TIGAR

Tracking the Impact of Gestational Age on health, educational and economic outcomes: a longitudinal Record linkage study

Tracking the **I**mpact of **G**estational **A**ge on Health, Educational and Economic outcomes: a Longitudinal **R**ecord Linkage Study

PROTOCOL

Ethics ref: 15/SW/0294
15th May 2017, Version 1.3

Chief Investigator:

Professor Maria Quigley
National Perinatal Epidemiology Unit (NPEU)
Nuffield Department of Population Health
University of Oxford Old Road Campus
Oxford OX3 7LF

Co-investigators:

Dr Elaine Boyle[1]
Dr Claire Carson[2]
Dr Victoria Coathup[2]
Ms Nirupa Dattani[3]
Dr Samantha Johnson[1]
Professor Jenny Kurinczuk[2]
Professor Alison Macfarlane[3]
Professor Stavros Petrou[4]
Dr Oliver Rivero Arias[2]
Professor Pamela Sammons[5]
Dr Kati Toth[5]

1 Department of Health Sciences, University of Leicester
2 NPEU, University of Oxford
3 Centre for Maternal and Child Health Research, City University
4 Warwick Medical School, University of Warwick
56 Department of Education, University of Oxford

Sponsor: University of Oxford
Funder: Medical Research Council

Chief Investigator Signature:

No potential conflicts of interest declared.

Table of Contents

1. SYNOPSIS

Study Title	Tracking the Impact of Gestational Age on Health, Educational and Economic outcomes: a Longitudinal Record Linkage Study	
Internal ref. no. / short title	TIGAR	
Study Design	Population-based retrospective cohort study	
Study Participants	All live births in England during 2005 and 2006	
Planned Sample Size	Approximately one million live births	
Planned Study Period	15th October 2015 to 31st June 2018	
	Objectives	**Outcome Measures**
	Objective 1 To create a linked dataset of a child's hospital episodes and school data up to age 11 years for children born in England in 2005-6. **Objective 2** To estimate the effect of gestational age, across the full gestation spectrum, on hospital admissions and educational outcomes in children up to age 11 years. **Objective 3** To describe the trajectory of these outcomes in different gestational age-groups of children using 'growth curve modelling' methods. **Objective 4** To determine whether these effects vary according to markers of socio-economic deprivation and whether there is a gestational age threshold beyond which the impact of gestational age is outweighed by the effects of socio-economic factors. **Objective 5** To estimate the economic costs for hospital and education services in different gestational age-groups during the first 11 years of life, and the incremental costs associated with varying levels of prematurity. **Objective 6** To create a methodological resource to enable further research on this dataset or an updated version of it.	**Primary** • a linked dataset of a child's hospital episodes and school data up to age 11 years for children born in England in 2005-6. **Secondary** Hospital episodes: • all-cause admission rate (all ages; age-specific) • cause-specific admission rate • length of stay School assessments • at age 4-5 (Early Years Foundation Stage Profile: overall score and scores for specific areas and whether achieved 'good level of development') • at age 6-7 (Key Stage 1: overall score, and point score/level for reading and maths) • at age 10-11 (Key Stage 2: overall score, and point score/level for English/reading and maths) • Prevalence of special educational needs (SEN), overall and by type of SEN, prevalence of SEN

		statement, whether SEN support was received, and type of SEN support • Type of school attended (special vs mainstream) • School attendance

2. ABBREVIATIONS

A&E	Accident and emergency
CAG	Confidential Advisory Group
DfE	Department for Education
DMAP	Data Management Advisory Panel
EDD	Education Data Division
EYFSP	Early Years Foundation Stage Profile
HES	Hospital Episode Statistics
HRA	Health research Authority
HSCIC	Formerly Health and Social Care Information Centre (now NHS Digital)
KS1	Key Stage 1
KS2	Key Stage 2
MRC	Medical Research Council
NHS	National Health Service
NHSNO	National Health Service Number
NN4B	NHS Numbers for Babies
NPD	National Pupil Database
NPEU	National Perinatal Epidemiology Unit
ONS	Office for National Statistics
PPI	Patient and Public Involvement
PPPI	Patient, Parent and Public Involvement
PTB	Preterm Birth
REC	Research Ethics Committee
RR	Risk Ratio
SAT	School Assessment Test
SEN	Special Educational Needs
SGA	Small for Gestational Age
VML	Virtual Microprocessor Laboratory

3. BACKGROUND AND RATIONALE

3.1 Clinical significance of preterm birth

Preterm birth (PTB) is defined as birth before 37 completed weeks of gestation. Preterm deliveries occur for a variety of reasons, which may be broadly classified as either medically indicated or spontaneous.[1] Medically indicated deliveries occur following induction of labour or planned caesarean section before 37 weeks' gestation, because of obstetric concerns about maternal or fetal wellbeing. However, most PTBs occur spontaneously due to spontaneous onset of preterm labour or following prelabour premature rupture of membranes. While there are known risk factors for spontaneous PTB (e.g. multiple pregnancy, infection, maternal age, diabetes, hypertension), no cause is identified in up to 50% of cases.[1]

PTB is the second largest cause of death in children under five years.[2] It is estimated that globally in 2010, 14.9 million babies were born preterm, representing 11.1% of all live births worldwide.[1-2] The development and implementation of interventions aimed at preventing PTB have had limited success. Therefore, research also needs to focus on understanding the health and developmental consequences in the large number of survivors of PTB. In 2012, 52,909 babies in England and Wales were born preterm (7.0% of all live births) and 97.6% of these babies survived infancy.[3] The annual cost to the public sector in England associated with children born preterm until age 18 is around £1.24 billion and total societal costs (including parental costs and lost productivity) are around £2.48 billion.[4]

3.2 Consequences of PTB

PTB is a major cause of long term loss of human potential worldwide.[1] Survivors of PTB, particularly those born extremely preterm (<28 weeks' gestation), have an increased risk of short-term neonatal morbidities some of which may have long lasting effects, including bronchopulmonary dysplasia, retinopathy of prematurity, necrotising enterocolitis and neonatal sepsis. PTB is also associated with neonatal brain injuries and aberrant neurodevelopment. As such, preterm babies have an increased risk of physical, neurosensory, cognitive and behavioural problems in childhood and adulthood compared with their term-born counterparts. The risk of adverse outcomes is inversely related to gestational age at birth and extends across the full spectrum of PTB. Recent studies have identified an increased risk of adverse outcomes in childhood[5-10] and adulthood[11-15], even among those born late preterm (34-36 weeks) and early term (37-38 weeks), compared with those born at full term (39-41 weeks).

