BMJ Open

Assessing the medium-term impact of a home-visiting programme on child maltreatment through routine data linkage to a trial cohort: The protocol for the Building Blocks 2-6 study.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-015728
Article Type:	Protocol
Date Submitted by the Author:	23-Dec-2016
Complete List of Authors:	Lugg-Widger, Fiona; Cardiff University, Centre for Trials Research Cannings John, Rebeca; Cardiff University, Centre for Trials Research Channon, Sue; Cardiff University, Centre for Trials Research Fitzsimmons, Deborah; Swansea University, Swansea Centre for Health Economics Hood, Kerenza; Cardiff University, Centre for Trials Research Jones, Kerina; Swansea University, College of Medicine Kemp, Alison; Cardiff University School of Medicine, Division of Population Medicine Kenkre, Joyce; University of South Wales, 5. Faculty of Life Sciences and Education Longo, Mirella; Cardiff University School of Medicine, Marie Curie Palliative Care Research Centre McEwan, Kirsten; University of Derby, Department of Psychology Moody, Gwenllian; Cardiff University, Centre for Trials Research Owen-Jones, Eleri; Cardiff University, Centre for Trials Research Sanders, Julia; Cardiff University, School of Healthcare Sciences Segrott, Jeremy; Cardiff University, Centre for Trials Research; Cardiff University, DECIPHer Centre, Cardiff School of Social Sciences Robling, Michael; Cardiff University, Centre for Trials Research
Primary Subject Heading:	Public health
Secondary Subject Heading:	Health informatics, Paediatrics, Research methods
Keywords:	Child maltreatment, Family Nurse Partnership, Early years prevention, Medical Record Linkage, Follow-Up Studies

Assessing the medium-term impact of a home-visiting programme on child maltreatment through routine data linkage to a trial cohort: The protocol for the Building Blocks 2-6 study.

Corresponding author

Fiona Lugg-Widger[1] LuggFV@cardiff.ac.uk Telephone: 02920687519

Address: 7th Floor Neuadd Meirionnydd, Heath Park, Cardiff. CF14 4YS.

Rebecca Cannings-John[1] CanningsRL@cardiff.ac.uk
Sue Channon[1] ChannonS2@cardiff.ac.uk
Deborah Fitzsimmons[2] D.Fitzsimmons@swansea.ac.uk
Kerry Hood[1] HoodK1@cardiff.ac.uk
Kerina Jones[3] k.h.jones@swansea.ac.uk
Alison Kemp[4] KempAM@cf.ac.uk
Joyce Kenkre[5] joyce.kenkre@southwales.ac.uk
Mirella Longo[6] LongoM1@cardiff.ac.uk
Kirsten McEwan[7] k.mcewan@derby.ac.uk
Gwenllian Moody[1] MoodyG@cardiff.ac.uk
Eleri Owen-Jones[1] Owen-JonesCE@cardiff.ac.uk
Julia Sanders[8] SandersJ3@cardiff.ac.uk
Jeremy Segrott[1,9] SegrottJ@cardiff.ac.uk
Michael Robling[1] RoblingMR@cardiff.ac.uk

Author Affiliations

1. South East Wales Trials Unit, Centre for Trials Research, Cardiff University, Cardiff, UK.
2. Swansea Centre for Health Economics, Swansea University, Swansea, UK
3. College of Medicine, Swansea University, Swansea, UK
4. Division of Population Medicine, School of Medicine, Cardiff University, Cardiff, UK.
5. Faculty of Life Sciences and Education, University of South Wales, Pontypridd, UK
6. Marie Curie Palliative Care Research Centre, Cardiff University, Cardiff, UK
7. Department of Psychology, University of Derby, UK
8. School of Healthcare Sciences, Cardiff University, Cardiff, UK.
9. DECIPHer Centre, Cardiff School of Social Sciences, Cardiff University, Cardiff, UK.

