Family Nurse Partnership
National Unit
Changing the world one baby at a time™

Using linked administrative data for monitoring and evaluating the Family Nurse Partnership in England: a scoping report

Katie Harron, London School of Hygiene and Tropical Medicine

Louise McGrath-Lone, Administrative Data Research Centre – England, UCL

Samantha Mason, Family Nurse Partnership National Unit

Ruth Gilbert, Administrative Data Research Centre – England, UCL Institute of Child Health

London School of Hygiene & Tropical Medicine

UCL

Administrative Data Research Centre England

Contents

Rationale for this report

Randomised controlled trials (RCTs) are considered a 'gold standard' for evaluating health and social care interventions, largely due to excellent internal validity. However, the generalisability of RCT results or their applicability to routine practice can be limited due to differences in trial participants and patients in routine practice. Population-based observational studies may mitigate or overcome these issues, can have increased statistical power to detect small differences in outcomes, and can be less expensive to administer (though they are subject to their own limitations). As such, well-designed population-based analyses can complement RCTS in evaluating complex interventions. This report presents a scoping review of how population-based linkage between data from the Family Nurse Partnership (FNP) in England and administrative datasets from other services could be used to generate evidence for commissioning, service evaluation and research.

A recent randomised controlled trial of the Family Nurse Partnership (FNP) provided limited evidence of effectiveness of the programme in relation to short-term maternal and child outcomes. Further follow-up is underway to evaluate the impact of FNP on longer-term outcomes among trial participants. This scoping report addresses the methodological considerations, permission pathways and technical challenges of using data from the FNP linked with routinely collected, administrative data from other public services for population-based analyses, at a national and local authority level. The aim of this report is to assess the feasibility of linking individual-level data to key administrative datasets related to health, education and children's social care, and to establish how these datasets could be used to monitor and evaluate the FNP and its impact on child and maternal outcomes. This report does not consider use of linked data for direct patient care, nor does it not consider the economic impact of FNP, which is currently being addressed elsewhere.

Wider linkage between FNP data and administrative data sources is needed for two main reasons. Firstly, the FNP National Unit (FNP NU) is leading a programme of rapid development and testing to improve and adapt the FNP programme for the UK context. There is a need for these adaptations to be monitored and evaluated in terms of implementation and outcomes. Secondly, funding for children's public health services (including the FNP) transferred from NHS England to Local Authorities (LAs) in October 2015. LAs require evidence (beyond a single randomised controlled trial) on the implementation and outcomes of FNP, to monitor the service and support commissioning decisions in their local area.

Key Messages

FNP National Unit

1. Linkage between FNP and administrative data sources, such as Hospital Episode Statistics (HES) and the National Pupil Database (NPD), would provide benefits for monitoring and evaluating maternal and child health and educational outcomes of the FNP at national and sub-national levels. However, there are important technical and methodological considerations that need to be addressed, and significant investment of time and resources will be required.
2. Linkage between FNP and HES could guide evaluation and adaptation of the programme's eligibility criteria through i) identifying other vulnerable groups of women who have similar characteristics and outcomes to FNP mothers and ii) exploring variation in uptake of FNP between sub-groups of eligible women (e.g. teenage mothers who have been in care).
3. The FNP NU should actively promote wider use of FNP data for commissioning, service evaluation and research and exploit linkage with national data for monitoring and evaluating short- and longer-term outcomes. The FNP NU should develop a clear strategy for identifying research and evaluation priorities, considering which questions are best addressed using linked administrative data.
4. Linked administrative data, combined with ongoing monitoring of modifications to the FNP and recording of details of local service as usual, offer the potential for control group comparisons for evaluating outcomes of the FNP.
5. Analyses at the Local Authority level is limited by small sample sizes, and a lack of appropriate comparison groups. Sub-national analyses using groups of local authorities or regions would be feasible.
6. The feasibility of monitoring and evaluating social care outcomes is currently limited due to a lack of identifiers recorded in the relevant administrative datasets at the national level.
7. Commissioners need to allow adequate timescales for permissions to access and analyse linked administrative data to evaluate and monitor the FNP.
8. The FNP NU should consider investing in data scientist/analyst support for complex data access applications, linkage, and analyses of linked data. Working with academic communities would facilitate independent and rigorous evaluation of the programme utilising such methods.
9. Support under Section 251 should be obtained for processing and linking FNP data on mothers and their children with national administrative data on health, education and social services, as seeking consent to linkage is impractical given the complexity, cost, and potential bias in contacting >24,000 FNP participants. A formal request for advice

on the circumstances in which identifiable data can be released to commissioners within local authorities should be submitted to the Confidentiality Advice Group (CAG).

10. Any secondary uses of FNP data should ensure compliance with fair processing of data and clearly define the public interest in terms of benefits to health. Privacy notices on the FNP website and other relevant forums, such as FNP participation leaflets, should provide participants with information about the use of their data and FNP NU should establish a system for opting out of personal identifiable data being used for secondary purposes.

11. To ensure accuracy of FNP data for secondary uses, tracing of FNP participants through the Personal Demographic Service (PDS) should be conducted before linkage is performed, and both original and updated identifiers should be retained in FNP national data collection. Flagging of FNP families in the Mother and Children's Data Set (MCDS) could facilitate routine evaluation and monitoring (prospectively from mid-2015) but feasibility has not yet been tested.

12. As FNP families mature, linkage with additional datasets (such as data on mental health services, employment, benefits, youth justice, family courts and crime) would further establish the inter-sectoral benefits of the FNP. However, the legal bases and approvals processes necessary to access and link these data for population-based analyses are not currently well defined.

Other government bodies

13. NIHR: In addition to funding the Building Blocks trial follow up, the Public Health Research Programme should commission rigorous evaluation of the FNP programme using linked administrative data to assess ongoing effectiveness.

14. PHE: PHE should consider how linked administrative data could be used to evaluate children's public health services.

15. DfE: National evaluation of the impact of FNP on interventions by children's social care services should be a priority. To facilitate accurate linkage with FNP and comparator data for pre-school children, DfE should collect postcode and name for all children in the Child Looked After return and Child in Need Census.

16. NHS Digital: To facilitate evaluation of FNP and comparator populations, flagging of FNP participation could be considered for mothers and children captured in the Maternity and Children's Data Set (MCDS). In addition, longitudinal information in maternal and baby hospital records should be routinely combined using the relational link captured in PDS. The PDS link should be used to add information on key maternal and child health attributes to education and children's social care datasets held by DfE. Regular updates of these datasets should be done to enable timely evidence for national commissioning, service evaluation and research.

Abbreviations

A&E: Accident and Emergency
ADRC-E: Administrative Data Research Centre for England
ALSPAC: Avon Longitudinal Study of Parents and Children
APC: Admitted Patient Care
BMI: Body Mass Index
CAFCASS: Children and Family Court Advisory and Support Service
CAG: Confidentiality Advisory Group
CHI: Community Health Index
CINC: Children in Need Census
CLA: Children Looked After return
CPRD: Clinical Practice Research Datalink
DAAG: Data Access and Approvals Group
DARS: Data Access Request Service
DBS: Demographic Batch Service
DfE: Department for Education
DMAP: Data Management Advisory Panel
DPA: Data Protection Act
DWP: Department of Work and Pensions
ELSA: English Longitudinal Study of Ageing
EYFSP: Early Years Foundation Stage Profile
FNP: Family Nurse Partnership
FNPIS: FNP Information System
FNP NU: FNP National Unit
GCSE: General Certificate of Secondary Education
gFNP: Group Family Nurse Partnership
GP: General Practitioner
HCP: Healthy Child Programme
HES: Hospital Episode Statistics
HESA: Higher Education Statistics Agency
HMRC: Her Majesty's Revenue and Customs
HRA: Health Research Authority
HSCIC: Health and Social Care Information Centre
ICD-10: International Classification of Disease, version 10
ICO: Information Commissioner's Office
ID: Identifier
IG: Information Governance
IGARD: Independent Group Advising on the Release of Data
IMD: Index of Multiple Deprivation
IRAS: Integrated Research Application System
ISAC: Independent Scientific Advisory Committee
LA: Local Authority
MCDS: Maternity and Children's Data Set
MHRA: Medicines and Healthcare products Regulatory Agency
MoJ: Ministry of Justice
NHS: National Health Service
NICE: National Institute for Clinical Excellence
NICU: Neonatal Intensive Care Unit
NN4B: NHS Numbers for Babies
NPD: National Pupil Database
OHC: Out-of-Home Care
ONS: Office for National Statistics
OP: Outpatient
OPCS-4: Office of Population, Censuses and Surveys Classification of Surgical operations and procedures, version 4
PEARL: Project to Enhance ALSPAC through Record Linkage
PDS: Personal Demographic Service
PMR: Pupil Matching Reference
SAIL: Secure Anonymised Information Linkage Databank
SCBU: Special Care Baby Unit
SDQ: Strengths and Difficulties Questionnaire
UCL: University College London
UPN: Unique Pupil Number

1. Introduction

Randomised controlled trials (RCTs) are considered a ‘gold standard’ for evaluating health and social care interventions, largely due to excellent internal validity. However, the generalisability of RCT results or their applicability to routine practice can be limited due to differences in trial participants and patients in routine practice.[1-3] Population-based observational studies may mitigate or overcome these issues, can have increased statistical power to detect small differences in outcomes, and can be less expensive to administer (though they are subject to their own limitations). As such, well-designed population-based analyses can complement RCTS in evaluating complex interventions.[1] This report presents a scoping review of how population-based linkage between data from the Family Nurse Partnership (FNP) in England and administrative datasets from other services could be used to generate evidence for commissioning, service evaluation and research.

The FNP is a voluntary preventive and early help home visiting programme for first time young mothers from early pregnancy until their child’s second birthday. FNP offers intensive and structured home visiting delivered by specially trained nurses, aiming to support young mothers (and fathers) in providing competent and sensitive care for their children, adopting healthy behaviours and planning for their own futures. The three main aims of the programme are to improve pregnancy outcomes, improve child health and development, and improve parents’ economic self-sufficiency.[4]

FNP is a licensed programme developed over 30 years in the US, which has been shown to have a positive impact on outcomes for both the mother and the child, notably improved prenatal health, decreased child maltreatment and associated injuries, improved family planning, increased maternal employment and improved school readiness.[5-7] Family Nurses aim to build trusting and supportive relationships with families, working with young parents on attachment, relationships and psychological preparation for parenthood. Behaviour change methods are used to encourage healthy lifestyles, good care for babies, and planned futures, through a series of structured home visits (weekly, fortnightly or monthly) lasting between one and one and a half hours.

There are currently 16,000 FNP places in approximately 125 different LAs in England, covering 25%-30% of the eligible population of first time teenage mothers. There are two main rationales for this scoping report. Firstly, the FNP National Unit (FNP NU) is leading a programme of rapid development and testing to improve and adapt the FNP programme for the UK context. There is a need for these adaptations to be monitored and evaluated in terms of delivery and impact. Secondly, funding for children’s public health services (including the FNP) transferred from NHS England to Local Authorities (LAs) in October 2015. LAs require evidence on the implementation and effectiveness of FNP in their local area (beyond RCT results), to monitor the service and support commissioning decisions.

This report considers linkage between FNP Information System (FNPIS) data and administrative data from other services for population-based analyses of the FNP programme. It does not consider use of linked data for direct patient care, nor impact on fathers' outcomes (no paternal identifiers are recorded in FNPIS).

Aims

We aimed to assess the feasibility of linking individual-level FNP Information System data to key administrative datasets related to health, education and children's social care, and to establish how these datasets could be used to monitor and evaluate the FNP and its impact on child and maternal outcomes. This report addresses three main aspects of feasibility:

1. Methodological considerations
2. Permission pathways
3. Technical requirements

Report structure

Section 2 provides further background on the FNP in an English context and an overview of the FNPIS. Section 3 describes the key administrative data sources considered for evaluating and monitoring the FNP, namely:

a) National Pupil Database (NPD), including Children in Need Census (CINC) / Children Looked After return (CLA)
b) Hospital Episode Statistics (HES)
c) Maternity and Children's Dataset (MCDS)
d) Clinical Practice Research Datalink (CPRD)

In Section 4 we describe the permission pathways for linkage between FNP and these key administrative datasets, including consent and legal and ethical requirements, for both, i) service evaluation and commissioning on behalf of the Department of Health and/or Local Authorities (LAs) and ii) population-based research to evaluate the effectiveness of the FNP.

Section 5 assesses the technical requirements for linking FNP data with administrative data by establishing the availability and completeness of identifiers common to FNP and administrative datasets. This section also considers deterministic and probabilistic linkage techniques and methods for evaluating linkage quality.

Section 6 highlights the methodological considerations for evaluating and monitoring FNP using linked administrative data, including identification of key indicators for short- and long-term outcome measures and appropriate comparison groups. This section also addresses the use of locally linked data, highlighting potential benefits and challenges and key considerations.

Findings from this report are summarised in Section 7 alongside recommendations for FNP National Unit and administrative data providers. A series of case studies (including potential exemplar linkages of FNP to national administrative data sources and at a local level) are presented in Appendix 5 to illustrate the feasibility of linkage based on the principles in this report and highlight opportunities, barriers and important considerations.

2. Background and evidence

Introduction of the FNP in England

The Nurse-Family Partnership programme (as FNP is known in the US) was developed by the University of Colorado, who licensed the programme to ensure it was delivered in accordance with the original model that demonstrated significantly improved outcomes for vulnerable young families.[8] The Department of Health holds the licence in England. The license stipulates that replication of the Nurse-Family Partnership should follow four stages: i) adaptation to local contexts and populations whilst ensuring fidelity to model elements; ii) assessment of feasibility and acceptability in a pilot study; iii) evaluation of the adapted program in a randomised controlled trial; iv) expansion if the evaluation shows significant positive outcomes.

In line with these requirements, FNP was originally introduced in England in 2007 in 10 pilot sites.[9] A formative evaluation of the pilot sites found that FNP families and staff were positive about and accepting of the programme. By the second wave of pilot sites, the FNP was being delivered successfully in terms of the number and content of visits and adherence up the child's second birthday.[10]

A randomised controlled trial of the effectiveness of FNP in England compared to usual services was then commissioned by Department of Health and undertaken by investigators at Cardiff University.[11] Findings from the Building Blocks trial were published in 2015 based on short-term outcomes from pregnancy to the child's second birthday for a cohort of 1618 families recruited in 2009 and 2010.[12] Results from the trial showed that FNP had no effect on the study's four primary outcomes (smoking at late pregnancy, birth weight, subsequent pregnancy by 24 months, child A&E and hospital attendances in first two years of life) but did have a positive effect on a number of secondary outcomes, including children's early language and cognitive development. However, this phase of the trial was based on short-term outcomes only (from pregnancy to the child's second birthday) for a sample of families recruited in 2009 and 2010. While further follow-up of trial participants is underway to evaluate the impact of FNP on longer-term outcomes past the age of two,[13] there is a need for evidence that takes into account changes to programme delivery and content as FNP has evolved in subsequent cohorts during national roll-out.

There are some similarities between outcomes observed in the Building Blocks trial in the UK and the original US trials.[14-16] For example, neither of the two US trials that included birth weight as an outcome observed any differences between groups. However, there were also major differences between the English and US trials, which may in part be explained by the different contexts.[17] Firstly, eligibility criteria differed across trials, with the US trials finding greatest impact for the most vulnerable groups of mothers. Secondly, the US trials were conducted decades before the UK trial. Thirdly, support services for young first-time mothers in England are freely available and more comprehensive than those that are offered in the US. The most convincing evidence for the effectiveness of the FNP in the US derives from long-term follow-up of maternal employment and welfare dependency, childhood developmental and behavioural outcomes, and maltreatment[5, 6, 18-21]; follow-up of the UK trial participants is now underway to evaluate impact of the FNP on child maltreatment and neglect up to six years of age. For a summary of evidence from pilot and trial sites in England, and the US trials, see Appendix 1.

Existing services outside FNP

The following section describes how the FNP fits within the extensive antenatal and health visiting services currently available for mothers and families in the UK. The National Institute for Clinical Excellence (NICE) recommends that women expecting their first baby receive a minimum of 10 antenatal check-ups.[22] In addition, teenage mothers should be offered age-appropriate services, information about help with transportation to and from appointments, antenatal care in the community, and the opportunity for the partner/father of the baby to be involved in antenatal care.[23] NICE guidance recommends that teenage mothers should be given a named midwife, from whom the majority of antenatal care is provided, and that commissioners should consider a specialist antenatal service for teenage mothers, including antenatal care and education in peer groups.

Antenatal and postnatal support is also provided under the Healthy Child Programme (HCP): midwives and health visitors offer support to parents during pregnancy and up to the child's first birthday, in the form of home or clinic visits,[24] and health visitors provide additional support up to school age (five years). Responsibility then passes to school nurses. The HCP operates a proportionate universalism model, meaning that while support is universal, scale and intensity is proportionate to the level of disadvantage.[25] Under this model, there are four levels of intervention: Community, Universal Services, Universal Plus, and Universal Partnership Plus.[24] The frequency of home visits and contact under the HCP varies locally, and according to individual need, but there are a core set of mandated visits that health visitors must complete, including the new baby review (around 14 days after birth) and the two year check (between 24-30 months of age.)[26]. Depending on local systems, extra support is also offered for vulnerable mothers, such as those with mental health issues or drug and alcohol problems.[26] Child health clinics, run by health visitors and GPs, offer baby health and development reviews and vaccinations.

The FNP works with families to deal with complex problems and is an example of the highest tier of the HCP (Universal Partnership Plus). In the control arm of the Building Blocks trial, mothers received an average 11 visits from midwives and 16 visits from health visitors between enrolment (on average at 18 weeks gestation) and 24 months post-partum.[12] Mothers in the intervention arm received an average 39 visits from Family Nurses.

Moving forward

The FNP currently provides 16,000 places at any one time in approximately 125 different LAs in England, covering 25%-30% of the eligible population. FNP is now commissioned by LAs as part of children's public health services, and delivered by health service providers. The FNP NU, based in the Tavistock and Portman NHS Trust, has been commissioned by Public Health England (PHE) to undertake a programme of work to support the effective implementation of the FNP in England. The FNP NU support quality improvement through working directly with local FNP teams and monitoring implementation quality. In addition, the FNP NU aim to support local FNP teams with ongoing quality improvement, evaluation of impact, and commissioning decisions as part of children's public health services. To achieve these purposes, detailed information in addition to data currently captured through delivery of the FNP is required, including information on appropriate comparison groups not participating in FNP.

Administrative data has played a pivotal role in capturing maternal and child outcomes in the original Building Blocks trial, the ongoing follow-up study, and a trial of a group version of FNP (gFNP).[27, 28] In the main trial, maternity records were used to derive antenatal, birth and neonatal data including birth weight. Linkage was performed for HES and GP records, abortions data and immunisation data. The follow-up study outcomes will be captured in administrative data, with no further data collection from participants currently planned. Linkage with hospital, education and social care data will be performed for 1570 families with access to linked data controlled through a trusted third party (the Secure Anonymised Information Linkage Databank; SAIL) who will anonymise the data before returning to the study team. Finally, for gFNP, hospital activity during the period between randomisation and 12 month postpartum was captured in HES.[28] An assessment of options for evaluating the FNP in Scotland, conducted by NHS Scotland, recommended that administrative data sources could be used in a natural experimental study capturing routinely-collected data on pregnancy, birth and maternal and child health for FNP participants and young first-time mothers giving birth outside of the recruitment period for FNP.[29]

The Family Nurse Partnership Information System

The Family Nurse Partnership Information System (FNPIS) is a bespoke system that supports the implementation of the FNP programme in England, provided by NHS Digital (formerly the Health and Social Care Information Centre, HSCIC) under contract to the FNP NU on behalf of the Department of Health and PHE. The FNPIS exists so that local FNP practitioners can access data collected from their clients as part of the programme delivery process. FNPIS has been operational since April 2009 and data are thought to be of reliable quality from 2010. 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 FNPIS 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, referrals to other services, etc.). Appendix 2 provides a full list of FNP variables.

FNPIS contains also contains maternal and child identifiers such as name, date of birth and NHS number (described in Section 5, Table 4); however, identifiable data are accessible only to relevant staff in the FNP team with whom the individual client is registered, via the NHS secure N3 network. Furthermore, these data are only accessible locally whilst a mother is participating in the programme; when mothers graduate from the FNP, data are pseudonymised and identifiers are retained solely on secure NHS Digital servers. Access to information on FNP participation captured in local systems through clinical care records varies according to local practice. Aggregate or pseudonymised individual-level data are accessed by the FNP NU to help conduct a range of routine and ad hoc analyses.

The data recorded by FNP teams during delivery of the programme, combined with detailed information on both maternal and child outcome indicators captured in administrative data, could be used to evaluate the effectiveness of any local or national adaptations to the FNP. There are two main aspects of the programme that could be adapted: delivery and eligibility, which are discussed in the following sections.

Adapting delivery

The FNP, whilst highly prescribed, also gives Family Nurses room to adapt their curriculum during individual visits, i.e. to adjust content of visits in response to mothers' needs, preferences and goals. Evaluation of the pilot sites recommended that quantitative information derived from routine documentation on visit content should be monitored, to ensure a balance between the need to deliver the FNP with fidelity and the need to respond to practical considerations such as preventing clients from leaving the programme or dealing with immediate crises.[30] There has been significant adaption of the programme model, method of delivery and implementation of the programme in the period since the FNP was introduced in England.

Currently, FNP visits ideally begin at 14-16 weeks gestation. Visits are conducted weekly for the first month, followed by fortnightly visits until the baby is born, weekly visits for six weeks post-partum, fortnightly visits up to 20 months post-partum, and monthly visits to 24 months post-partum.

Alternative, systematic adaptations to the delivery of FNP, in either content (e.g. increased focus on smoking cessation) or practice (e.g. altering nurse caseloads)[31], are being planned for a sample of 11 "ADAPT" sites. It is important that such modifications are carefully monitored and evaluated through comparisons with relevant control groups, to understand impact on outcomes. Such evaluations could be conducted using linked administrative data.

One variation already being evaluated through a randomised controlled trial is the group version of FNP (gFNP). The gFNP includes teenage mothers with one or more previous live births, or mothers aged 20 to 24 with low educational qualifications and no previous live births. Results are currently being analysed, to determine impact on child abuse, parent/child interaction, maternal depression, parenting stress, health related quality of life, social support, and use of services.[28]

Adapting eligibility

Participation in FNP is voluntary. Currently, eligible mothers are:

- First time mothers aged 19 or under at conception (including those with previous pregnancies ending in miscarriage, termination or still-birth)
- First time mothers aged 20-24 with identified vulnerabilities
- Mothers living within an agreed catchment area (usually an LA, but sometimes sub-areas)
- No more than 28 weeks pregnant
- Not planning to have their child adopted.

The eligibility criteria for FNP in England have recently been extended as part of the FNP Next Steps programme so that women up to age 24 can be recruited if they have identified vulnerabilities. Alternatively, sites can focus their FNP resources on first time mothers under 19 with vulnerabilities only. Linked administrative data offers the potential to evaluate such modifications.

During the pilot phase of FNP in England, some of the smaller sites also recruited first-time mothers aged 20-22 who had at least one risk factor relating to lack of income, education/employment, or absence of a partner. However, enrolment of mothers in this group was less successful, mainly due to problems identifying women with these risk factors early in pregnancy. Pilot study investigators noted that the inclusion of additional criteria was likely to slow down the recruitment process, and was heavily influenced by the level of involvement and accessibility of data from midwifery services. Different systems exist within NHS Scotland, where the programme will be expanded to eligible 20-24 year olds (www.gov.scot/Topics/People/Young-People/early-years/parenting-early-learning/family-nurse-partnership/training).

There may be additional approaches for targeting women at similar disadvantage to those eligible for FNP. For example, evidence suggests that the effects of adolescent motherhood persist not only for the first child.[32] It would be feasible to extend eligibility criteria to all teenage mothers who were not previously enrolled in the programme (i.e. women who were teenagers at their first birth, irrespective of their current age). To inform or evaluate such changes in mothers receiving FNP, outcomes for young multiparous mothers could be compared with those of young primiparous mothers using data captured in administrative sources.

3. Data sources

Access to and use of health and educational datasets for research and evaluation in the UK is well established. Four key datasets could be linked with FNPIS data to provide evidence nationally and locally for commissioning, service evaluation and research:

1. National Pupil Database (NPD): school achievement, special educational needs, exclusions and absences, children looked after and children in need
2. Hospital Episode Statistics (HES): hospital inpatient and outpatient admissions, A&E attendance
3. Clinical Practice Research Datalink (CPRD): primary care consultations, prescribing and referrals
4. Maternity and Children's Data Set (MCDS): data collected from maternity information systems and child health surveillance.

Although these data are collated at a national level, data initially flows from within LAs (e.g. NPD) or health providers (e.g. HES). The extent to which FNP mothers and/or their children are captured in these administrative datasets varies (Figure 1), as does the ability to identify maternal-child dyads within them. All FNP mothers and their children should be captured at birth in MCDS and HES (excluding a small proportion (~3%), where delivery takes place in a private hospital and the mother/baby are never admitted to an NHS hospital and not captured in HES). Approximately 10% of mothers and their children will be captured in CPRD. Mothers who are at school, and their children (once at school-going age), should be captured in NPD in separate data years. The DfE holds data on looked after children and children in need for England and these are linked to the NPD for children receiving these services while attending school.