The neurodevelopmental sequelae of PTB span multiple developmental domains, but are particularly associated with deficits in working memory, attention, executive function, visuo-spatial processing, anxiety and peer relationship problems[16-17]. These difficulties have a major impact on a child's performance and integration at school. As such, children born preterm tend to have poorer academic attainment and a higher prevalence of learning difficulties and special educational needs (SEN) than their term-born classmates[6,17]. Children born preterm are also less likely to attend university than their term-born peers and are more likely to have reduced employment prospects and earning potential later in life[18]. Academic attainment is an important predictor of subsequent health and wellbeing[19-20]. For example, a 'strong dose-response' effect is observed where lower levels of educational achievement have a negative effect on physical and mental health in adulthood, as well as affecting income, employment and quality of life[21]. Thus, educational

outcomes are not only important markers of the impact of PTB on children's functional outcomes, but are key predictors of an individual's lifelong health and well-being and broader life chances.

Some studies have assessed the effect of PTB in separate socio-economic groups[7,13,22]. These suggest that PTB and social disadvantage represent a double jeopardy, and that social disadvantage may be an effect modifier of PTB, since some childhood outcomes are disproportionately worse if both PTB and socio-economic deprivation are present. Exploring socio-economic factors that may be protective or can attenuate the adverse effects of PTB is an important goal to identify potential targets for intervention. However, large sample sizes are needed to disentangle the separate effects of PTB and social disadvantage.

3.3 Limitations of previous studies

Studies of long term outcomes following PTB need to be based on large numbers. This is because, either the expected effects are relatively small (the odds ratios for outcomes in children born late preterm or early term are typically <1.3) or the gestational age group is relatively small in number (0.44% of all live births in England and Wales are <28 weeks' gestation[3]). Large cohorts can be created from good quality routine datasets, provided that they can be linked together. Several studies have used record linkage to create large birth cohorts (>100,000) to assess the long term effects of PTB using national (UK[6,10,23] and Scandinavia[11-14,18,24-25]) or regional[5,26-28] population-based data (Table 1). The UK studies include one from Scotland, which has a long history of linking high quality medical data, and one from Wales, based on the Wales Electronic Cohort for Children. The study from England linked routinely available maternity data with infant hospital admissions, both obtained using Hospital Episode Statistics (HES).

The studies in Table 1, however, have limitations which may affect their generalisability to a UK population. First, the babies in most studies were born in the 1960s-90s. Since this time, substantial advances in obstetric and neonatal care, and changing practice with respect to medical indications for early delivery, have led to improved survival rates following PTB. Thus, cohorts from earlier years have a smaller proportion of children born extremely preterm compared with the numbers who survive today. Second, the effects observed in the non-UK studies may vary in a UK population owing to differences in demographics (e.g. social disadvantage) and services (health and education).

In addition, the studies in Table 1 do not provide a complete picture of the impact of gestational age. For example, some studies have focused on particular gestational age-groups and, hence, do not measure effects across the whole gestational age spectrum. The UK studies in Table 1 each focused on only one outcome: SEN in 2005 in Scotland; emergency respiratory hospital admissions in the under-fives in Wales; or hospital admissions for bronchiolitis in infants in 71 hospitals in England. Other UK studies of long term outcomes following PTB have been much smaller (<15,000 in the UK Millennium Cohort Study[7,29] and Avon Longitudinal Study of Parents and Children[8]) or have focused only on extremely preterm babies, such as EPICure[16-17].

Table 1: Previous large studies (>100,000) of long term outcomes following PTB

Source	Country	Year of Birth	N	Outcomes	Gestation in weeks
National, population-based (UK)					
MacKay[6]	Scotland	1987-2000	407,503	SEN at one time point (age 5-18)	all weeks
Paranjothy[10]	Wales	1998-2008	318,613	Emergency respiratory hospital admission < age 5	all weeks
Murray[23]	71 hospitals in England	2007-8	296,618	Hospital admission for RSV bronchiolitis in infancy	term (37-41) versus preterm (<37)
National, population-based (Scandinavia)					
Eide[11]	Norway	1967-79	317,761	Adult IQ	all weeks
Moster[14]	Norway	1967-83	903,402	Medical & social disability adulthood	all weeks
Lindstrom[12] Lindstrom[24] Crump[15,25]	Sweden	1973-9	522,310 545,628 630,090	Disability & income, adults Psychiatric adolescence Adult health e.g. epilepsy, diabetes, asthma	all weeks all weeks 23-31, 32-34, 35-37, 37-42
Ekeus[13]	Denmark	1973-6	119,664	Cognitive test adult	all weeks
Mathiasen[18]	Denmark	1988-9	118,281	Completing school, adult	all weeks
Regional studies (within the US)					
Morse[5]	US	1996-7	161,804	Pre-school, age 5	34-41 weeks
Petrini[26]	US	2000-04	141,321	Neurological early child	30-33, 34-36, 37-41, 42+
Lipkind[27]	US	1994-8	215,138	School 3rd grade	32-33, 34-36, 37-42
Noble[28]	US	1988-92	128,050	School age 8	37, 38, 39, 40, 41
Proposed study					
Proposed study	England	2005-6	~1 m	Cause-specific hospital episodes (<1, 1-4, 5-9 years) School tests, prevalence & type of SEN, attendance (age 5, 7, 11)	all weeks

age 5-18 in 2005

The economic cost of PTB birth during childhood has also received attention in the literature over the past decade[30,31]. Published studies have shown that the largest contribution to the economic impact of preterm birth is through hospital inpatient services during the birth admission, accounting for 92% of the incremental cost per preterm survivor in England and Wales.[30] They also show that

economic consequences for the health services, and more broadly for society, are substantial and sustained through childhood[31]. These studies, however, were either based on historical region-specific data,[31] or synthesised evidence from several sources using decision modelling frameworks and are therefore prone to substantial parameter uncertainty[30]. Therefore, there is a need to generate current estimates of economic costs of preterm birth using individual-level data contained in up-to-date record linkage datasets that can both inform clinical and service delivery, and validate long-term cost estimates reported in previous modelling exercises.

Hence there is a need for a large UK study to assess health, educational and economic outcomes across the whole gestational age spectrum, in a population-based cohort of babies born in the 21st century.

3.4 Our proposed study: added value and beneficiaries

Since the introduction of NHS numbers for babies (NN4B) in 2002, high quality data on gestational age at birth has been available nationally in England. A series of record linkage studies have cross-sectionally linked NN4B with birth and death registrations and hospital maternity records for all babies born in England and Wales in 2005-7[32-33]. This work was conducted by two of the co-investigators of the present study (AJ, ND) and part-funded by MRC. They are now working to link this to records of subsequent admissions of the children for in-patient and day case care up to March 2015 in an NIHR-funded project to analyse the outcome of pregnancy by time of day and day of the week. We propose to further enhance these data by adding in data about the most recent hospital admissions plus data from the National Pupil Database to create a longitudinal record linkage study: a national cohort of all live births in England during 2005-6, which links maternity information, including gestational age at birth, with subsequent hospital admissions and educational outcomes up to 11 years of age.

