digital.nhs.uk/services/demographics.

34. Office for National Statistics. 2021 [Available from: www.ons.gov.uk.

35. NHS Digital. Hospital Episode Statistics (HES) 2021 [Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics.

36. Secure Anonymised Information Linkage Databank. SAIL Databank 2021 [Available from: www.saildatabank.com.

37. Paediatric Intensive Care Audit Network (PICANet)

2021 [Available from: www.picanet.org.uk

38. NHS Digital. Mental Health Service Datasets (MHSDS) 2021 [Available from: www.digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/mental-health-services-data-set.

39. Department for education. National Pupil Database, [Available from: https://www.gov.uk/government/collections/national-pupil-database.

40. Downs JM, Ford T, Stewart R, Epstein S, Shetty H, Little R, et al. An approach to linking education, social care and electronic health records for children and young people in South London: a linkage study of child and adolescent mental health service data. BMJ Open. 2019;9(1):e024355.

41. University College London. ECHILD project website [Available from: https://www.ucl.ac.uk/child-health/research/population-policy-and-practice-research-and-teaching-department/cenb-clinical-20.

42. South London and Maudsley (SLaM). Clinical Record Interactive Search (CRIS) 2021 [Available from: www.maudsleybrc.nihr.ac.uk/facilities/clinical-record-interactive-search-cris/.

43. Imperial College London. Small Area Heath Statistics Unit (SAHSU) 2021 [Available from: www.imperial.ac.uk/school-public-health/epidemiology-and-biostatistics/small-area-health-statistics-unit/.

44. Neonatal Data Analysis Unit. Information for parents and carers [Available from: https://www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data-analysis-unit/information-for-patients-and-carers/.

45. NHS Digital. Delivering linked datasets to support health and care delivery and research 2018 April [Available from: https://digital.nhs.uk/services/data-access-request-service-dars/linked-datasets-supporting-health-and-care-delivery-and-research.

46. Secure Anonymised Information Linkage. SAIL datasets 2019 [Available from: https://saildatabank.com/saildata/sail-datasets/.