A mother-child link is available within FNPIS and CPRD, and is planned for MCDS, but is not routinely available in HES or NPD. The periods for which corresponding data are available in FNPIS and administrative datasets also varies (Table 1). Nonetheless, these data could be used to quantify key indicators for FNP coverage, as well as short- and long-term outcome measures for mothers, their children and appropriate comparison groups.

There are a number of other administrative datasets in England that have been used for research and evaluation purposes that offer potential for monitoring and evaluation the FNP programme in England. For example, HMRC data on income tax and benefits, higher education and student loan data,[33] criminal justice data (from the Ministry of Justice; MoJ) and data from the Children and Family Courts Advice and Support Service (CAFCASS).[34] However, the use of these datasets in research and evaluation has been limited to date by data access restrictions.

Figure 1: Overlap between FNP and administrative datasets.

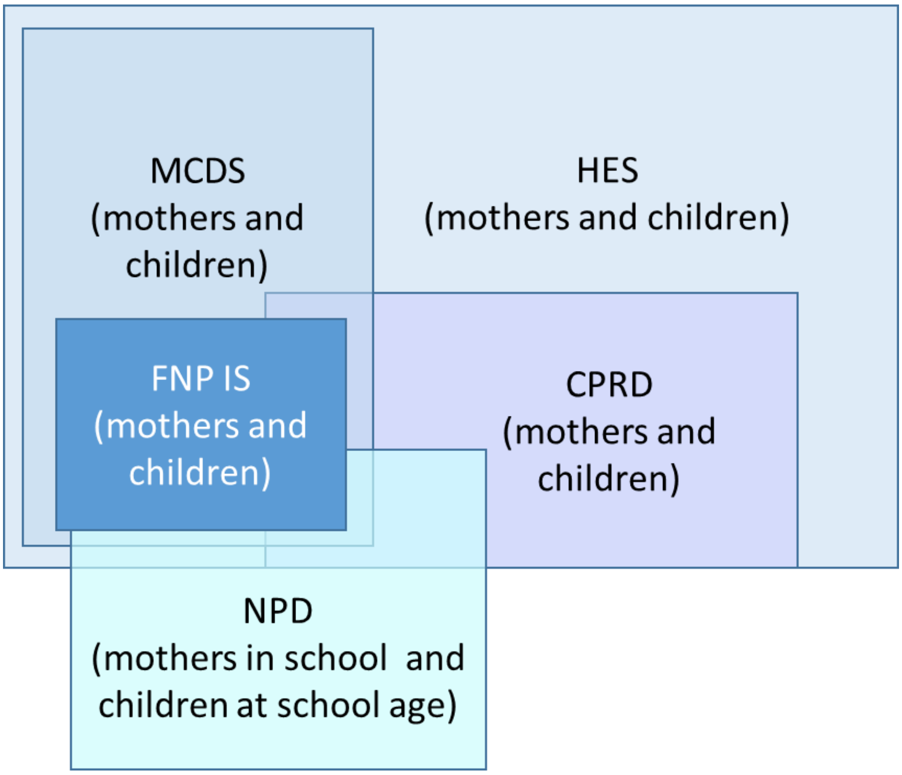


Table 1: Overview of data sources by year of birth and year of data collection. Numbers in boxes correspond to age of mother (blue) and child (orange). The first cohort of FNP children with good quality data in FNPIS should enter school in 2014/15 (children born from 2010 onwards).

		Year of FNP data								
		2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Availability / age of individual	**NPD: School Census: age 2-19**	, 13-19	13-19	2, 13-19	2,3, 13-19	2-4, 13-19	2-5, 13-19	26, 13-19	2-7, 13-19	2-8, 13-19
	NPD: Early Years Census: age 2-4			2	2,3	2-4	2-4	2-4	2-4	2-4
	NPD: Key Stage 1,2,4: age 7,11,16	16	16	16	16	16	16	16	7,16	7,,16
	NPD: CLA	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NPD: CINC	✓	✓	✓	✓	✓	✓	✓	✓	✓
	HES-ONS	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MCDS	✓	-	-	-	-	-	✓	✓	✓
	CPRD	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of birth	1991	19								
	1992	18	19							
	1993	17	18	19						
	1994	16	17	18	19					
	1995	15	16	17	18	19				
	1996	14	15	16	17	18	19			
	1997	13	14	15	16	17	18	19		
	1998		13	14	15	16	17	18	19	
	1999			13	14	15	16	17	18	19
	2000				13	14	15	16	17	18
	2001					13	14	15	16	17
	2002						13	14	15	16
	2003							13	14	15
	2004								13	14
	2005									13
	2006									
	2007									
	2008									
	2009									
	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
	2018									0

National Pupil Database

The National Pupil Database (NPD) was first produced in 2002. It contains a range of information about pupils and students, derived from a number of data sources with some retrospective data going back to 1995/96. Information from the NPD is used to promote education and well-being of children in England and support school improvement, through research and analysis, statistics, and advice and guidance.

Information on test results, 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. Further information on attainment data for pupils in non-maintained special schools, sixth form and further education is also available. Data in the NPD are obtained from schools, LAs and awarding bodies, and processed by the Department for Education (DfE) Education Data Division. A full list of data items is available at: www.gov.uk/government/uploads/system/uploads/attachment_data/file/465939/NPD_Data_Tables.xlsx; the main tables are as follows:

- **The School Census (formerly PLASC)** contains pupil-level information from 2002 for pupils aged 2-19+. The Census is carried out termly (three times a year) in nurseries, primary schools, secondary schools and special schools. Data from Pupil Referral Units, schools that are established and run by local authorities for pupils who cannot attend mainstream or special schools, were collected in a separate census from 2010 and incorporated into the main census from 2014.
- **The Early Years Census** contains data from 2008 for three and four year olds, and two year olds from 2014. Data are collected annually from all English providers of funded Early Years education.
- **The Early Years Foundation Stage Profile (EYFSP)** contains data collected annually on teacher assessments.
- **Key Stages 1-5** data contain teacher assessments and/or test results taken in Year 2, 6, 9, 11 and 12/13 (at ages 7, 11, 14, 16, and 17-18).

The DfE also hold other databases that can be linked to NPD: the Children Looked After return and the Child in Need Census.

Children Looked After return

The Children Looked After (CLA) return is a national individual-level dataset held by the Department for Education (DfE) that contains information on all looked after children and recent care leavers in England. Data collection began in 1992 and is ongoing via an annual online census of local authorities. Initially, data collection was mandated for all children in England who were looked after, but between 1998 and 2003 it was restricted to a one-third sample (selected as children with a day of birth divisible by three) before reverting to include all looked after children in 2004. CLA is used to monitor the care and outcomes of looked after children and to enable evaluation of the potential effects of government policy initiatives.[35]

The CLA contains detailed care histories for looked after children including the start and end dates of each episode of out-of-home care (OHC). Approximately 3% of all children are placed in OHC at least once by the age of 18 years.[36] For children in continuous care for 12 months or more, information on health-related outcomes such as immunisations, health checks and Strengths and Difficulties Questionnaire (SDQ) scores has been collected since 2009. In 2002, collection of data on the activity and accommodation of care leavers at age 19 began with further follow-up at other ages introduced in later years. The CLA does not include private fostering arrangements in which a child is cared for by an adult who is not a close relative (i.e. someone other than a parent, grandparent, sibling, aunt or uncle).[37]

A Unique Pupil Number (UPN) is used for linkage 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, children who are looked after only before they enter nursery/school will not have a UPN recorded in the CLA dataset (approximately 20% of children recorded in CLA).[38] Adopted children are provided with a new UPN, and any subsequent episodes of care cannot be tracked within CLA. The only identifiers available at a national level for these children are sex, date of birth, and ethnicity (for children in care for a continuous period of 12 months or more, postcode is also collected). Linkage for this group of children is therefore likely to be incomplete.

Children in Need Census

The Children in Need Census (CINC) has collected information on all referrals to children's social care, assessments carried out upon those children, and whether the children became the subject of a child protection plan, since October 2008.[39] Approximately 5-6% of all children are referred each year. The CINC includes looked after children, those supported in their families or independently, and children who are the subject of a child protection plan. Children in need are those assessed to be in need by children's social care services through an initial assessment.[40] Data items collected include indicators and dates for child protection plans, details of abuse, and service provision. As with the CLA, linkage between CINC data and NPD is performed using the UPN, and cannot easily be performed for children captured in CINC only before entering nursery/school or adopted children.

Data access

Access to NPD is subject to approval from the DfE. Data are divided into 4 tiers based on sensitivity of data: tier 1 contains sensitive personal information (e.g. date of birth, postcode, looked after status); tier 2 contains less sensitive versions of tier 1 data (e.g. month and year of birth, lower layer super output area (LSOA), special education needs depending on the level of detail); tier 3 contains school-level data only; and tier 4 contains other pupil-level data (e.g. attainment, absence and exclusions). Identifiable or sensitive data items can only be accessed with approval from the Data Management Advisory Panel.

The NPD holds name, postcode, and date of birth, as well as unique identifiers (UPN and Pupil Matching Reference (PMR)), to enable the same pupil to be tracked across different data years. NPD is routinely linked with data from further and higher education sectors (the individualised learner record and the Higher Education Statistics Agency (HESA) student record). Linkage with cohort and survey data (e.g. the Longitudinal Study of Young People in England and Understanding Society)[41] has also taken place. The Buildings Block follow-up trial will use linkage with NPD (to be performed by SAIL).

Costs

Currently, DfE do not charge for access to the NPD, CLA, CINC or datasets routinely linked with the NPD (e.g. HESA), or for linkage with external datasets.

Hospital Episode Statistics

Hospital Episode Statistics (HES) is a data warehouse containing details of all admissions, outpatient appointments and A&E visits at NHS hospitals in England. HES was originally developed as an administrative dataset to support financial management, and is not used for clinical care, but for secondary purposes including service evaluation and research. HES has collected inpatient data since 1989, but data are routinely provided from April 1997. Outpatient data have been collected since 2003 and A&E data since 2007. Each year, approximately 19 million inpatient records, 90 million outpatient records and 18 million A&E records are added to the dataset.

The 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. Data items collected in HES are available with a full description at http://digital.nhs.uk/hesdatadictionary.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).[42, 43]

Data access

Pseudonymised, individual-level HES data are available by application to the NHS Digital Data Access Request Service (DARS; http://digital.nhs.uk/dars). Applications are submitted to the Data Access Advisory Group (DAAG) who review the legal basis for accessing data, the security of data handling and storage systems, technical feasibility and whether the purpose for using the data meets the criteria of delivering benefits to the health and social care system.

Linkage between HES and other data sources is performed within NHS Digital. For example, HES does not contain information on deaths occurring out of hospital , but linkage with mortality data from the ONS (obtained from death certificates) can be requested, and is conducted based on deterministic linkage of NHS number, sex, date of birth and postcode. Where identifiers are complete, this linkage is straightforward. Linked HES-ONS data provide information on date, cause and place of death, but do not include stillbirths. Since only deaths that have been registered can be linked, delays in registration (e.g. whilst a coroner's investigation is completed) can result in a difference between the year of registration and the year of death.[44] Linked HES data are also used to support a number of cohort studies, trials, audits and disease registries. Typically, cohort data are initially "traced" using the Personal Demographics Service (PDS), which contains up-to-date identifiers for NHS patients. This tracing aims to provide complete identifiers, including NHS number, date of birth and postcode, before linkage takes place. PDS tracing can be used by NHS Digital to provide information on date of death (but not cause of death), thereby removing the need for additional applications for linked HES-ONS data.

Costs

As NHS Digital is a publicly funded body, the Data Linkage and Extract Service operates on a cost-recovery basis. This means that there are charges for processing and delivering the service, but no data costs. Charges vary according to what is required, but linkage between FNP and HES-ONS would come under a "bespoke data linkage" charge, of approximately £2000 (set up charge), £1000 (new application), £300 per dataset for per year, and £400 for auto-matching of up to 50,000 records. Any manual matching would be costed at £250 for 100 records.

Maternity and Children's Data Set

The Maternity and Children's Data Set (MCDS) was developed to provide comparative data on mothers, babies and children for commissioning and quality improvement purposes. As with HES, MCDS is re-use of clinical and operational data, for purposes other than direct patient care. MCDS data from April 2015 has been collected from three sources:

1. The Maternity Services Data Set (MSDS)
2. The Children and Young People's Health Services Data Set (CYPHS)
3. The Child and Adolescent Mental Health Services Data Set (CAMHS).

Information on the same mother and baby are linked using a local patient identifier (from local maternity information systems) and include details on booking and diagnosis, screening and tests, admissions, labour/foetus outcomes, postpartum details, baby screening, and sexual health and infectious diseases. Full technical specifications are available from digital.nhs.uk for both maternity and child data.

Although the MCDS is a mandatory system, not all providers are currently registered or submitting data. Services that operate on paper are currently exempt; however, all services will become obliged to implement electronic systems and to submit the data to NHS Digital in the future. The maternity data system became live in June 2015, but only around two thirds of providers are submitting maternity data, with regional variations. As of October 2015, about half of providers in London and in the Midlands were submitting good data, compared with three-quarters of providers in the North and less than half in the South. The child data system became live in October 2015, and around a quarter of providers are currently submitting child data.

A unique pregnancy ID, derived from NHS Number (reported to be 100% valid in MCDS, Table 2) and estimated delivery date which will link pregnancies across hospitals and time, is being planned (but is not yet available). Although aggregate data from the MSDS have been reported by NHS Digital (http://digital.nhs.uk/maternityandchildren/maternityreports), the MCDS data are still under development. MCDS records will be routinely linked to HES records within NHS Digital; however, requirements for access to individual-level data and costs for extracts are not yet clear (though it is likely that these will follow requirements for linkage with HES and ONS data). Access to MCDS data would be beneficial for evaluation and monitoring of the FNP, as it would allow comparator populations to be identified through linkage with FNPIS data or the addition of an FNP flag to individual child records in MCDS. As no retrospective data have been collected, complete data will only be available for future FNP mothers and their children from mid-2015 onwards.

Table 2: Maternity and Children's Data Set completeness and validity for 58.2% of Trusts contributing data (June 2015)[45]. *of those completed

Data Item	Complete	Valid*
Weekly alcohol units (mother at booking)	36%	100%
Care professional type code (pregnancy first contact)	57%	83%
Support status (mother at booking)	58%	82%
Status of folic acid supplement (mother at booking)	61%	83%
Employment status (partner at booking)	63%	86%
Pregnancy total previous stillbirths	64%	100%
Pregnancy first contact date	64%	100%
Pregnancy total previous live births	66%	100%
First language English indicator (mother at booking)	69%	89%
Pregnancy previous caesarean sections	69%	100%
Pregnancy total previous losses less than 24 weeks	69%	100%
Last menstrual period date	71%	100%
Employment status (mother at booking)	71%	89%
Person height (mother at booking)	72%	100%
Substance use status (mother at booking)	73%	84%
Cigarettes per day (mother at booking)	73%	100%
Mental health prediction and detection indicator (mother)	73%	97%
Person weight (mother at booking)	73%	99%
Organisation code (code of commissioner)	74%	86%
Estimated date of delivery method (agreed)	75%	100%
Complex social factors indicator (mother at booking)	76%	100%
Smoking status (mother at booking)	79%	94%
Physical disability status indicator (mother at booking)	82%	100%
Organisation code (residence responsibility)	84%	99%
Ethnic category (mother)	93%	90%
NHS number status indicator code (mother)	96%	86%
Postcode of usual address (mother)	99%	100%
NHS number (mother)	100%	100%
Person birth date (mother)	100%	100%
General Medical Practice code (patient registration (mother))	100%	99%
Organisation code (local patient identifier (mother))	100%	83%
Person death date time (mother)	100%	100%

Clinical Practice Research Datalink

The Clinical Practice Research Datalink (CPRD) is a primary care database containing information from routinely recorded GP data in England, Scotland, Wales and Northern Ireland. The dataset was originally established in 1987, became known as the General Practice Research Database (GPRD) in 1993, and expanded to the CPRD in 2012 following a

£60 million investment from the UK Government's plan for a secure data service to strengthen the international competitiveness of UK life sciences research.[46]

CPRD contains data on >11 million patients from 674 practices, with an active population of >4 million (approximately 7% of the UK population). Individual patients are assigned an encrypted unique identifier 'patid' and an encrypted practice identifier 'pracid'. Historic data from patients who have died or who are no longer registered with a CPRD practice are retained in the dataset although no further data are collected for those who de-register (unless they re-register with a CPRD practice, in which case they would be assigned a different patient identifier). Patients included in CPRD are broadly representative of the UK population in terms of age, sex and ethnicity, and the median duration of follow-up in CPRD is five years.

CPRD collates monthly extracts of anonymised electronic health record data from GPs. Participation in CPRD is at the practice level, so all patients registered with a participating practice are included in the dataset (unless an individual chooses to opt out). Data contained in CPRD include clinical data from consultations and referrals, demographic information, prescriptions, immunisations and test results. Data are coded using Read clinical classification codes.[47] A full list of data items is available in Appendix 3. Identification of patients, practices or clinicians in CPRD is not permitted.

Data access

Access to patient-level data is provided for research studies approved by the MHRA Independent Scientific Advisory Committee (ISAC). Linkage between CPRD and other data sources (including HES, ONS mortality data, and a number of disease registries) is available for a subset of approximately 75% of consenting practices in England (covering ~5% of the population).[48, 49] Bespoke linkage is available for individual research studies.[50] Linkage of CPRD data is performed via NHS Digital (acting as a trusted third party), based on NHS number, date of birth, sex and postcode. Although CPRD do not collect patient identifiers, these are sent by GPs to NHS Digital for linkage purposes (see www.cprd.com/researchpractice/researchgppractice.asp).

Costs

A license for the whole CPRD database currently costs approximately £155,000 per year, as the same costs are charged for public and commercial sector users. The costs for bespoke CPRD extracts vary according to the number of records to be linked and the complexity of the data. For linked CPRD-HES data on 10,000-50,000 individuals, discounted fees (available to projects funded by academic, NIHR or charitable sources) would be around £17,500, plus £3000-£5000 extraction costs and £5000-£10,000 for the linkage itself (personal communication, CPRD Knowledge Centre).

Other potential administrative data sources and future linkages

Beyond the key administrative data sources described in detail in this report, there are a number of other data sources that could help establish the cross-sectoral benefits of the FNP in the future, particularly as FNP participants mature. Within health, it could be possible to link FNP participants with data on mental health services held by NHS Digital (http://digital.nhs.uk/mhsds). However, routine access to mental (and sexual) health data is outside the remit of CAG, so alternative approvals would need to be explored. Outside health, access to and linkage with data sources on crime, employment, benefits and courts (Table 3) will require investment in time working with legal teams from different government departments, since there is a lack of experience and understanding of the legal bases in these areas. This work should build on experiences of consented linkage between survey and cohort data:

ELSA: Participants in the English Longitudinal Study of Ageing (ELSA) were asked for consent to linkage with health records, National Insurance records from the Department for Work and Pensions (DWP), and benefits and tax records from HMRC.[51] Approximately 80% of ELSA respondents (aged >50 years) provided consent.[52] ELSA linkage with HMRC, DWP and HES data is available for researchers in approved enclaves.[53]

ALSPAC: As part of the Project to Enhance ALSPAC through Record Linkage (PEARL), ALSPAC participants were asked for consent to linkage with education, economic and criminal data.[54] [55] Respondents were generally positive (87-96% depending on the data), but there were high rates of non-response and non-response bias mirrored that seen more generally in ALSPAC. Alternative legal routes are being explored for the situation in which consent would not be obtained: privacy notices may be appropriate for health and education data; it is less clear whether this is an appropriate pathway for criminal data; legislation makes this inappropriate for economic data.

Table 3: Future impact measures

	Maternal variable	Dataset	Child variable	Dataset
Employment	- Unemployment and benefits	- DWP/HMRC		
Crime	- Offenders / youth custody	- MoJ	- Offenders / youth custody	- MoJ
Family courts	- Divorce and separation	- CAFCASS		

DWP=Department for Work and Pensions; HMRC=Her Majesty's Revenue and Customs; MoJ=Ministry of Justice; CAFCASS=Children and Family Courts Advice and Support Service.

4. Permission pathways

The following section relates to the processing of personal data that is required for FNPIS linkage with administrative data. The FNP NU is the data controller for the FNPIS, meaning that the FNP NU is responsible for determining the purposes for and manner in which any personal data are processed.

Currently, identifiers collected through the FNP programme are held in local FNPIS on a NHS Digital server until four months after graduation (24 months post-partum) or until a participant leaves the programme (i.e. due to a still birth, miscarriage, a child being removed by social services or a family moving out of the catchment area). Approval for access to FNPIS is required from both the FNP NU and the local health provider's Caldicott Guardian. For participants who have graduated from FNP, access to the identifiable data is restricted to the NHS Digital Systems and Service Delivery section in Exeter, with approval from the data controller.

Processing of personal data

Under the Data Protection Act 1998 (DPA), one or more "conditions for processing" must be satisfied for the processing of personal data. Personal data are defined as relating to living individuals who can be identified from those data, or from data or other information in the possession of the data controller.[56] The most relevant of the conditions for processing FNP data for linkage with administrative data are 1) individual consent, or 2) "legitimate interests".

Consent

The Information Commissioner's Office (ICO) suggests that consent should cover the specific processing details, the type of information to be processed, and the purposes of the processing.[56] 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. Participants have the right to refuse consent. Programme management guidance on consent procedures do not explicitly state that consent for linkage with other datasets should be requested.

The "legitimate interests" condition may therefore be more appropriate for the use of retrospective administrative data for monitoring and evaluating the FNP. This condition would also cover use of administrative data for comparison groups (non-FNP mothers and babies), which are required for measuring effectiveness and for monitoring coverage of the programme.

Legitimate interests

To satisfy this condition, i) the personal information must be required for the purposes of the legitimate interests; ii) legitimate interests must be balanced against the interests of the individual(s) concerned, and iii) processing must be fair and lawful and comply with all the data protection principles. According to the ICO, fairness requires data processers to:

- be open and honest about their identity;
- tell people how any personal data collected about them will be used;
- usually handle personal data only in ways the individual would reasonably expect; and
- not use information in ways that unjustifiably have a negative effect on the individual.

The ICO recommends that "Privacy notices" are used to give information to individuals whose personal data are processed, and should state:[56]

- the data processor's identity;
- the purposes for which information will be processed; and
- any extra information required to enable information to be processed fairly (including mechanisms for opting out).

There are additional requirements for the processing of *sensitive* personal data.[57] Sensitive personal data includes an individual's ethnic origin or information on physical or mental health. There are two main requirements applicable to personal data relating to FNP: firstly that the processing is necessary for medical purposes, and is undertaken by a health professional or by someone who is subject to an equivalent duty of confidentiality; or secondly that the processing i) is in the substantial public interest; ii) is necessary for research purposes; iii) does not support measures or decisions with respect to any particular data subject and iv) does not cause, nor is likely to cause, substantial damage or substantial distress to the data subject or any other person.

Decisions and approvals

The use of personal (identifiable) data without consent for purposes other than direct patient care needs support from the Confidentiality Advice Group (CAG). In England, the Health Research Authority (HRA) is responsible for the regulation and governance of health and social care research under the Care Act 2014.[58] CAG is an independent group appointed to provide advice on uses of data set out in The Care Act 2014 and on accessing confidential patient information for research and non-research projects.

Under The Health Service (Control of Patient Information) Regulations 2002 (known as section 251 support), CAG provides advice on whether applications to process confidential patient information without consent should be approved. Section 251 recognises that essential activities of the NHS and important medical research sometimes require the use of personal data without consent. This regulation enables the common law duty of confidentiality to be overridden for defined medical purposes where it is not possible to use anonymised data and where seeking consent is impractical due to available funds or technology.[59] Section 251 is relevant for FNP data, since seeking additional consent from approximately 24,000 mothers already enrolled in FNP would be problematic in terms of accurate contact details, and would involve a disproportionate financial cost. In terms of gaining consent for linkage from prospective FNP participants, identifiable data are needed only to establish linkage with administrative data, and there is no need for individuals to be identified once the link has been established. A requirement for consent for linkage could introduce substantial bias into analyses, as vulnerable mothers would be likely to have a high non-response rate. Use of data without consent, through Section 251 support, may be the most appropriate approach for prospective participants (given appropriate processes as described below and privacy notices allowing opt-out) and was recommended in Scotland (see below).

CAG support is provided on a project-specific basis (and applies for the duration of the study subject to annual review) but is not equivalent to permission – final decisions on data access lie with either the HRA (for research applications), or the Secretary of State for Health (non-research applications, such as service evaluation or audit). Data providers, including NHS Digital and CPRD, consider CAG advice when reviewing applications for data. Full review of an application by CAG is expected to take up to 60 days; amendments can take another 30 days.