We will use the cohort to estimate: the association between gestational age and health and educational outcomes; trajectories of these outcomes in different gestational age-groups; and whether these effects vary according to markers of socio-economic deprivation. We will also measure the costs associated with PTB to health and education services. The study will be the largest UK study of the consequences of birth before full-term and will be the only one to concurrently study health, educational and economic outcomes. This large representative cohort will allow us to estimate more precisely the large effects expected in the relatively small group of babies born at less than 28 weeks of gestation and the smaller effects expected in the large group of babies born late preterm (34-36 weeks' gestation) or at early term (37-38 weeks' gestation).

Results on the risks, risk trajectories, effect modifiers and costs will be made available to inform clinical decision-making, for example, in the scenario of induction of labour for intrauterine fetal growth restriction. The results will also inform education professionals about the nature of learning difficulties associated with birth before full-term and what kind of SEN support is needed. A recent survey suggests that ~90% of teachers feel ill-equipped to support preterm children in the classroom[34] highlighting the pressing need for data to inform educational policy and practice. The results will also assist in counselling of parents about the types of problems that may occur, which will enable them to be more prepared about what to expect and when to seek help from professionals. The results will inform those involved in resource planning, policy and service provision (such as neonatal and paediatric healthcare services, public health and education services) and the development of interventions in terms of populations targeted and the content

and timing of delivery (for example, whether it is more important to focus on all children born very preterm, or socio-economically deprived preterm children). Finally, we will create a methodological resource describing the process and methods for accessing and analysing these datasets, which will be made available to other researchers to facilitate further work in this area.

4. OBJECTIVES AND OUTCOME MEASURES

The overall aim of the study is to investigate the effect of gestational age on health, educational and economic outcomes up to age 11 years.

4.1 Objectives

1. To create a linked dataset of a child's hospital episodes and school data up to age 11 years for children born in England in 2005 and 2006.

2. To estimate the effect of gestational age, across the full spectrum, on hospital admissions and educational outcomes in children up to age 11 years.

3. To describe the trajectory of these outcomes in different gestational age-groups of children using 'growth curve modelling' methods.

4. To determine whether these effects vary according to markers of socio-economic deprivation and whether there is a gestational age threshold beyond which the impact of gestational age is outweighed by the effects of socio-economic factors.

5. To estimate the economic costs for hospital and education services in different gestational age-groups during the first 11 years of life, and the incremental costs associated with varying levels of prematurity.

6. To create a methodological resource to enable further research on this dataset or an updated version of it.

4.2 Outcome measures

The study will include a wide range of health and educational outcomes which are directly relevant to UK health and education systems (Table 2). The educational data are particularly relevant in the context of PTB as they measure not only academic performance (which is an important marker of lifelong attainment and future health), but also the presence and *type* of special educational need, and what support is given, during the child's primary schooling (at age 4-5, 6-7 and 10-11 years).

Table 2 Study outcomes, exposure, confounders, mediators and moderators

Outcomes	**Data source**
Hospital episodes (inpatients, A & E and outpatients): • all-cause admission rate (all ages; age <1, 1-4, 5-9 years) • cause-specific admission rate (e.g. asthma, bronchiolitis, other respiratory infections/conditions, gastroenteritis, seizures). • length of stay (for neonatal and other admissions)	Hospital Episode Statistics (HES)
School assessments • EYFSP (age 4-5) (overall score and scores for specific areas and whether achieved 'good level of development') • KS1 (age 6-7) (overall score, and point score/level for reading and maths) • KS2 (age 10-11) (overall score, and point score/level for English/reading and maths) • Prevalence of special educational needs (SEN), overall and by type of SEN (behaviour, social, emotional problem, academic, specific learning difficulty, vision impairment etc), type of statement (none, school action, school action plus, school action plus and statutory assessment, statement), whether SEN support was received, and type of SEN support (e.g. 1-2-1 classroom support, small group special needs tuition; outreach teacher etc) • Type of school attended (special vs mainstream; state vs private) • School attendance*	National Pupil Database (NPD)
Exposure (and related variables)	
Gestational age Birth weight and small for gestational age (derived from birth weight, gestation, gender, ethnicity, mother's country of birth) Admitted to neonatal unit Mode of delivery Induction of labour	Baby-Cohort
Confounders, mediators, moderators at delivery#	
Sex of baby Maternal age at delivery (derived from date of birth) Multiple birth and birth order Mother's and father's country of birth and baby's ethnicity Registration status (married, joint/sole registration outside marriage, same/different address) Index of multiple deprivation score (derived from postcodes at birth reg. and school data) Month of birth e.g. for infectious morbidity and age within the school year	Baby-Cohort
Confounders, mediators, moderators at school	
Gender Chronological age and age within the school year Ethnicity (20 groups) Language (10 groups) Area deprivation (derived from postcode) Free school meals (ever and eligibility) Disability (and type)* Absences* Exclusions (and reasons) Whether part-time Children looked after Children in need (i.e. referred to social services)	NPD

* these are potential outcomes and confounders/mediators/moderators

EYFSP Early Years Foundation Stage Profile KS1 Key Stage 1 KS2 Key Stage 2

5. STUDY DESIGN AND METHODS

5.1 Study design, data sources and record linkage

The TIGAR study will create a national longitudinal population-based cohort of all live births in England during 2005-2006 (~1 million) with follow-up of children through record linkage of several routine datasets (Table 3). This will provide long-term hospital data on approximately 967,806 children and school data on 860,272 children (see Section 8.2, Sample Size), and as such will be the largest cohort of babies born in the 21st century.

Table 3: Data sources for record linkage

		Linkage already done	Linkage to HES in 2015	Linkage to school data in early 2016 and early 2017 (and possibly early 2018)		
Year of birth	**No. of live births in England**	**No. in master dataset #**	**No. of HES years to be linked**	**Year of EYFSP* assessment (age 5)**	**Year of KS1* Assessment (age 7)**	**Year of KS2* Assessment (age 11)**
2005	613,028	561,380	10	May 2010, 2011	May 2012, 2013	May 2016, 2017
2006	635,748	580,178	9	May 2011, 2012	May 2013, 2014	May 2017, 2018

number with known gestational age

*EYFSP Early Years Foundation Stage Profile KS1 Key Stage 1 KS2 Key Stage 2

Linkage will be performed following the necessary approvals which are described in detail in Section 10.3.

The master dataset (referred to as the **Baby-Cohort**) will be an extract of an existing dataset which is held at ONS. In brief, birth registration data for all births in England and Wales in 2005-7 has already been linked with the NN4B and HES maternity datasets by two of the co-investigators (AM, ND)[32-33]. As part of an NIHR-funded project, AM and ND are updating this dataset to include births in 2008-2012 and linkage to subsequent HES data about admissions of the children to hospital for in-patient or day case care. Linkage will be done by NHS Digital (formerly the Health and Social Care Information Centre). The Section 251 approval for this NIHR-funded project (CAG 9-08(b)/2014) includes creating a research database which could be accessed (with CAG approval) by other researchers using the ONS Virtual Microprocessor Laboratory (VML). The TIGAR team will apply to access a subset of that dataset i.e. data on children born in England in 2005-6, referred to here as the **Baby-Cohort**).

