

Evaluating the real-world implementation of the Family Nurse Partnership in England: a data linkage study

Protocol Version 1.11

July 2019

IRAS ID: 230351

Chief investigator: Katie Harron

Killen

Date: 08.07.19

Co-investigators: Ruth Gilbert (University College London), Jan van der Meulen (London School of Hygiene & Tropical Medicine), Eilis Kennedy (Tavistock and Portman NHS Foundation Trust), Emma Howarth (University of Cambridge), Sally Kendall (University of Kent)

Collaborators: Ailsa Swarbrick (FNP National Unit), Mike Robling (University of Cardiff), Rebecca Cannings-John (University of Cardiff), Fiona Lugg-Widger (University of Cardiff), Erin Walker (UCL Partners)

Study Steering Committee chair: Jane Barlow (University of Oxford)

Funding: This study/project is funded by the National Institute for Health Research (NIHR) Health Services Delivery & Research (project reference 17/99/19). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Contents

1	Sun	nmary	. 3
2	Bac	kground	.4
	2.1	What is the problem being addressed?	.4
	2.2 patier	Why is this research important in terms of improving the health of the public and/or to its and the NHS?) .4
3	Aim	s and Objectives	.5
4	Stu	dy design, study period, population and data sources	.6
	4.1	Hospital records: Hospital Episode Statistics (HES)	.8
	4.2	Education and social care records: National Pupil Database	.8
	4.3	FNP Information System (FNP IS)	.9
5	Ana	Ilysis	.9
	5.1	Outcomes	.9
	5.2	Analysis strategies	10
	5.3	Sample size	11
	5.4	Limitations	11
6	Par	ticipant involvement	11
7	Gov	vernance	12
	7.1	Study oversight	12
	7.2	Approvals	12
	7.3	Data flows	13
	7.4	Data Protection	16
8	Diss	semination and output	17
	8.1	Plans for disseminating the findings of the research	17
	8.2	Expected outputs	17
	8.3	Expected measureable benefits to health and/or social care	17
	8.4	Promoting the education and well-being of children in England	18
	8.5	Benefits to researchers	18
9	Ref	erences	19

1 Summary

More than 20,000 babies are born to teenage mothers in England each year (~4% of births). Teenage mothers face a number of challenges in pregnancy (including lower levels of education, less stable careers and lower income than older mothers), which can lead to worse outcomes for mothers and their children. Whilst teenage motherhood can be a positive experience for some, early pregnancy is also related to domestic violence, less engagement with education and employment, and rapid repeat pregnancies. Unhealthy behaviours during pregnancy and inadequate prenatal care can lead to adverse neonatal outcomes, and children born to teenage mothers are at greater risk of maltreatment and associated long-term consequences, including adverse physical, social, emotional and cognitive outcomes, and depression, anxiety and suicidal behaviour.

Pregnancy and the postnatal period therefore provide important opportunities for supporting teenage mothers and their children. In the US, a home visiting intervention was shown to benefit mothers and their children, notably through improved health during pregnancy, decreased child maltreatment and injuries, improved school readiness, and greater maternal workforce participation. In England, the Family Nurse Partnership (FNP) was introduced in 2007, aimed primarily at first time teenage mothers. A trial of 1600 families in England found no evidence of benefit from FNP on birth weight, maternal smoking, subsequent pregnancy or A&E attendances. However, the trial did find evidence of benefit on language and cognitive development up to age 2, and, as benefits are recognised locally, the programme is still delivered across the UK. By 2016, 25,000 mothers had been enrolled in FNP across 125 Local Authorities (covering around 25% of eligible mothers), and, learning from the trial, the service is evolving and adapting over time. Understanding the context in which the intervention is now delivered, and the factors that may influence how well it works (e.g. the local setting, and the types of families receiving FNP), is key to understanding its value.

This study aims to evaluate how the real-world implementation of FNP varies across England. The research will complement results from the trial, by providing a more detailed assessment of whether there are particular settings in which FNP works well. To do this, we will use electronic records that are routinely collected as part of health, education, and social care services to compare outcomes for FNP participants with similar families who did not take part in FNP. We discussed our proposal with mothers participating in FNP, and they were supportive of the study.

We will look at outcomes for children (e.g. emergency hospital admissions for possible neglect or abuse, development at school-age, referrals to social services) and mothers (e.g. continuing education after birth, subsequent pregnancies, hospital admissions due to violence or injuries). Exploring whether FNP works better for some families (e.g. the youngest teenagers) than others will help improve targeting of resources and highlight groups in need of alternative support. Findings from the study will help policy-makers decide whether FNP should be offered to families in their local setting. Evidence generated by this study will support commissioners in providing improved services for mothers and children who could benefit most, and lead to increased efficiency through more effective targeting of resources.

2 Background

2.1 What is the problem being addressed?

The Family Nurse Partnership (FNP) is a home visiting programme from early pregnancy until the child's second birthday that aims to improve outcomes of teenage pregnancies. Following evidence of benefit on a number of maternal and child outcomes in the US, FNP was introduced in the UK in 2007.[1-3] An RCT of 1618 families recruited in 2009-2010 in England (Building Blocks; BB) showed no benefit of FNP on early outcomes (smoking at late pregnancy, birth weight, subsequent pregnancy, or child A&E and hospital attendances).[4] However, trial results were hotly debated due to a high number of usual care visits in the control arm, the choice of early child development as a secondary outcome (for which a positive effect was observed), and strong support locally for the value of FNP on parent-child relationships.[5-7] There remains significant commitment to delivery of FNP as part of the Healthy Child Programme.

MRC guidance on evaluating complex interventions highlights the importance of understanding the context in which an intervention is delivered.[8] Understanding the mechanisms and drivers of change is also critical to determining why effects of complex interventions differ across contexts and populations.[9] These issues are particularly relevant to evaluating FNP.

First, the comparison of usual care may be inconsistent across areas (and between trial and target populations), due to differences in organisation of local services.[7] This may partly explain different results of trials conducted in the US, Holland, and England, but is also relevant to comparisons over time and between LAs within England.[5] Service providers also report variation between LAs and over time in how FNP is delivered and targeted.[10] Understanding this variation is key for quality improvement.[9]

Second, research examining drivers of change in child development in the FNP population demonstrates that effects are concentrated in mothers with low education, but highlights the need for further research to examine other behaviours such as drug/alcohol abuse and domestic violence.[11] This supports evidence from the US showing that the youngest, most disadvantaged mothers are most likely to benefit from FNP.[5] Subgroup analyses in RCTs are insufficient for capturing how mediating factors vary according to local context.[7]

Our proposal addresses variation in the delivery and targeting of FNP in practice by generating evidence on the real-world implementation of FNP, and the factors that may influence effectiveness (including setting, provider, participants and programme delivery).[12, 13] Longitudinal observational data for the whole population eligible for FNP will be used to evaluate outcomes for eligible families who did or did not receive FNP, and to explore variation in the populations and the intervention across the country, and compared with BB. Understanding the context in which the programme is now delivered and the factors that may influence results, is key to understanding who might benefit most from FNP and how service delivery may be improved.