Relevant experiences

Applications to relevant approvals boards should build on the experiences of the Building Blocks trial and the FNP in Scotland. As described below, obtaining explicit consent for linkage from prospective FNP participants may not be required.

FNP Scotland: Advice given to the FNP NU following the Scottish Government Ethics Peer Review process recommended that FNPIS data should be treated in the same way as all other NHS Scotland data, and that further explicit consent for linkage with NHS data was not required. It would be impractical to obtain consent for past FNP participants in both England and Scotland given the large numbers involved. A requirement for consent would also limit the usability of the data: introduction of bias would be a major concern due to consent being more difficult to obtain from the most vulnerable mothers, who would then be excluded from analyses. Analysis of FNPIS data is in line with the purposes for which consent was originally obtained (service delivery and research). In addition, mothers who originally dissented are not included in the FNP cohort.

Building Blocks: The trial team recognised that obtaining consent for follow-up would be problematic for several reasons: the sensitivity of the maltreatment focus of the study, the relative mobility of participants, the introduction of non-response bias and the cost and logistical requirements of securing high levels of consent if contact were possible. Support to link with administrative health data for the follow-on trial was therefore obtained under Section 251 from CAG and approval for linkage with education data was obtained under Schedule 2 of the DPA. To satisfy the legal bases, participants have been contacted via post, email and/or mobile, informed of the study, and provided with a two-month period in which to opt out before any data are accessed. Participants can also opt out at any other time. Contact details collected at enrolment were updated from GP registration data using the NHS Wales Shared Services Partnership (NWSSP). Identifiers will be accessed only by NHS Digital (health data) and DfE (education data); linked anonymised data are then held within a secure haven (SAIL databank in Swansea).[60]

General requirements

In addition to the legal basis, the following conditions apply to the use of personal data under the DPA:

- Appropriate security arrangements are in place to process the data;
- Data are used only for the specified purpose;
- Data are kept only for the specified length of time; and
- Data are not further disclosed.

Approvals panels consider several aspects of applications to ensure that requests meet the above requirements:

Security: Evidence of minimum security standards, e.g. up to date reviewed Information Governance Toolkit score or ISO27001:2013. The Tavistock and Portman NHS Foundation Trust hold a satisfactory IG Toolkit assessment (94%) for 2014/15 (scores available at www.igt.hscic.gov.uk). Detailed information on security requirements are available from NHS Digital.[61] Increasingly, data users are expected to store and access data only within approved secure settings, known as 'Data Safe Havens'.[62] Secure settings can be physical or virtual.[63] For example, the Administrative Data Research Centre for England (ADRC-E) provides a physical safe room at the Farr Institute in London and allows access for approved projects (https://adrn.ac.uk/protecting-privacy/secure-environment). In addition, UCL provides access to a safe haven through a virtual environment (www.ucl.ac.uk/isd/itforslms/services/handling-sens-data/tech-soln).

Data Protection Registration: The Tavistock and Portman NHS Foundation Trust has Data Protection registration (Z1595854, expires January 2017, available at https://ico.org.uk) which covers the use of personal information for research and for providing healthcare services for patients.

Data destruction: Data cannot be held in perpetuity, but only for a specified length of time defined by data controllers, and evidence of data destruction is required when this timeframe expires.

Ethics approval: Additional approval from the National Research Ethics Committee is required for using identifiable data in research studies that do not fall under service evaluation. Applications are submitted via the Integrated Research Application System (IRAS) and considered by a panel of lay and expert committee members.

Data specific requirements

For each of the key administrative datasets described in this report, it is recommended that individuals or organisations intending to apply for data contact the relevant data team to discuss their application before it is submitted. CAG can also provide initial advice for specific studies before full applications are submitted. Decisions on whether a specific study is research or non-research (e.g. service evaluation) can be informed by the HRA Decision Tool (www.hra-decisiontools.org.uk/research).

One key element common to all applications is a clear data flow diagram, which maps out flows of identifiable and de-identified data to all relevant parties. Direction of data flows should be justified by proportionality. For example, release of FNPIS identifiers to DfE for linkage with NPD would involve a smaller number of records being transferred than if identifiers for all records in NPD were released to NHS Digital for linkage with FNPIS. Example data flow diagrams are provided for the case studies in Appendix 5.

NPD/CLA/CINC

Access to the NPD is considered for organisations who are conducting research or analysis, producing statistics, or providing information, advice or guidance for the purpose of promoting the education or well-being of children in England. Application processes depend on the sensitivity of data. Access to tier 1 data requires justification for each sensitive item, and approval by the DfE Data Management Advisory Panel (DMAP). Requests for tier 3-4 data are dealt with by the NPD Data Request team in the Education Data Division. For requests relating to tier 1-2 data, data users must also provide a 'basic disclosure' certificate (available at www.disclosurescotland.co.uk/basicdisclosureonline/index.htm). Data users must sign a license agreement and individual declarations before data are released, and the timeframe for which data are held should be specified. For NPD, data applications for standard, non-sensitive extracts are expected to be processed within 6-8 weeks. Applications to the DMAP are assessed at monthly meetings and may take considerably longer to be processed.

HES-ONS

Applications for data held by NHS Digital are submitted to the Data Access Request Service (DARS). DARS review the legal basis for accessing the data, suitability of data handling and storage systems, technical feasibility of the request and the purpose of the application, before submitting the application to the Data Access Advisory Group (DAAG). Importantly, DAAG consider whether the request is compliant with The Care Act 2014: NHS Digital may only disseminate information for the purposes of a) the provision of health care or adult social care, or b) the promotion of health. DAAG require that applicants are specific about how these benefits will be achieved by the research or service evaluation.

Currently, DAAG makes a final decision on data access and if approved, a Data Sharing Agreement and Contract are signed by both the applicant and NHS Digital. However, DAAG is currently being transformed and will be succeeded by the Independent Group Advising on the Release of Data (IGARD). Permissions are reviewed annually and subject to audit. For NHS Digital data, expected processing times are up to 16 days for applications to be processed, up to 30 days for approval, and up to 14 days for access. In reality, the total time taken can be much longer, as these are NHS Digital in-house timelines and do not include applicant time or time taken to coordinate other approvals (e.g. CAG/ONS/ethics).

Additional permissions are required for use of ONS mortality data. There are two main routes for this: projects conducted within the FNP NU as part of the Tavistock and Portland NHS Foundation Trust are covered under s42(4) of the Statistics and Registration Service Act 2007, or researchers can apply for Approved Researcher status on a project-specific basis. Alternatively, date of death (but not cause of death) can be provided from PDS by NHS Digital, without the requirement for separate ONS approval.

MCDS

As MCDS is still under development, no requests to date have been made to use these data. However, requirements for access will follow those for HES as described above.

CPRD

Applications for CPRD data are considered by the MHRA Independent Scientific Advisory Committee (ISAC), alongside a detailed protocol and investigator CVs. Access to CPRD data is only approved for medical research benefiting the public and is granted on a study-specific basis. ISAC considers security measures, information governance, scientific standards, and patient or user group involvement. The involvement of investigators with primary care experience is viewed as essential for the majority of studies using CPRD. The average first response time for initial applications to ISAC is five days.[64]

Anonymised FNPIS-linked data

Many linkage studies follow the 'separation principle', which means that identifiers are held separately from attribute data (e.g. clinical or demographical variables of interest).[65] This principle is typically upheld using a trusted third party (e.g. NHS Digital) to link the identifiers, which are held separately from attribute data.

Following linkage, personal data may not need to be processed further if data can be fully de-identified; meaning that wider use of properly anonymised, linked FNPIS data would be possible without the same level of approval. Data anonymisation or de-identification relies on principles of statistical disclosure control, a technical process that limits the opportunity for finding an individual within a dataset (identification) or finding out something about an individual (attribution). Statistical disclosure control assesses and mitigates these risks, for example by supressing small cell sizes or only using high levels of geography.[66]

Linkage with multiple administrative datasets

The permission pathways described above relate to the linkage of FNPIS with each administrative dataset in turn, with the exception of CPRD, which can provide linkage with both CPRD and HES for a subset of contributing GPs. Linkage of FNPIS with multiple datasets simultaneously could potentially result in a more disclosive linked dataset, even if identifiers were stripped out. For example, bringing together data on health and education of teenage mothers could result in a greater risk of re-identification.[67] Linkage between administrative datasets (in addition to between FNPIS and single administrative datasets) would also increase the complexity of approvals processes.

Depending on the questions being addressed, simultaneous linking of multiple administrative datasets may not be necessary. However, if such linkage were justified for the purposes of monitoring or evaluating the FNP, one approach would be to create a dataset containing only the minimal number of variables necessary that defined purpose. Such a dataset would need to be functionally anonymous, i.e. ensuring that both the risk of re-identification is sufficiently small and the environment within which the data are accessed is adequately controlled through governance, secure settings and sanctions.[63] For example, rather than raw data variables, derived variables (e.g. indicators of health status derived from CPRD) or categorisation (e.g. of school achievement or attendance) could be used to ensure that the risk of re-identification is minimised. The ICO Anonymisation Code of Practice provides examples of anonymisation techniques including removal of records, sampling, data reduction, and data perturbation.[68]

5. Technical requirements

This section describes potential methods for linking FNPIS data with administrative data based on an assessment of the availability, accuracy and completeness of identifiers common to FNPIS and administrative datasets.

The FNPIS holds complete data for both the mother and baby on name and date of birth, almost complete postcode, and a reasonable proportion of NHS numbers (~99% for mothers, ~80% for babies, see Table 4). Since the unit of interest in FNP is the family, identifiers for mothers and babies are held together. It therefore makes sense to use FNPIS as the primary linkage file, i.e. the file to which other datasets are linked in turn. To maximise the potential for linkage, FNPIS identifiers could be updated and completed using the NHS Tracing Service, which uses information from the Personal Demographics Service (PDS) coordinated by NHS Digital.

PDS

The PDS is a master index of patient records containing name, address, date of birth and NHS number for patients receiving treatment in an NHS setting in England, Wales and the Isle of Man (see Appendix 4 for full list of data items). No clinical or sensitive information is retained in PDS. PDS is used by authorised healthcare staff to identify patients (for example, to look up an NHS number for an individual presenting in A&E). National data are held by NHS Digital and can be used to obtain additional identifiers to facilitate data linkage, for example to obtain contact details for cohort study members or to notify a GP of a patient's participation in a study.[69, 70]

PDS was launched in 2003, originally populated from primary care data via the NHS Strategic Tracing Service. Patients without a record on PDS are generally those with no previous contact with the NHS (and therefore no NHS number). New PDS records are created when a patient makes first contact with the NHS, either as a newborn baby, when registering for a GP, or at a number of secondary care organisations. For newborn babies, NHS numbers are allocated by PDS after the midwife registers the birth on PDS (replacing the NN4B system that was decommissioned in 2015). Where systems are not compliant with the PDS, births are registered via the Birth Notification System. This NHS number is then made available to organisations providing ongoing care for the baby (child health services, the NHS Newborn Hearing Screening Service and the ONS). Stillbirths are assigned NHS numbers, provided gestational age is at least 24 weeks. PDS also receives updates from birth and death registrations (either from the General Register Office where a formal death certificate has been issued, or from the NHS where a Medical Certificate of the Cause of Death has been issued, see http://systems.digital.nhs.uk/demographics/pds/contents/deaths).

Birth notifications submitted to PDS include information on time of birth, birthweight, gestational age, place of birth, birth order and mother's NHS number. These data are used to identify duplicate birth notifications and are only retained for six months, with the exception of time and place of birth, and mother's NHS number.[71]

Access to the PDS 'tracing' service is facilitated through the Demographic Batch Service (DBS) at NHS Digital.[72] The DBS allows batches of records to be checked against PDS, and returned with the latest PDS data for each matching patient (including NHS number). DBS can be accessed by NHS organisations for direct health or social care or for performance management purposes, through the DBS client software on a secure N3 server.[73] Additional approvals are required for using the DBS for research purposes (e.g. section 251 approval from the Confidentiality Advice Group; CAG).[73] As FNP identifiers are already held at NHS Digital, access to the tracing service would likely be done within NHS Digital.

Table 4: Availability and completeness of identifiers across data sources.

Only FNPIS reliably contains the link between mother and child; for the remaining data sources, presence of identifiers is for individuals only. For example, date of birth would be available in CPRD for the mother and baby as separate individuals.

Identifier	FNPIS	PDS	HES-ONS	CPRD	NPD	CLA / CINC	MCDS
Mother's NHS Number	✓ ~99%[$]	✓ 100%	✓ ~97-99% APC/OP; ~85-94% A&E[74]	✓	X	X	✓ 100%
Mother's first name	✓ 100%	✓[£]	X	X	✓	X[&]	X
Mother's last name	✓ 100%	✓[£]	X	X	✓	X[&]	X
Mother's date of birth	✓ 100%	✓	✓ 99%	✓	✓	✓ 100%	✓ 100%
Mother's postcode	✓ ~ 99%[$]	✓[£]	✓ 99%	✓	✓	✓ 58%**	✓ 99%
Mother's GP code	✓ ~ 94%[$]	✓	✓ 100%	✓	X	X	✓ 99%
Child's NHS Number	✓ ~80%	✓ 100%	✓	✓	X	X	✓
Child's first name	✓ 100%	✓[£]	X	X	✓*	X[&]	X
Child's last name	✓ 100%	✓[£]	X	X	✓	X[&]	X
Child's sex	✓ 100%	✓	✓ 100%	✓	✓ 100%^	✓ 100%	✓
Child's date of birth	✓ 100%	✓	✓ 99%	✓	✓ 100% ^	✓ 100%	✓
Child's postcode	X	✓[£]	✓ 20%	✓	✓ (LSOA 100%)^	✓ 58%**	X
Child's ethnicity	X	X	✓ 92%	✓ 78%[75]	✓^	✓ 99%	X
Multiple birth flag	✓	✓	✓ (unreliable)	X	X	X	X
Unique Pupil Number	X	X	X	X	✓	✓***	X

[$] since 2011, most recent postcode; ^since 2001; *middle name collected from 2004/5; **postcode is collected for children in continuous care for at least 12 months but DfE have raised concerns about data quality; [£]historical postcodes/names available; ***children in care at school-age; [&]not in national data; LSOA=lower layer super output area; OP=outpatient; APC=admitted patient care.

PDS for linkage

Since PDS contains both NHS number and name, it is an important link between datasets from different sources. For example, PDS could be used to complete NHS Number for individuals in FNPIS for whom NHS number is missing. Historical postcodes and names (e.g. where an individual moved addresses or changed their name through marriage) are also captured in PDS, which means it could also be used to obtain NHS Number for records captured in the NPD (postcodes and names recorded within FNPIS may have changed by the time a child reaches school age and is recorded within NPD).

Importantly, PDS also provides a method for linking records for mothers and babies. Since 2006, the Mother's NHS number has been held on a relationship record allowing the mother and the baby to be linked together, although this is not routinely available outside of NHS Digital systems. Although CPRD does contain a database-specific mother-baby link, such a link is not currently available in HES. Using PDS to create this link would avoid the need for the more complex linkage methods that would be required for linking based on common characteristics captured in the HES maternity data (e.g. date of delivery, birth weight, and gestational age).

Linkage methods

In data linkage, records from different datasets are brought together with the aim of linking information belonging to the same individual. For the purposes of monitoring and evaluating the FNP, the aim of the linkage will firstly be to flag FNP participants in administrative data sources, using identifiers for the mother and baby held in each dataset. A second purpose of the linkage is to bring together information from FNPIS and administrative datasets, e.g. to validate outcomes captured in both sources. High quality linkage will find the same individual in more than one dataset; linkage error occurs either when no match can be found (missed-match) or when records belonging to different people are linked together (false-match). There are two broad fields of linkage methods: deterministic and probabilistic.

Deterministic linkage

Deterministic linkage is a relatively straightforward linkage method, typically requiring agreement on a unique identifier (such as NHS number) or on a specified set of common identifiers (e.g. name, sex and postcode).[76-82] Deterministic methods are useful when records have highly discriminative or unique identifiers that are well completed and accurate. For example, the community health index (CHI) is used for much of the linkage in the Scottish Record Linkage System.[83] Modifications of strict deterministic linkage allow for small differences in identifiers, by using a succession of rules. For example, the deterministic algorithm used to link hospital admission records for the same individual in Hospital Episode

Statistics is based on a sequential set of rules looking for agreement on a combination of NHS number, date of birth, sex, local patient identifier, and postcode.[84]

Deterministic methods are designed to avoid false-matches, since it is unlikely that different individuals will share the same set of identifiers. On the other hand, deterministic methods requiring exact agreement on identifiers are prone to missed-matches, as any recording errors or missing values can prevent identifier agreement.[85, 86]

Probabilistic linkage

Probabilistic methods were originally proposed as a means to allow linkage in the presence of recording errors or for data without a unique identifier.[87, 88] In probabilistic linkage, a match weight is created to represent the likelihood that two records are a true match, given agreement or disagreement on a set of common identifiers.[89] Each identifier contributes separately to the overall match weight, where agreement on an identifier contributes positively to the weight, and disagreement contributes a penalty. The size of the contribution depends on the discriminatory power of the identifier, so that agreement on name makes a larger contribution than agreement on sex.[90]

In probabilistic linkage, links are determined by comparing match weights to a threshold or cut-off match weight. Thresholds are typically chosen through manual inspection of pairs of records: records with weights above an upper threshold are classified as links and records below a lower threshold are classed as non-links. Any record pairs between the thresholds are classed as uncertain and subjected to manual review, where feasible. Alternatively, an automated single threshold is chosen depending on a maximum tolerated level of error.

Probabilistic linkage is more computationally intensive than deterministic linkage, and requires some user-input in choosing thresholds and manual review, but it typically results in a greater number of matches than deterministic linkage alone.[91, 92] In practice, linkage studies often use a combination of deterministic and probabilistic methods, using initial deterministic steps to reduce the number of comparison pairs for probabilistic linkage.[93]

Designing a linkage algorithm

A combination of deterministic and probabilistic methods would seem an appropriate approach for linkage of the FNPIS with administrative data sources. There is a high level of completeness of FNPIS identifiers (particularly if traced in PDS), which lends itself towards deterministic linkage using a combination of name or NHS number, postcode, sex and date of birth. However, identifier quality in administrative data sources can be variable over time and for different groups of individuals.[94] Evidence shows that vulnerable or minority groups of individuals often have poorer data quality and are therefore more difficult to link: probabilistic linkage may be helpful for capturing additional links in these groups.[95]

One group, for whom linkage will be extremely difficult (even with probabilistic linkage), is children captured in CINC or CLA only before nursery or school age; national data do not contain sufficient identifiers to link outcomes for these children.

Although probabilistic linkage is used routinely with administrative sources in Wales and Scotland, there is a lack of experience in this approach within trusted third parties in England, where deterministic linkage has traditionally been favoured. The following sections provide examples of methods used previously to link HES, CPRD and NPD datasets.

NPD

- Linkage between the NPD and the Millennium Cohort Study was performed deterministically, with manual review, based on name (first and last), sex, date of birth, postcode of residence, name of school, reference number of current school, Local Education Authority of current school, school number of current school.[96]
- Linkage between the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort and the NPD was conducted deterministically, using name, date of birth, sex and address.[97]
- Linkage between NPD and Understanding Society (the UK household longitudinal study) was performed deterministically using names, gender, date of birth and postcode. Additional "fuzzy" matching steps were incorporated, comparing the first four, three and two characters of names and postcodes manual checking.[98]

HES

- Internal linkage within HES (linking episodes of care for the same individual over time) is performed using a 3-pass deterministic algorithm:[86]
    - NHS number, sex, date of birth
    - Postcode, sex, date of birth, local patient identifier
    - Postcode, sex, date of birth.
- Linkage between HES and the Mental Health and Learning Disabilities Data Set is performed using a more comprehensive 9-step algorithm using NHS number, date of birth, postcode and sex and incorporating steps that allow matches on partial date of birth.[99]

CPRD

- Linkage between CPRD and the Myocardial Ischaemia National Audit Project (MINAP) is conducted based on deterministic linkage of NHS number, date of birth, postcode and sex.[100]
- Linkage with HES is conducted based on an 8-step deterministic algorithm (Table 5).

Table 5: Deterministic linkage algorithm for linkage between CPRD and HES

Step	NHS Number	Date of Birth	Sex	Postcode
1	Exact	Exact	Exact	Exact
2	Exact	Exact		Exact
3	Exact	Partial	Exact	Exact
4	Exact	Partial		Exact
5	Exact			Exact
6	Exact	Exact		Exact
7	Exact	Exact		Exact
8		Exact		

Evaluating linkage quality

Linkage error occurs when records belonging to different people are linked together (false-matches) or when a link cannot be made at all (missed-matches). These errors typically occur when identifiers are not sufficiently discriminative, or when available identifiers are prone to missing values, recording errors, or changes over time (such as postcodes or married names).

The impact of linkage error can be substantial. For example, incorrect linkage of an FNP child with a CINC record for a non-FNP child (a false-match) could results in an overestimation of child protection plans for FNP participants. False-matches can also dilute true relationships and tend to lead to bias towards the null.[101] On the other hand, if we failed to link a FNP child with their corresponding record in CINC (a missed-match), we could underestimate the number of child protection plans. Missed-matches can reduce the number of records available for analysis and so result in a loss of statistical power. In addition, if particular groups are systematically less likely to link (due to poorer data quality) there is a risk of selection bias, with specific groups of mothers/children being excluded from analysis.[102-105]

Assessment of linkage quality will be vital, to highlight any limitations of the linked FNPIS data and to allow these to be considered within choice of study design and analyses.[106] The most useful methods for evaluating linkage quality are:

- Comparing linked data with reference or ‘gold-standard’ datasets where the true match status is known;[107]
- Structured sensitivity analyses where a number of linked datasets are produced using different linkage criteria;[102]
- Comparisons of characteristics of linked and unlinked data to identify any potential sources of bias;[108] and
- Statistical methods (e.g. for missing data).[109]

For FNP linkage, approaches for assessing linkage quality will need to be discussed with the trusted third party. One straightforward approach for data linkers would be to create a gold-standard dataset through manual review of a sub-sample of records, on which to test linkage algorithms and quantify linkage error using different linkage criteria. Users of the linked data could then compare the characteristics of FNP participants, who have successfully linked, with those of participants for whom linkage was not possible. This comparison would help to assess the representativeness of the linked data for providing accurate monitoring and evaluation of the FNP.

6. Methodological considerations

This report aims to establish how linked FNPIS-administrative data could be used for two purposes; firstly, to monitor the implementation of FNP at national and local level and secondly, to evaluate its impact on child and maternal outcomes. This section discusses important methodological considerations for both monitoring and evaluation in turn. These considerations (relating to data validity, statistical power or study design, for example) will vary according to the data being used and the purposes of the analyses; detailed case studies of linkage to key administrative datasets using exemplar questions relevant to FNP are provided in Appendix 5. Considerations for sub-national analyses are also addressed in this section.

Monitoring

Monitoring of the FNP based on data collected as part of the service has traditionally been guided by US license requirements relating to the following five aspects of the programme:

- Recruitment and enrolment: Enrolment early in pregnancy with at least 60% of mothers enrolled before 16 weeks of pregnancy and 100% before 28 weeks. High level of uptake of places with at least 75% of mothers who are approached enrolling in the programme.
- Retention: Minimised attrition preferably <40%: at most 10% during pregnancy, 20% during infancy, and 10% during toddlerhood.
- Dosage: Maximised number of visits delivered with at least 80% of expected visits during pregnancy, at least 65% in infancy, and 60% in toddlerhood.
- Programme content: The content of visits as recorded by Family Nurse, which is expected to be consistent with broad goals for each stage of the programme.
- Short-term outcomes: A number of short-term outcomes are currently captured and monitored for FNP mothers and their children. However, as no comparable data are collected on non-FNP mothers or children, the utility of these measures is limited.

FNP practitioners review data based on these indicators at the individual client, nurse and site level on a regular basis and use this to inform their quality improvement inquiries and actions. Local Advisory Boards and commissioners also use the aggregate data to reflect on progress and identify improvements, priorities and actions. A similar process is replicated nationally. All these indicators are available to local FNP sites and the FNP NU on a real-time basis through the FNPIS. Reflecting and responding to such data is an integral part of the FNP programme.

National administrative data - unlinked

Unlinked administrative data at a population-level could support monitoring by providing additional information on the underlying population and variation in uptake and achievement of fidelity targets in different subgroups of mothers, or across LAs. For example, geographical information on the rate of births to teenage mothers (overall and by ethnic group and/or quintile of deprivation) could be used to explore variation in uptake of the FNP programme. Table 6 provides illustrative numbers of teenage mothers for the whole of England based on an extract of HES data for 2012/13. These data could be compared with FNPIS data at a national and local level to identify potential barriers to reaching particular groups of mothers by highlighting groups and/or particular areas that do not appear to be accessing FNP. This could inform the development of new approaches to enrolment or commissioning. However, as lower rates of uptake may be due to lack of access to FNP (i.e. women were not offered the programme) or due to refusal to participate, such assessments should take into account local information on FNP integration within existing services. The number of nurses available and their client capacity, as well as information on the number of refusals to participate in FNP (which is not currently monitored nationally but is recorded by local FNP teams) should also be considered.