Word Count: 4608

Keywords: Child maltreatment, Family Nurse Partnership, Early years prevention, Medical Record Linkage, Follow-Up Studies

ABSTRACT

Introduction: Child maltreatment involves acts of omission (neglect) or commission (abuse) often by caregivers that results in potential or actual harm to a child. The Building Blocks trial [ISRCTN 23019866] assessed the short-term impact of an intensive programme of antenatal and postnatal visiting by specially trained nurses to support young pregnant women in England. The Building Blocks: 2-6 study will assess the medium-term impacts of the programme for mothers and children (n=1562), through the linkage of routinely collected data to the trial data, with a particular emphasis on the programme's impact upon preventing child maltreatment.

Methods and analysis: We have developed bespoke model of data linkage whereby outcome data for the trial cohort will be retrieved by linked anonymous data abstraction from NHS Digital, Office for National Statistics and the Department for Education's National Pupil Database. Participants will be given reasonable opportunity to opt-out of this study prior to data transfer. The information centres will match participants to the information held in their databases using standard identifiers, and send extracts to a third party safe haven. The study will have 80% power to detect a 4% difference (4% vs 8%) for the binary primary outcome of Child in Need status (from birth to key stage one) at a two-sided 5% alpha level by following up 602 children in each trial arm. Analysis will be by intention to treat using logistic multilevel modelling. A cost and consequences analysis will extend the time-frame of the economic analysis from the original trial.

Ethics and Dissemination: The study protocol has been approved by NHS Wales Research Ethics Committee and the Health Research Authority's Confidentiality Advisory Group. Methods of innovative study design and findings will be disseminated through peer-review journals and conferences, results will be of interest to clinical and policy stakeholders in the UK.

Strengths and limitations of this study

- This study aims to provide much needed evidence about the medium-term benefits of the Family Nurse Partnership programme in England. This study has the capacity to either confirm the current perspective on the value of the intervention or demonstrate clinically meaningful benefits to children in vulnerable young families.
- There are distinct benefits associated with using routine data including a reduction in cost and participant burden over prospectively collected data, and relative completeness and therefore minimisation of bias over self-report, particularly for such sensitive outcomes.
- The establishment of a regulatory secure research database for this cohort of trial participants also offers the prospect of further data being added over the longer term and of broadening the scope of the dataset to other outcome domains relevant to this intervention, such as criminal justice and welfare benefits.
- The extent of this benefit will be balanced by our ability to adequately access the data from information centres in a timely fashion, the quality of matching conducted as well as the quality of the data ultimately retrieved.

INTRODUCTION

Maltreatment

Child maltreatment involves acts of omission (neglect) or commission (abuse) often by caregivers that inflict harm, or fail to act to prevent harm to a child.[1] Abuse may be physical, emotional or sexual. Neglect represents persistent failure to meet basic physical or psychological needs, often resulting in serious impairment of the child's health and/or development.[1] Neglect may involve failing to: protect a child from physical and emotional harm or danger; provide adequate supervision; or ensure access to appropriate medical care. In the year ending 31st March 2015 in England there were 635,600 referrals to children's social care services, 403,400 children starting an episode of need (an overall rate

of 348.0 per 10,000) and 62,200 children became subject of a child protection plan.[2] Of children who became subject of a child protection plan, the most common initial category of abuse was neglect (43.2%) followed by emotional abuse (33.7%).