Following the necessary approvals (see Section 10.3), the linkage process will involve several stages (see the flow chart in **APPENDIX A**):

1. ONS will create Baby-Cohort (data from birth registration, NN4B, age at death in months and HES maternity, births and readmissions for all children born in 2005-6) and will create two new identifiers or pseudonyms (e.g. random numbers) for each child:

- *tigarid* will be included in the pseudonymised extracts of the data to be released for analysis in the NPEU's facilities in Oxford. ONS will hold the code which links *tigarid* to patient identifiers (e.g. NHS number).
- *linkid* will be used as a pseudonym to enable the sharing of identifiers and linkage between ONS and NHS Digital (Phase 1 linkage at NHS Digital), then between NHS Digital and the EDD (Phase 1 linkage at EDD), and then between the EDD and ONS (Phase 2 linkage at NHS Digital), as described below.

Therefore, ONS will hold the code which enables *linkid* to be linked with the patient identifiers it already holds (e.g. NHSNO, DoB), NHS Digital will hold the code which enables *linkid* to be linked with the patient identifiers it already holds (NHSNO, DoB, name, postcode) and EDD will hold the code which enables *linkid* to be linked with the patient identifiers it already holds (DoB, name, postcode).

2. In order that the TIGAR team can check the matching between birth registration/NN4B and HES in Baby Cohort, and clean the data, we will apply to access a version of Baby-Cohort which will include some identifiers (e.g. NHS number, date of birth, postcode, age at death in months), but no names or addresses, using the secure Virtual Microdata Laboratory (VML) facilities at ONS (see Section 10.3). ONS will then produce a data extract of Baby-Cohort which will include *tigarid*, but will be stripped of any identifiers (i.e. it will not contain names, addresses, postcodes, dates of birth, NHS numbers). The data will include month and year of birth (which is required for analyses such as age within the school year), and age at death in months (which is required for censoring in survival analysis in hospital admissions). The TIGAR team will apply for a copy of this pseudonymised extract of the data (no. 1 on left hand side of flow chart) to be released for analysis in the NPEU's facilities in Oxford (See Section 10.3). Importantly, this extract will contain no identifiers and will be effectively be anonymous from the point of view of the TIGAR researchers, but ONS will hold the code which will enable further linkage via ONS. This extract will be analysed by the TIGAR team at Oxford while the subsequent linkages to **Baby-Cohort** are ongoing. All publications will follow disclosure control guidance as described in Section 10.5.

3. **For the Phase 1 linkage,** ONS will create a file which includes the pseudonym (*linkid*) and the main identifiers (including the baby's NHS number, date of birth and sex) from **Baby-Cohort** and send this to NHS Digital. Then NHS Digital will link this file with the Personal Demographics Service (PDS) in order to generate a file of identifiers, including the child's name and most recent postcode (ideally at primary school age). These identifiers will be needed for linkage to the children's records in the National Pupil Database as described in the Flow chart (Phase 1 linkage at NHS Digital).

4. Then NHS Digital will send a file which includes *linkid* and the identifiers required for linkage (name, postcode, date of birth, sex) to the Education Data Division (EDD) of the Department for Education. The EDD will link this file with the National Pupil Database (NPD) as described in the Flow chart (Phase 1 linkage at EDD).

5. Then EDD will send a file which includes *linkid* and the NPD school data to ONS. ONS will link the school data to the existing data in Baby-Cohort using *linkid* and the new data (**Baby-School-Cohort)** will be held hold on one of its own secure servers.

6. In order that the TIGAR team can check the matching and clean the data, we will apply to access a version of **Baby-School-Cohort** which will include some identifiers (e.g. NHS number, date of birth, postcode), but no names or addresses, using the secure Virtual Microdata Laboratory (VML) facilities at ONS (see Section 10.3). Once the matching has been checked, ONS will be notified of any corrections via the VML. In addition, the file of identifiers to be used for linkage with the NPD will be crosschecked for completeness by ONS. ONS will produce a table showing the number of children who have complete data on name, postcode and date of birth, overall and by gestational age in weeks.

7. The ONS will make the necessary corrections (from point 6) to Baby-School-Cohort and will place the corrected version of **Baby-School-Cohort** on one of the secure ONS servers. ONS will then produce a data extract of Baby-School-Cohort which will include *tigarid*, but will be stripped of any identifiers (i.e. it will not contain names, addresses, postcodes, dates of birth, NHS numbers, but will include month and year of birth, and age at death in months). Following the same process as in point 2, the TIGAR team will apply for a copy of this pseudonymised extract of the data (no. 2 on left hand side of flow chart) to be released for analysis in the NPEU's facilities in Oxford (See Section 10.3).

8. **For the Phase 2 linkage,** ONS will create a file which includes the pseudonym (*linkid*) and the main identifiers (including the baby's NHS number, date of birth and sex) from **Baby-School-Cohort** and send this to NHS Digital. Then NHS Digital will link this file with HES (primarily based on the baby's NHS number) in order to obtain further hospital in-patient and day case episodes since March 2015 up to approximately age 11 years.

9. NHS Digital will send a file which includes *linkid* and the HES data from NHS Digital to ONS. ONS will link the HES data to the existing data in Baby-School-Cohort using *linkid* and the new data will be called **Baby-School-HES-Cohort** and will be held hold on one of its own secure servers.

10. In order that we can check the matching and clean the data, we will apply to access a version of Baby-School-HES-Cohort which includes some identifiers (e.g. NHS number, date of birth, postcode), but no names or addresses. This will be done using the same process described in 5), via the secure VML facilities at ONS. Once the matching has been checked at the VML, ONS will be notified of any corrections via the VML.

11. ONS will make the necessary corrections (from 9) to **Baby-School-HES-Cohort** and will place the corrected version of **Baby-School-HES-Cohort** on one of the secure ONS servers. ONS will then produce a data extract of Baby-School-HES-Cohort which will include *tigarid*, but will be stripped of any identifiers (i.e. it will not contain names, addresses, postcodes, dates of birth, NHS numbers, but will include month and year of birth, and age at death in months). Following the same process as in point 2, the TIGAR team will apply for a copy of this pseudonymised extract of the data (no. 3 on left hand side of flow chart) to be released for analysis in the NPEU's facilities in Oxford (See Section 10.3).

Hence, the process will result three pseudonymised extracts (see left hand side of Flow chart):

- **Pseudonymised extract number 1**: existing health data which can be analysed by the TIGAR team while Phase 1 of the linkage is ongoing.

- **Pseudonymised extract number 2**: existing health data linked with school data, which can be analysed by the TIGAR team while Phase 2 of the linkage is ongoing.
- **Pseudonymised extract number 3**: existing and updated health data, linked with school data.