Unlinked administrative data could also be used to inform the eligibility criteria for FNP. For example, HES data could be used to compare characteristics and outcomes of mothers eligible for FNP (i.e. first time teenage mothers) with other groups of vulnerable women in the wider population who are not currently eligible for FNP, such as teenage mothers with previous births or women with a history of adversity-related hospital admissions. The potential for addressing this question (and the relevant technical and methodological considerations) are discussed in detail in Appendix 5, as part of Case Study A (Q3).

Table 6: Illustrative number of teenage mothers in England captured in HES during 2012/13 by ethnic group, year of age and deprivation quintile (N=26,490)

Ethnicity	N (%)	Maternal age (years)	N (%)	Deprivation* quintile	N (%)
White	22,532 (85)	*13*	11 (0)	*Most deprived*	8,725 (33)
Mixed	547 (2)	*14*	110 (0)	*2*	6,514 (25)
Asian	464 (2)	*15*	581 (2)	*3*	4,993 (19)
Black	690 (3)	*16*	1,864 (7)	*4*	3,747 (14)
Other	472 (2)	*17*	4,563 (17)	*Least deprived*	2,489 (9)
Unknown	1,785 (7)	*18*	7,756 (29)		
		19	11,605 (44)		

* Index of Multiple Deprivation

National administrative data - linked

Linked administrative data could provide further detailed information on FNP coverage among first time teenage mothers with additional risk factors. For example, linkage with HES and CPRD could provide comparable data on the proportion of eligible mothers with underlying physical or mental health conditions. As an illustration, almost 4% of teenage girls in CPRD have an indicator of depression each year (0.7% have a diagnosis of depression, 1.5% have depressive symptoms, and 1.8% have been prescribed antidepressants).[110] Such linked data would provide information on the prevalence of specific conditions in the wider population of mothers, and the relative proportion of these women captured by FNP. Flagging of FNP participation in CPRD would allow assessment of whether teenage mothers with mental health problems are being appropriately targeted and enrolled in FNP. Since CPRD does not disclose data at the regional level, such analyses would be performed nationally (unless CPRD were commissioned to provide aggregate outputs at LA level).

Similarly, linkage with NPD and CLA could identify vulnerable groups in which there is a high burden of teenage pregnancy. As an illustration, approximately 35,000 (0.02%) girls aged 13-17 were placed in out-of-home care during 2012 in England (Table 7). The ability to identify FNP participants within these data, and compare placement characteristics with non-participants, could help to highlight groups of mothers who are systematically missed from the programme. The potential for addressing this question (and the relevant technical and methodological considerations) are discussed in detail in Appendix 5, as part of Case Study C (Q9). Table 8 provides maternal characteristics captured in administrative data sources, which could be used to monitor coverage of the FNP.

Table 7: Number of female teenagers in out-of-home care in England during 2012, N=35,118 (unpublished CLA data, personal communication with L Mc Grath-Lone)

	Category of care	N	%
Placement type	Placed for adoption	1,327	3.8
	Foster care (kin)	3,337	9.5
	Foster care (stranger)	22,658	64.5
	With parents	1,440	4.1
	Secure unit	104	0.3
	Children's home	3,257	9.3
	Residential care home	650	1.8
	Residential school	336	1.0
	Other residential accommodation	671	1.9
	Independent living	1,193	3.4
	Other	146	0.4

Table 8: Additional maternal variables required for monitoring coverage of the FNP in different groups and for comparing effectiveness between FNP participants and control groups.

	Maternal variable	Dataset
Pregnancy health behaviours and birth	- Gestational age at first antenatal assessment - Hospital admissions and attendances - Length of stay in hospital at delivery - Primary care - Prior pregnancies (still births / miscarriages)	- HES - HES - HES - CPRD - HES
Maternal health	- Physical health - Mental health	- CPRD/HES - CPRD/HES
Education / School attainment	- Free School Meals / Special Educational Needs - Attendance and exclusions	- NPD - NPD
Neglect and abuse, safeguarding / social care status and service use	- Out-of-home care - Child protection plan - Referrals to social services (Child in Need)	- CLA - CINC - CINC
Social status	- Index of Multiple Deprivation	- HES/NPD*
Demographics	- Ethnicity - Age	- HES/NPD - HES/NPD

*via lower layer super output area (LSOA)

Evaluation

Administrative data provide the opportunity to evaluate a number of routinely-collected, relevant short- and long-term outcome indicators for mothers and their children (Table 9). Compared with RCTs, population-based analyses provide additional statistical power (e.g. for evaluating small changes and/or rare outcomes) and a complete representation of the FNP population over time (taking into account evolving implementation). For example, death is a key measure for both the mother and the child: child mortality rates for vulnerable mothers using drugs or alcohol during pregnancy have been reported to be as high as 8% by nine years after their child is born.[111] It is important to capture information on rare outcomes, such as death or serious injury, on a national level. Population-based analyses also provide statistical power for evaluation relevant sub-groups of families, some of whom have relatively small populations.

Comparison groups

Although choice of study design will ultimately be informed by the question being addressed, one of the most important considerations for evaluating the effectiveness of the FNP is the use of a robust comparator group. Linkage with administrative data will allow flagging of FNP families, and could include a cohort of non-FNP families, whose outcomes can be compared. It is important to consider which comparison group is the most relevant for answering a particular question:

- Eligible mothers within FNP areas: not able to participate (i.e. first time teenage mothers in an LA before FNP began)
- Eligible mothers within FNP areas: who did not participate (i.e. first time teenage mothers in an LA during FNP period who were not invited to participate, due to lack of capacity or issues with recruitment, or who refused the offer to participate)
- Eligible mothers in non-FNP areas
- Non-eligible mothers within FNP areas (e.g. aged 20-24 or multiparous teenage mothers)

Sample size considerations

There are two options for creating comparison groups: an entire population captured in administrative data or a defined sample. The size of the comparison group required depends on the event rate of the outcome being assessed and the expected difference between groups. For example, evaluating the effect of FNP on numbers of GP consultations may only require a sample of non-FNP mothers captured in CPRD; evaluating the impact on numbers of children placed in out-of-home care up to age five (a rare outcome) would require larger numbers of non-FNP children captured in CLA data. The potential for addressing this question (and the relevant technical and methodological considerations, including required sample size) are discussed in detail in Appendix 5, as part of Case Study C (Q8).

Sample sizes will also need to take into account the geographical level and period of analyses. In total, approximately 24,000 families were enrolled in FNP in England between 2010 and 2015. However, at the LA level, the median annual number of enrolments from 2010-2015 was 32 (interquartile range 0-56). Most analyses will therefore need to be undertaken using groupings of LAs with similar characteristics, or focussing on LAs with larger numbers of participants and/or combining data from all available years. Decisions on sample sizes will need to be made on a question-specific basis, taking into account the outcome of interest and expected impact of the intervention.

Validity of outcomes in administrative data

Although administrative data provide a valuable resource of detailed information on a population-level, there are concerns about the quality of data collected for purposes other than research. Linkage between FNPIS and administrative data sources provides an important opportunity for validating outcomes captured routinely in both sources. For example, flagging of FNP participants in HES and CPRD will allow validation of indicators of pregnancy and birth outcomes in both these sources. Comparison between indicators captured in FNPIS and administrative data sources will allow validation and measurement of data quality. Variables shaded in Table 9 are those where quality of administrative data is unclear and where validation is required before these variables are used:

- Smoking in pregnancy/BMI: Validity of these variables in primary care data is reasonable for adults aged >16 years but there is no evidence on quality of these data for teenagers.[112, 113]
- Illegal drug and alcohol use in pregnancy: Only the more severe cases would be captured in hospital admissions during pregnancy with diagnoses relating to drug and alcohol use. A&E attendance data do not include sufficient information on diagnoses to capture these indicators.
- Domestic violence: Severe cases would be captured in HES; the validity of coding of domestic violence within primary care data is unclear. Emotional abuse would not be captured.
- NICU / SCBU: Available in the HES maternity tail, but not 100% complete.
- Nutrition: Hospital admissions for jaundice or feeding problems would be captured in HES.
- Contraceptive use: Available in CPRD if prescribed by a GP.

Evaluating and monitoring FNP at sub-national level

Funding for children's public health services transferred from NHS England to LAs in October 2015. As a result, LAs increasingly require evidence of the implementation and effectiveness of services (such as FNP) in their local area, to monitor service provision, coverage and delivery and to support informed decision making regarding commissioning. More granular analyses (relevant to local service providers and commissioners) could be achieved using linked or unlinked administrative data held at a LA level. However, analyses by LA, group of LAs or region would also be possible using national administrative data that contain geographical variables such as LSOA or postcode.

Using local data to monitor or evaluate FNP has some advantages in comparison to conducting sub-national analyses of national administrative data. Firstly, more timely linkage and/or access to data may be possible at LA level. National administrative data are typically collected quarterly or annually and there is a time lag before it is made available for request by researchers and other organisations. In comparison, data can flow in an ongoing manner at LA level data (once data sharing agreements have been established and technical pathways developed). Local datasets may also contain information that is not routinely available nationally. While national administrative datasets are created from data returns from local services, additional data that is not part of these mandated returns may also be collected locally (for example, details of social care interventions that are not included in CLA). As clinical record keeping continues to move away from paper-based systems, the volume of data available electronically at a local level is likely to increase.

However, local data presents a number of challenges. Development of local data linkage systems is a non-trivial undertaking that requires significant time and investment to develop the legal framework, permission pathways and relationships with data providers to access the data, and to develop the technical solutions and expertise required for its secure transfer, linkage, storage and analysis. These challenges are highlighted in Case Study D in Appendix 5 that explores data linkage in Kent, an LA that is considered an early implementer of integrated intelligence. As processes and pathways for data linkage would be established separately within individual LAs, a lack of consistent standards may hinder the ability to conduct comparative analyses between LAs or regions using locally linked data. For example, comparisons of statistics derived from local datasets may be unfair if the methods and sources of data used to create linked datasets differ (e.g. variation in linkage quality, data quality, matching algorithms or in measurement of outcomes). Longitudinal and comparative analyses within an LA may also be problematic: depending on the size of the population (i.e. the number of mothers enrolled in FNP) and the frequency of the outcomes of interest, comparative analyses within an LA may have insufficient power to detect significant changes. For example, to detect a significant decrease of 20% in the rate of injury-related admissions by age five would require a sample size of 13,500 children (6,750 FNP vs. 6,750 non-FNP), which would not be possible for a single LA (see Case Studies A-C in Appendix 5 for further discussion of required sample sizes). Furthermore, without information related to maternal age and parity, appropriate comparison groups will not be identifiable in local datasets. Finally, as local linkage would only include prospective data, the immediate utility and potential for analyses of longer-term FNP outcomes at a local level will be limited.

Taking these strengths and limitations into consideration, analyses at the level of individual local authorities do not appear to be feasible (due small sample sizes, insufficient statistical power and a lack of appropriate comparison groups). Linkage of national FNPIS and administrative data with sub-national analyses using groups of LAs or regions would be preferable and of greater utility in terms of evaluating and monitoring the FNP.

Table 9: Key impact measures for evaluating the FNP. Italics indicate uncertainty in data quality; shaded items are collected to 24 months after birth in FNPIS and could be used for validation of outcomes indicated in administrative data. NICU=Neonatal Intensive Care Unit; SCBU=Special Care Baby Unit.

	Maternal variable	Dataset	Child variable	Dataset
Pregnancy health behaviours and birth	- *Smoking in pregnancy / at delivery* - *Illegal drug & alcohol use in pregnancy* - Pregnancy complications - Subsequent pregnancies - *Contraceptive use post-partum*	- CPRD/MCDS - HES - HES/MCDS - HES - CPRD	- Birth outcome (still or live), *NICU / SCBU* - Birth weight / length of gestation - Neonatal death	- HES/MCDS - HES/MCDS - HES/MCDS
Maternal health	- Mental health (during/post-pregnancy) - *BMI* - *Smoking post-partum* - *Illegal drug & alcohol use post-partum*	- CPRD - CPRD - CPRD - HES		
Education / School attainment / Cognitive & social development	- Attainment (GCSEs and above) - Further/Higher education	- NPD - NPD	- Age at school start - Developmental assessments - Early education (FSP, KS1) - Attendance and exclusions - Free School Meals / Special Educational Needs	- NPD - NPD/MCDS - NPD - NPD - NPD
Physical health and development			- Breastfeeding to discharge - Disability - Height & weight (BMI) - Immunisations - Accidents and injuries - *Nutrition*	- MCDS - HES/MCDS - MCDS - MCDS - HES - HES
Neglect and abuse, safeguarding / social care status and service use	- *Domestic violence*	- CPRD	- Out-of-home care - Child protection plan - Referrals to social services (Child in Need)	- CLA/MCDS - CINC/MCDS - CINC
Use of services	- Hospital admissions and attendances - Primary care contact	- HES - CPRD	- Hospital admissions and attendances - Primary care contact - Funded early years education	- HES - CPRD - NPD

7. Summary and recommendations

This report has assessed the feasibility of linking individual-level FNPIS data to administrative datasets, and suggests that such linkage would provide benefits for monitoring and evaluating the FNP programme in England, particularly for capturing maternal and child health and educational outcomes at a national level, and guiding adaptation of the programme's eligibility criteria. The FNP NU should actively promote wider use of FNP data for these purposes.

There are important technical and methodological considerations that need to be addressed, and a significant investment of time and resources would be required. Importantly, demonstrating the effectiveness of the FNP requires robust control group comparisons, complex analyses and methodological rigour combined with ongoing monitoring of modifications to the FNP and recording of details of local service as usual. Analyses at the LA level are limited by small sample sizes and a lack of appropriate comparison groups though sub-national analyses using groups of local authorities or regions would be feasible. The feasibility of linkage with social care outcomes is also currently limited due by a lack of identifiers recorded in the relevant administrative datasets for preschool children.

Priorities for data linkage to evaluate and monitor FNP in England

When deciding on which data linkage and analyses to undertake as a priority, the FNP NU should consider the utility of the proposed analyses (in terms of the questions relevant to monitoring and evaluation of FNP) and the investment required to achieve linkage and complete the analyses. Collaboration with academic communities and researchers experienced in using administrative data would help to prioritise evaluation questions while taking into account the strengths and limitations of the data.

While it is not possible to provide an exhaustive list, Appendix 6 provides a summary of potential linkages between FNPIS and administrative data and relevant indicators that could be explored using the linked data. Analyses that are most readily achievable and those most difficult to achieve, taking into account the most important aspects of data access, technical requirements and methodological considerations (as outlined in this report) are summarised in Figure 2. Given the limitations of linking and analysing data at an LA level outlined in this report, the FNP NU should focus on linkage of FNPIS and administrative data at a national level (which may also be analysed at a sub-national level, if appropriate for the question of interest).

Exemplars of relevant national linkages and a review of local data linkage processes within one case study LA presented Appendix 5 highlight additional considerations for identifying the best approaches for data linkage and analysis. The remainder of this section focusses on recommendations in terms of permission pathways, technological requirements and methodological considerations, as well as general recommendations for FNP NU in relation to using administrative data for monitoring and evaluation in the future.

Figure 2: Priorities for data linkage to evaluate and monitor FNP

HES-ONS **High priority**	+ Could be used to produce a range of relevant health indicators from 1997 onwards + Deterministic linkage is possible + Processes for data access are well established; costs are reasonable + Robust comparators are identifiable via information on maternal age and previous pregnancies + Mother-child link is possible - Data access timelines are variable
NPD **High priority**	+ Could be used to produce relevant developmental indicators for mothers from 2001 onwards (currently limited by age profile for FNP children) + Processes for data access are well established; there are currently no costs involved + Timelines for accessing data are reasonable - Deterministic linkage is not possible as NHS number is not included - Robust comparators are not directly identifiable as maternal age is not captured
MCDS **Medium priority**	+ Could be used to produce a range of indicators relevant to FNP, but up to age five only + Deterministic linkage is possible + Robust comparators are identifiable via information on maternal age and previous pregnancies + Mother-child link is possible - Processes, timelines and costs for data access are not yet established - Only prospective analyses from 2015 are possible
CPRD **Medium priority**	+ Could be used to produce relevant health indicators from 1987 onwards + Deterministic linkage is possible + Processes for data access are well established - Timelines for accessing data are variable and the associated costs are very high - Does not have universal coverage, which will limit the sample size and statistical power of analyses

CLA / CINC **Low priority**	+ Could be used to produce relevant social care indicators for mothers and children from 1992 (CLA) and 2008 (CINC) + Processes for data access are well established; there are currently no costs involved + Timelines for accessing data are reasonable - Deterministic linkage is not possible for pre-school children and probabilistic linkage is limited due to the lack of name - Robust comparators are not identifiable as maternal age is not captured

CAFCASS **Low priority**	+ Could be used to produce relevant family court indicators from 2001 onwards - Deterministic linkage is not possible - Processes, timelines and costs for data access are unknown - Robust comparators are not identifiable as maternal age is not captured

DWP / HMRC **Low priority**	+ Could be used to produce relevant economic / employment indicators for mothers - Deterministic linkage is not possible - Processes, timelines and costs for data access are unknown - Robust comparators are not identifiable as information on maternal age is not captured

MoJ **Low priority**	- Could be used to produce relevant crime indicators for mothers - Deterministic linkage is not possible - Processes, timelines and costs for data access are unknown - Robust comparators are not identifiable as maternal age is not captured

Permission pathways

Since consent to process personal data is not practical, given the cost and complexity of contacting >24,000 FNP participants and the risk of bias due to lower consent rates in more vulnerable or transient populations, the legal basis for processing personal data without consent is likely to be Section 251 support through an application to CAG. (This applies to national data: any LAs or local health providers currently accessing and linking FNP data fall outside the remit of CAG, as the legal basis for this activity is held within LAs). An application to CAG should emphasise:

- Fair processing: Privacy notices informing participants of how and why their data are being processed, and by whom, should be made available on the FNP website and any other relevant forums. Applicants could also consider contacting all FNP families directly with information about the study (as with the Building Blocks follow-up), and we recommend that any such information should firstly be reviewed by NHS Digital. Participants should be given the option to opt out.

- Public interest: The specific benefits of the project to the provision of care and/or the promotion of health should be strongly emphasised. Applications need to assure approvals boards that the processing of personal data is in the public interest and poses minimal risk to individuals.
- Justification for the data: Applications should demonstrate that the minimum amount of personal data are being processed, and should justify the variables requested. For sensitive variables such as dates and locations, the level of detail required should be carefully considered (e.g. assessing whether month and year would be sufficient, rather than exact date). The length of time for which the data will be retained also needs to be justified.
- Data flows: Applications should clearly show who will be processing any identifiable data, emphasise the separation of identifiers from attribute data (e.g. clinical or education data) and justify the direction of data flows in terms of minimising the transfer of identifiable data. The role of any trusted third parties (e.g. for conducting the linkage of identifiers or holding the linked anonymised data) should be clearly stated.

Technological requirements

The identifiers captured in FNPIS should facilitate good quality linkage with administrative data sources, particularly if kept up-to-date via the PDS:

- Tracing of FNP participants through the PDS should be conducted before linkage is performed. An additional benefit of linking with PDS would be the ability to update the FNPIS automatically when an FNP baby is born. This would provide timely and accurate birth information and avoid the need for FNP teams to obtain this information on a manual basis (which can take up to six weeks).
- Currently, identifiers collected at enrolment are updated as information is collected during FNP visits. Original identifiers should be retained separately in FNPIS, to facilitate linkage with retrospective data (e.g. GP contact or hospital admissions during pregnancy).

Periods for which data are available, time scales in which data can be accessed and the costs associated with data access and linkage should be considered alongside the technical requirements for linkage.

Suggestions for data controllers

DfE: Based on identifiers currently collected, outcomes for children in out-of-home care or with child protection plans recorded in CLA or CINC can only be evaluated if the child was captured in these datasets from school age. Approximately 20% of children who have ever been in OHC by age 18 years will only have been in care prior to school-entry, and will only have sex, date of birth and ethnicity available for linkage at a national level. In order for this important group of children to be captured, additional identifiers (postcode and name) would need to be collected in CLA / CINC data.

PDS: Collaboration with data linkage methodologists would help to exploit fully data that is captured in PDS but not routinely used for linkage, including the relational link between mothers and their babies, and postcode histories.

All: The Building Blocks trial found that NHS numbers recorded in FNPIS did not always match with NHS numbers recorded in GP registration data. Therefore, trusted third parties performing linkage should consider investing in capacity for using probabilistic linkage in order to capture participants with missing or incorrect identifiers. Close collaboration between data linkers and data users will enable appropriate assessment of linkage quality, including estimates of linkage error rates and comparisons of linked and unlinked data characteristics.[114]

Methodological considerations

Decisions relating to the most appropriate comparison groups, outcome measures, covariates and sample size will depend on the evaluation or monitoring question being addressed; detailed examples are provided in the case studies in Appendix 5. However, general considerations are:

- Monitoring of FNP adaptations: Modifications to the programme at LA or group of LA level should be carefully monitored and assessed using information on changes to the programme (e.g. differences in delivery, number of sessions, staff expertise, visit durations and content) and on local service as usual (e.g. local data on children's centres, midwifery services and health visitor programmes).
- Validity of recorded outcomes: Outcomes derived from administrative data should be validated as much as possible through comparisons with data captured in FNPIS. The limitations of administrative datasets to capture outcomes of interest must also be considered e.g. HES does not capture all pregnancies; early miscarriages and terminations may be missed.
- Data quality: Sufficient understanding of data quality and completeness is required before data are used to evaluate and monitor the FNP. For example, whilst data collected in MCDS could provide an enhanced level of detail and outcomes not captured elsewhere, the dataset is still under development and is not yet complete.
- Identification of appropriate comparison groups: To evaluate the effectiveness of the FNP a robust comparator group is required, ideally first time teenage mothers not enrolled in FNP. Linkage of FNPIS and HES data would allow this group to be identified, and facilitate comparisons of outcomes between FNP and non-FNP mothers. The addition of an FNP flag to individual child records in MCDS (or linkage

between FNPIS and MCDS) could also provide a robust comparator group, but would be limited to prospective cohorts of mothers.

Next steps for FNP NU

- Linkage of administrative data holds great potential for providing ongoing monitoring and evaluation of the FNP programme in England. Linkage between FNP and HES could also guide evaluation and adaptation of the programme's eligibility criteria through i) identifying other vulnerable groups of women who have similar characteristics and outcomes to FNP mothers and ii) exploring variation in uptake of FNP between sub-groups of eligible women (e.g. teenage mothers who have been in care). The FNP NU should actively promote wider use of FNP data for commissioning, service evaluation and research and exploit linkage with national data for monitoring and evaluating short- and longer-term outcomes. To capitalise on this potential, the FNP NU should:
- Set priorities: A clear strategy for identifying research and evaluation priorities should be developed, considering which questions are best addressed using linked administrative data. Pre-requisites for facilitating linkage are:
    - Establishing a link with PDS to update identifiers and provide birth information
    - Seeking advice from CAG on the use of FNP data for secondary purposes and providing privacy notices to establish fair processing of personal data
- Set realistic timeframes: The time taken for data access applications should be incorporated into plans for evaluation and monitoring. Sufficient time for preparation of applications and contingency for delays in the approval process should be factored into any future project timelines.
- Build sufficient capacity and expertise: Working with academic communities would facilitate independent and rigorous evaluation of the programme and provide access to additional resources (such as Data Safe Havens) and expertise. This would build on existing experience with deterministic and probabilistic methods to enable methods for complex linkages to be developed, data management expertise in appropriate anonymisation and extraction of linked data, and strong statistical support for analyses of the resulting linked data. The FNP NU should also consider investing in data scientist/analyst support.
- Acknowledge the strengths and limitations of administrative data: The use of administrative data can be a cost-effective and timely method of monitoring and evaluation. However, it is important to acknowledge that administrative datasets are not collected primarily for the purposes of evaluation and monitoring (e.g. HES is collected to reimburse health care providers for services provided) and that it will not be possible to address all questions of interest using this approach.[114] A sound understanding of administrative datasets and clear planning of analyses is required at an early stage.