Family Nurse Partnership home-visiting programme

There has been increasing emphasis upon the primary prevention of child maltreatment, including interventions directed at general populations and those targeting high-risk groups.[3] One such intervention is the Family Nurse Partnership (FNP) programme (developed in the US as the Nurse Family Partnership or NFP) – a home-visiting approach with three overarching goals: to improve birth outcomes, optimise child health and development - including reducing maltreatment - and promote economic self-sufficiency of mothers.[4]

In three US trials (in Elmira, Memphis and Denver),[5-7] the NFP has demonstrated improvements in prenatal health behaviours and birth outcomes, sensitive child care, maternal life course (e.g. greater workforce participation) and child and adolescent functioning. It has also shown positive effects in relation to reductions in rates of child injuries, abuse and neglect. In the first US trial in 1977, a sub-group analysis of poor unmarried teens (54 families) revealed that by age 2 there was verified abuse / neglect in 19% of control children compared to 4% in the group in receipt of NFP in both pregnancy and infancy (treatment difference of 0.15, 95% confidence interval of -0.01 to 0.31) and 56% relative reduction in emergency department encounters for injuries and ingestions during the second year of life.[5] Amongst the sub-group of children (56 families) with a state-verified report of maltreatment by age 4, the NFP group of children exhibited fewer risks for harm than the control group (e.g. fewer attendances with injuries / ingestions, safer home environment) at follow-up points between 25 and 50 months of life.[8] This was considered to be due to the earlier and more comprehensive detection of maltreatment by nurse-visited families.

The NFP programme was adapted for implementation as the Family Nurse Partnership and was introduced in England in 2007. Our Building Blocks trial (ISRCTN23019866) was the first trial of FNP in England and evaluated short-term outcomes to age 2 – the duration of the FNP programme.[9] The trial reported no difference for four primary outcomes: biomarker-calibrated self-reported tobacco use by the mother at late pregnancy, birth weight of the baby, the proportion of women with a second pregnancy within 24 months post-partum, and emergency attendances and hospital admissions for the child within 24 months post-partum.[10] We observed some differences for secondary child development outcomes including the rate of safeguarding events reported in primary care records. While the current evidence does not support continuation of the programme in England, previous evaluations have demonstrated benefit over the longer-term (e.g. up to 15 years of age).[11] For maltreatment outcomes this benefit has been increasingly evident after age 4 years,[12] therefore, the current study will establish whether FNP has moderated maltreatment outcomes over a medium-term period of follow-up (i.e. to the point where the child is aged six years old).

METHODS AND ANALYSIS

Research objective

The Building Blocks: 2-6 Study will use data linkage of routinely collected national datasets to assess the medium-term impact of the FNP intervention upon child maltreatment outcomes and key indicators of neglect.

Study design

This is a data linkage study, which will generate a linked anonymised database hosted by an independent Trusted Third Party (TTP). Participant mothers and children from Building Blocks: 0-2 (BB:0-2) will be followed up for a further four years using routine data only. Data from various routine public sector sources will be retrieved and linked to the trial data to

enable children and mothers to be followed until the child reaches key stage one (the two years of schooling when pupils are aged between 5 and 7). A summary of the data sources is provided in Table 1 and the time points for each dataset are shown in Figure 1. Study outcomes are summarised in Table 2.

Table 1. Summary of data sources

BB: 0-2	BB: 2-6	Provided by	Dataset		Time period*	Eligibility / Coverage		Mother	Child	Indicative / key data items
✓		Trial participants maternal self-report	Baseline		2009-2013	Trial participants		Yes	No	Socioeconomic; Maternal health & well-being; Health behaviour; pregnancy complications, Neonatal outcomes; Feeding & development.
✓			Late Pregnancy						Yes	
✓			6 month	Post-birth						
✓			12 month							
✓			18 month							
✓			24 month							
✓		Maternity records	Maternal outcomes		2009-2010	UK		Yes	Yes	Maternal health & well-being; Neonatal outcomes
✓		GP records	GP consultations		2009-2013	UK		Yes	Yes	Immunisations; safeguarding
✓		PCTs	Immunisation		2009-2013	England		No	Yes	Immunisations
✓	✓	DoH	Abortions		2009-2013	England and Wales All abortions performed in the NHS or an approved independent sector		Yes	No	Abortions
✓	✓	ONS	Mortality records		2009 - 2017	UK		Yes	Yes	Mortality data
✓	✓	NHS Digital / HES	Inpatient;		2009 - 2017	Any NHS hospital in England		Yes	Yes	Injuries and ingestions; subsequent pregnancies;
✓	✓		Outpatient;							
✓	✓		A&E							
	✓	Dept. for Education / NPD	CIN;		2009 - 2017	< 18 years Registered with social services in England		Yes	Yes	Child in need status and child looked after status
	✓		CLA							
	✓		EYFSP		2013-2017	Public Schools in England	4 yrs	No	Yes	Indicators of maltreatment; educational development and attainment; eligible for free school meals;
	✓		Census				2-19 yrs			
	✓		Alt Provision				2-19 yrs			
	✓		PRU				2-19 yrs			
	✓		Key stage One		2016-2017		5-7 yrs	No	Yes	