2.2 Why is this research important in terms of improving the health of the public and/or to patients and the NHS?

Adverse maternal and child outcomes of teenage pregnancy, due to social adversity, disruption to education and employment, and child-rearing practices, are of major importance to the NHS.[14, 15] In England, approximately 4% of babies each year (22,465 in 2016) are born to mothers aged <20. These mothers are more likely to experience domestic violence, be less engaged with education and employment, and have rapid repeat pregnancies. Young maternal age is associated with a higher burden on services throughout childhood, and a greater risk of child maltreatment and associated

long-term consequences including adverse physical, social, emotional and cognitive outcomes.[16, 17] Programmes such as FNP are therefore likely to remain a priority for the NHS.[18]

Understanding how best to target services to the most vulnerable mothers is key to improving health for mothers and their children. Study results will help improve targeting of support during and beyond pregnancy, by providing evidence on who FNP currently targets and who is most likely to benefit. This evidence is being called for by service providers who need to understand the value of interventions in the context of their target populations and local services, in order to inform commissioning and justify spending.[13] Findings will help divert resources to those most in need by identifying where gaps in coverage exist.

There is a need to establish a scalable system for monitoring and evaluation of intensive support for young and vulnerable mothers, taking into account changes over time and any cuts to services within LAs. Linkage of existing administrative records provides a cost-efficient means of evaluating services as they are implemented in the real world.

This research is timely, coinciding with reporting of results from the BB follow-up trial (BB2-6) in 2018. Our population-based study will help generalise results from BB to practice, but BB results will also be used to validate our observational evidence. Evaluating outcomes for up to 25,000 FNP families will provide increased statistical power to detect smaller differences, differences in rarer outcomes, and subgroup differences for which BB is underpowered. Our study will help inform ongoing adaptations to the FNP, and inform commissioning by generating evidence on whether the real-world implementation of FNP benefits those most in need.

3 Aims and Objectives

This study aims to generate evidence on the real-world, ongoing implementation of FNP in England. The study will build upon results from BB and BB2-6, and a planned evaluation of FNP in Scotland using linked administrative data, by answering the following questions:

1. Are mothers who are most in need of support receiving FNP in all LAs?

Objectives:

- Describe variation in delivery of FNP and usual care across LAs to identify support that vulnerable families receive, including information from ADAPT sites that have changed eligibility/delivery of the programme
- Describe variation in health, education and social care characteristics of families participating in FNP over time and by LA; compare characteristics of FNP participants with families who met eligibility criteria but did not enrol, to understand which mothers are prioritised across different LAs
- Explore individual and LA-level predictors of engagement (number of valid visits), to determine whether some families are more likely to engage with FNP than others
- 2. Which families does the real-world implementation of FNP benefit?

Objectives:

- Evaluate the effectiveness of FNP on a broad range of health, education and social care outcomes for both children and mothers, selected according to logic models for FNP and restricted to those captured in administrative data [4, 19, 20]
- Determine which families stand to benefit from FNP using detailed information on maternal trajectories prior to pregnancy (e.g. special educational needs (SEN), out-of-home care, chronic conditions)

- Evaluate outcomes for groups now eligible for FNP but excluded from BB (e.g. mothers up to age 24)
- 3. What are the contextual factors associated with benefiting from FNP?

Objectives:

- Explore the effect of contextual factors such as usual care models, nurse characteristics and programme content covered, to understand the impact of how the programme is delivered
- Determine how the effect of FNP differs between LAs and compared with the BB population, to help generalise trial results

4 Study design, study period, population and data sources

We will establish a linked dataset of hospital, education and social care records for all mothers aged ≤24 years, and their children born in England between 2010-2017. We will link to FNP programme data to identify participants. The cohort will be created through linkage of hospital records (Hospital Episode Statistics; HES), education and social care records (National Pupil Database; NPD), and FNP programme data. Our approach builds on existing experience of linking education and health records, and validated methods of linking hospital records for mothers and babies.[21, 22]

An example of linked data is presented in Figure 1. The time periods covered by the data sources are described in Figure 2. Differing lengths of follow-up (e.g. for hospital records) and look-back periods (e.g. for social care records; CLA and CIN) will be taken into account within analyses.



Figure 1: Example linked data for FNP mother and child

	Data									Year	of data								
	source	2001	2002	203	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
	Births																		
	FNP																		
	HES																		
	NPD																		
	CLA																		
	CIN																		
	1986	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
	1987	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
	1988	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	1989	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
	1990	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
	1991	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
	1992	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
	1993	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
	1994	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
	1995	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	1996	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	1997	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	1998	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	1999	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Year of	2000	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
birth of	2001	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
cohort	2002		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
member	2003			0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	2004				0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	2005					0	1	2	3	4	5	6	7	8	9	10	11	12	13
	2006						0	1	2	3	4	5	6	7	8	9	10	11	12
	2007							0	1	2	3	4	5	6	7	8	9	10	11
	2008								0	1	2	3	4	5	6	7	8	9	10
	2009									0	1	2	3	4	5	6	7	8	9
	2010										0	1	2	3	4	5	6	7	8
	2011											0	1	2	3	4	5	6	7
	2012												0	1	2	3	4	5	6
	2013													0	1	2	3	4	5
	2014														0	1	2	3	4
	2015															0	1	2	3
	2016																0	1	2
	2017																	0	1

Figure 2: Time periods covered by data sources. Shaded boxes indicate cohort member age during study intake (blue); look-back period (grey); follow-up (light blue).

4.1 Hospital records: Hospital Episode Statistics (HES)

Hospital records for mothers aged ≤24 years and their children in England will be extracted from records of births and deliveries in HES. HES is a data warehouse containing details of all admissions (from 1997), outpatient appointments (from 2003) and A&E visits (from 2010) at NHS hospitals in England. HES data have been used extensively in research. Information captured in HES includes administrative data (e.g. admission dates, NHS trust, GP code), demographic information (e.g. age, sex, ethnicity), and clinical information (diagnoses and procedures). A unique 'HESID' is assigned to enable episodes of care for the same individual to be combined. Diagnoses are coded by professional coders in hospitals using ICD-10 codes (International Classification of Disease, version 10); procedures are coded using OPCS-4 codes (Office of Population, Censuses and Surveys Classification of Surgical operations and procedures, version 4).[23, 24] Date of death will be obtained via linkage between HES and ONS mortality data.