8. References

1. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. Br J Cancer. 2014. 110(3):551-5

2. Cartwright N, Munro E. The limitations of randomized controlled trials in predicting effectiveness. J Eval Clin Pract. 2010. 16(2):260-6

3. Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the Randomized Controlled Trial in Evaluating Population-Based Health Interventions. Am J Prev Med. 2007. 33(2):155-61

4. Family Nurse Partnership National Unit. The Famiy Nurse Partnership Programme: Information leaflet. 2012. Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/216864/The-Family-Nurse-Partnership-Programme-Information-leaflet.pdf [Accessed 08/08/16]

5. Olds DL, Holmberg JR, Donelan-McCall N, Luckey DW, Knudtson MD, Robinson J. 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):114-21

6. Olds DL, Kitzman H, Knudtson MD, Anson E, Smith JA, Cole R. 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):800-6

7. Eckenrode J, Campa M, Luckey DW, 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):9-15

8. MacMillan HL, Wathen CN, Barlow J, Fergusson DM, Leventhal JM, Taussig HN. Interventions to prevent child maltreatment and associated impairment. Lancet. 2008. 373(9659):250-66

9. Barnes J, Ball M, Meadows P, McLeish J, Belsky J, FNP Implementation Research Team. Nurse-Family Partnership Programme: First year pilot sites implementation in England. Pregnancy and the Post-partum Period. Research Report DCSF-RW051. 2009. Available from: www.iscfsi.bbk.ac.uk/projects/files/Year-1-report-Barnes-et-al.pdf [Accessed 08/08/16]

10. Barnes J, Ball M, Meadows P, Howden B, Jackson A, Henderson J, et al. The Family-Nurse Partnership Programme in England: Wave 1 implementation in toddlerhood & a comparison between Waves 1 and 2a of implementation in pregnancy and infancy. 2011. Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/215837/dh_123366.pdf [Accessed 08/08/16]

11. Owen-Jones E, Bekkers M-J, Butler C, Cannings-John R, Channon S, Hood K, et al. The effectiveness and cost-effectiveness of the Family Nurse Partnership home visiting programme for first time teenage mothers in England: a protocol for the Building Blocks randomised controlled trial. BMC Pediatr. 2013. 13(1):1-13

12. Robling M, Bekkers M-J, Bell K, Butler CC, Cannings-John R, Channon S, et al. Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial. Lancet. 2015. 387(10014):146-55

13. South East Wales Trials Unit. Building Blocks: 2-6. Evaluating the long-term effectiveness, and the cost and consequences of the Family Nurse Partnership parenting support programme in reducing maltreatment in young children 2014. Available from: www.nets.nihr.ac.uk/__data/assets/pdf_file/0015/123423/PRO-11-3002-11.pdf [Accessed 08/08/16]

14. Olds DL, Henderson CR, Chamberlin R, Tatelbaum R. Preventing child abuse and neglect: a randomized trial of nurse home visitation. Pediatrics. 1986. 78(1):65-78

15. Kitzman H, Olds DL, Henderson CR, Jr, et al. Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing: A randomized controlled trial. JAMA. 1997. 278(8):644-52

16. Olds DL, Robinson J, O'Brien R, Luckey DW, Pettitt LM, Henderson CR, et al. Home visiting by paraprofessionals and by nurses: a randomized, controlled trial. Pediatrics. 2002. 110(3):486-96

17. Olds D. Building evidence to improve maternal and child health. Lancet. 2015. 387(10014):105-7

18. Olds D, Henderson, Jr CR, Cole R, et al. Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. JAMA. 1998. 280(14):1238-44

19. Olds DL, Eckenrode J, Henderson CR, Jr, et al. Long-term effects of home visitation on maternal life course and child abuse and neglect: Fifteen-year follow-up of a randomized trial. JAMA. 1997. 278(8):637-43

20. Olds DL, Kitzman H, Cole R, Robinson J, Sidora K, Luckey DW, et al. Effects of nurse home-visiting on maternal life course and child development: Age 6 follow-up results of a randomized trial. Pediatrics. 2004. 114(6):1550-9

21. Olds DL, Kitzman H, Hanks C, Cole R, Anson E, Sidora-Arcoleo K, et al. Effects of nurse home visiting on maternal and child functioning: age-9 follow-up of a randomized trial. Pediatrics. 2007. 120(4):e832-e45

22. National Institute of Clinical Excellence. Antenatal care for uncomplicated pregnancies: NICE guidelines CG62. 2008. Available from: www.nice.org.uk/guidance/cg62/chapter/Aim [Accessed 08/08/16]

23. National Institute of Clinical Excellence. Pregnancy and complex social factors: a model for service provision for pregnant women with complex social factors. NICE guidelines CG110. 2010. Available from: www.nice.org.uk/guidance/CG110/chapter/1-Guidance#young-pregnant-women-aged-under-20 [Accessed 08/08/16]

24. Public Health England. Best start in life and beyond: Improving public health outcomes for children, young people and families. Guidance to support the commissioning of the Healthy Child Programme 0-19: Health Visiting and School Nursing services. 2016. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/518657/Service_specification_CG2_FINAL_12_Feb.pdf [Accessed 09/08/16]

25. Marmot M. Fair Society, Healthy Lives. The Marmot Review. . 2010. Available from: http://www.instituteofhealthequity.org/projects/fair-society-healthy-lives-the-marmot-review [Accessed 09/08/16]

26. Department of Health. Healthy Child Programme. Pregnancy and the first five years of life. 2009. Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/167998/Health_Child_Programme.pdf [Accessed 08/08/16]

27. Robling M. The Building Blocks Trial. Evaluating the Family Nurse Partnership Programme in England: A Randomised Controlled Trial. 2015. Available from: http://medicine.cardiff.ac.uk/media/filer_public/6c/a9/6ca9945b-ecc5-461e-bc22-5f0f7d48f9b0/bb_final_report_full_121015.pdf [Accessed 09/08/16]

28. Barnes J, Aistrop D, Allen E, Barlow J, Elbourne D, Macdonald G, et al. First steps: study protocol for a randomized controlled trial of the effectiveness of the Group Family Nurse Partnership (gFNP) program compared to routine care in improving outcomes for high-risk mothers and their children and preventing abuse. Trials. 2013. 14:285-

29. Wimbush E, Craig P, R J. Evaluability assessment of the Family Nurse Partnership in Scotland. 2015. Available from: www.healthscotland.com/documents/26102.aspx [Accessed 09/08/16]

30. Ball M, Barnes J, P M. Issues emerging from the first 10 pilot sites implementing the Nurse-Family Partnership home-visiting programme in England. 2012. Available from:

www.gov.uk/government/uploads/system/uploads/attachment_data/file/212894/3-Birkbeck-Final-Issues-Evaluation-Report-For-Publication-July-2012.pdf [Accessed 09/08/16]

31. Jack S, Catherine N, Gonzalez A, MacMillan H, Sheehan D, Waddell C. Adapting, piloting and evaluating complex public health interventions: lessons learned from the Nurse--Family Partnership in Canadian public health settings. Health Promot Chronic Dis Prev Can. 2015. 35

32. Jutte DP, Roos NP, Brownell MD, Briggs G, MacWilliam L, Roos LL. The Ripples of Adolescent Motherhood: Social, Educational, and Medical Outcomes for Children of Teen and Prior Teen Mothers. Acad Pediatr. 2010. 10(5):293-301

33. Institute of Fiscal Studies. How English domiciled graduate earnings vary with gender, institution attended, subject and socio-economic background. 2016. Available from: www.ifs.org.uk/uploads/publications/wps/wp201606.pdf [Accessed 08/08/16]

34. Broadhurst K, Alrouh B, Yeend E, Harwin J, Shaw M, Pilling M, et al. Connecting events in time to identify a hidden population: Birth mothers and their children in recurrent care proceedings in England. Br J Soc Work. 2015. online first:doi: 10.1093/bjsw/bcv130

35. Department for Education. Children looked after by local authorities in England: Guide to the SSDA903 collection - 1 April 2016 to 31 March 2017. 2015. Available from: www.gov.uk/government/publications/children-looked-after-return-2016-to-2017-guide [Accessed 09/08/16]

36. Mc Grath-Lone L, Dearden L, Nasim B, Harron K, Gilbert R. Changes in first entry to out-of-home care from 1992 to 2012 among children in England. Child Abuse Negl. 2016. 51:163-71

37. Department for Education. Notifications of private fostering arrangements: year ending 31 March 2014. 2014. Available from: www.gov.uk/government/statistics/notifications-of-private-fostering-arrangements-in-england-2014 [Accessed 09/08/16]

38. Grath-Lone LM, Woodman J, Gilbert R. Safeguarding children and improving their care in the UK. Lancet. 2015. 386(10004):1630

39. Department for Education. Statistical First Release: Characteristics of children in need in England, 2013-14. 2014. Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/367877/SFR43_2014_Main_Text.pdf [Accessed 09/08/16]

40. Department for Education. National pupil database: user guide and supporting information 2015. Available from: www.gov.uk/government/publications/national-pupil-database-user-guide-and-supporting-information [Accessed 09/08/16]

41. Schoon I. Parental worklessness and the experience of NEET among their offspring. Evidence from the Longitudinal Study of Young People in England (LSYPE). Longit Life Course Stud. 2014. 5(2):129-50

42. ICD. International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM). 2011 28/10/13. Report No. Available from: www.cdc.gov/nchs/icd/icd10cm.htm [Accessed 09/08/16]

43. Health and Social Care Information Centre. OPCS-4 Classification of Interventions and Procedures. Available from: http://www.datadictionary.nhs.uk/web_site_content/supporting_information/clinical_coding/opcs_classification_of_interventions_and_procedures.asp?shownav=1 [Accessed 09/08/16]

44. Hardelid P, Dattani N, Gilbert R. 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)

45. Health and Social Care Information Centre, editor Data Quality. Maternity Providers & System Suppliers Stakeholder User Group; 2016; HSCIC, Leeds.

46. Directorate RD. The Government plan for a secure data service: Strengthening the international competetiveness of UK life sciences research 2011. Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/215337/dh_131242.pdf [Accessed 09/08/16]

47. Chisholm J. The Read clinical classification. BMJ. 1990. 300(6732):1092-

48. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015. 44(3):827-36

49. Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. BMJ. 2013. 346

50. Cornish RP, Boyd A, Van Staa T, Salisbury C, Macleod J. Socio-economic position and childhood multimorbidity: a study using linkage between the Avon Longitudinal study of parents and children and the general practice research database. Int J Equity Health. 2013. 12(1):1-13

51. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort Profile: The English Longitudinal Study of Ageing. Int J Epidemiol. 2012. 42(6):1640-8

52. Tetlow G. Difficulties and rewards of linking to administrative data: experience from ELSA. 2012. Available from: www.share-project.org/fileadmin/pdf_documentation/SHARE-RV/2_2_Tetlow.pdf [Accessed 09/08/16]

53. Bozio A, Crawford R, Emmerson C, Tetlow G. Retirement outcomes and lifetime earnings: descriptive evidence from linked ELSA-NI Data. 2010. Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/214382/WP81.pdf [Accessed 09/08/16]

54. Audrey S, Brown L, Campbell R, Boyd A, Macleod J. Young people's views about consenting to data linkage: findings from the PEARL qualitative study. BMC Med Res Methodol. 2016. in press

55. Boyd A, Davies A, Cornish RP, Shannon B, Onslow J, Macleod J. Project to Enhance ALSPAC through Record Linkage. 2013. Available from: www.closer.ac.uk/wp-content/uploads/Boyd_CLOSER_20130702.pdf [Accessed 09/08/16]

56. Information Commissioner's Office. The Guide to Data Protection 2016. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/ [Accessed 09/08/16]

57. Home Office. The Data Protection (Processing of Sensitive Personal Data) Order 2000. 2000. Available from: www.legislation.gov.uk/uksi/2000/417/pdfs/uksi_20000417_en.pdf [Accessed 09/08/16]

58. Department of Health. The Care Act. 2014. Available from: www.legislation.gov.uk/ukpga/2014/23/contents/enacted [Accessed 08/03/16]

59. Health Research Authority. Principles of Advice. 2015. Available from: www.hra.nhs.uk/documents/2015/03/195369.pdf [Accessed 09/08/16]

60. Ford D, Jones K, Verplancke JP, Lyons R, John G, Brown G, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. BMC Health Serv Res. 2009. 9(1):157

61. Health and Social Care Information Centre. Guidance notes on security. 2015. Available from: http://digital.nhs.uk/media/15698/DARS---Guidance-Notes-on-Security/pdf/Guidance_Notes_on_Security_v4.0_(1).pdf [Accessed 09/08/16]

62. Burton PR, Murtagh MJ, Boyd A, Williams JB, Dove ES, Wallace SE, et al. Data Safe Havens in health research and healthcare. Bioinformatics. 2015.

63. Dibben C, Elliot M, Gowans H, Lightfoot D. Chapter 3: The data linkage environment. In: Harron K, Dibben C, Goldstein H, editors. Methodological Developments in Data Linkage. London: Wiley; 2015.

64. Medicines & Healthcare products Regulatory Agency. Independent Scientific Advisory Committee for Medicines and Healthcare products Regulatory Agency (MHRA) database research (ISAC) Annual report 2014. 2015. Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/463822/ISAC_Annual_Report_January_2014_to_December_2014.pdf [Accessed 09/08/16]

65. Kelman C, Bass A, Holman C. Research use of linked health data - a best practice protocol. Aust N Z J Public Health. 2002. 26(3):251 - 5

66. Longhurst J TN, Young C, Office for National Statistics,. Statistical disclosure control for the UK Census. 2011. Available from: www.ons.gov.uk/census/2011census/howourcensusworks/howwetookthe2011census/howweplannedfordatadelivery/protectingconfidentialitywithstatisticaldisclosurecontrol [Accessed 09/08/16]

67. Linking Registry Data With Other Data Sources To Support New Studies. In: Gliklich RE, Dreyer NA, Leavy MB, editors. Registries for Evaluating Patient Outcomes: A User's Guide 3rd Edition. Rockville: Agency for Healthcare Research and Quality; 2014.

68. Information Commissioner's Office. Anonymisation: managing data protection risk code of practice. 2012. Available from: https://ico.org.uk/media/1061/anonymisation-code.pdf [Accessed 09/08/16]

69. Raynor P. Born in Bradford, a cohort study of babies born in Bradford, and their parents: Protocol for the recruitment phase. BMC Public Health. 2008. 8(1):1-13

70. Swerdlow AJ, Jones ME, Schoemaker MJ, Hemming J, Thomas D, Williamson J, et al. The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer aetiology. Br J Cancer. 2011. 105(7):911-7

71. Information Standards Board for Health and Social Care. ISB 1555 Birth Notifications Specification. 2013. Available from: http://webarchive.nationalarchives.gov.uk/+/http://www.isb.nhs.uk/documents/isb-1555/amd-10-2012/index_html [Accessed 09/08/16]

72. Health and Social Care Information Centre. Fact Sheet on access methods to trace and use the NHS number of local records. 2014. Available from: http://systems.digital.nhs.uk/nhsnumber/staff/guidance/factsheet.pdf [Accessed 09/08/16]

73. Health and Social Care Information Centre. Fact Sheet on access methods to trace and use the NHS number on local records. 2014. Available from: http://systems.hscic.gov.uk/nhsnumber/staff/guidance/factsheet.pdf [Accessed 10/02/16]

74. Hippisley-Cox J. Validity and completeness of the NHS Number in primary and secondary care: electronic data in England 1991-2013. University of Nottingham 2013. Available from: http://eprints.nottingham.ac.uk/3153/ [Accessed 09/08/16]

75. Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. J Public Health. 2014. 36(4):684-92

76. Abrahams C, Davy K. Linking HES maternity records with ONS birth records. Health Stat Q. 2002. 13:22-30

77. Mears GD, Rosamond WD, Lohmeier C, Murphy C, O'Brien E, Asimos AW, et al. A link to improve stroke patient care: a successful linkage between a statewide emergency medical services data system and a stroke registry. Acad Emerg Med. 2010. 17(12):1398-404

78. Maso L, Braga C, Franceschi S. Methodology used for software for automated linkage in Italy (SALI). J Biomed Inform. 2001. 34(6):387-95

79. Muse A, Mikl J, Smith P. Evaluating the quality of anonymous record linkage using deterministic procedures with the New York State AIDS registry and a hospital discharge file. Stat Med. 1995. 14(5-7):499-509

80. Tromp M, Ravelli A, Bonsel G, Hasman A, Reitsma J. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. J Clin Epidemiol. 2011. 64(5):565-72

81. Yu N, Donnan PT, Leese GP. A record linkage study of outcomes in patients with mild primary hyperparathyroidism: The Parathyroid Epidemiology and Audit Research Study (PEARS). Clin Endocrinol. 2011. 75(2):169-76

82. Poluzzi E, Piccinni C, Carta P, Puccini A, Lanzoni M, Motola D, et al. Cardiovascular events in statin recipients: impact of adherence to treatment in a 3-year record linkage study. Eur J Clin Pharmacol. 2011. 67(4):407-14

83. Kendrick S, Clarke J. The Scottish Record Linkage System. Health Bull (Edinb). 1993. 51(2):72-9
84. Health and Social Care Information Centre. Methodology for creation of the HES Patient ID (HESID). 2014. Available from: www.hscic.gov.uk/media/1370/HES-Hospital-Episode-Statistics-Replacement-of-the-HES-patient-ID/pdf/HESID_Methodology.pdf [Accessed 09/08/16]
85. Grannis S, Overhage J, McDonald C. Analysis of identifier performance using a deterministic linkage algorithm. Proc AMIA Symp. 2002.305–9
86. Hagger-Johnson G, Harron K, Fleming T, Gilbert R, Goldstein H, Landy R, et al. Data linkage errors in hospital administrative data when applying a pseudonymisation algorithm to paediatric intensive care records. BMJ Open. 2015. 5(8)
87. Newcombe H, Kennedy J, Axford S, James A. Automatic linkage of vital records. Science. 1959. 130(3381):954-9
88. Fellegi IP, Sunter AB. A theory for record linkage. J Am Stat Assoc. 1969. 64(328):1183-210
89. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic record linkage. Int J Epidemiol. 2015.
90. Zhu V, Overhage M, Egg J, Downs S, Grannis S. An empiric modification to the probabilistic record linkage algorithm using frequency-based weight scaling. J Am Med Inform Assn. 2009. 16(5):738-45
91. Grannis S, Overhage J, Hui S, McDonald C. Analysis of a probabilistic record linkage technique without human review. AMIA Annu Symp Proc. 2003. 2003:259-63
92. Méray N, Reitsma J, Ravelli A, Bonsel G. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. J Clin Epidemiol. 2007. 60(9):883-91
93. Jamieson E, Roberts J, Browne G. The feasibility and accuracy of anonymized record linkage to estimate shared clientele among three health and social service agencies. Method Inform Med. 1995. 34(4):371-7
94. Bohensky M. Chapter 4: Bias in data linkage studies. In: Harron K, Dibben C, Goldstein H, editors. Methodological Developments in Data Linkage. London: Wiley; 2015.
95. Bohensky M, Jolley D, Sundararajan V, Evans S, Pilcher D, Scott I, et al. Data linkage: A powerful research tool with potential problems. BMC Health Serv Res. 2010. 10(1):346-52
96. Institute of Education. MCS4: A guide to the linked education administrative datasets. 2015. Available from: www.cls.ioe.ac.uk/page.aspx?&sitesectionid=1099&sitesectiontitle=Data+Linkage [Accessed 09/08/16]
97. Cornish RP, Tilling K, Boyd A, Davies A, Macleod J. Using linked educational attainment data to reduce bias due to missing outcome data in estimates of the association between the duration of breastfeeding and IQ at 15 years. Int J Epidemiol. 2015.
98. Institute for Social and Economic Research. Linked Understanding Society - National Pupil Database. Wave 1 Linkage user manual. 2015. Available from: www.understandingsociety.ac.uk/documentation/mainstage [Accessed 09/08/16]
99. Health and Social Care Information Centre. HES-MHLDDS Linkage: data matching quality report for February 2016. 2015. Available from: http://digital.nhs.uk/media/20081/HES-MHLDDSDataMatchingQualityReport/pdf/HES-MHLDDS_Data_Matching_Quality_Report.pdf [Accessed 09/08/16]
100. Denaxas SC, George J, Herrett E, Shah AD, Kalra D, Hingorani AD, et al. Data Resource Profile: Cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). Int J Epidemiol. 2012. 41(6):1625-38
101. Krewski D, Dewanji A, Wang Y, Bartlett S, Zielinski J, Mallick R. The effect of record linkage errors on risk estimates in cohort mortality studies. Surv Methodol. 2005. 31(1):13-21
102. Lariscy JT. Differential record linkage by hispanic ethnicity and age in linked mortality studies. J Aging Health. 2011. 23(8):1263-84

103. Coeli CM, Barbosa Fdos S, Brito Ados S, Pinheiro RS, Camargo KR, Jr., Medronho Rde A, et al. Estimated parameters in linkage between mortality and hospitalization databases according to quality of records on underlying cause of death. Cad Saude Publica. 2011. 27(8):1654-8

104. Adams MM, Wilson HG, Casto DL, Berg CJ, McDermott JM, Gaudino JA, et al. Constructing reproductive histories by linking vital records. Am J Epidemiol. 1997. 145(4):339-48

105. Boyle D, Cunningham S. Resolving fundamental quality issues in linked datasets for clinical care. Health Informatics J. 2002. 8(2):73-7

106. Harron K, Wade A, Muller-Pebody B, Goldstein H, Gilbert R. Opening the black box of record linkage. J Epidemiol Commun H. 2012. 66(12):1198

107. Harron K, Goldstein H, Wade A, Muller-Pebody B, Parslow R, Gilbert R. Linkage, evaluation and analysis of national electronic healthcare data: application to providing enhanced blood-stream infection surveillance in paediatric intensive care. PLoS One. 2013. 8(12):e85278

108. Ford JB, Roberts CL, Taylor LK. Characteristics of unmatched maternal and baby records in linked birth records and hospital discharge data. Paediatr Perinat Epidemiol. 2006. 20(4):329-37

109. Harron K, Wade A, Gilbert R, Muller-Pebody B, Goldstein H. Evaluating bias due to data linkage error in electronic healthcare records. BMC Med Res Methodol. 2014. 14(1):36

110. Wijlaars LPMM. Early comorbid parental depression and its effects on child outcomes. PhD Thesis: University College London; 2014.

111. Kahila H, Gissler M, Sarkola T, Autti-Rämö I, Halmesmäki E. Maternal welfare, morbidity and mortality 6–15 years after a pregnancy complicated by alcohol and substance abuse: A register-based case-control follow-up study of 524 women. Drug Alcohol Depend. 2010. 111(3):215-21

112. Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, White IR, et al. Smoker, ex-smoker or non-smoker? The validity of routinely recorded smoking status in UK primary care: a cross-sectional study. BMJ Open. 2014. 4(4)

113. Bhaskaran K, Forbes HJ, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). BMJ Open. 2013. 3(9)

114. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Med. 2015. 12(10):e1001885

115. Eckenrode J, Campa M, Luckey DW, 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. JAMA Pediatr. 2010. 164(1):9-15

116. Conde-Agudelo A, Rosas-Bermúdez A, Kafury-Goeta AC. Birth spacing and risk of adverse perinatal outcomes: a meta-analysis. JAMA. 2006. 295(15):1809-23

117. Conde-Agudelo A, Rosas-Bermudez A, Castaño F, Norton MH. Effects of birth spacing on maternal, perinatal, infant, and child health: a systematic review of causal mechanisms. Stud Fam Plann. 2012. 43(2):93-114

118. Wolfe I, Macfarlane A, Donkin A, Marmot M, Viner R. Why children die: death in infants, children and young people in the UK Part A. 2014. Available from: http://ncb.org.uk/media/1130496/rcpch_ncb_may_2014_-_why_children_die__part_a.pdf [Accessed 09/08/16]

119. Public Health England. Reducing unintentional injuries in and around the home among children under five years. 2014. Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/322210/Reducing_unintentional_injuries_in_and_around_the_home_among_children_under_five_years.pdf [Accessed 08/08/16]

120. Department of Health. Eligibility for the Family Nurse Partnership programme: Testing new criteria. 2012. Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/212893/Eligibility-for-the-Family-Nurse-Partnership-programme-Testing-new-criteria.pdf [Accessed 08/08/16]

121. Royal College of Obstetricians and Gynaecologists. Patterns of maternity care in English NHS Hospitals. 2013 24/11/13. Report No. Available from: https://www.rcog.org.uk/globalassets/documents/guidelines/research--audit/patterns-of-maternity-care-in-english-nhs-hospitals-2011-12_0.pdf [Accessed 09/08/16]

122. Murray J, Saxena S, Modi N, Majeed A, Aylin P, Bottle A, et al. Quality of routine hospital birth records and the feasibility of their use for creating birth cohorts. Journal of Public Health. 2012.