All phases of linkage will be piloted using algorithms that have previously resulted in successful matches using similar matching variables and datasets[32-33,35]. The first stage linkage is currently being piloted as part of the current NIHR-funded project. The linkage with the National Pupil Database records will be piloted and then checked by accessing the data in the ONS VML. The algorithm will be modified as necessary and the linkage repeated, until the resulting linkage is deemed successful. The linkage of birth registration (including NN4B data) to HES delivery records had a successful linkage to a recorded gestational age for 99.3% of records and a recorded birth weight for ~99% of babies. There was good agreement between NN4B data and HES maternity data for gestational age (e.g. 89% exact agreement and 96% agreement within 1 week in 2005) and birth weight (e.g. 99% agreement within 500g in 2005[32]). Studies that have linked to school data using matching on child's name, date of birth, gender and postcode have obtained a successful match in 85-95% of children[35].

5.2 Health and educational outcomes

See Section 4.2 and Table 2.

5.3 Gestational age and birth weight

Data on gestational age are complete for 99% of births and have good agreement with HES[32]. We will analyse gestational age in weeks (with 40 weeks as the comparison group) although we will also present some results according to the following well-defined groups: extremely preterm (<28 weeks), very preterm (28-31 weeks), moderately preterm (32-33 weeks), late preterm (34-36 weeks), early term (37-38 weeks) and full term (39-41 weeks; the comparison group). Those with a gestation of 42+ weeks will be analysed as a separate group or excluded, depending upon data quality. We will also analyse the effect of small for gestational age (SGA) using centiles that will be derived from gestational age, birth weight and sex. Data will not be available on underlying cause of PTB although we will separate out medically-indicated and spontaneous PTB using data on induction of labour and mode of delivery.

5.4 Confounders, mediators and moderators

The potential confounders, mediators and moderators are listed in Table 2.

6. PARTICIPANTS

6.1 Study participants

The TIGAR study will create a national longitudinal population-based cohort of all live births in England from 1st January 2005 until 31st December 2006.

6.2 Exclusion criteria

Births for which there are no available data on gestational age will be excluded from the analyses.

7. STUDY PROCEDURES AND INFORMED CONSENT

Individual consent will not be sought from parents or children whose data are included in this study. The size of the study and the retrospective nature of the design (about one million children born in 2005-2006) mean that it is not possible to obtain informed consent. In addition, all identifiers (e.g. names, postcodes, dates of birth, NHS numbers) will be removed from the dataset before it is provided to the TIGAR researchers who will do the analysis, hence it will be anonymised data. Therefore, it will not be possible to remove records for any individuals who might not wish their data to be used in this way.

We consider our research to be in the public good, as it will yield information about the situation and needs of children born preterm and inform the provision of services to support them.

Fair processing will be addressed by the provision of information on the NPEU website (www.npeu.ox.ac.uk\tigar). Following the guidance of the Information Commissioner we have taken a layered approach to fair processing and the provision of privacy notices although we acknowledge that this approach is not ideal, given the historical nature of the cohort we are planning to assemble.

We considered the provision of information via antenatal clinics in the form of posters, as are often used in this situation. However, given the historical nature of the cohort (births in 2005-2006) it is unlikely that such posters would be seen by relevant parents. Information provision via schools is not a practicable option given the number of schools involved (in 2012, there were 16,818 state-funded primary schools in England, not including special schools and independent schools). Our approach therefore is via our website (www.npeu.ox.ac.uk\tigar) and dissemination via relevant third sector organisations.

8. STATISTICS AND ANALYSIS

8.1 Statistical methods

First, a descriptive analysis will be conducted on the exposures, outcomes, confounders, mediators and moderators. Internal and external validation of the data will be performed e.g. plausible birthweight for gestational age and comparison of frequency distributions and cross-tabulations of most data items with published reports by ONS, DfE and relevant journal articles[6,10,23,32-33]. We will assess the extent of missing data and factors associated with 'missingness', and determine the most appropriate method for addressing this in the analysis (e.g. complete case analysis, multiple imputation, non-response weights).

For objective 2, the health and educational outcomes will be compared across gestational age in weeks (with 40 weeks as the comparison), gestation groups (<28; 28-31; 32-33; 34-36; 37-38 and 39-41 weeks), and small for gestational age (SGA) percentiles (e.g. less than10%). Standard statistical methods will be used to estimate the effect of these exposures on the outcomes with adjustment for confounders (Table 2): logistic regression will be used to estimate odds ratios for

rare binary outcomes (e.g. cause-specific hospital admission in infancy, 1-4 years or 5-10 years); modified Poisson regression will be used to estimate risk ratios for common binary outcomes (e.g. achieving an appropriate level in a school assessment); Cox regression will be used to take account of censoring if appropriate (due to variable length of follow-up, for example, due to death); and linear regression will be used for continuous outcomes (e.g. school assessment score).

For objective 3, the analysis will utilise the longitudinal nature of the data to describe the trajectory of these outcomes in different gestational age-groups. This will be done using multilevel modelling (e.g. for modelling repeated outcomes over time) or structural equation modelling of latent 'growth curves' (e.g. for modelling different outcomes over time). Our models will include separate trajectories for health and educational outcomes, but also a particular novel aspect which is a trajectory incorporating both health and educational outcomes, and their impact on each other.

For objective 4, we will add interaction terms for markers of socio-economic deprivation to the models fitted for objectives 2 and 3. The markers will include area deprivation index, ethnicity, whether the parents were born in the UK, mother's age and marital status at delivery, and eligibility for free school meals (the children in our study will not have been affected by the policy of free school meals for all children in reception and KS1 from September 2014). If there is evidence that the impact of birth before full term is disproportionately worse in the presence of socio-economic disadvantage (i.e. effect modification), then the models will be stratified according to this socio-economic factor(s). The effect sizes will be compared for gestational age and the socio-economic factors. In particular, we will aim to identify the gestational age cut-off beyond which the impact of socio-economic factors is significantly stronger than gestational age, if in reality such an effect exists. Unlike most previous studies which have explored the separate effects of gestation and socio-economic deprivation[7,13,22,36], our study will have sufficient statistical power to disentangle these effects in the context of contemporary measures of socio-economic deprivation, health and education in England.

For objective 5, the costs during the first 11 years of life associated with hospital services, encompassing inpatient, day case and outpatient hospital services, and education services will be estimated. Hospital admissions will be grouped into healthcare resource groups and associated costs calculated using the most up-to-date reference cost grouper[37] whereas education services will be costed using information from the literature and nationally-applicable unit costs. Economic costs will be estimated for each gestation group (<28; 28-31; 32-33; 34-36; 37-38 and 39-41 weeks). For each category of costs, the time-path for each baby in the linked database will be divided into costs incurred between <5 years, 5-11 years and overall to understand cost drivers. Panel data regression models, adjusting for similar confounders as in objective 2, using random and fixed effects will be estimated to obtain measures of variability over time within and between individuals and a better understanding of individual heterogeneity[38]. We will develop models to estimate incremental resource use consumption (e.g. length of stay, frequency of use of educational services) and costs associated with varying level of prematurity.

Analysis will be conducted in Stata [Stata Corporation, College Station, USA] and MPlus.