The HES cohort will include mothers aged ≤24 years and their children, based on birth and delivery records within an existing HES extract held by the research team. This existing dataset will be used to derive a number of variables that will subsequently be used to match FNP participants to similar families, or used as variables in the analysis. A full list of variables is provided in Appendix 1. The use of this existing dataset avoids the release of any new data, but will require a change of purpose in the original permissions for data processing.

Inclusion criteria:

• All mothers aged 13-24 years delivering a live birth in England between April 01 2010 and March 31 2017 (and their children)

4.2 Education and social care records: National Pupil Database

The HES cohort will be linked to education and social care records from the National Pupil Database (NPD). Information on assessments, attainment and progression at each key stage is available for all pupils in state schools in England, alongside eligibility for free school meals, information about special educational needs and information about absences and exclusions. NPD data, including social care data, have been used extensively in research.[25]

For both mothers and their children, we will link to:

- The School Census (formerly PLASC), the Pupil Referral Unit (PRU) Census, and Alternative Provision: pupil-level information from 2002 for pupils aged 2-19+ (i.e. for both FNP mothers and children). Includes special educational needs, free school meal eligibility and information about absences and exclusions.
- The Children Looked After (CLA) return: information on all looked after children and recent care leavers in England, from 2005.[26, 27]
- The Children in Need Census (CINC): information on all referrals to children's social services, assessments carried out upon these children, and whether the children became the subject of a child protection plan, from 2008.[28]

For mothers only, we will link to:

• Key Stages 4-5 data: teacher assessments and/or test results taken in Year 11 and 12/13 (i.e. ages 16, and 17-18).

For children only, we will link to:

• The Early Years Census and Early Years Foundation Stage Profile (EYFSP): teacher assessments for three and four year olds (from 2008), and two year olds (from 2014).

A Unique Pupil Number (UPN) is used for linkage of CLA and CINC with the NPD. The UPN is usually assigned at first entry to a maintained school or nursery, typically around the age of four or five. Therefore it is not possible to link the NPD to CLA data for children who are looked after only before they enter nursery/school, or adopted children (who are provided with a new UPN).[29]

4.3 FNP Information System (FNP IS)

The HES cohort will be linked to the FNP IS to obtain information on participation in FNP. The FNP IS is a bespoke system that supports the implementation of the FNP programme in England, provided by NHS Digital under contract to the FNP NU on behalf of the Department of Health and PHE. Data are reported in real time and are used locally by FNP teams and nationally by FNP NU to monitor programme delivery and support quality improvement.

Data collected in the FNP IS include information from the mother and child collected at programme intake (e.g. mother's age, marital status, living arrangements, education, employment, social care), 36 weeks pregnancy (e.g. maternal health, alcohol, drugs and smoking), birth (e.g. baby's birth weight, gestational age) and at regular intervals until 24 months after birth (including child health and development, social care and other maternal baseline variables). Information on each visit is also collected (e.g. date, length of visit, family nurse seen, referrals to other services, etc.). The FNP IS become functional in 2009, and data quality is reported to be sufficiently high from 2010 onwards. FNP data have been used in previous research.[11]

FNP IS also contains maternal and child identifiers at enrolment/birth: name, sex, date of birth, postcode, GP code and NHS number. When mothers graduate from the FNP, data are pseudonymised and identifiers are retained solely on secure servers at NHS Digital Exeter. Since some identifiers might have changed since enrolment (e.g. mother's name, postcode), identifiers will first be updated using the Personal Demographic Service within NHS Digital, so that the most relevant set of identifiers can be used for linkage.[30, 31]

5 Analysis

5.1 Outcomes

Children up to age 6:

HES:

- Child abuse or neglect: the percentage of children with unplanned hospital admissions for any injury or maltreatment-related diagnosis (using ICD10 code lists)[32, 33]
- Unplanned hospital admissions
- A&E visits
- Referrals to outpatient departments

NPD:

- A "good level of development", as defined by the Department for Education
- School absences
- Child ever in care
- Child in need status

Mothers:

HES:

- A&E visits and/or unplanned hospital admissions, including for violence, self-harm, or drug/alcohol abuse

- Subsequent live or still births resulting from pregnancies within 18/24 months of index birth
- NPD:
 - Qualifications before versus after birth (GCSE, GNVQ, AS/A level)

5.2 Analysis strategies

We will use our cohort data to provide robust measures of coverage, by identifying the proportion of eligible families participating in FNP by LA and over time. Predictors of participation will be derived from HES and NPD data. Since recruiting practices vary locally, we will compare data from high- and low-coverage LAs, to explore whether FNP families in low-coverage LAs are more highly selected than those within LAs offering the programme to more mothers. This will enable us to identify which groups are prioritised as coverage decreases. For example, it may be that nurses in some areas target the most advantaged mothers, whereas in other areas, a first-come-first-served model is used. We will take into account local data on the number of mothers receiving recommended health visits within each LA, and on health visits specifically for teenage mothers.[34]

We will compare outcomes for mothers ever enrolled in FNP versus those who were never enrolled. Two analysis strategies will be used to take account of measured confounders related to both participation in FNP and outcomes: i) propensity score matching; ii) adjusted analyses.

i) Propensity scores

To derive propensity scores, we will regress FNP participation on all available maternal characteristics, e.g. pre-pregnancy chronic conditions, looked after status, educational achievement. Matched groups will be formed based on the propensity of participation.[35] Effects will be estimated as the difference in outcomes between matched groups. Statistical models will allow for clustering of families within LAs, and multiple imputation will be used to account for missing data. The main analysis will restrict matching within the same LA and within the time periods in which FNP was offered within that LA. Secondary analyses aiming to achieve more closely matched groups (with potentially smaller numbers) will match i) within the same LA but in different time periods, comparing outcomes for eligible families before vs. after FNP was offered; and ii) within the same time period but in different LAs, comparing outcomes for eligible families in LAs that did and did not offer FNP.

ii) Adjusted analyses

This analysis will be an unmatched comparison, adjusting for maternal variables (e.g. pregnancy complications, ethnicity, Index of Multiple Deprivation), neonatal variables (e.g. gestational age, birthweight, length of postnatal hospital stay, season of birth, congenital anomalies, admission to NICU), and education variables (e.g. SEN, free school meals, type of school).

Sensitivity analyses will determine the strength of unmeasured confounding required to invalidate results.[36, 37] To further assess the robustness of findings to the analysis approach and to evaluate any potential differences in results due to the use of real-world data, we will use our cohort to replicate findings observed in BB. For each analysis strategy, we will derive trial outcomes for a group of families in the administrative data cohort with the same aggregate baseline characteristics as trial participants.