123. Cromwell DA, Knight HE, Gurol-Urganci I. Parity derived for pregnant women using historical administrative hospital data: Accuracy varied among patient groups. J Clin Epidemiol. 2014. 67(5):578-85

124. Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P, et al. School readiness and later achievement. Dev Psychol. 2007. 43(6):1428

125. Kitzman HJ, Olds DL, Cole RE, Hanks CA, Anson EA, Arcoleo KJ, et al. Enduring effects of prenatal and infancy home visiting by nurses on children: follow-up of a randomized trial among children at age 12 years. Arch Ped Adol Med. 2010. 164(5):412-8

126. Berlin LJ, Appleyard K, Dodge KA. Intergenerational continuity in child maltreatment: Mediating mechanisms and implications for prevention. Child Dev. 2011. 82(1):162-76

127. Hagger-Johnson G, Harron K, Gonzalez-Izquierdo A, Cortina-Borja M, Dattani N, Muller-Pebody B, et al. Identifying false matches in anonymised hospital administrative data without patient identifiers Health Serv Res. 2014. (online first)

128. NHS Improving Quality. Using integrated intelligence to drive service change: supporting people with multiple morbidities in Kent. Available from: www.nhsiq.nhs.uk/media/2632000/kent_using_integrated_intelligence_to_drive_service_change.pdf [Accessed 08/08/16]

9. Acknowledgements

The authors would like to thank the following people for very helpful advice, support and contributions to this report:

- Bharat Ayer, Stephen Clarke, Philip Howlin, Sue Hilsden, Adele Jones, Alex Stevenson & Ailsa Swarbrick (FNP NU)
- Wilhelmine Meeraus (CPRD)
- Helen Duncan (PHE)
- Hannah Morris & Efrosini Setakis (ADRC-E)
- Rebecca Cannings-John & Fiona Lugg (Building Blocks)
- Franca MacLeod (Scottish Government)
- Natasha Dunkley (CAG)
- Jo Tonkins, Abraham George & Dave Smith (Kent County Council)
- Susan Gough (Kent FNP Supervisor)

10. Appendices

Appendix 1: Evidence for the effectiveness of FNP

Evidence for the US-licensed Nurse-Family Partnership originates from three main US trials. The first of these recruited 400 mothers between 1978-1980 in New York county.[14] Eligible participants were nulliparous pregnant woman, less than 26 weeks gestation, with at least one risk factor (aged <19 years, single parent, or low socio-economic status). The trial observed a reduction in smoking in the intervention group but no overall differences in birth weight or gestational age (except for 14-16 year olds, whose babies were an average 395 grams heavier in the intervention group). Emergency department visits in the first and second year of life were lower in the intervention group.

The second trial recruited 1139 mothers between 1990-1991 in Memphis.[15] Eligible participants were low-income nulliparous women, less than 29 weeks gestation, with at least two risk factors (unmarried, <12 years of education, or unemployed). The majority of participants were African American. No differences in birth weight or length of gestation were observed, although there were fewer healthcare encounters and admissions for injuries/ingestions, and subsequent pregnancies by 24 months post-partum in the intervention group. Breastfeeding initiation was higher in the intervention group, although no differences in duration were observed.

The third trial recruited 735 mothers between 1994-1995 in Denver.[16] Eligible participants were nulliparous pregnant women who were eligible for Medicaid or who did not have private health insurance (at any gestation). Reductions in cotinine for baseline smokers, fewer subsequent pregnancies by 24 months, and improved child development were observed in the intervention group.

The US trials continued to observe positive outcomes during long-term follow-up: at ages six and nine, greater levels of employment and less welfare dependency in mothers, and lower levels of behaviour problems and better developmental outcomes in the children were observed in the intervention groups.[20, 21] At age 15, mothers had fewer arrests or substance-abuse impairment, there were lower levels of maltreatment, and children had fewer sexual partners, arrests and convictions, and were less likely to run away from home.[18, 19] Differences between groups at age 19 were mostly restricted to females, who had fewer children and less Medicaid use.[115]

Pilot Sites in England

In 2006, 10 pilot sites were chosen to test the FNP in England: County Durham and Darlington, Manchester, Barnsley, Derby City, Walsall, South East Essex, Slough, Somerset, Southwark and Tower Hamlets. An initial evaluation of the first 1303 FNP participants was conducted to document, analyse and interpret the feasibility of FNP implementation during pregnancy, and to establish grounds for a randomised controlled trial.[9] The primary focus of the formative evaluation was fidelity targets chosen to promote ongoing performance based on evidence from the US. These targets covered recruitment, attrition, and delivery of the programme.

Recruitment

The evaluation found that the pilot sites met targets for 75% of eligible referrals being enrolled and 100% of those enrolled being first time mothers. The sites did not meet targets for enrolling 60% of pregnant women by 16 weeks gestation (women were on average 17 or 18 weeks gestation) but were close to meeting the target of full-time nurses having a caseload of 25 within 8-9 months. The evaluation also found that 78% of participants were enrolled during their first pregnancy.

Attrition

Attrition varied across pilot sites: one site achieved the target of <10% attrition for the pregnancy phase, but others ranged from 11-24%. The evaluation noted that two reasons for attrition (miscarriage or moving out of the area) should be assessed separately.

Delivery

The average proportion of expected visits received in the pilot sites was 53% (compared with a target of 80%). However, the target duration of visits (60 minutes) was exceeded (average of 73 minutes). Pilot sites were close to achieving targets for content of home visits.

Overall, FNP mothers and staff in the pilot sites were positive about and accepting of the programme.

A second evaluation, relating to the infancy period, focussed on factors affecting delivery and attrition. Approximately one third of participants received the target 65% of scheduled visits during infancy, and overall, the target of a maximum 20% attrition during infancy was achieved, although there was substantial variation between sites.

Building Blocks

The Building Blocks randomised controlled trial enrolled 1618 participants from 18 NHS sites in England, between 2009 and 2010. There were no significant differences in any of the four primary outcome measures:

1. Self-reported prenatal tobacco use at late pregnancy: 55.6% (intervention) versus 56.1% (control)
2. Birth weight: An adjusted mean difference of 21 grams (97.5% CI -47, 89)

3. Proportion of women with a second pregnancy by two years post-partum: 66.3% (intervention) versus 66.1% (control).
4. Emergency attendances and hospital admissions for the child within two years of birth: 81.0% (intervention) and 76.6% (control).

There were also a large number of secondary outcome measures relating to pregnancy and birth, child health and development, parental life course, maternal health-related outcomes. There were some observed differences in secondary outcomes:

1. Developmental concern: At 24 months, the proportion of children with a concern were 8.1% (intervention) and 12.6% (control), corresponding to an odds ratio of 0.61 (95% CI 0.40-0.90).
2. Language: Maternally reported developmental delay was significantly lower in the intervention arm at 12 months and 18 months. By 24 months, Early Language Milestone percentiles were significantly higher in the intervention group.
3. Breastfeeding: The proportion of mothers intending to breastfeed was significantly higher in the intervention (54.9%) compared with the control arm (50.4%), although there were no differences in the proportion initiating breastfeeding or the median duration of breastfeeding between groups.
4. Social services referral / Safeguarding: The proportion of participants reporting that their child had ever been referred to Social Services was significantly greater in the intervention (20.5%) compared with control arm (16.8%). The proportion of children with a safeguarding event recorded in GP data was significantly greater in the intervention (13.6%) compared with control group (8.0%).
5. Social support/self-efficacy: The maximum level of social support at 18 months was higher in the intervention compared with control arm at 18 and 24 months. A small increase in relationship quality and self-efficacy was also observed in the intervention arm.

The trial’s economic analysis determined that there was no significant difference between trial arms and that the FNP could not be considered cost-effectiveness based on the evidence of impact on the four primary outcomes, i.e. smoking in pregnancy and subsequent pregnancy for the mother, and birthweight emergency hospital attendance and admission for the child. The process evaluation, aiming to establish fidelity of the intervention, found that a high proportion of women (89%) offered FNP enrolled. The mean number of visits received was lower than that specified as a fidelity goal, but the average duration was almost 20 minutes longer. Attrition was approximately 20%.

Appendix 2: FNPIS data items

Table A2.1: Maternal data elements

	Programme intake	36 weeks pregnancy	Birth	6 weeks	6 months	12 months	18 months	24 months
Date of birth	x							
Age at intake	x							
Ethnicity	x							
Primary language	x							
Marital status	x				x	x	x	x
Current partner	x				x	x	x	x
Current partner baby's biological father	x				x	x	x	x
Contact with baby's biological father	x				x	x	x	x
Living arrangements	x				x	x	x	x
Housing - tenure	x				x	x	x	x
Housing - other occupants	x				x	x	x	x
Education - school completion yes/no	x				x	x	x	x
Education - age completed school	x				x	x	x	x
Education - last school year completed	x				x	x	x	x
Education - GCSEs or equivalent - yes/no	x				x	x	x	x
Education - GCSEs - how many	x				x	x	x	x
Education - GCSEs - number grade c or above	x				x	x	x	x
Education - qualifications other than GCSEs - yes/no	x				x	x	x	x
Education - qualifications other than GCSEs - what	x				x	x	x	x
Education - currently in education - yes/no	x				x	x	x	x
Education - currently in education - what	x				x	x	x	x
Employment - ever had a paid job	x							

	Programme intake	36 weeks pregnancy	Birth	6 weeks	6 months	12 months	18 months	24 months
Employment - currently working - yes FT, yes PT, No	x				x	x	x	x
Employment - stopped working due to pregnancy - yes/no	x							
Employment - paid work since birth of baby					x	x	x	x
Employment - months paid work since birth of baby					x	x	x	x
Employment - current salary					x	x	x	x
Income - household or personal income, by category	x				x	x	x	x
Income - all income from benefits	x				x	x	x	x
Income - client unable/unwilling to give information on income	x				x	x	x	x
Income - benefits received - list	x				x	x	x	x
Income - money from baby's biological father for baby					x	x	x	x
Services currently received - list	x							
Lived away from parents before 18	x							
Lived away from parents before 18 - who with	x							
Social care - currently Child in Need	x							
Social care - currently on a child protection plan	x							
Social care - date CPP started	x							
Social care - reason for CPP	x							
Social care – pre-birth assessment planned for child	x		x					
Social care - parenting assessment			x					
Baby's biological father takes care of child					x	x	x	x
Contraception - use of birth control yes/no					x	x	x	x
Contraception - if no, reasons					x	x	x	x
Contraception - how often use					x	x	x	x
Contraception - types					x	x	x	x
Subsequent Pregnancy - pregnant since birth of child					x	x	x	x
Subsequent Pregnancy - pregnancy details and outcome					x	x	x	x

	Programme intake	36 weeks pregnancy	Birth	6 weeks	6 months	12 months	18 months	24 months
Subsequent Birth - child details (dob, gender, birthweight, time in SCBU)					x	x	x	x
Service received - mother			x		x	x	x	x
Services received - baby					x	x	x	x
Multi-agency plan in place			x		x	x	x	x
Use of children's centres			x		x	x	x	x
Services used at children's centres			x		x	x	x	x
Social care - pre-birth assessment planned, completed, outcome	X	X						
Social care - CPP in place in last 6 months - mother, baby			x		x	x	x	x
Social care - date CPP Stopped			x		x	x	x	x
Social care - date CPP started			x		x	x	x	x
Social care - reason for CPP			x		x	x	x	x
Social care - CIN in last 6 months - mother, baby			x		x	x	x	x
Social care - date CIN Stopped			x		x	x	x	x
Social care - date CIN started			x		x	x	x	x
Social care - interim care order in place / living with or away from mother			x		x	x	x	x
Social care - mother referred to social services in last 6 months - yes/no					x	x	x	x
Social care - mother of any referrals, date, reason, outcome					x	x	x	x
Abuse - ever, by someone close - yes/no	X							
Abuse - physical in last 12 months, yes/no, whom	X							
Abuse - physical in last 12 months, type and frequency	X							
Abuse - physical abuse since pregnant, yes/no, whom	X							
Abuse - physical abuse since pregnant, type , frequency	X							
Abuse - injuries from physical abuse in last 12 months	x	x				x		
Abuse - forced sexual relations - yes/no, frequency	x	x				x		
Abuse - afraid of current or previous partner or someone else important - yes/no, whom	x	x				x		

	Programme intake	36 weeks pregnancy	Birth	6 weeks	6 months	12 months	18 months	24 months
Abuse - use of weapon	x	x				x		
Abuse - physical, emotional since enrolled on FNP		x				x		
Abuse - nature and frequency of any abuse since enrolled on FNP		x				x		
Abuse - ever abused since birth of baby, by partner, someone important						x		
Abuse - since birth of baby, nature, frequency						x		
Maternal health - how many times previously pregnant		x						
Maternal health - weeks pregnant when started receiving ante-natal care		x						
Maternal health - estimated date of delivery		x						
Maternal health - health history		x						
Maternal health - pregnancy health - UTIs, how many times treated since pregnant		x	x					
Maternal health - pregnancy health - STIs, how many times treated since pregnant		x	x					
Maternal health - prescription or antibiotics in pregnancy for infection, frequency			x					
Maternal health - height		x						
Maternal health - usual weight (pre-pregnancy)		x						
Maternal health - usual BMI (calculated by Family Nurse)		x						
Breastfeeding - plan to breastfeed		x						
Breastfeeding - breastfed as baby		x						
Breastfeeding - ever breastfed or expressed milk		x		x				
Breastfeeding - continuing				x				
Breastfeeding - age exclusive breastfeeding				x				
Breastfeeding - age stopped any breastfeeding				x				
Maternal health - positive affect		x						
Maternal health - personal beliefs		x						
Cigarettes smoked in last 48 hours, number		x	x	x		x		
Smoked at all during pregnancy		x	x	x		x		
Alcohol use - on how many days in last 14 days		x	x			x		

	Programme intake	36 weeks pregnancy	Birth	6 weeks	6 months	12 months	18 months	24 months
Alcohol use - usual units per day		x	x			x		
Marijuana use - how many days in last 14 days		x	x			x		
Marijuana use - how much per day		x	x			x		
Cocaine use - how many days used in last 14 days		x	x			x		
Cocaine use - how many times per day		x	x			x		
Other drugs - how many days used in last 14 days		x	x			x		
Other drugs - how many times per day		x	x			x		
Maternal health - anxiety (HADS)		x		x				
Maternal health - depression (HADS)		x		x				

Table A2.2: Child data elements

	Birth	6 weeks	4 months	6 months	10 months	12 months	14 months	18 months	20 months	24 months
Date of birth (month, year)	x									
Gender	x									
Birth weight (grams)	x									
Gestational age at birth	x									
Any time in SCBU	x	x		x						
Number of days in SCBU	x	x		x						
Child development - ASQ3			x		x		x		x	
Child development - ASQ-SE				x		x		x		x
Location of child check ups				x		x		x		x
Immunisations up to date - yes/no				x		x		x		x
Immunisations - source of info				x		x		x		x
Child current weight, development percentile				x		x		x		x
Hospital attendances for injury ingestion - number since birth				x		x				
Hospital attendances for injury ingestion - number since first birthday								x		x
Date, age of child and reason for each attendance				x		x		x		x
Hospital admissions for injury ingestion - number since birth				x		x				
Hospital admissions for injury ingestion - number since first birthday								x		x
Date, age of child and reason for each admission				x		x		x		x
Breastfeeding - yes/no				x		x		x		x
Breastfeeding - currently				x		x		x		x
Breastfeeding - age exclusive breastfeeding				x		x		x		x
Breastfeeding - age any breastfeeding				x		x		x		x
Social care - child referred to social services in last 6 months -				x		x		x		x

	Birth	6 weeks	4 months	6 months	10 months	12 months	14 months	18 months	20 months	24 months
yes/no										
Social care - details of any referrals, date, reason, outcome				x		x		x		x
Social care - CPP in place in last 6 months - baby				x		x		x		x
Social care - reason for CPP				x		x		x		x
Social care - date CPP started				x		x		x		x
Social care - date CPP Stopped				x		x		x		x
Social care - CIN in last 6 months - baby				x		x		x		x
Social care - date CIN started				x		x		x		x
Social care - date CIN Stopped				x		x		x		x
Social care - interim care order in place / living with or away from mother				x		x		x		x

Table A2.3: Programme activity data elements

	Each (completed) visit	As applicable
Date of visit	x	
Home visitor ID	x	
Who present at visit	x	
Time of visit	x	
Length of visit	x	
Cancelled visit		x
Attempted visit		x
Location of visit	x	
DANCE observation completed at visit - yes/no	x	
Interpreter present - yes	x	
Involvement of those present at visit	x	
Proportion of visit time spent on each of programme domains	x	
Proportion of planned content covered	x	
Referrals to other services - yes/no	x	
Details of referrals		x
Telephone encounter		x
Client left programme		x
Date left programme		x
Reason for leaving programme		x
Client inactive on FNP (6 months after no contact		x
Date inactive		x
Client returns to programme (after leaving)		x
Reason for return		x
Date for return		x
Change of client status*		

*amended estimated delivery data, new Family Nurse, transferred to a different site, change of address, father main carer, child into care, child on CPP or CIN, child returned to mother, infant born.

Appendix 3: CPRD data items

Table A3.1: Patient variables

Variable	Description
Patient Identifier	Encrypted unique identifier given to a patient in CPRD GOLD (current version of CPRD)
VAMP Identifier	Old id for the patient when the practice was using the previous VAMP system
Patient Gender	Patient’s gender
Birth Year	Patient’s year of birth
Birth Month	Patient’s month of birth (for those aged under 16). 0 indicates no month set
Marital Status	Patient’s current marital status
Family Number	Family ID number
CHS Registered	Value to indicate whether the patient is registered with Child Health Surveillance
CHS Registration Date	Date of registration with Child Health Surveillance
Prescription Exemption	Type of prescribing exemption the patient has currently (e.g. medical or maternity)
Capitation Supplement	Level of capitation supplement the patient has currently (e.g. low, medium, or high)
Socio-Economic Status	Patient’s socio-economic status. Currently 0; to be populated in future
First Registration Date	Date the patient first registered with the practice. If patient only has ‘temporary’ records, the date is the first encounter with the practice; if patient has ‘permanent’ records it is the date of the first ‘permanent’ record (excluding preceding temporary records)
Current Registration Date	Date the patient’s current period of registration with the practice began (date of the first ‘permanent’ record after the latest transferred out period). If there are no ‘transferred out periods’, the date is equal to First Registration date
Registration Status	Status of registration detailing gaps and temporary patients
Registration Gaps	Number of days missing in the patients registration details
Internal Transfer	Number of internal transfer out periods, in the patient’s registration details
Transfer Out Date	Date the patient transferred out of the practice, if relevant. Empty for patients who have not transferred out
Transfer Out Reason	Reason the patient transferred out of the practice. Includes 'Death' as an option
Death Date	Date of death of patient – derived using a CPRD algorithm
Acceptable Patient Flag	Flag to indicate whether the patient has met certain quality standards: 1 = acceptable, 0 = unacceptable

Table A3.2: Practice variables

Variable	Description
Practice identifier	Encrypted unique identifier given to a specific practice in CPRD GOLD
Region	Value to indicate where in the UK the practice is based. The region denotes the Strategic Health Authority for practices within England, and the country i.e. Wales, Scotland, or Northern Ireland for the rest
Last Collection Date	Date of the last collection for the practice
Up To Standard Date	Date at which the practice data is deemed to be of research quality. Derived using a CPRD algorithm that primarily looks at practice death recording and gaps in the data

Table A3.1: Staff variables

Variable	Description
Staff Identifier	Encrypted unique identifier given to the practice staff member entering the data
Staff Gender	Staff's gender
Staff Role	Role of the member of staff who created the event

Table A3.5: Clinical variables

Variable	Description
Patient Identifier	Encrypted unique identifier given to a patient in CPRD GOLD
Event Date	Date associated with the event, as entered by the GP
System Date	Date the event was entered into Vision (software system currently in use)
Consultation Type	Code for the category of event recorded within the GP system (e.g. diagnosis or symptom)
Consultation Identifier	Identifier that allows information about the consultation to be retrieved, when used in combination with Practice ID
Medical Code	CPRD unique code for the medical term selected by the GP
Staff Identifier	Identifier of the practice staff member entering the data. A value of 0 indicates that the Staff ID is unknown
Text Identifier	Identifier that allows free text information on the event to be retrieved, when used in combination with Practice ID and event type 'Clinical'. A value of 0 indicates that there is no free text information for this event
Episode	Episode type for a specific clinical event
Entity Type	Identifier that represents the structured data area in Vision where the data was entered
Additional Details Identifier	Identifier that allows additional information to be retrieved for this event, when used in combination with Practice ID. A value of 0 signifies that there is no additional information associated with the event.

Table A3.6: Additional clinical details

Variable	Description
Patient Identifier	Encrypted unique identifier given to a patient in CPRD GOLD
Entity Type	Identifier that represents the structured data area in Vision where the data was entered
Additional Details Identifier	Identifier that allows information about the original clinical event to be retrieved, when used in combination with Practice ID
Data 1-7	Depends on Entity Type

Table A3.7: Referral variables

Variable	Description
Patient Identifier	Encrypted unique identifier given to a patient in CPRD GOLD
Event Date	Date associated with the event, as entered by the GP
System Date	Date the event was entered into Vision
Consultation Type	Code for the category of event recorded within the GP system (e.g. management or administration)
Consultation Identifier	Identifier that allows information about the consultation to be retrieved, when used in combination with Practice ID
Medical Code	CPRD unique code for the medical term selected by the GP
Staff Identifier	Identifier of the practice staff member entering the data. A value of 0 indicates that the Staff ID is unknown
Text Identifier	Identifier that allows free text information on the event to be retrieved, when used in combination with Practice ID and event type 'Referral'. A value of 0 indicates that there is no free text information for this event
Source	Classification of the source of the referral e.g. GP, Self
NHS Speciality	Referral speciality according to the National Health Service (NHS) classification
FHSA Speciality	Referral speciality according to the Family Health Services Authority (FHSA) classification
In Patient	Classification of the type of referral, e.g. Day case, In patient
Attendance Type	Category describing whether the referral event is the first visit, a follow-up etc.
Urgency	Classification of the urgency of the referral e.g. Routine, Urgent

Table A3.8: Immunisation variables

Variable	Description
Patient Identifier	Encrypted unique identifier given to a patient in CPRD GOLD
Event Date	Date associated with the event, as entered by the GP
System Date	Date the event was entered into Vision
Consultation Type	Code for the category of event recorded within the GP system (e.g. intervention)
Consultation Identifier	Identifier that allows information about the consultation to be retrieved, when used in combination with Practice ID
Medical Code	CPRD unique code for the medical term selected by the GP
Staff Identifier	Identifier of the practice staff member entering the data. A value of 0 indicates that the Staff ID is unknown
Text Identifier	Identifier that allows free text information on the event to be retrieved, when used in combination with Practice ID and event type 'Immunisation'. A value of 0 indicates that there is no free text information for this event
Type	Individual components of an immunisation, e.g. Mumps, Rubella, Measles
Stage	Stage of the immunisation given, e.g. 1, 2, B2
Status	Status of the immunisation e.g. Advised, Given, Refusal
Compound	Immunisation compound administered – may be a single or multi-component preparation, e.g. MMR
Source	Location where the immunisation was administered, e.g. In this practice
Reason	Reason for administering the immunisation, e.g. Routine measure
Method	Route of administration for the immunisation, e.g. Oral, Intramuscular
Batch Number	Immunisation batch number

Table A3.9: Test variables

Variable	Description
Patient Identifier	Encrypted unique identifier given to a patient in CPRD GOLD
Event Date	Date associated with the event, as entered by the GP
System Date	Date the event was entered into Vision
Consultation Type	Code for the category of event recorded within the GP system (e.g. examination)
Consultation Identifier	Identifier that allows information about the consultation to be retrieved, when used in combination with Practice ID
Medical Code	CPRD unique code for the medical term selected by the GP
Staff Identifier	Identifier of the practice staff member entering the data. A value of 0 indicates that the Staff ID is unknown
Text Identifier	Identifier that allows free text information on the event to be retrieved, when used in combination with Practice ID and event type 'Test'. A value of 0 indicates that there is no free text information for this event
Entity Type	Identifier that represents the structured data area in Vision where the data was entered
Data 1-7	Qualifier, normal ranges, operator, value

Table A3.10: Therapy variables

Variable	Description
Patient Identifier	Encrypted unique identifier given to a patient in CPRD GOLD
Event Date	Date associated with the event, as entered by the GP
System Date	Date the event was entered into Vision
Consultation Identifier	Identifier that allows information about the consultation to be retrieved, when used in combination with Practice ID
Product Code	CPRD unique code for the treatment selected by the GP
Staff Identifier	Identifier of the practice staff member entering the data. A value of 0 indicates that the Staff ID is unknown
Text Identifier	Identifier that allows free text information (dosage) on the event to be retrieved, when used in combination with Practice ID and event type 'Therapy'. A value of 0 indicates that there is no free text information for the event. Use the Common Dosages Lookup (constituting ~ 95% of dosage strings in data) to interpret values < 100,000
BNF Code	Code representing the chapter & section from the British National Formulary for the product selected by GP
Total Quantity	Total quantity entered by the GP for the prescribed product
Numeric Daily Dose	Numeric daily dose prescribed for the event. Derived using a CPRD algorithm on common dosage strings (represented by textid < 100,000). Value is set to 0 for all dosage strings represented by a non-numeric textid
Number of Days	Number of treatment days prescribed for a specific therapy event
Number of Packs	Number of individual product packs prescribed for a specific therapy event
Pack Type	Pack size or type of the prescribed product
Issue Sequence Number	Number to indicate whether the event is associated with a repeat schedule. Value of 0 implies the event is not part of a repeat prescription. A value = 1 denotes the issue number for the prescription within a repeat schedule

Appendix 4: PDS fields

Table A4.1: PDS variables

PDS field	Description of data
NHS Number*	The unique patient identifier.
Patient name	Including any previous names, aliases and preferred name.
Date of birth	The patient's date of birth.
Place of birth	The patient's place of birth.
Additional birth information	The delivery time and birth order for multiple births.
Date of death	The patient's date of death.
Death notification status	Indicates a formal death certificate has been issued for the patient and the death has been registered.
Gender	Administrative gender.
Address	Includes main, temporary and correspondence addresses.
Alternative contacts	The patient's legal guardian, proxy, family/close contact.
Telecommunication details	Contact details such as telephone number, fax number and email address.
Preferred contact times	Patient's preferred contact times.
Preferred contact method	The patient's preferred contact method, e.g. telephone contact by proxy, no telephone contact, sign language required in face-to-face contact or minicom.
Preferred written communication format	Specialised patient contact requirements, e.g. large print, Braille, audio tape.
Preferred language	Information on patient's preferred language of communication.
Interpreter required	Indicates that the patient requires an interpreter.
NHS Care Record consent to share status	Indicates that the patient has agreed to share their health record.
Nominated dispensing contractor	The patient's nominated dispensing contractor that could include a community pharmacy, dispensing appliance contractor and a dispensing doctor.
Reason for removal	Indicates that the patient is no longer registered with a GP, e.g. if a patient has moved from England or Wales.
Previous NHS contact indicator	Indicates that the patient confirms they have had previous NHS treatment. This allows the PDS National Back Office to check for a duplicate of a record.
Patient call-back consent status	Indicates that the patient is willing to be called back from a Choose & Book call-centre.
Shared secret	An encrypted password used to validate a patient's identity when contacted from a Choose & Book call-centre.
Sensitive record indicator	Indicates either that the record is not accessible to PDS users, or that the content of the record is being reviewed to ensure the data is correct.
Primary care	The GP Practice with whom the patient is registered.
Date of registration	The date the patient registered with the GP Practice.
Serial change number	The mechanism for synchronising local and national records.