8.2 Sample size

The number of births from 2005-6 with known gestational age in the pre-existing database is about 1.1 million (Table 3). After allowing for gestation-specific infant mortality (0.5% overall), further

deaths beyond infancy (<1%), and inability to match to the child's HES data or school data (estimated at no more than 10% and 20% respectively), we estimate obtaining HES follow-up data on 967,806 children and school data on 860,272. As an example, this would yield a sample size of 1701 (for HES) and 1512 (for school) for those born at 28 weeks gestation, 58,275 (for HES) and 51,800 (for school) for those born at 37 weeks gestation, and 272,115 (for HES) and 241,880 (for school) for those born at 40 weeks gestation (comparison group).

These numbers will result in at least 90% power to detect even smaller risk ratios than those observed for similar outcomes in UK studies. For example, when comparing those born at 28 weeks and 40 weeks, we will have 90% power to estimate the following risk ratios: ≥1.58 for hospitalisation for bronchiolitis in infancy (estimated RR is 3.9[10]), ≥1.77 for hospitalisation for asthma (estimated RR is above 2.6[10]) and ≥1.08 for having good school performance at age 5 (estimated RR is above 1.19[7]). For gestational ages greater than 28 weeks, we will have at least 90% power to detect much smaller effects than these, and those observed in the literature[7, 9-10].

9. DATA MANAGEMENT

9.1 Access to data

The full database from this study will be held on one of the secure servers at ONS along with the files of identifiers used to generate it. At various stages, this dataset will be made available in the ONS VML temporarily for checking and cleaning and the derivation of a pseudonymised dataset. Access to the individual-level data in the VML will be restricted to members of the TIGAR team who have ONS Approved Researcher Status. The pseudonymised data generated from this study will be held within the NPEU and, during the study, it will be accessed only by the study team and NPEU IT staff to generate analyses for the team and co-applicants. Access to the data held at the NPEU will be restricted to the following individuals:

NPEU Investigators:	Prof Maria Quigley (Chief investigator) Dr Claire Carson (Co-investigator) Prof Jenny Kurinczuk (Co-investigator; data custodian as Director of NPEU) Dr Oliver Rivero-Arias (Co-investigator)
3 NPEU Researchers	Dr Victoria Coathup and two researchers to be appointed
NPEU IT Security:	NPEU Head of IT and Information Security (currently Brian Hicks) NPEU Network Manager (currently Paul Semple)

As part of this project, we will create a linked dataset (objective 1) and a methodological resource to enable further research on this dataset or a similar one (objective 6). At the end of the study, the linked dataset created by the study will be made available to other researchers on application to the data custodian at ONS and subject to gaining the necessary ethics and other approvals from the CAG and Data Management Advisory Panel (DMAP) at the Education Data Division. We will help facilitate this process by creating a resource which has information on the process and methods for accessing and analysing these datasets. For example, this will document information about the approvals process, linkage methods and results, data validity and statistical methods.

The resource will include a short report which will be made available to interested researchers on request and will be disseminated through a workshop at the end of our study.

9.2 Data recording and record keeping

Once the data are linked, the TIGAR team will apply to the ONS Microdata Release Panel to release a pseudonymised extract of the data for the TIGAR researchers (see Section 5.1). This will be stored on NPEU secure servers, which are backed-up daily, and back-ups will be stored only within the University backup system, which is only accessible via fully authenticated, traceable means. Version control will be enforced, both for the data and for Stata 'do' files used in the analysis, and new iterations of the dataset will be saved with a clear filename and date time stamp.

Data retention will be in line with the agreements between ONS, NHS Digital and EDD for the use of their secure data. We will request that ONS, as data custodian, hold the linked database, as it does the pre-existing linked database so that other researchers can apply for access to the data, subject to obtaining the necessary permissions. We also envisage that the pseudonymised dataset may be held by the NPEU research team on secure servers. The exact terms and conditions will be negotiated during the course of the project.

10. ETHICAL AND REGULATORY CONSIDERATIONS

10.1 Declaration of Helsinki

The Investigator will ensure that this study is conducted in accordance with the principles of the Declaration of Helsinki.

10.2 Good Clinical Practice

The Investigator will ensure that this study is conducted in accordance with relevant regulations and with Good Clinical Practice.

10.3 Approvals

The protocol and lay summary have been approved by the Southwest - Frenchay Research Ethics Committee (REC) (reference number 15/SW/0294). The study will also require personal identifiable information for matching by a third party. Approval has also been granted by the Health Research Authority's Confidentiality Advisory Group (CAG), under Section 251 of the NHS Act 2006 to allow the common law duty of confidentiality to be set aside without individual patient consent for defined medical purposes which are in the public interest (CAG reference: 15/CAG/0196).

Following REC and CAG approvals, several data access and data sharing agreements will be required for the different stages of linkage (by a third party), and access to the data for checking and analysis (by the TIGAR researchers). Application will be made to ONS (the data custodian of the pre-existing dataset containing linked birth registration, NN4B and HES data) to use it subject to the necessary approvals. Application will also be made to NHS Digital for access to data from HES and the PDS, and to the Data Management Advisory Panel of the Education Data Division, which is the custodian of the school data.

Members of the TIGAR team who will be accessing individual record level data at the ONS VML will apply to ONS for Approved Researcher status. Two of the co-investigators (AM, ND) already have this status. Application will also be made to the ONS Microdata Release Panel to access the data in the VML and for release of pseudonymised datasets for analysis in secure facilities in Oxford.

The Investigator will submit and, where necessary, obtain approval from the above parties for any necessary substantial amendments to the original approved documents.

10.4 Reporting

The Chief Investigator shall submit throughout the study, or on request, an Annual Progress report to the REC Committee, the Confidentiality Advisory Group and to ONS plus the host organisation, Sponsor or Funder. In addition, an End of Study notification and final report will be submitted to the same parties.

10.5 Participant confidentiality

Data will be stored securely and accessible only by study staff and authorised personnel, as described in Section 9. The pseudonymised study datasets used for analysis will not contain any identifying information (names, addresses, postcodes, dates of birth, NHS or hospital numbers). It may be theoretically possible, with multiple cross-tabulations, to try to identify an individual child or family that has a rare combination of characteristics (e.g. gestation 24 weeks and multiple birth and minority ethnic group), but the TIGAR team will never seek to do this. All output tables will be checked prior to publication to ensure that they comply with applicable disclosure control guidance and policies.[39-43]

It has been the policy of the NPEU, as a Department-of-Health-funded Policy Research Unit since 1978, to hold research data in perpetuity where this is permitted. On completion of the study and following publication of the study results, datasets will be encrypted and archived in a secure electronic archive. Future access will be controlled by the data custodian (Director of the NPEU; currently Prof Jenny Kurinczuk) and would be subject to further regulatory approvals should access be required for any purpose other than that outlined in this protocol. The NPEU is covered by the University of Oxford Data Protection Act registration (registration number: Z575783X). However, the study data will be anonymised and non-identifiable within the terms of the Data Protection Act 1998.