The following variables will be explored through stratified analyses or interaction terms:

Maternal (based on previous evidence suggesting the most disadvantaged and youngest mothers are most likely to benefit):

- Year of age
- Health care use prior to pregnancy (unplanned admissions and A&E visits; admissions for drug/alcohol abuse, violence, or self-harm)[38]
- Chronic conditions (including mental health conditions captured in hospital data)[39]
- Educational achievement
- SEN
- Out-of-home care
- Subsequent children
- Deprivation (Index of Multiple Deprivation of maternal postcode)

Programme:

- Engagement (e.g. length and number of visits)
- Delivery (e.g. programme content covered)
- Number of referrals to other services
- Nurse characteristics
- Year
- LA
- Models of usual care

5.3 Sample size

Exploratory analysis of HES data suggests that approximately 9.5% of children born to first time teenage mothers are admitted to hospital for injuries by age 6. Assuming a baseline rate of 9.5% for the percentage of children with the primary outcome up to age 5, a total of 6798 children (3399 in each group) would be required to observe a significant reduction of 20% in the primary outcome, with 80% power and a 5% significance level. In our cohort, birth outcomes will be available for around 25,000 FNP births; 19,000 FNP families will have a full 3 years of follow-up and 6000 will have a full 6 years of follow up.

5.4 Limitations

We do not have individual-level data on primary care, or the number of health visits that families receive outside of FNP, but will obtain local aggregate data on health visitors and usual care models over time. We will explore whether outcomes in families receiving usual care have changed over time, and whether there were any knock-on effects to other families, of the FNP being introduced/decommissioned within an LA. For example, the FNP might divert resources away from other families, or alternatively, knowledge of the programme may influence practice for other health visitors.

Data on mothers who were offered FNP but declined is not routinely collected. To address this, we will obtain local data, where available, on the rate and characteristics of declining mothers.

6 Participant involvement

We held a structured workshop to engage with FNP participants and FNP nurses, run by an experienced facilitator and incorporating visual presentations, open discussion, and interactive sessions. Four mothers and one family nurse attended, at a children's centre in Southwark. We discussed the use of administrative data for research, which involves concepts that mothers were

not previously familiar with (such as de-identification and data linkage). We also discussed linkage of health and education data without explicit consent, and the use of these data specifically for evaluating the FNP.

Workshop participants were supportive of the proposed study and fed-back that they thought that linking health and education data for mothers and their children was a good idea. We had a long discussion about the distinction between NHS data and commercial data. Participants voiced strong opinions about wanting to share their data so that services could be improved and so that future mothers could benefit. However, there was also a strong feeling that participants wanted to know how their data had been used to benefit others - and we discussed different methods of dissemination, e.g. through social media.

Several of the participants were keen to continue to be involved in the research study, and we are planning to include at least one mother on the Study Steering Committee. We also plan that participants will help develop information on the study for the FNP and institutional websites, and help with interpretation of results and planning the most appropriate methods of dissemination.

We also conducted a public engagement event with mothers who did not participate in FNP (but who are representative of mothers who may be included in the study). Participants included mothers who had their child(ren) at a range of ages from 21 to 30. We discussed the FNP as it is currently delivered, the research project, and the data, outcomes, and need for evidence to underpin service provision. Mothers expressed a range of opinions about the likely effectiveness of FNP: some thought it was common sense that more contact with health visitors would lead to better education of mothers and improved outcomes, whilst another (based on their experience with a sister who was a teenage mother) thought that being given additional advice from a healthcare professional wouldn't make any difference to mothers' behaviours. There were strong opinions about the importance of taking into account maternal demographics such as education and area, and whether or not the wider family was on board with the programme, in order to understand whether the programme worked. These opinions have been incorporated into our proposal: we will evaluate whether the effect of FNP differs by area, and by maternal characteristic.

7 Governance

Funding applications are currently in progress for this study. The NIHR Research Design Service has independently reviewed the study protocol. The study is sponsored by UCL, who provide indemnity insurance.

7.1 Study oversight

The study will be overseen by a Study Steering Committee (SSC), which will include an independent chair and two parent representatives. The SSC will provide overall supervision and ensure the study is conducted to rigorous standards, and will meet at the project start and then at least once per year. The day-to-day management of the study will be the responsibility of the Principal Investigator, who will lead a separate Project Management Group that will meet at least 6 times per year.

7.2 Approvals

Access to identifiable data will be required for the purposes of linkage. As described in Figure 2, identifiers will be separated from attribute data: linkers will have access only to identifiers, and UCL researchers will only have access to the attribute data within a linked, pseudonymised dataset.

At enrolment to FNP, participants are asked for explicit (but not written) consent. This consent covers the collection of personal identifiable data to be used for programme management and research purposes, and storage on a national database and within clinical records. Mothers who

originally dissented to the use of their data are not included in the FNP cohort. However, consent for linkage with other datasets is not currently sought explicitly. Seeking consent for linkage from approximately 25,000 mothers enrolled in FNP since 2010 would not be feasible, would not always be possible without further disclosure (a need to obtain up to date address details), and would involve a disproportionate financial cost. Furthermore, a requirement for consent could introduce substantial bias into analyses, as the most vulnerable mothers may have a high non-response rate.

Therefore, we will seek support from the Confidentiality Advisory Group for the processing of identifiable data for this study (Section 251). Approval for linkage with education data will be sought from DfE under Schedule 2 of the Data Protection Act. Support for this study has been obtained from the National Research Ethics Council (ref 18/EM/0014).

7.3 Data flows

The initial data flows described here relate to linkage of FNP IS and HES data. These data flows will be updated in a revised protocol to cover linkage with NPD.

De-identified FNP programme data along with a study ID will be transferred to the research team and stored within a secure setting - the UCL Data Safe Haven

(https://www.ucl.ac.uk/isd/itforslms/services/handling-sens-data/tech-soln). HES data required for analysis will be extracted from an existing HES cohort held by the research team to a new server within the UCL Data Safe Haven, with separate access rights. Identifiers for FNP participants will be updated by NHS Digital prior to linkage with HES. A link-key will be created and returned to the research team, to allow merging of de-identified HES and FNP data within the secure setting (Figure 3). The data flows are summarised as follows:

- A. FNP data
 - 1. Identifiers for FNP mothers and their first child are currently held by NHS Digital via Open Exeter. These identifiers will be updated via PDS within NHS Digital.
 - 2. The pseudonymised FNP study ID, plus identifiers (NHS number, GP code, name, sex, date of birth, postcode), will be made available for linkage within NHS Digital (FNP programme data will be held separately from the identifiers).
 - 3. FNP programme data with the FNP study ID but no identifiers will be transferred separately to the secure setting.
- B. HES data
 - 1. A HES cohort of mothers and babies will be prepared by the researchers based on an existing dataset held by the research team at UCL. HESIDs for these records will be transferred to NHS Digital.
 - 2. A limited version of the HES cohort, containing HESID and a number of analysis variables, will be extracted from an existing HES cohort held at UCL (DSA NIC-393510-D6H1D) to a new server within the secure setting.
 - 3. NHS Digital will extract identifiers for the list of HESIDs (sex, date of birth, NHS number) and updated identifiers from PDS (name, postcode, GP code) and use these for linkage with the FNP cohort.
 - 4. A pseudonymised link-key will be transferred from NHS Digital to the secure setting.
- C. Secure setting

1. The link-key will be used to merge the de-identified FNP programme data with the HES analysis variables within the secure setting. Identifiers will not be held in the secure setting.