*Trial started 2009; 2 year follow up ended 2013; 6 year follow up ends in 2017. PCTs – Primary Care Trusts ONS- Office for National Statistics; HES-Hospital Episode statistics; NPD-national pupil database; CIN-child in need; CLA-child looked after; PRU-pupil referral unit; EYFSP-Early Years Foundation Stage Profile DoH – Department of Health

Table 2. Study Outcomes

Domains	Outcomes	HES	ONS	NPD
Primary: Child in Need status recorded at any time during the follow-up period.	CIN status as of 31 March each year			✓
Secondary: (i) Objective measures of maltreatment	Child Protection registration			✓
	Details of a child protection plan			✓
	CIN categorisation			✓
	CIN duration			✓
	Looked after status			✓
	CLA period of care			✓
	Legal status of CLA			✓
	Cause of death		✓	
(ii) Associated measures of maltreatment	DNA appointments	✓		
	Injuries and ingestions	✓		
(iii) Intermediate FNP programme outcomes	Subsequent pregnancies	✓		
(iv) Costs	Health and Social Care resource use	✓		✓
(v) Child health, developmental and educational outcomes	Special Educational Needs			✓
	Disability	✓		✓
	Day care attendance			✓
	Early Years assessment			✓
	School attendance			✓
	Key stage one attainment			✓

CIN – Child in Need; CLA – Child Looked After; DNA – Did not attend; FNP – Family Nurse Partnership; HES – Hospital Episode Statistics; NPD – National Pupil Database; ONS – Office for National Statistics

Data providers and datasets

The BB:0-2 Trial Data

Data collected for the initial trial will be used in the present study.[9-10] A baseline home assessment was conducted upon trial entry using Computer-Assisted Personal Interview (CAPI). Follow-up was by computer-assisted telephone interview (CATI) at 34-36 weeks gestation and 6, 12 and 18 months postnatal. A final home-based CAPI was conducted at 2 years after birth. Several routinely collected datasets were accessed and data obtained from the following sources: maternity records (medical and obstetric history items, antenatal

attendances and maternal and neonatal outcomes), primary care notes for each mother and child dyad (consultations, immunisations, pregnancies, safeguarding), abortions data from the Department of Health Abortions Statistics Team, and immunisation data via COVER (Cover of Vaccination Evaluated Rapidly) contacts.

NHS Digital

The Hospital Episode Statistics (HES) datasets hold records on over 125 million hospital admissions, outpatient and accident and emergency episodes each year. Data can be requested from NHS Digital (formerly known as the Health and Social Care Information Centre), the executive non-departmental public body established under the Health and Social Care Act 2012.[13] All available records belonging to cohort members (mothers and children) will be obtained from study entry of the mother, which occurred between June 2009 and July 2010 until the date the child turns six. The data requested include diagnoses, procedures, length of episode and external causes of injuries coded according to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems [ICD-10] codes.[14]

NHS Digital has responsibility for collecting these data from across the health and social care system to allow NHS hospitals to be paid for the care they deliver. At the end of the financial year (March) a final dataset is collated. This dataset is cleaned and validated before being available for research at the end of the year (December).