10.6 Other ethical considerations

As previously mentioned in Section 7, it will not be possible for individual consent to be sought from parents or children whose data are included in this study. It will also not be possible to remove records for any individuals who might not wish their data to be used in our study. For these reasons, special approvals will be sought, as described in Section 10.3.

10.7 Other research governance

The project will be managed by the Co-investigator Group (CIG) which will include the chief investigator and co-investigators (as listed previously) together with the three researchers to be appointed. Three further groups have been formed to advise on the project:

- An **Advisory Group** includes relevant experts, voluntary organisations, (e.g. the support group 'BLISS for babies born too soon, too small, too sick') and representatives of families affected by preterm birth.

- A Patient, Parent and Public Involvement Group (**PPPI Group**) will provide service users' perspectives on all aspects of the project. Members will include families affected by preterm birth (e.g. parents of preterm children, and young adults who were born early).

- Another PPI group will include parents who are both affected by premature birth and who are also professionals with an interest in this area (**Parent-Professional Group**).

11. FINANCE AND INSURANCE

11.1 Funding

The study is funded by the UK Medical Research Council from April 2015 until June 2018 (Funder's reference MR/M01228X/1).

11.2 Sponsorship

The University of Oxford maintains Public Liability and Professional Liability insurance, which will operate in this respect.

12. PUBLICATION POLICY

The Investigators will be involved in reviewing drafts of the manuscripts, abstracts, press releases and any other publications arising from the study. Authors will acknowledge that the study was funded by the Medical Research Council. Authorship will be determined in accordance with the ICMJE guidelines and other contributors will be acknowledged.

The study PPPI Group and Parent Professional Group, and the Advisory Group (see Section 10.7) will comment on documents such as manuscripts and press releases, as appropriate, and will be involved in dissemination.

13. REFERENCES

1. Born too soon: the global action report on preterm birth. *WHO* 2012. http://www.who.int/ maternal_child_adolescent/documents/born_too_soon/en/.
2. Blencowe H, Cousens S, Oestergaard MZ et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet* 2012; 379: 2162–72

3. Gestation-specific Infant Mortality in England and Wales, 2012. Office of National Statistics Statistical Bulletin, October 2014.
4. Annual report of the chief medical officer 2012. Our children deserve better: prevention pays. October 2013.
5. Morse SB, Zheng H, Tang Y, Roth J. Early school-age outcomes of late preterm infants. *Pediatrics*. 2009;123:e622-629.
6. MacKay DF, Smith GC, Dobbie R et al. Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren. *PLoS Med* 2010;7:e1000289.
7. Quigley MA, Poulsen G, Boyle E et al. Early term and late preterm birth are associated with poorer school performance at age 5 years: a cohort study. *Arch Dis Child Fetal Neonatal Ed* 2012;97:F167–F173.
8. Peacock PJ, Henderson J, Odd D et al. Early school attainment in late-preterm infants. *Arch Dis Child* 2012;97:118-120.
9. Boyle EM, Poulsen G, Field D, Kurinczuk JJ, Wolke D, Alfirevic Z, Quigley MA. Population-based cohort study of the effects of gestational age at birth on health outcomes at three and five years of age. *BMJ* 2012;344:e896.
10. Paranjothy S, Dunstan F, Watkins WJ et al. Gestational age, birth weight, and risk of respiratory hospital admission in childhood. *Pediatrics* 2013;132;e1562
11. Eide MG, Oyen N, Skjaerven R et al. Associations of birth size, gestational age, and adult size with intellectual performance: evidence from a cohort of Norwegian men. *Pediatr Res* 2007;62:636-42.
12. Lindström K, Winbladh B, Haglund B et al. Preterm infants as young adults: a Swedish national cohort study. *Pediatrics* 2007;120:70–7.
13. Ekeus C, Lindström K, Lindblad F et al. Preterm birth, social disadvantage, and cognitive competence in Swedish 18- to 19-year-old men. *Pediatrics* 2010;125;e67-e73.
14. Moster D, Lie RT, Markestad T. Long-term medical and social consequences of preterm birth. *N Engl J Med* 2008;359:262–73.
15. Crump C, Sundquist K, Winkleby MA et al. Preterm birth and risk of epilepsy in Swedish adults. *Neurology* 2011;77:1376-82.
16. Johnson S, Hollis C, Kochhar P et al. Psychiatric Disorders in Extremely Preterm Children: Longitudinal Finding at Age 11 Years in the EPICure Study. J Am Acad Child Adolesc Psychiatry 2010;49:453-463.e1.
17. Johnson S, Hennessy E M, Smith RM et al. 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:*F283-F289.
18. Mathiasen R, Hansen BM, Nybo Andersen A-M N et al. Gestational age and basic school achievements: a national follow-up study in Denmark. *Pediatrics* 2010;126:e1553–1561.
19. Feinstein L, Duckworth K. Development in the early years: its importance for school performance and adult outcomes. London: Centre for Research on the Wider Benefits of Learning 2006.
20. Duncan GJ, Dowsett CJ, Claessens A et al. School readiness and later achievement. *Dev Psychol* 2007;43:1428-46.
21. Fair Society, Health Lives. The Marmot Review. Strategic review of health inequalities in England post-2010.
22. Potijk MR, Kerstjens JM, Bos AF et al. Developmental delay in moderately preterm-born children with low socioeconomic status: risks multiply. *J Pediatr* 2013;163:1289-95.
23. Murray J, Bottle A, Sharland M et al. Risk factors for hospital admission with RSV bronchiolitis in England: a population-based birth cohort study. *PLoS ONE* 9:e89186.