Figure 3: Data flow diagram. Data items in italics are identifiable.



7.4 Data Protection

Our study will adhere to the eight data protection principles (Table 1).

Table 1: Data Protection Principles

1. Fair processing Privacy notices will be used to give information to individuals whose personal data may be processed, stating the purpose for which data be processed, and including a mechanism for opting out. For FNP participants, notices will be posted on the FNP website, and information to account the study will also be discominated via social media forward.	se :a will
personal data may be processed, stating the purpose for which data be processed, and including a mechanism for opting out. For FNP participants, notices will be posted on the FNP website, and informa- chaut the study will also be discominated via social media forums.	a will
be processed, and including a mechanism for opting out. For FNP participants, notices will be posted on the FNP website, and informa- about the study will also be discominated via social media forums.	ation
participants, notices will be posted on the FNP website, and information of the study will also be discominated via social modia forums.	ation
about the study will also be discominated via social modia forums	lation
about the study will also be disseminated via social media forums	
currently used to engage with FNP participants. An email address w	vill be
provided for opt-out, which will be administered by NHS Digital Exer	eter
who currently provide storage of FNP identifiers. For individuals who	10
have not participated in the ENP it will not be possible to make dire	ect
contact. However, we will publicise the study on our institutional	
websites (e.g. www.ucl.ac.uk/child-health/research/nonulation-noli	licy-
and-practice/research/research-approaches/electronic-health-reco	ards-
and practice/research/research approaches/electronic reactive received and practice/research approaches/electronic received and practice/	
and). Information posted on these websites will include details of the	10 W
patients within England can opt out of their personal confidential	
Information being shared by NHS Digital for purposes other than the	ieir
own direct care (Type 2 opt-out).	
2. Used for specified The data requested will only be used for the purposes set out in this	IS
purposes protocol.	
	mited
3. Minimum necessary We request only those variables necessary for analysis, and have lin	
3. Minimum necessary We request only those variables necessary for analysis, and have lin for the purpose the sensitivity of our request by asking for month and year rather the	han
3. Minimum necessary We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fire full date of birth or death and for IMD rather than postcode.	han inal
3. Minimum necessary for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather th full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see	han inal
3. Minimum necessary We request only those variables necessary for analysis, and have line for the purpose for the purpose the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fire analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data	han inal ataset.
3. Minimum necessary for the purposeWe request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather th full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the date4. AccuracyWe will request identifiers to be updated via PDS prior to linkage.	han inal ataset.
3. Minimum necessary We request only those variables necessary for analysis, and have line the purpose for the purpose We request only those variables necessary for analysis, and have line the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fire analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data 4. Accuracy We will request identifiers to be updated via PDS prior to linkage. 5. Kept for minimum Identifiable data will be destroyed following linkage. In accordance or linkage.	han inal ataset. with
3. Minimum necessary for the purposeWe request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather th full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data 4. Accuracy4. AccuracyWe will request identifiers to be updated via PDS prior to linkage. Identifiable data will be destroyed following linkage. In accordance of University research data policy, research data will be kept for ten year	han inal ataset. with ears
3. Minimum necessary for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fin analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data 4. Accuracy 4. Accuracy We will request identifiers to be updated via PDS prior to linkage. 5. Kept for minimum time necessary Identifiable data will be destroyed following linkage. In accordance of after the end of the study.	han inal ataset. with ears
 Minimum necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data Accuracy We will request identifiers to be updated via PDS prior to linkage. Identifiable data will be destroyed following linkage. In accordance with Individuals will not be identified in the anonymised linked dataset updated via PDS prior to the study. 	han inal ataset. with rears used
 Minimum necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data Accuracy We will request identifiers to be updated via PDS prior to linkage. Identifiable data will be destroyed following linkage. In accordance with rights of data subject In accordance with rights of data subject 	han inal ataset. with ears used
3. Minimum necessary for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data 4. Accuracy We will request identifiers to be updated via PDS prior to linkage. 5. Kept for minimum time necessary Identifiable data will be destroyed following linkage. In accordance with rights of data subject 6. In accordance with rights of data subject Individuals will not be identified in the anonymised linked dataset u for analysis. 7. Security and Data transfers from NHS Digital and DfE will be encrypted and mana	han inal ataset. with ears used aged
 Minimum necessary for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fin analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data Accuracy Kept for minimum time necessary Identifiable data will be destroyed following linkage. In accordance with rights of data subject In accordance with rights of data subject Security and confidentiality Me request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fin analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data Kept for minimum time necessary Identifiable data will be destroyed following linkage. In accordance with rights of data subject Individuals will not be identified in the anonymised linked dataset u for analysis. Security and confidentiality Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a 	han inal ataset. with ears used aged a
 Minimum necessary for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fin analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data 4. Accuracy We will request identifiers to be updated via PDS prior to linkage. Kept for minimum time necessary Identifiable data will be destroyed following linkage. In accordance with rights of data subject In accordance with rights of data subject Security and confidentiality protection Me request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fin analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data university research data policy, research data will be kept for ten yea after the end of the study. In dividuals will not be identified in the anonymised linked dataset u for analysis. Security and confidentiality protection Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NHS 	han inal ataset. with ears used aged a IS IG
 Minimum necessary for the purpose We request only those variables necessary for analysis, and have line the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fire analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data 4. Accuracy We will request identifiers to be updated via PDS prior to linkage. Kept for minimum time necessary Identifiable data will be destroyed following linkage. In accordance of university research data policy, research data will be kept for ten yea after the end of the study. In accordance with rights of data subject Security and confidentiality protection Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NHS Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au 	han inal ataset. with ears used a aged a IS IG udited
 Minimum necessary for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather th full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the dat Accuracy We will request identifiers to be updated via PDS prior to linkage. Kept for minimum time necessary Identifiable data will be destroyed following linkage. In accordance of University research data policy, research data will be kept for ten yea after the end of the study. In accordance with rights of data subject Security and confidentiality protection Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NHS Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au and certified (IS 612909). All data backups for the safe haven are 	with ears used a IS IG udited
 <i>Minimum necessary</i> for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the dat <i>Accuracy</i> <i>We will request identifiers to be updated via PDS prior to linkage.</i> <i>Kept for minimum</i> <i>time necessary</i> <i>Identifiable data will be destroyed following linkage. In accordance of</i> <i>university research data policy, research data will be kept for ten yea after the end of the study.</i> <i>In accordance with</i> <i>rights of data subject</i> <i>Security and</i> <i>confidentiality</i> <i>protection</i> <i>Data transfers from NHS Digital and DfE will be encrypted and mana</i> according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NHS Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au and certified (IS 612909). All data backups for the safe haven are encrypted. Access to the safe haven is restricted to authorised users 	han inal ataset. with rears used a aged a IS IG udited rs who