NHAIS information	The PDS holds certain information to allow it to interact with the NHAIS system that administers primary care. This information is not routinely accessible by the NHS, but NHAIS demographic information is regularly synchronised with PDS demographic information.

* A relational link is available within PDS that allows mother’s NHS number to be retained within the baby record (for births since 2006). This link is used for the summary care record and demographic spine within NHS Digital but is not routinely available for linking with external datasets.

Appendix 5: The feasibility of using linked data to monitor and evaluate the FNP programme: case studies at national and local level

This appendix presents four case studies that explore the feasibility of monitoring and evaluating FNP using linked data at a national and local level. It aims to illustrate further the important technical and methodological considerations described in Section 5 and 6. Firstly, we describe three case studies of potential analyses that could be carried out using national administrative data. For each case study, we identify key exemplar questions (relevant to FNP service development leads, supervisors, providers and commissioners) and consider the following aspects of study design:

- Population
- Exposure
- Comparator populations
- Outcome indicators
- Sample size
- Variables required

We also discuss specific considerations for analysis, along with data flows, approvals processes and linkage methods required to undertake these projects using national administrative data, as discussed in Section 4. We finally present a fourth case study to describe the development of data linkage systems in one local authority, and explore the potential advantages and challenges of linking and analysing FNPIS data at this level.

Case Study A: Measuring maternal and child health outcomes using linkage with Hospital Episode Statistics

This case study explores how the FNP programme could be monitored and evaluated using HES data and uses three exemplar questions to highlight important considerations related to study design. Addressing these questions requires linkage between FNPIS and HES to flag families who had enrolled (as well as appropriate comparison groups, where necessary); eligible pregnancies and births in HES could then be analysed.

Key exemplar questions

Question 1: Does FNP reduce repeat pregnancies within one year?

One of the main goals of the FNP programme is to help parents develop a vision for their future. A relevant outcome suggested in the FNP Logic Model is increased birth intervals between children, which indicates reproductive planning. Short inter-pregnancy intervals have been associated with adverse outcomes for both mothers and children (e.g. uterine rupture and low birth weight).[116, 117]

Question 2: Does FNP reduce the risk of injury-related hospital admissions in children aged <five years?

Injuries and ingestions are a major cause of disability and mortality among children aged <five in the UK; for example, they are the most common cause of death among children aged 1-4 accounting for up to 15% of deaths.[118] Hospital admissions for injuries and ingestions also incur significant healthcare costs.[119] The FNP programme aims to help parents provide sensitive and competent caregiving and as a result prevent both non-accidental and accidental injuries and ingestions.

Question 3: Could vulnerable women benefit from the FNP programme during a subsequent pregnancy?

The FNP programme is designed to provide additional support to first-time mothers aged <20 but the eligibility criteria can be adapted locally to include other vulnerable women; for example, first-time mothers aged <25.[120] However, women who are not first time mothers may also benefit from additional support; for example, the one in four women who have another child after their first child is removed by the courts.[34] By comparing characteristics and outcomes of women eligible for FNP and those who are vulnerable but are not eligible, it may be possible to identify other groups who could potentially benefit from the programme and guide adaption of the eligibility criteria.

Population

Questions 1 and 3

The extract can be defined based on deliveries identified within HES. There are a number of methods for this, and we recommend starting with records containing ICD-10 codes Z37-Z38 (outcome of delivery, live born infant), OPCS codes R14-R27 (delivery procedures), or two or more valid maternity tail fields.[121, 122]

Question 2

The population can be defined by either:

- Linking children to mothers (as defined in Question 1) in HES
- Identifying all relevant births in HES using ICD-10 codes and other fields relating to maternal age and parity as captured on the maternity tail (approximately 55% complete)

Statistical methods, such as propensity score matching, could be used to improve similarity between comparison groups.

Table A5.1: Key study design characteristics for exemplar questions 1-3

	Question 1 Does FNP reduce repeat pregnancies in the next year?	**Question 2** Does FNP reduce the risk of injury-related hospital admissions in children aged <five years?	**Question 3** Which other groups of women have similar vulnerabilities to first time mothers eligible for FNP?
Population	First time mothers aged <20 (in FNP and non-FNP LAs)	Index child for first time mothers aged <20 (in FNP and non-FNP LAs)	Full pregnancy histories for mothers aged <25
Exposure	a. Enrolment in FNP* b. Engagement in FNP (e.g. number of visits completed)	a. Enrolment in FNP* b. Engagement in FNP (e.g. number of visits completed)	First birth aged <20 years
Outcome indicators	- Rate of second pregnancy within 12 months - Time to second delivery	Rate of unplanned injury-related admission within five years of birth.	Proportion of women: - whose child has low birth-weight - who have a repeat pregnancy within 12 months Proportion of children with injury admissions
Comparison(s)	a. Women enrolled in FNP vs. similar women not enrolled* b. Association between number of visits and outcome among enrolled women	a. Children of women enrolled in FNP vs. similar women not enrolled* b. Association between number of visits and outcome among children of enrolled women	First birth to mothers aged <20 vs. births to all other mothers <25
Sample size	3874 women would be required to observe a significant reduction of 20% in the rate of repeat pregnancies in 12 months, or 16,168 women for a 10% reduction (assuming a baseline rate of 20%)	13,500 children would be required to observe a significant reduction of 20% in the rate of injury-related admissions by age five, or 28,400 women for a 10% reduction (assuming a baseline rate of 5%)	*Not applicable*

* see 'Considerations' box for details relating to exposure groups and comparisons

Considerations

Outcome indicators

Question 1 and 3

- HES does not capture all pregnancies; early miscarriages and terminations may be missed.
- Rates of repeat pregnancies during the FNP programme are recorded in FNPIS and could be used to validate outcomes in HES.

Question 3

- MCDS contains richer data related to risk factors and health behaviours such as smoking, alcohol consumption and substance use at booking that may be useful for identifying other groups of women who may benefit from FNP. However, pregnancy data has only been collected in MCDS since 2015; these data do not contain pregnancy history.

Comparisons

Question 1 and 2

- The comparison between similar mothers could be defined as:
    - i. Eligible mothers in the same LA, enrolled vs. not enrolled (assumes no selection bias, i.e. that mothers who were not enrolled were randomly missed, rather than more difficult to reach)
    - ii. All eligible mothers in a FNP LA vs. non-FNP LA (irrespective of actual enrolment), or in a non-FNP time period vs. non-FNP period (could produce bias towards the null as only uptake not 100%; methods to account for coverage would be required)
    - iii. Propensity-score matched mothers, enrolled vs. eligible in a non-FNP LA or non-FNP time period)

Question 1-3

- Analyses need to account for clustering of births for the same mother, and varying length of follow up.

Variables required

Table A5.2: Variables required for exemplar questions 1-3

	Variables required	Question
Defining the population	Maternal age	1, 2, 3
	Parity** (based on the number of previous pregnancies as captured in the maternity tail, or linking pregnancies across years)	1, 2, 3
	LSOA of residence*	1, 2, 3
	Month and year of delivery (based on baby's date of birth captured in maternity tail, or procedure codes and dates)	1, 2, 3
	Mother-baby link	2, 3
Outcomes	Subsequent pregnancy within 12 months, identified from subsequent admissions with the same HESID, using the above criteria to identify deliveries	1, 3
	Time to second delivery	1, 3
	Details of unplanned injury admissions in children up to age five (e.g. number of admissions, length of stay per admission, reason for admission)	2
	Birthweight	3
	Maternal / childhood deaths, including cause of death	1, 2, 3
Exposures	FNP participation and engagement, as defined through linkage	1, 2
	Maternal age and parity	3
Additional covariates	Mother: Pregnancy complications, previous birth outcomes, gestational age at first assessment, length of stay in hospital, ethnicity, IMD, etc.	1, 2, 3
	Child: Gestational age, birthweight, length of stay in hospital, IMD, ethnicity, season of birth, neonatal conditions (including congenital anomalies), postnatal stay, admission to NICU, etc.	1, 2, 3

*LSOA can be mapped to LA to identify LSOAs within FNP catchment areas (e.g. using data from https://census.edina.ac.uk). LSOA is approximately 99% complete in HES. **Parity is approximately 75% complete for mothers in HES; pregnancies can be mapped across previous years using HESID (this would require additional years of data).[123]

Case Study B: Measuring education outcomes for mothers and children through linkage to the National Pupil Database

This case study considers exemplar questions relevant to the long-term evaluation of educational outcomes of children and their mothers enrolled in FNP and to monitoring the educational characteristics of mothers enrolled. The case study requires linkage between FNPIS and NPD.

Key exemplar questions

Question 4: What is the impact of FNP on children's personal, social and emotional development at age five?

School readiness (including socioemotional development at school-entry) has been associated with later achievement in maths and English.[124] By helping parents to provide sensitive and competent caregiving, FNP seeks to improve children's personal, social and emotional development. The recent RCT evaluating FNP in England identified positive effects on early cognitive and language development at 24 months, and the US version of the programme has been associated with persistent benefits in child development and academic achievement up to age 12.[12, 125]

Question 5: What is the educational attainment of FNP mothers at enrolment?

Question 6: What is the impact of FNP engagement on a mother's subsequent educational attainment up to two years following programme completion?

Teenage mothers face numerous social, practical and financial barriers to returning to education.[126] One aim of the FNP programme is to improve the parental life course, which includes continuing their education. However, the information related to education that is routinely collected in FNPIS at baseline and throughout the programme is minimal.

Table A5.3: Key study design characteristics for exemplar questions 4-6

	Question 4 What is the impact of FNP on child development at age five?	**Question 5** What is the educational attainment of FNP mothers at enrolment?	**Question 6** What is the impact of FNP engagement on a mother's subsequent educational attainment?
Population	Children in Early Years education (in FNP and non-FNP LAs)	FNP mothers aged 16 or older on enrolment to FNP	FNP mothers who should have completed the programme more than two years ago
Exposure	a. Enrolment in FNP b. Engagement in FNP (e.g. number of visits completed)	*Not applicable*	Engagement in FNP (e.g. number of visits completed)
Outcome indicators	Proportion of children with a "good level of development" as defined by the DfE (i.e. meeting or exceeding the expected level of development in three prime areas of learning; communication and language, physical development and Personal, Social and Emotional (PSE) development).	Proportion of FNP mothers who have attained: - at least five GCSEs (at A*-C) including Maths and English - any other NVQ level 2 qualification (e.g. BTEC, Functional Skills, Skills for Life) - an NVQ level 3 or higher qualification (e.g. A level).	Proportion of FNP mothers who gained a qualification while enrolled in the programme or in the two years after it ended.
Comparison(s)	a. Children of women enrolled in FNP vs. similar women not enrolled b. Association between number of visits and outcome among children of enrolled women	*Not applicable*	Association between number of visits and outcomes among enrolled women
Sample size	1,500 children would be required to observe a significant increase of 20% in the proportion of children with a "good level of development" by age five, or 8,300 children for a 10% increase (assuming a baseline rate of 52%)	Between 2007 and 2014, 90% of FNP mothers were aged 16 or older on enrolment giving a sample size of approximately 11,300 women.	Approximately 8,000 mothers who enrolled in 2007-11 should have completed the FNP programme by 2013 (most recent year of graduation with two year follow-up)

* see 'Considerations' box for details relating to exposure groups and comparisons

Considerations

Outcome indicators

Question 4

- EYFSP scores for individual domains have a narrow distribution (range: 2-6 or 3-9) and would not be informative indicators of development: an alternative indicator of global development is a child's overall development score (although a high score can be achieved while not meeting the expected level of development in one or more domains). The domains of development that are included in the DfE's "good level of development" align closely with the FNP logic model and are functionally important.
- As EYFSP is assessed by classroom teachers, potential biases (based on ethnicity, deprivation or maternal age, for example) must be considered.
- Comparing the proportion of FNP and non-FNP children who achieve a "good level of development" will be subject to confounding by maternal age, which is not measured in NPD.
- The impact of FNP on child development could also be explored by comparing development within the population of FNP children by their exposure level (e.g. by the number of visits that were completed).
- Data for EYFSO is not comparable before 2012/13 due to a change in the EYFSP assessment

Question 5

- Information on educational attainment at enrolment (GCSEs and other qualifications) is recorded in FNPIS, and could be used to validate outcomes recorded in administrative data.
- Linked administrative data could also provide more information about the baseline educational level of FNP mothers, without the need for additional data collection. For example, FNPIS currently records only a binary indicator for qualifications other than GCSEs.
- Linkage with NPD will provide information about the schools that FNP mothers attended, which may inform broader public health interventions.
- Linkage to more sensitive data could be considered, for example, on special educational needs or school absenteeism, with strong justification for health and other benefits.

Question 6

- This analysis could be stratified by educational attainment or age at enrolment to explore different pathways through education that FNP mothers follow.
- Absence of gaining additional qualifications must be carefully interpreted as there may be missed matches between FNPIS and NPD data or a woman may have no match because she:
  - moved outside of England or died
  - is in employment
  - is in education or training with a provider that does not return data to NPD (i.e. who are not funded by Skills Funding Agency or Education Funding Agency).

Comparisons

Question 4

The most appropriate comparison group to explore the impact of FNP on children's development would be other children born to mothers aged <20. However, as maternal age is not recorded in NPD it is not possible to identify these children.

- Linkage of HES to NPD, and subsequently to FNPIS data, would allow this comparator group to be identified; however, it would require the sharing of identifiable data between government departments (NHS Digital, DfE and FNP at PHE).
- An alternative approach would be to use the relational link in PDS to identify children born to mothers <20. This approach would avoid the explicit use of health data but it has not been used before.

Questions 5 and 6

- These questions aim to monitor the educational attainment of FNP mothers at enrolment and their subsequent progress during the programme and after it is completed. Comparison to other groups (e.g. women of the same age) is not currently possible due to absence of information in NPD related to pregnancies; appropriate comparator groups could be achieved through linkage between held by DfE and NHS Digital.

Sample size

Question 4

- The completeness of data recorded in FNPIS for children born pre-2010 is likely to limit linkage to NPD for these years.

Population

As FNPIS and NPD data do not share a unique identifier, the population of FNP children can be defined via a set of common partially-identifying variables; for example, first name, last name, sex and date of birth.

Variables required

Table A5.4: Variables required for exemplar questions 4-6

	Variables required	Question
Defining the population	• First & last name	4, 5, 6
	• Sex	4, 5, 6
	• Date of birth	4, 5, 6
	• Postcode*	4, 5, 6
	• LSOA†	4
Outcome	• Score for all domains of Early Years Foundation Stage Profile (EYFSP)	4
	• KS4 and KS5 attainment	5, 6
	• Participation in further or higher education	5, 6
Exposure	• FNP participation, as defined through linkage	4
Additional covariates	• Factors associated with educational outcomes such as ethnicity, age within school year, special education needs (SEN) provision, etc.	4, 5, 6
	• Measures of deprivation such as free school meals (FSM) eligibility Income Deprivation Affecting Children Indices (IDACI) rank for child's home LSOA, etc.	4, 5, 6
	• Characteristics of school such as type, number of pupils, proportion of FSM/SEN, etc.	4, 5
	• Previous measures of child development recorded during FNP i.e. Ages & Stages Questionnaire (ASQ) scores	4

*A child's postcode is not recorded in FNPIS, but it could be assumed to be the same as their mother. †LSOA is recorded in NPD and can be mapped to local authority to identify a comparator population of children.

Considerations

- Requests for sensitive pupil-level data must be clearly justified when applying for an NPD extract and reasons why less sensitive information is not sufficient must be given. Tier 1/2 data (see Section 3 – Data sources) related to ethnicity, SEN status and FSM eligibility are required as they are associated with educational outcomes.
- Postcode could be used in addition to name, sex and date of birth when linking FNPIS and NPD data. However, the transience of the population should be considered: it is likely that some mothers will have moved address in the time between sitting their GCSEs and having their baby. Similarly, children may move address between the end of the programme and the time they start school. Therefore, postcodes should be updated via PDS prior to linkage.

Case Study C: Measuring social care outcomes for mothers and children through linkage to the Children Looked After and Children In Need datasets

This case study considers exemplar questions about the relationship between children's social care and FNP. Addressing these questions requires flagging of FNP mothers or children within NPD linked to CLA and CIN. Where appropriate, comparison groups will also be identified.

Key exemplar questions

Question 7: What is the impact of FNP on children's contact with social services by age two?

Question 8: What is the impact of FNP on children's placement in out-of-home care (OHC) by age five?

The development and promotion of secure attachment between parent and child is central to the FNP programme. By helping parents to provide sensitive and competent caregiving, it aims to reduce incidents of abuse and neglect during and after the programme. Interventions such as Child Protection Plans (CPP), Child In Need (CIN) status or placement in out-of-home care (OHC) are considered indicators of early adversity and have been associated with multiple adverse outcomes, in childhood and in later life.

Question 9: Do first-time mothers who are in OHC or have previously been in OHC participate in FNP?

Children whose parents have a history of maltreatment are more likely to be in contact with social services than other children (e.g. 7.3% vs. 0.4% by age one).[126] Participating in the FNP programme may help first-time mothers, who are in or have left care, to break the intergenerational transmission of maltreatment.

Table A5.5: Key study design characteristics for exemplar questions 7-9

	Question 7 What is the impact of FNP on children's contact with social services during the programme (by age two)?	**Question 8** What is the impact of FNP on children's placement in out-of-home care (OHC) by age five?	**Question 9** Do first-time mothers who are in OHC or care leavers participate in FNP?
Population	Children aged five or older (in FNP and non-FNP local authorities)	Children aged five or older (in FNP and non-FNP local authorities)	FNP mothers
Exposure	a. Enrolment in FNP* b. Engagement in FNP (e.g. number of visits completed)	a. Enrolment in FNP* b. Engagement in FNP (e.g. number of visits completed)	Placement in OHC
Outcome indicators	By age two: - Referred to social services - Recorded as CIN - Placed on CPP - Placed in OHC - Characteristics of care episodes (e.g. length, category of need, reason episode ceased)	By age six: - Placed in OHC - Characteristics of care episodes (e.g. length, category of need, reason episode ceased)	- Attrition from FNP at each phase - Engagement in each phase of FNP (e.g. number of visits completed)
Comparison(s)	a. Children of mothers enrolled in FNP vs. similar mothers not enrolled* b. Association between number of visits and outcomes among children of enrolled women	a. Children of women enrolled in FNP vs. similar children not enrolled* b. Association between number of visits and outcomes among children of enrolled women	a. FNP mothers who were in care or recently in care vs. those who were not in care in past few years
Sample size	2,900 children would be required to observe a significant reduction of 20% in the proportion of children referred to social services by age two, or 12,100 children for a 10% increase (assuming a baseline	31,400 children would be required to observe a significant reduction of 20% in the proportion of children placed in OHC by age six, or 132,700 children for a 10% reduction (assuming a	Between March 2009 and March 2014, there were approximately 7,500 mothers enrolled in FNP. Details of placement in OHC are not routinely

	rate of 20%)	baseline rate of 2.2%)	collected in FNPIS.

* see ‘Considerations’ box for details relating to exposure groups and comparisons

Considerations

Outcome indicators

Question 7 and 8

- Referrals and episodes of care that occur before school entry only will not have a UPN recorded in CLA/CIN. It will not be possible to link these episodes to a child’s NPD record or to FNPIS if they have not been in care after school entry. This will lead to an underestimation of contact with social care service and placement in OHC.
- Data on contact with social care services during the FNP programme is recorded in FNPIS. This could be used to assess the scale of underestimation due to missing UPN or matching errors in the linked dataset.

Question 9

- Data linkage between CLA and NPD is only possible post-2005 (when collection of UPN in CLA began); therefore, it will not be possible to ascertain ‘in care/recent care leaver’ status for earlier cohorts of FNP mothers.

Comparisons

Question 7 and 8

- Comparative analyses should take into account factors (such as ethnicity) which are associated with referrals to social care services and placement in OHC.
- Comparisons with non-FNP populations will be subject to bias due to confounding by maternal age not measured in NPD. This could be overcome through linkage between data held by DfE and NHS Digital to identify similar women and their pregnancies.
- The comparison between similar mothers/children could be defined as:
    i. Eligible mothers in the same LA, enrolled vs. not enrolled
    ii. All eligible mothers in a FNP LA vs. non-FNP LA (irrespective of actual enrolment), or in a non-FNP time period vs. non-FNP period
- Comparisons of contact with social care services (including placement in OHC) needs careful interpretation, even after potential confounding factors have been taken into account. There may be more contact with social care services among FNP families initially because of increased professional surveillance during the programme and increased parental awareness of risk.
- It may also be useful to explore children’s trajectories of care over time, to determine the impact of FNP on reducing escalation in interventions and improving engagement with social care services.

Population

CLA and CINC datasets do not contain NHS number or name, but do contain UPN which can be used to link to NPD. However, FNPIS and NPD data do not share a unique identifier; therefore, the population of FNP children will be defined via a set of common partially-identifying variables (for example, first name, last name, sex and date of birth).

Variables required

Table A5.6: Variables required for exemplar questions 7-9

	Variables required	Question
Defining the population	• First & last name	7, 8, 9
	• Sex	7, 8, 9
	• Date of birth	7, 8, 9
	• Postcode*	7, 8, 9
	• LSOA†	9
Outcome	• Details of referral to social services	7
	• Details of CPP status (number of episodes, length, category of need)	7
	• Details of CIN status (number of episodes, length, category of need)	7
	• Details of placement in OHC (number of episodes, length, reason looked after)	7, 8, 9
Exposure	• FNP participation, as defined through linkage	8, 9
Additional covariates	• Factors associated with outcomes such as ethnicity, age within school year, special education needs (SEN) provision, etc.	8, 9
	• Measures of deprivation such as free school meals (FSM) eligibility Income Deprivation Affecting Children Indices (IDACI) rank for child's home LSOA, etc.	8, 9

*A child's postcode is not recorded in FNPIS, but it could be assumed to be the same as their mother. †LSOA is recorded in NPD and can be mapped to local authority to identify a comparator population of children.

Considerations

- All variables from CLA and CINC extracts are tier 1 (see Section 3 – Data sources). Requests for such sensitive child-level data must be clearly justified when applying for an NPD extract.

Data flows and approval processes for linking FNPIS to national administrative data

Data access requirements for HES, NPD, CLA and CINC are described in detail in the main report. Ethics approval may be required for research studies and initial discussions with each of these bodies are recommended prior to submission of applications. In brief, access to HES requires applications to NHS Digital DARS, CAG and ONS (for mortality data, if required) and for NPD, CLA and CIN, applications to DfE are required. Applications that request CLA, CINC and sensitive NPD data (or linkage to other datasets) must be approved by the Data Management Advisory Panel (DMAP) who meet monthly. While applications for standard NPD extracts are expected to be processed within 6-8 weeks, applications that require DMAP approval can take considerably longer.

When making an application for HES, NPD, CLA or CINC data that will be linked to other data, the data flows for processing and linking must be clearly described. For example, the data flows for processing and linking FNP and HES data are as follows (see Figure A5.1):

1. Identifiers for mother and babies participating in FNP are currently held by NHS Digital in Exeter. These identifiers will be updated via PDS within NHS Digital before linkage.
2. FNP identifiers only, with the pseudonymised FNP study ID, will be transferred to the NHS Digital linkage service for all mothers and babies participating in FNP.
3. A HES extract of mothers and babies including a pseudonymised HESID and other requested variables will be prepared by NHS Digital.
4. NHS Digital will link FNP identifiers to HES identifiers (trusted third party linkage) and create a link-key to map between the pseudonymised HESID and the FNP study ID for both mother and baby.
5. Once satisfactory linkage has been achieved, FNP identifiers held by the NHS Digital linkage team will be destroyed.
6. The pseudonymised HES extract and HESID will be transferred with the link-key from NHS Digital to the secure setting.
7. The pseudonymised FNP extract and FNP study ID will be transferred from the FNP NU to the secure setting.
8. The link-key will be used to link FNP data with HES within the secure setting. Identifiable data will not be held in the secure setting – FNP identifiers will remain with NHS Digital in Exeter.