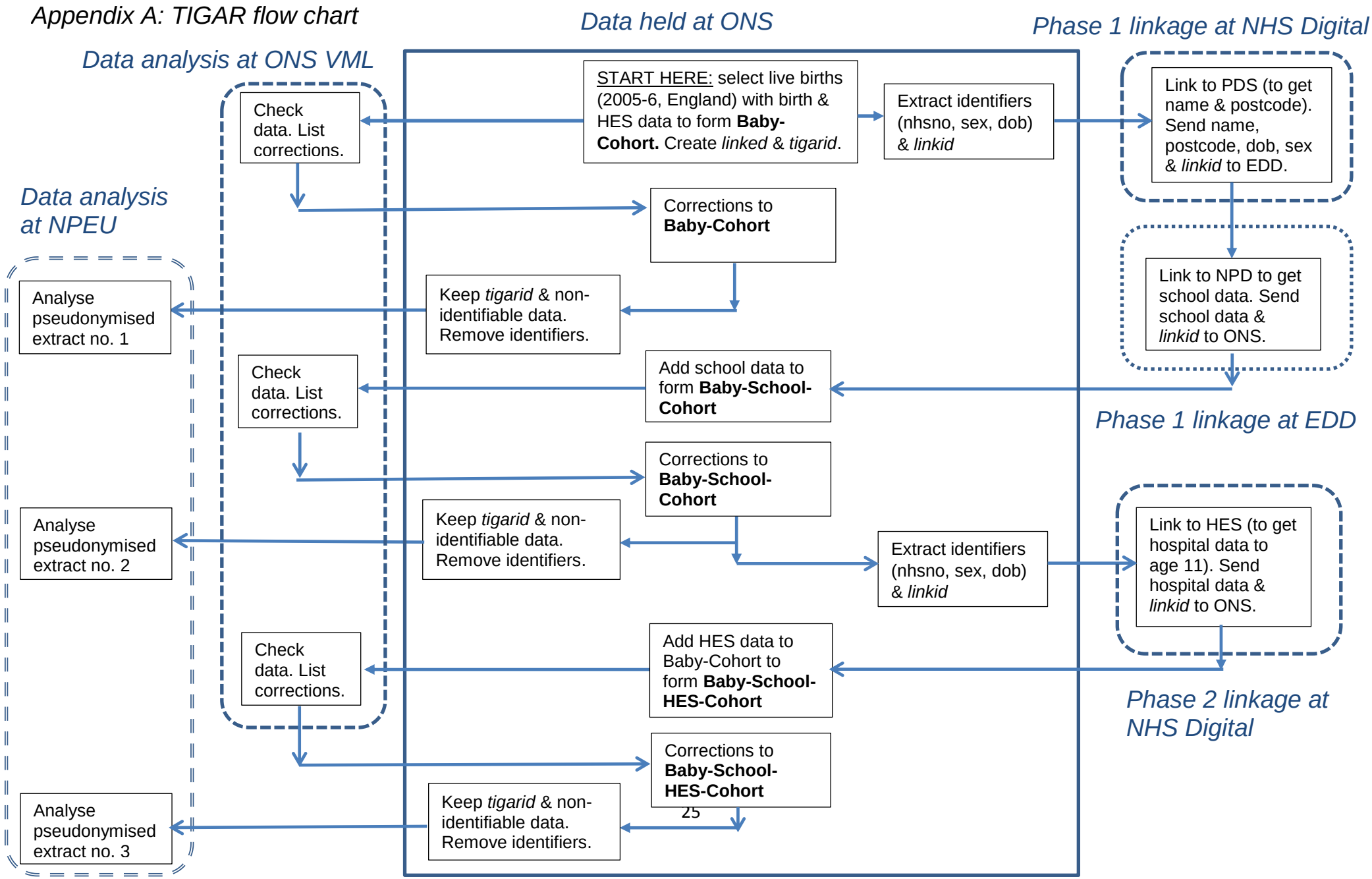
24. Lindström K, Lindblad F, Hjern A. Preterm birth and attention-deficit/hyperactivity disorder in school children. *Pediatrics* 2011;127:858–65.
25. Crump C, Winkleby MA, Sundquist K et al. Risk of diabetes among young adults born preterm in Sweden. *Diabetes Care* 2011;34:1109-13.
26. Petrini JR, Dias T, McCormick MC et al. Increased risk of adverse neurological development for late preterm infants. *J Pediatr* 2009;154:169-176.e3.
27. Lipkind HS, Slopen ME, Pfeiffer MR et al. School-age outcomes of late preterm infants in New York City. *Am J Obstet Gynecol* 2012;206:222.e1–6.
28. Noble KG, Fifer WP, Rauh VA et al. Academic achievement varies with gestational age among children born at term. *Pediatrics* 2012;130:e257–264.
29. Poulsen G, Wolke D, Kurinczuk JJ et al. Gestational age and cognitive ability in early childhood: a population-based cohort study. *Paedr Perinat Epidemiol 2013;27:371-9.*
30. Mangham LJ, Petrou S, Doyle LW, Draper ES, Marlow N. The cost of preterm birth throughout childhood in England and Wales. *Pediatrics* 2009;123:e312-327.
31. Petrou S. The economic consequences of preterm birth during the first 10 years of life. *BJOG* 2005;112 Suppl 1:10-15.
32. Dattani N, Datta-Nemdharry P, Macfarlane A. Linking maternity data for England 2005-6: methods and data quality. Health Statistics Quarterly 2011;49:1-27.
33. Dattani N, Datta-Nemdharry P, Macfarlane A. Linking maternity data for England 2007: methods and data quality. Health Statistics Quarterly 2013;53:1-18.
34. Henderson D, Beer C, Wolke D et al. Supporting the schooling of very preterm children: education professionals' opinions and information needs. *Arch Dis Child* 2012; 97:Suppl 2:A353.
35. Hansen K, Joshi H, Dex S [Ed]. Children of the 21st century. The first five years. Ed Hansen K, Joshi H, Dex S. The Policy Press 2010.
36. Brown HK, Speechley KN, Macnab J, Natale R, Campbell MK. Mild prematurity, proximal social processes, and development. Pediatrics. 2014;134:e814-24.
37. Health and Social Care Information Centre. (2013). HRG4 2013/14 Local Payment Grouper. Available from http://www.hscic.gov.uk/article/2580/HRG4-201314-Local-Payment-Grouper. [Accessed 19 May 2013]
38. Jones, A. Panel Data Methods and Applications to Health Economics, in Palgrave Handbook of Econometrics Volume II: Applied Econometrics, T.C. Mills and K. Patterson, Editors. 2011, Palgrave MacMillan: Basingstoke.
39. Office for National Statistics. Briefing Note: ONS policy on protecting confidentiality within birth and death statistics (revised 2010). Office for National Statistics 2010.
40. Office for National Statistics. Disclosure control guidance for birth and death statistics. Briefing Note (revised Jan 2014) on the publication of tabular data. Office for National Statistics 2014.
41. Government Statistical Service. GSS/GSR Disclosure Control Guidance for Tables Produced from Administrative Sources. Government Statistical Service 2014.
42. Information Standards Board for Health and Social Care. Anonymisation Standard for Publishing Health and Social Care Data 2013 (http://www.isb.nhs.uk/library/standard/128).
43. Information Commissioners Office. Guide to data protection. Anonymisation: managing data protection risk code of practice. 2015 (http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation.)

14. APPENDICES

AMENDMENT HISTORY

Amendment No.	Protocol Version No.	Date issued	Author(s) of changes	Details of Changes made
1	1.1	24th Feb 2016	Maria Quigley	Staff changes: added Alex Lloyd (new investigator) and Brian Hicks (new Head of IT at NPEU). Changed order of linkage and simplified the process. Added references for approvals from REC and CAG.
2	1.2	4th Mar 2016	Maria Quigley	When describing record linkage, have clarified that we will access 'postcode' and not 'address'.
3	1.3	15th May 2017	Maria Quigley	Have specified that we will access data on age at death (in months). Staff changes: removed Alex Lloyd (no longer involved) and added Victoria Coathup (new investigator). Have changed linkage process to clarify that data will be checked at ONS VML at 3 separate times (not 2). Have replaced all occurrences of 'HSCIC' with 'NHS Digital'.

Appendix A: TIGAR flow chart



EDD Education Data Division; NHS Digital was formerly HSCIC; NPEU National Perinatal Epidemiology Unit; ONS Office for National Statistics; PDS Personal Demographics Services