 <i>Minimum necessary</i> <i>Minimum necessary</i> <i>for the purpose</i> <i>for the purpose</i> <i>for the purpose</i> <i>for the purpose</i> <i>the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the date analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the date analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the date analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the date analysis dataset will request identifiers to be updated via PDS prior to linkage.</i> <i>Kept for minimum time necessary</i> Identifiable data will be destroyed following linkage. In accordance with rights of data subject <i>In accordance with rights of data subject</i> <i>Individuals will not be identified in the anonymised linked dataset u for analysis.</i> <i>Security and confidentiality protection</i> Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NHS Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au and certified (IS 612909). All data backups for the safe haven are encrypted. Access to the safe haven is restricted to authorised users are required to have certified training in its use and data governance and the procedure is an are equired to have certified training in its use and data governance and the procedure is an are required to have certified training in its use and data governance and the proce	with ears used a lS IG udited rs who ce.
 Minimum necessary for the purpose We request only those variables necessary for analysis, and have in the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data Accuracy Kept for minimum time necessary Identifiable data will be destroyed following linkage. In accordance or University research data policy, research data will be kept for ten ye after the end of the study. Individuals will not be identified in the anonymised linked dataset u for analysis. Security and confidentiality protection Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NHS Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au and certified (IS 612909). All data backups for the safe haven are encrypted. Access to the safe haven is restricted to authorised users are required to have certified training in its use and data governance. Authorised users require a PIN, password, and dual authentication, 	han inal ataset. with rears used a 1S IG udited rs who ce. , and
 Minimum necessary for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather the full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the data Accuracy Kept for minimum time necessary Identifiable data will be destroyed following linkage. In accordance or University research data policy, research data will be kept for ten ye after the end of the study. In accordance with rights of data subject Security and confidentiality protection Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NHS Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au and certified (IS 612909). All data backups for the safe haven are encrypted. Access to the safe haven is restricted to authorised users are required to have certified training in its use and data governanc Authorised users require a PIN, password, and dual authentication, will sign a data confidentiality declaration before using the data. All 	han inal ataset. with ears used a 1S IG udited rs who ce. , and I
 Minimum necessary for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather th full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the date 4. Accuracy We will request identifiers to be updated via PDS prior to linkage. Identifiable data will be destroyed following linkage. In accordance or University research data policy, research data will be kept for ten yea after the end of the study. In accordance with rights of data subject Security and confidentiality protection Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NHS Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au and certified (IS 612909). All data backups for the safe haven are encrypted. Access to the safe haven is restricted to authorised users are required to have certified training in its use and data governance Authorised users require a PIN, password, and dual authentication, will sign a data confidentiality declaration before using the data. All exports of data from the safe haven are logged and audited, and 	han inal ataset. with rears used a aged a 1S IG udited rs who ce. , and I
 <i>Minimum necessary</i> for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather th full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the date Accuracy We will request identifiers to be updated via PDS prior to linkage. Identifiable data will be destroyed following linkage. In accordance of University research data policy, research data will be kept for ten yea after the end of the study. <i>In accordance with rights of data subject</i> Security and confidentiality protection Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NHS Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au and certified (IS 612909). All data backups for the safe haven are encrypted. Access to the safe haven is restricted to authorised users are required to have certified training in its use and data governance Authorised users require a PIN, password, and dual authentication, will sign a data confidentiality declaration before using the data. All exports of data from the safe haven are logged and audited, and authorised by the principal investigator. Only aggregate data, subject 	han inal ataset. with rears used a aged a 1S IG udited rs who ce. , and l
 Minimum necessary for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather th full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the dat Accuracy We will request identifiers to be updated via PDS prior to linkage. Kept for minimum time necessary Identifiable data will be destroyed following linkage. In accordance of University research data policy, research data will be kept for ten yea after the end of the study. In accordance with rights of data subject Security and confidentiality protection Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NHS Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au and certified (IS 612909). All data backups for the safe haven are encrypted. Access to the safe haven is restricted to authorised users are required to have certified training in its use and data governance Authorised users require a PIN, password, and dual authentication, will sign a data confidentiality declaration before using the data. All exports of data from the safe haven are logged and audited, and authorised by the principal investigator. Only aggregate data, subjec statistical disclosure control (e.g. avoidance of small cell sizes). will 	han inal ataset. with ears used a aged a 1S IG udited rs who ce. , and l ect to be
 <i>Minimum necessary</i> <i>Me</i> request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather th full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the dat <i>Accuracy</i> We will request identifiers to be updated via PDS prior to linkage. <i>Kept for minimum</i> <i>time necessary</i> Identifiable data will be destroyed following linkage. In accordance of University research data policy, research data will be kept for ten yea after the end of the study. <i>Individuals will not be identified in the anonymised linked dataset u for analysis.</i> <i>Security and confidentiality protection</i> Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NHS Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au and certified (IS 612909). All data backups for the safe haven are encrypted. Access to the safe haven is restricted to authorised users are required to have certified training in its use and data governance Authorised users require a PIN, password, and dual authentication, will sign a data confidentiality declaration before using the data. All exports of data from the safe haven are logged and audited, and authorised by the principal investigator. Only aggregate data, subject statistical disclosure control (e.g. avoidance of small cell sizes), will b exported from the safe haven. No potentially disclosive outputs will 	han inal ataset. with ears used a sed a 1S IG udited rs who ce. , and I sect to be II be
 Minimum necessary for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather th full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the dat Accuracy We will request identifiers to be updated via PDS prior to linkage. Kept for minimum time necessary Identifiable data will be destroyed following linkage. In accordance vith rights of data subject Individuals will not be identified in the anonymised linked dataset u for analysis. Security and confidentiality protection Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NH Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au and certified (IS 612909). All data backups for the safe haven are encrypted. Access to the safe haven is restricted to authorised users are required to have certified training in its use and data governance. Authorised users require a PIN, password, and dual authentication, will sign a data confidentiality declaration before using the data. All exports of data from the safe haven are logged and audited, and authorised by the principal investigator. Only aggregate data, subjec statistical disclosure control (e.g. avoidance of small cell sizes), will t exported from the safe haven. No potentially disclosive outputs will shared or published. 	han inal ataset. with rears used a sused a 1S IG udited rs who ce. , and l ect to be ll be
 Minimum necessary for the purpose We request only those variables necessary for analysis, and have lin the sensitivity of our request by asking for month and year rather th full date of birth or death, and for IMD rather than postcode. The fir analysis dataset will contain derived analysis variables only (see Appendix 1). No additional data will be gathered or linked to the dat Accuracy We will request identifiers to be updated via PDS prior to linkage. Identifiable data will be destroyed following linkage. In accordance vith rights of data subject In accordance with rights of data subject Security and confidentiality protection Data transfers from NHS Digital and DfE will be encrypted and mana according to standard operating procedures. Data will be stored in a secure safe haven, which is a state of the art facility that is both NH Toolkit Level 2 compliant and ISO/IEC 27001:2013 independently au and certified (IS 612909). All data backups for the safe haven are encrypted. Access to the safe haven is restricted to authorised users are required to have certified training in its use and data governanc Authorised users require a PIN, password, and dual authentication, will sign a data confidentiality declaration before using the data. All exports of data from the safe haven are logged and audited, and authorised by the principal investigator. Only aggregate data, subjec statistical disclosure control (e.g. avoidance of small cell sizes), will the shared or published. Not disclosed outside Data will not be disclosed outside the EU. 	han inal ataset. with rears used aged a 15 IG udited rs who ce. , and I sect to be II be