Office for National Statistics (ONS)

The Office for National Statistics (ONS) collects information on cause of death from civil registration records. Mortality data can be accessed through NHS Digital. For registered deaths, the underlying cause of death is derived from the sequence of conditions leading directly to the death and is recorded on the death certificate. Deaths are subsequently coded in line with the ICD-10.

Department for Education (DfE)

The Department for Education (DfE) holds information on pupils throughout the different phases of education. Records are sourced from publicly funded schools, local authorities and awarding bodies and held in the National Pupil Database (NPD). Datasets are available on various aspects of education such as school census data, absence data, and school attainment.[15] All available records for the children in the cohort will be obtained from the various datasets held. The data requested includes the number of hours attended, early educational development, eligibility for free school meals, and special educational needs (SEN) provision type. Datasets are collated throughout the year and are available at set time points annually.

Social Care Data

Social care data from local authorities is available through the NPD via two datasets, Child in Need (CIN) and Child Looked After (CLA). The CIN census captures individual level information on children referred to and assessed by children's social care services within each 12month period.[16] CLA is collected in the SSDA903 return – an annual statutory data collection for all local authorities.[17] Any child in the cohort who is in one of these datasets will be identified. Mothers who were <18 years at the time of participation in the BB:0-2 trial will also be identified in these datasets.

Study participants: Inclusion and exclusion criteria

Eligible participants are those mothers and children exiting the BB:0-2 trial. Women were recruited as nulliparous women aged 19 or under, living in one of 18 local authority FNP catchment areas; recruited by 24^{+6} weeks gestation, have conversational level of English and were able to consent to research.[10]

Children in medium-term foster placements or adopted within the six year study period can be linked up to the date of adoption. Maternal or child death will be captured as an outcome.

Recruitment / Dissent

Participants previously consented to enter into the BB:0-2 trial and provide self-report and access to their routine records for the period up to two years postpartum. In order to obtain an unbiased estimate of the medium-term effect of FNP on objective and associated maltreatment outcomes we have received section 251 (s251) support of the 2006 NHS Act approval from the Health Research Authority's Confidentiality Advisory Group (HRA CAG) to pass identifiable participant data legally held by Cardiff University to the information centres (IC) to link to routine data. This is without obtaining further consent from participants, instead using an opt-out/dissent model.

Justification of approach

Consent for longer-term follow-up (i.e. beyond 24 months postpartum) was originally proposed in the BB:0-2 trial. However, upon ethical review it was considered that greater specificity about exact outcomes than could be provided at recruitment was required. Additionally, providing meaningful consent for much longer follow-up was also challenging, particularly on behalf of yet to be born children.

Developing the opt-out approach was necessary due to (i) the child protection focus of the study and the consequent sensitivity and impracticality in asking directly for consent, (ii) the mobility and relative difficulty in ongoing direct access to these participants (iii) the consequent introduction of non-ascertainment bias on sample representativeness – resulting in a non-random sample, and (iv) the likely cost and logistical requirements of securing even modest levels of additional consent.

Methods of notifying participants

We discussed the issue of dissent and fair processing with the HRA CAG and have subsequently attempted to contact all mothers recruited to the original BB:0-2 trial to inform them that medium-term follow-up using anonymised records will be undertaken.

Details of participants' residential addresses were updated using their most recent address registered with their GP. Where available, mobile number and email addresses collected for the trial were used to send SMS and emails to participants. All three modes of contact were used over a two-day period and participants were provided with a two-month window in which to contact the project team to discuss the project and opt-out if they wished. A website was also available with the same information which directed participants to contact the project team if they wished.

Development of opt-out letter

A group of care-experienced young people (CASCADE Voices)[18] advised on the layout, wording and tone of a letter to be sent to all participants. A key consideration was to communicate the focus of this follow-on study in a sensitive manner. The final letter was approved by both an NHS Research Ethics Committee and CAG committee as part of overall governance approval for the study. The letter contained both information on the trial, the follow-on study and a flowchart for what to do if women wished to discuss the project and/or opt out.