The data flows for processing and linking FNP and NPD, CLA or CINC data are similar (see Figure A5.2), except the trusted third party will link FNP identifiers to NPD identifiers to create a link-key to map between the pseudonymised Pupil Matching Reference (which is based on UPN) and the FNP study ID.

Considerations

- An appropriate secure setting, meeting IG requirements, will be needed to hold the anonymised linked FNP-HES data (e.g. FNP NU at Tavistock and Portland NHS Foundation Trust, or the UCL Data Safe Haven).

Figure A5.1: Data flow diagram for HES-FNPIS linkage for mothers and babies

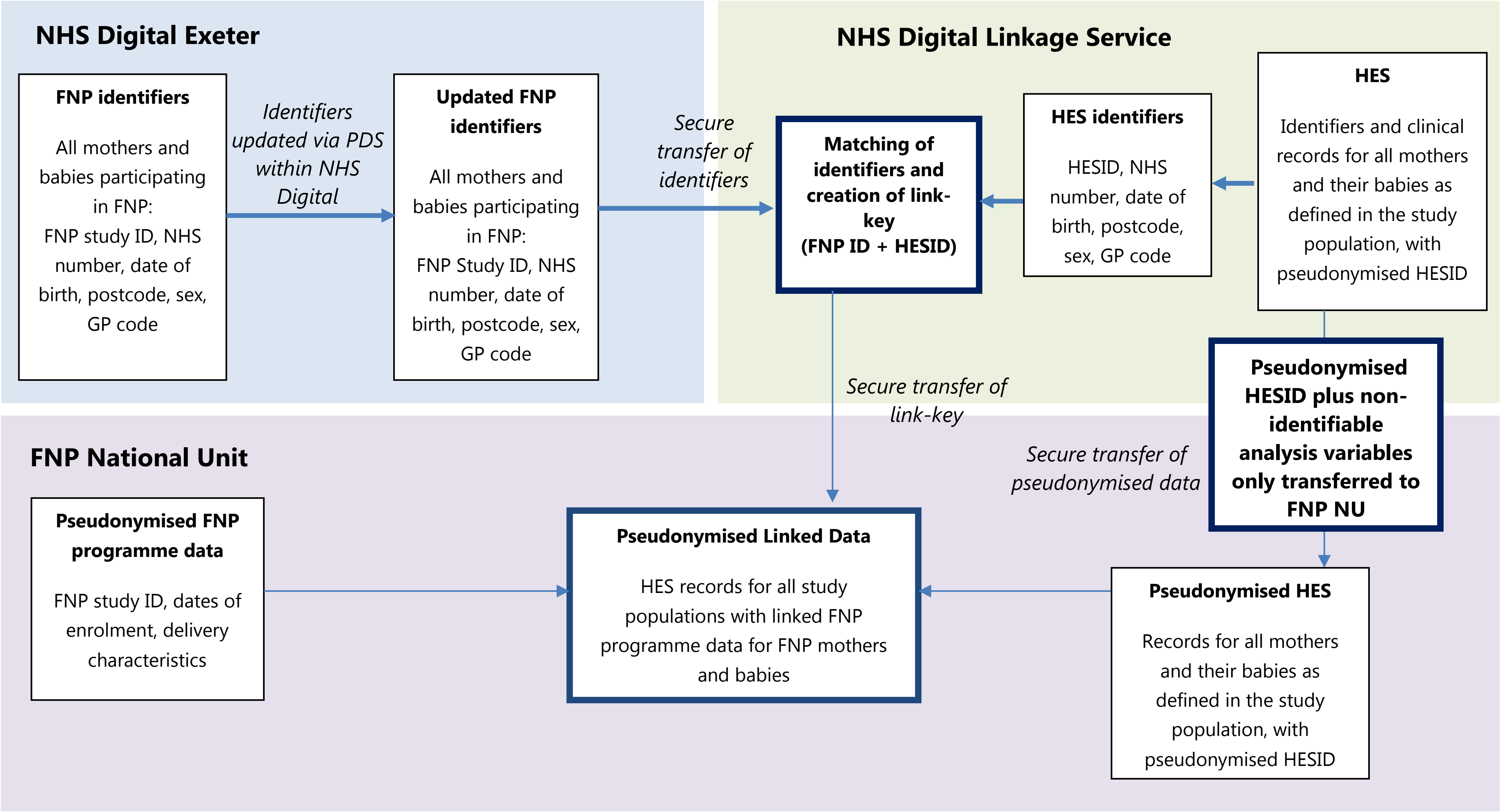
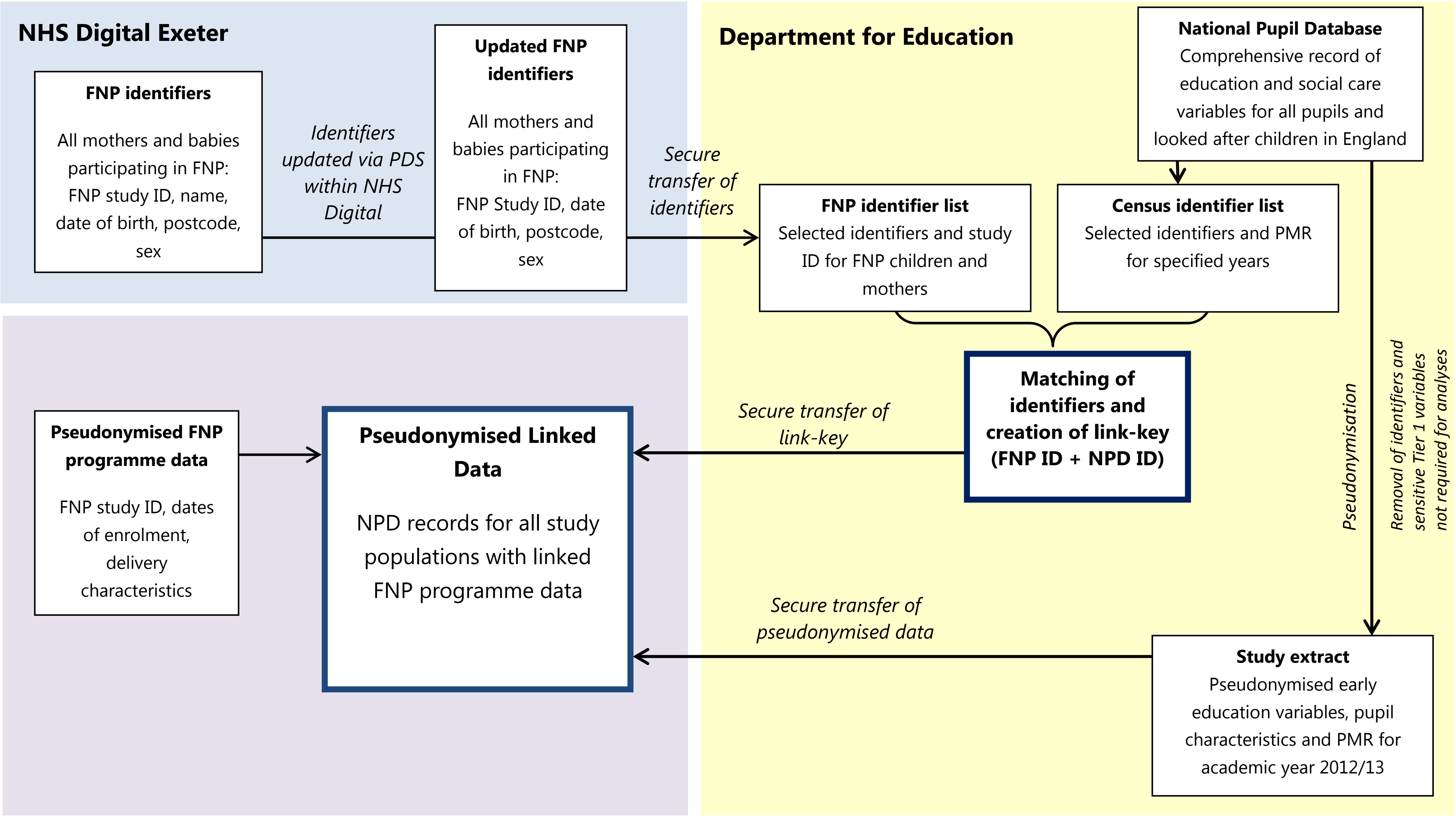


Figure A5.2: Data flow diagram for NPD-FNPIS linkage



Linkage methods

The majority of linkage between FNPIS and HES or NPD should be possible using deterministic (exact) matching on combinations of PDS-updated identifiers: name, date of birth, postcode, sex and (for HES only) NHS number and GP code. As previously described, linkage of CLA/CINC is only possible through linkage with NPD, and depends on UPN being recorded.

To answer specific questions about the relationship between FNP, maternal characteristics and child outcomes, linkage of mothers and their children within comparison cohorts will be required. This is not routinely done within HES, but could be facilitated by the mother-baby relational link in PDS, which contains NHS numbers for both the mother and the baby. An alternative method would be to use probabilistic linkage based on common identifiers in both maternal and baby HES records (gestational age, birthweight, LSOA, etc.). Accuracy of identifiers may differ between datasets; depending on the number of unlinked FNP records, probabilistic linkage and/or manual review may be required. The quality of linkage could be assessed by:

- Quantifying the number of unlinked FNP participants (missed-matches)
- Using validation methods to identify any false-matches (e.g. activity after death, multiple HESIDs or PMRs per FNP ID, implausible clinical scenarios)[127]
- Comparing the characteristics of unlinked and linked records to identify any potential sources of bias.

Considerations

- Manual review may be feasible with a small sample size; however, it would need to be negotiated with data providers, may incur additional costs and is not sustainable in the long-term.
- To enable routine linkage of FNPIS data to HES, NPD, CLA and CINC a robust matching algorithm that is not reliant on manual review must be developed.
- The creation of a synthetic dataset that recreates the quality of FNPIS identifiers would be useful when developing an optimum matching algorithm.
- The time required to develop a matching algorithm should be factored into project timelines.
- Information on the algorithm or linkage criteria used to match each record should be provided by the trusted third party performing the linkage.

Case Study D: Using the locally linked data to monitor and evaluate FNP

Funding for children's public health services transferred from NHS England to LAs in October 2015. As a result, LAs increasingly require evidence on the implementation and effectiveness of services (such as FNP) in their local area, to monitor service provision, coverage and delivery and to support commissioning decisions. This case study explores the feasibility of monitoring and evaluating the FNP programme at a local authority level, focusing on the experiences of Kent. Kent are an early implementer of local data linkage and considered a leader in integrated intelligence by NHS Improving Quality[128] due to their development of the Kent Integrated Dataset and its contribution to service change locally. This case study outlines the development of the Kent Integrated Dataset and its planned expansion to include children, and highlights key ethical, technical and methodological challenges in these processes, as previously discussed in detail in Sections 4-6. Given these considerations, this case study reflects on the feasibility of using local data linkage to monitor and evaluate FNP in Kent.

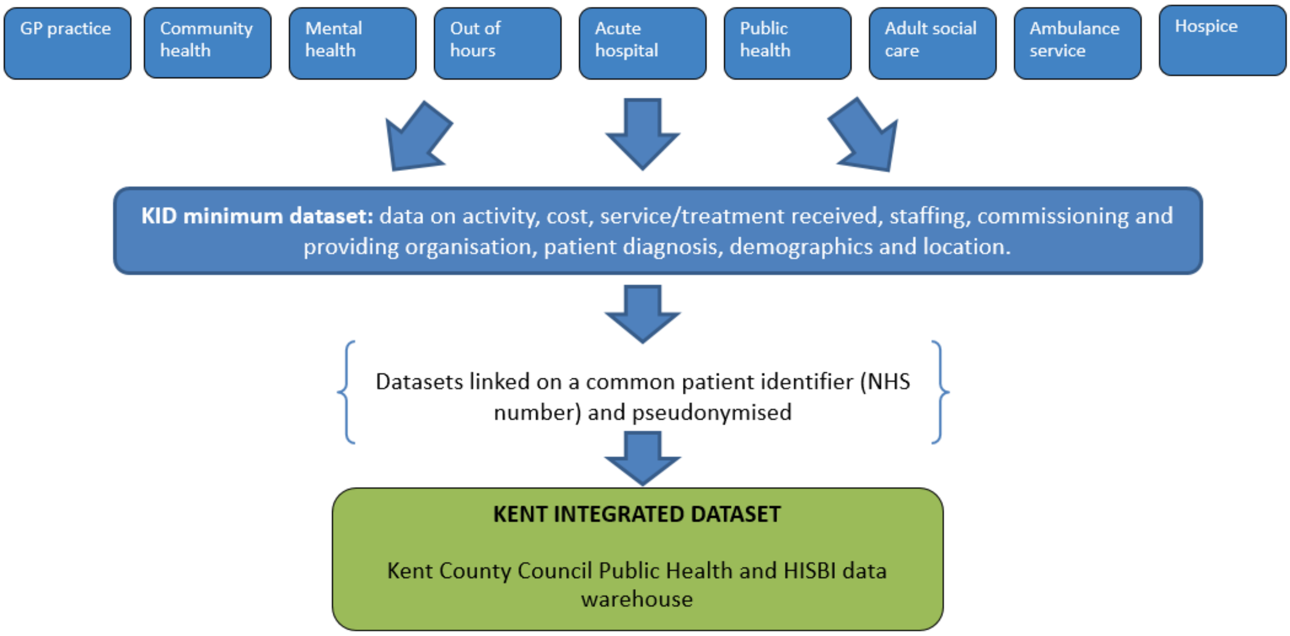
Development of the Kent Integrated Dataset

The Kent Integrated Dataset (KID) was developed in 2012 as part of the Year of Care Commissioning Programme, with the aim of creating a baseline profile of how people with long-term conditions in Kent made use of hospital and other services. It contains health and social care data for adults from multiple sources, including hospitals, community health, primary care and GP practices (Fig. A5.3). To achieve this integration of data, agreements were established between care providers and Kent County Council (KCC) to share and link data for the benefit of public health. Data are transferred to a trusted third party warehouse on an ongoing basis via a secure gateway and N3 connection. Data are linked and pseudonymised before being returned to the public health team at KCC and used to create dashboards and reports for care providers. KID was initially created to develop capitated budgets but it has also been used for service evaluation (e.g. to explore the impact of a community health intervention on fall-related hospital admissions among the elderly population). There are plans to expand KID further to include data related to probation, fire & rescue and education.

Expansion of KID to include children's data

The KID does not currently include data related to children (except for details of specialist children's services). 'KID for kids' is the planned expansion of KID to include a range of children's datasets. This integrated dataset would be used to explore service use and outcomes among children in Kent. This project is currently being trialled using a feasibility study that seeks to link KCC's integrated education dataset to Child and Adolescent Mental Health Services (CAMHS) data. The locally linked data would be used to explore children's use of mental health services and the impact on their emotional wellbeing (as measured by clinical tools such as the Strengths and Difficulties Questionnaires, for example) and educational outcomes, such as exam attainment, absences and exclusions. This project began in September 2015; by June 2016 data flows and data sharing agreements had been agreed, technical aspects of data transfer and storage had been decided and matching algorithms for the two datasets were being developed. Initial linkage and user acceptability testing are anticipated to be completed by September 2016 (however, this is dependent on final approval from the relevant Caldicott Guardian, which could introduce a delay).

Figure A5.3: Flow of data into Kent Integrated Dataset



Using local linked data to monitor and evaluate FNP in Kent

Depending on the outcome of the current 'KID for kids' feasibility study it may be possible to link FNPIS data to Kent's integrated education dataset. While the current feasibility project will have taken >12 months to negotiate data sharing and to develop technical solutions for transferring, processing, storing and linking data locally, linkage of FNPIS data may take less time as the technical aspects (e.g. data flows and matching algorithm) would already be established. This linkage would allow the educational outcomes of FNP children to be explored locally; however, as Kent's integrated education dataset does not contain early education data, it could only be used to explore educational outcomes after entry to foundation stage education (something that can already be achieved through linkage to NPD administrative data). Linkage would also be prospective only, which would restrict analyses to more recent cohorts of FNP children, delay evaluation of the programme until current cohorts reach school going age and limit sample size and the power to detect a significant effect of the programme. Furthermore, as Kent's integrated dataset does not contain maternal age, the comparator group for analyses would be all non-FNP children in the local authority (potentially diluting results).

An alternative approach for evaluating the effect of FNP programme in Kent may be to link data from its Community Information System (CIS) to other datasets, such as the integrated education dataset. CIS is an electronic dataset that records limited information related to the universal health-visiting programme. (Currently, clinical records are paper based, though there are plans to record these electronically in CIS in the future). CIS records details of health visits from the antenatal period to the age of two and began collecting data in October 2015. It primarily contains information related to Key Performance Indicators for the local authority's health visiting service (e.g. smoking, breastfeeding) but also contains limited data related to safeguarding that is shared by social services within the local authority. FNP visits are also recorded in CIS at time points that correspond with the mandated Healthy Child Programme visits (i.e. antenatal, 10-14 days post-birth, six weeks, one year and two years). This information could be used as an indicator for FNP children in the dataset. CIS also contains NHS number and other partially identifying factors for mothers and children that could be used to link to other datasets. Linkage of CIS data to other datasets would allow outcomes for FNP children to be compared locally with children of other young mothers. However, as data collection only began in October 2015 evaluation of longer-term outcomes could not yet be explored using CIS data.

Summary

Within Kent, there is potential to monitor and evaluate FNP by linking FNPIS data to existing local datasets, such as CIS; however, considerable technical and methodological barriers exist. Firstly, establishing permission pathways, negotiating data sharing agreements and developing algorithms for matching would require significant investment of time from local FNP teams (i.e. commissioners, providers, supervisors, Family Nurses). The costs of data linkage, storage and processing would also need to be met by local sites. Furthermore, the utility of analyses of linked datasets may take several years to realise, as data linkage is prospective only. While linkage to CIS would allow a comparator population teenage mothers not enrolled in FNP to be identified, there may be other underlying differences in the populations that cannot be accounted for. For example, the teenage mothers not enrolled in the programme may not have been eligible for FNP due to previous births or late engagement with midwifery services, or may have been offered and refused to participate. Finally, while Kent is a large local authority is likely that the power to detect significant differences between groups will be limited, particularly if coverage and uptake of FNP among teenage mothers is high.

Appendix 6: Detailed assessment of priorities for linkage between FNPIS and administrative data

Table A6.1: Assessment of achievability and utility of linkage between FNPIS and administrative data

Dataset	Indicators relevant to the monitoring and evaluation of FNP *	Data linkage	Time period available	Costs for accessing or linking data	Time line for data access and linkage	Comparator population
HES-ONS	**Pregnancy health behaviours and birth, e.g.:** - *Illegal drug & alcohol use in pregnancy* - Pregnancy complications - Subsequent pregnancies **Maternal health, e.g.:** - Accidents and injuries - *Illegal drug & alcohol use post-partum* **Child health and development, e.g.:** - Disability - Accidents and injuries - *Nutrition* **Use of services, e.g.:** - Maternal hospital admissions and attendances - Child hospital admissions and attendances	Both deterministic and probabilistic matching is possible. Datasets contain common identifiers (including NHS number, name and date of birth for both mother and child). However, there is no mother-child link within HES.	Inpatient: 1997→ Outpatient: 2003→ A&E: 2007→	£3,700 for set up, plus additional charges for manual matching	First decision within 8-10 weeks from submission of application. Data supplied within 4 weeks of approval. However, process can take up to 12 months from date of application. Time to complete application form should also be considered.	Robust comparators are identifiable as contains information on maternal age.
NPD	**Child and maternal education** - Age at school start - Early education attainment (FSP, KS1) - Attainment (KS2, KS4, KS5)	Only probabilistic matching is possible. Datasets contain common identifiers (including name	Varies depending on extract but most data are	No costs for data access or linkage.	First decision within 6-8 weeks from submission of application for less sensitive tier 3/4 data. Applications for tier 1/2	Robust comparators are not identifiable as contains

	- Attendance and exclusions - Free school meals eligibility - Special educational needs status **Child cognitive & social development, e.g.:** - Developmental assessments **Use of services, e.g.:** - Funded early years education uptake - Free school meals uptake	and date of birth for both mother and child). However, NHS number is not recorded in NPD and UPN is not recorded in FNPIS.	available from 2001, and some are available from 1995.		will take longer. Data supplied within 4 weeks of approval. However, process can take up to 12 months from date of application. Time to complete application form should also be considered.	no information on maternal age.
MCDS	**Pregnancy health behaviours and birth, e.g.:** - Smoking in pregnancy/at delivery - Pregnancy complications - Birth outcome (still or live) - *Stay in NICU or SCBU* - Birth weight / length of gestation - Neonatal death **Child cognitive & social development, e.g.:** - Developmental assessments **Child health and development, e.g.:** - Breastfeeding to discharge - Disability - Height & weight (BMI) - Immunisations **Contact with social care services, e.g.:** - Out-of-home care - Child protection plan	Both deterministic and probabilistic matching is possible. Datasets contain common identifiers (including NHS number, name and date of birth for both mother and child). There is also a mother-child link within MCDS.	2015 →	MCDS is currently in set-up phase and processes and costs for data access are yet to be determined. Likely to be comparable to HES.	MCDS is currently in set-up phase and processes and costs for data access are yet to be determined. Likely to be comparable to HES.	Robust comparators are identifiable as contains information on maternal age.
CPRD	**Pregnancy health behaviours and**	Both deterministic and	1987→	£155,000 for	Varies by study.	Robust

	birth, e.g.: - *Contraceptive use post-partum* - *Smoking in pregnancy/at delivery* **Maternal health, e.g.:** - Mental health (during/post-pregnancy) - *BMI* - *Smoking post-partum* - *Domestic violence* **Use of services, e.g.:** - Maternal primary care contact - Child primary care contact	probabilistic matching is possible. Datasets contain common identifiers (including NHS number, name and date of birth for both mother and child). There is a mother-child link within CPRD.		annual licence, plus £30,000+ for data linkage.	Initial response within five days of application. However, process can take up to 12 months from date of application. Time to complete application form should also be considered.	comparators are not readily identifiable as contains no information on maternal age.
CLA/CINC	**Contact with social care services, e.g.:** - Out-of-home care - Child protection plan - Referrals to social service - Child in need status	Only probabilistic matching is possible. Datasets contain limited common identifiers (date of birth and ethnicity for both mother and child). NHS number and name are not recorded in CLA/CINC.	CLA:1992→ CINC:2008→	No costs for data access or linkage.	CLA/CINC are tier 1 and approval will take longer than the standard 6-8 weeks. Data supplied within 4 weeks of approval. However, process can take up to 12 months from date of application. Time to complete application form should also be considered.	Robust comparators are not readily identifiable as contains no information on maternal age.
CAFCASS	**Contact with family courts, e.g.:** - *Family breakdown (divorce, separation)*	Only probabilistic matching is possible.	2001→	CAFCASS data has had limited use for	CAFCASS data has had limited use for research and evaluation purposes	Robust comparators are not

	- Custody arrangements	Datasets contain common identifiers (name, date of birth and address for both mother and child). NHS number is not recorded in CAFCASS.		research and evaluation purposes and the costs for data access are not established.	and the timeline for data access are not established.	readily identifiable as contains no information on maternal age.
DWP/HMRC	**Employment and benefits, e.g.:** *- Income* *- Employment/unemployment* *- Receipt of benefits*	Only probabilistic matching is likely to be possible. Datasets contain common identifiers (name, date of birth and postcode), though the availability of these identifiers is not established. NHS number is not recorded in DWP/HMRC data.	Unknown	DWP/HMRC data has had limited use for research and evaluation purposes and the costs for data access are not established.	DWP/HMRC data has had limited use for research and evaluation purposes and the timeline for data access are not established.	Robust comparators are not readily identifiable as contains no information on maternal age.
MoJ	**Contact with justice system, e.g.:** *- Placement in youth custody* *- Criminal offences committed*	Only probabilistic matching is likely to be possible. Datasets contain common identifiers (name, date of birth and postcode), though the availability of these identifiers is not established. NHS number is not recorded in MoJ data.	Unknown	MoJ data has had limited use for research and evaluation purposes and the costs for data access are not established.	MoJ data has had limited use for research and evaluation purposes and the timeline for data access are not established.	Robust comparators are not readily identifiable as contains no information on maternal age.

HES=Hospital Episode Statistics; ONS=Office of National Statistics; NPD=National Pupil Database; MCDS; Maternal and Children's Data Set; CPRD; Clinical Practice Research Datalink; CLA=Children Looked After return; CINC=Child in Need Census; CAFCASS=Children and Family Courts Advice and Support Service CAFCASS=Children and Family Court; DWP=Department for Work and Pensions; HMRC=Her Majesty's Revenue and Customs; MoJ=Ministry of Justice. *Italics indicate uncertainty in data quality.