8 Dissemination and output

8.1 Plans for disseminating the findings of the research

We will work with the FNP NU to disseminate results of the study to Local Authorities, policy makers and commissioners. We will also build on existing relationships between the FNP NU and family nurses to inform providers about the families most likely to benefit from participating in the programme. These networks will form a two-way process where stakeholders will help inform the research and ensure effective dissemination of results.

For parents, we will publish findings of the study on the FNP website, and via existing social media forums. We will also work with our parent representatives to identify any other relevant streams of communication, and to produce findings in an appropriate format. We will also write for the Conversation (a free, online, independent source of news and views from the academic community aimed at the broader public).

We also plan to deliver a ½ day symposium on findings, to include the FNP NU, commissioners, FNP parents and family nurses.

For academic beneficiaries and other researchers, we will publish peer review journal articles, present at conferences, and disseminate our findings via social media networks and institutional blogs.

8.2 Expected outputs

Our main output will consist of evidence on the effectiveness of FNP for different groups of families. We will prepare briefings of these results for policy makers and disseminate these using the FNP NU's existing networks. Our findings will be used by the FNP NU to inform ongoing research into the adaptation of the FNP in England (ADAPT sites). Characteristics of families participating in FNP, and those eligible but not participating, will be particularly useful for Local Authorities wishing to target the most disadvantaged families.

We will work with parent representatives to co-produce a range of outputs suitable for families participating in FNP, e.g. fact sheets about the impact of FNP from a parent perspective. We will also publish our main findings as peer review publications in high quality journals (e.g. Lancet Public Health, BMJ, JAMA Pediatrics). These research outputs are likely to have significant international impact, given the ongoing roll-out of the FNP internationally.

Secondary outputs will include methodological research on the accuracy and reliability of linkage of data from health, education and social care sectors. These subsidiary analyses will be published to inform data providers and other researchers on the use of these data for future and ongoing studies. We will target journals such as the International Journal of Epidemiology and PLoS One. All journal articles will be published with open access.

8.3 Expected measureable benefits to health and/or social care

Teenage birth rates in England have been decreasing over the last 20 years. However, approximately 4% of babies each year (22,465 babies in 2016) are born to mothers aged <20 years, and rates are much higher than in other countries. Children born to teenage mothers have higher rates of hospital admissions and healthcare needs throughout childhood, but teenage mothers are also less likely to seek out preventive care in the community, meaning that pregnancy and the postnatal period provide important opportunities for intervention in this vulnerable group. There is growing recognition of the need for evidence on the best ways to support young and vulnerable mothers.

This study aims to inform policy on the implementation of the FNP in England, and to provide evidence on the likely benefits to maternal and child health. This research will complement existing evidence from the Building Blocks trial of FNP (based on approximately 1600 families recruited between 2009-2010), by evaluating outcomes for the 25,000 families enrolled in FNP since 2010. The study will directly benefit the Health and Social Care sector by providing NHS managers, commissioners and policy makers with detailed and up-to-date evidence to aid decision making about the ongoing roll out and targeting of early interventions designed to support young mothers.

Outputs from this study will provide information on variation in health outcomes and healthcare use according to different maternal characteristics and differing engagement with FNP. This will help inform decision makers on those most likely to benefit from increased early support, and on the potential gains from reducing child maltreatment, and emergency use of hospital services. Our findings will also be used by the FNP NU to inform ongoing adaptations of the FNP in England. Identifying the characteristics of families participating in FNP, and those eligible but not participating, will be particularly useful for LAs wishing to target families most likely to benefit from the programme.

Targeted support for the most disadvantaged children and their families is recognised as a priority for research, and programmes such as the FNP are likely to remain a priority for services as understanding how best to provide early support to young mothers and their families could help improve maternal and childhood outcomes. Our research outputs are likely to have significant impact, given the ongoing roll-out of the FNP internationally.

The study team will work with the FNP National Unit to ensure that outputs are used to support policy makers and commissioners in their efforts to improve the quality of care for young mothers and their families in England.

8.4 Promoting the education and well-being of children in England

This study aims to measure the association between children's exposure to the FNP and early developmental outcomes. To achieve these aims, it is necessary to have a carefully characterised control group, whose outcomes can be compared to those of FNP participants. These data can only be generated by using individual-level data on pupil and care characteristics (including information on mothers' special educational needs, exclusions and absenteeism, educational achievements, and information on their child's development and contact with social services).

Our findings will benefit the well-being of children in England by contributing to the evidence base for policy on the use of early interventions such as FNP for reducing maltreatment and encouraging a good level of child development. This information will inform policy on the ongoing roll-out, adaptation, and targeting of the FNP, and ultimately help to improve services for a vulnerable population of children in England.

8.5 Benefits to researchers

Recent investments in e-health (MRC) and administrative data (ESRC) reflect recognition of the need to use administrative data more efficiently for informing policy and planning, and the potential for revolutionising health services and outcomes research. The rich datasets created through data linkage provide the opportunity to provide an evidence base for improving services and outcomes for children. However, cross-sectoral linkage of health, education and social data has not yet been fully exploited. Establishing how these data could be used for improving services is crucial at a time when evidence guiding reorganisation of services is limited to information within sectors.