Process to manage dissent

Women notifying the study team of their dissent will be recorded as "opted out", removed from all project datasets for this follow-up work and identifiable datasets to be sent to ICs. They will not be included in any of the datasets or analyses for this follow-on study.

Governance and compliance

Following Ethical approval (14/WA10062) and s251 support (CAG 10-08(b)/2014), data request applications were submitted to DfE, NHS Digital and ONS.

In order to satisfy the requirements of the s251 support and NHS Digital contract, the Information Governance (IG) Toolkit self-assessment[19] (commissioned by the Department for Health for NHS Digital to develop and maintain) was required. This organisation-level assessment provides reassurance of satisfactory information governance within the host

trials unit. Both the s251 support and IG Toolkit are assessed and renewed on an annual basis. The opt-out model was also required to satisfy s251 support as well as the DfE assessment of compliance with principle one of the Data Protection Act 1998. Governance and IC requirements prior to application approval are shown in Figure 2.

Data matching

Maternal and child identifiers will be sent to both NHS Digital and DfE for matching with their databases. Each IC holds differing identifiers including a unique identifier for each individual (NHS Number; Unique Pupil Number UPN).

Matching with HES data will be by exact matching on NHS Number; Date of Birth; Postcode and Gender. This was conducted for BB:0-2 and achieved a high match rate where 99.6% of mothers and babies' records were matched fully (i.e. matched on all identifiers provided) or partially (i.e. matched on a reduced, but acceptable number of identifiers provided). This will be repeated for this study. NHS Digital will then exact match with ONS using NHS number in order to obtain mortality data.

As the NPD does not include NHS numbers, initially exact matching on first name and surname, date of birth and postcode (of both mother and child for social care data; all other datasets just child) will be undertaken. Further matching required will be by fuzzy matching of first name. The CIN and CLA datasets do not contain names or postcodes therefore the matching will be in two phases: i) Participants will be matched with the NPD, the UPN added to all participants and ii) this will be used to identify individuals in the CIN and CLA datasets.

Data matching at DfE and NHS Digital/ONS are independent therefore match rates at the participant level are expected to vary (some may match to NPD but not HES). Educational records should be available for all children in the trial cohort whereas health and social care derived data will only exist where the child has received a relevant episode of care. Participants will be compared using trial baseline data to check for any bias between those

who are matched and not-matched for those datasets where they would all be expected to be present (e.g. school census for all children).

The Pseudonymised Dataset

A unique study ID will be attached to each participant's record prior to data transfer to ICs. Once ICs have matched records to their database, only the unique study ID is retained. Data extracts from both ICs plus data files from the trial (following a process of de-identification and standardisation in Cardiff to reduce risk of later unintentional participant level identification) will all be securely transferred to a data safe haven,[20] the Secure Anonymised Information Linkage (SAIL) databank, for linking and storage. The data flow is shown in Figure 3.

A SAIL data analyst will re-assign the study ID with a new anonymous linking field [ALF] and store the corresponding ID in a separate encrypted password protected file.[21]

Participants will not be identifiable to the study team, or to the SAIL analyst, but incoming datasets can be linked at the individual level using the ALF. The study team will have controlled remote access to these data thus ensuring the security of the pseudonymised database.[22] All data cleaning and analysis will be carried out via the remote portal by the study data manager and statistician.

Data from NHS Digital and NPD will be requested at two time points. The first data extract will confirm the data flow model, assess data quality and the suitability of data for answering key study analyses. The second data request will be made once all children in the study have reached key stage one (April 2017) and on which the study findings will be reported on (in 2018).