This study will evaluate the accuracy of linkage of data from different sectors, and the consistency of data captured for different purposes. Will we use FNP data to understand the reliability of data captured in administrative data sources (i.e. by comparing data collected directly from participants with that recorded in their health and education records). We will also compare outcomes derived from administrative data with published results from randomised controlled trials. This information on data quality will be valuable to other researchers using administrative data.

9 References

- 1. Olds, D.L., et al., *Effects of home visits by paraprofessionals and by nurses on children: Follow-up of a randomized trial at ages 6 and 9 years.* JAMA Pediatrics, 2014. **168**(2): p. 114-121.
- 2. Olds, D.L., et al., *Effect of home visiting by nurses on maternal and child mortality: Results of a 2-decade follow-up of a randomized clinical trial.* JAMA Pediatrics, 2014. **168**(9): p. 800-806.
- 3. Eckenrode, J., et al., Long-term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up of a randomized trial. Arch Ped Adol Med, 2010. **164**(1): p. 9-15.
- 4. Robling, M., et al., *Effectiveness of a nurse-led intensive home-visitation programme for firsttime teenage mothers (Building Blocks): a pragmatic randomised controlled trial.* Lancet, 2015. **387**(10014): p. 146-155.
- 5. Olds, D., *Building evidence to improve maternal and child health.* Lancet, 2015. **387**(10014): p. 105-107.
- 6. Barlow, J., et al., *Questioning the outcome of the Building Blocks trial*. Lancet, 2016. **387**(10028): p. 1615-1616.
- 7. Fearon, R.M.P., *Making clinical trials smarter (and more interesting).* J Child Psychol Psychiatr, 2017. **58**(2): p. 113-115.
- 8. Craig, P., et al., *Developing and evaluating complex interventions: the new Medical Research Council guidance.* BMJ, 2008. **337**.
- 9. Berwick, D.M., *The science of improvement*. JAMA, 2008. **299**(10): p. 1182-1184.
- 10. Harron K, et al., Using linked administrative for evaluating the Family Nurse Partnership in England: a scoping report. 2016.
- 11. Cattan S, et al., *The production of child development in the Family Nurse Partnership*. 2017.
- 12. Ogilvie, D., et al., Assessing the evaluability of complex public health interventions: five questions for researchers, funders, and policymakers. Milbank Q, 2011. **89**(2): p. 206-25.
- Booth, C.M. and I.F. Tannock, Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. Br J Cancer, 2014. 110(3): p. 551-555.
- 14. Wiggins, M., et al., *Teenage parenthood and social exclusion: a multi-method study*. 2005.
- 15. Lawlor, D., M. Shaw, and S. Johns, *Teenage pregnancy is not a public health problem*. BMJ, 2001. **323**(7326): p. 1428.
- 16. Stier, D.M., et al., *Are children born to young mothers at increased risk of maltreatment?* Pediatrics, 1993. **91**(3): p. 642-8.
- 17. Gonzalez, A. and H.L. MacMillan, *Preventing child maltreatment: an evidence-based update.* J Postgrad Med, 2008. **54**(4): p. 280-6.
- 18. Buck, D. and S. Gregory, *Improving the public's health: a resource for local authorities*. 2013: The King's Fund.

- Lugg-Widger, F.V., et al., Assessing the medium-term impact of a home-visiting programme on child maltreatment in England: protocol for a routine data linkage study. BMJ Open, 2017. 7(7).
- 20. Wimbush E, Craig P, and J. R, *Evaluability assessment of the Family Nurse Partnership in Scotland*. 2015.
- 21. Harron, K., et al., *Linking data for mothers and babies in de-identified electronic health data.* PLoS One, 2016. **11**(10): p. e0164667.
- 22. Harron K, et al., *A guide to evaluating linkage quality for the analysis of linked data.* Int J Epidemiol, 2017. **46**(5): p. 1699-1710.
- 23. ICD, International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM). 2011.
- 24. Health and Social Care Information Centre, *OPCS-4 Classification of Interventions and Procedures*.
- 25. Mc Grath-Lone, L., et al., *Data Resource Profile: Children Looked After Return (CLA).* Int J Epidemiol, 2016. **45**(3): p. 716-717f.
- 26. Department for Education, *Children looked after by local authorities in England: Guide to the SSDA903 collection 1 April 2016 to 31 March 2017.* 2015.
- 27. Mc Grath-Lone, L., et al., *Changes in first entry to out-of-home care from 1992 to 2012 among children in England*. Child Abuse Negl, 2016. **51**: p. 163-171.
- 28. Department for Education, *Statistical First Release: Characteristics of children in need in England, 2013-14.* 2014.
- 29. Grath-Lone, L.M., J. Woodman, and R. Gilbert, *Safeguarding children and improving their care in the UK.* Lancet, 2015. **386**(10004): p. 1630.
- 30. Health and Social Care Information Centre, *Fact Sheet on access methods to trace and use the NHS number of local records*. 2014.
- 31. Health and Social Care Information Centre, *Fact Sheet on access methods to trace and use the NHS number on local records*. 2014.
- 32. Schnitzer, P.G., et al., *Identification of ICD codes suggestive of child maltreatment*. Child Abuse Negl, 2011. **35**(1): p. 3-17.
- 33. Gilbert, R., et al., *Child maltreatment: variation in trends and policies in six developed countries.* The Lancet, 2012. **379**(9817): p. 758-772.
- 34. Public Health England, *Health visitor service delivery metrics 2016 to 2017*. 2017.
- 35. Austin, P.C., *An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies.* Multivariate Behavioral Research, 2011. **46**(3): p. 399-424.
- 36. Liu, W., S.J. Kuramoto, and E.A. Stuart, *An Introduction to Sensitivity Analysis for Unobserved Confounding in Non-Experimental Prevention Research.* Prev Sci, 2013. **14**(6): p. 570-580.
- 37. VanderWeele, T.J. and P. Ding, *Sensitivity analysis in observational research: Introducing the e-value.* Ann Intern Med, 2017. **167**(4): p. 268-274.
- 38. Herbert, A., et al., *Violence, self-harm and drug or alcohol misuse in adolescents admitted to hospitals in England for injury: a retrospective cohort study.* BMJ open, 2015. **5**(2): p. e006079.
- 39. Hardelid, P., N. Dattani, and R. Gilbert, *Estimating the prevalence of chronic conditions in children who die in England, Scotland and Wales: a data linkage cohort study.* BMJ Open, 2014. **4**(8).