Control of data

Cardiff University control under contract the identifiable trial data that are being transferred to the ICs and to the safe haven. Data held by NHS Digital, ONS and DfE, for which they are the controllers, are de-identified and then sent to SAIL to be linked and held (including the de-identified trial data) in a secure anonymised standalone database for use by nominated study team members. SAIL will control the safe haven environment, and will process the pseudonymised data for secure use by study team. Cardiff University will control the purposes to which the data are put in answering research questions as per the study protocol. Once linked in the data safe haven, the ability to submit queries to each IC about individual records will be more limited than if identifiable data were returned to the research team in Cardiff. Data cleaning will remain possible however as will generic queries about data provided in batch. The quality of matching conducted by NPD and NHS Digital/ONS will be a key factor in the success of the study.

Analysis

Power Calculation

Primary outcome (CIN status at any point between birth and six years): For CIN status, available UK data on rates are not specific to the age-range of interest, but the rate in the general population aged 5-9 years is 4.6% (for local authorities comprising study sites in BB:0-2). The rate would be expected to be greater in the specific study sample, and therefore we have assumed a rate of 8%. To detect a difference of 4% (FNP: 4% vs Usual Care: 8%) would require 602 children in each arm (1204 in total) using 80% power and a two-sided 5% alpha level.

BB:0-2 recruited 1645 women, with 1562 available for follow-up (i.e. excluding those subject to a mandatory withdrawal). Follow-up through medical records (assuming 10% loss in

tracking & linkage) would result in 1405 participants, thus securing enough data to test the primary outcome and the key secondary outcome

Main analysis

Analyses will be conducted on an intention-to-treat basis and due emphasis placed on confidence intervals for the between-arm comparisons. Descriptive statistics of demographic and baseline measures will be used to ascertain any marked imbalance between the trial arms. The primary comparative analysis on CIN status at any point between birth and six years will use logistic multilevel modelling to investigate differences between the groups, and odds ratios alongside 95% confidence intervals (CIs) will be reported. Multilevel modelling will allow for clustering of effect within a site and family nurse. Modelling the impact of key subgroups and different intervention elements (e.g. gestational age at programme entry, dosage) on outcome will be undertaken by extending the primary models and testing for interaction effects. The role of potential moderators of programme effect (e.g. domestic violence) will also be explored.

Although the study will be powered to examine a 4% difference in CIN status, secondary analyses will assess group differences in referral rates to CSC, maltreatment profile, and child protection outcomes. Levels of concern will be examined by looking at extent of action taken. A state transition model using Markov chains will be used to assess the probabilities of moving from one stage marker (states) to another.[23] The transition probabilities (the probability of the various state-changes) in our model will be derived from our data and compared between groups.

Bias in the followed-up BB:2-6 sample will be quantified by examining group differences (participants and non-participants) in baseline variables such as age, deprivation, gestational age, and education. Surveillance bias in detection of maltreatment during the child's infancy and toddlerhood can be assessed by examining subsequent reporting.[24] The duration between birth and the date of first referral to CSC will be calculated and group differences

examined using Cox regression analysis to calculate hazard ratios for referral, together with 95% CIs. Surveillance bias is most likely to occur during the intervention phase, although improved handover to other services at 2 years may lead to higher identification in the following year. Severity of the referral will also be compared between the two groups (an approach used in US trials of NFP to explore surveillance bias).

Health economics

The economic evaluation will consider costs and consequences of the FNP over the full follow-up period (BB:0-2 & BB:2-6). The current BB:0-2 study reported 1) a within trial cost utility analysis assessing NHS costs against quality adjusted life years (QALY) from the perspective of the mother, and 2) a within trial cost consequences analysis relating all costs (including those to the social care, education and criminal justice sectors as well as health) against the full range of effects.[10] Cost and consequences framework is deemed the most appropriate economic evaluation framework for public health interventions[26] and preferred by NICE[27] because it enables to capture equity consideration as well as intersectoral costs and consequences[28] yet applications are still limited.[26]

The absence of additional data on Health Related Quality of Life within the BB2 study means that it will not be possible to estimate QALYs beyond 24 months postpartum and hence extend the within trial cost utility analysis. However, the within trial cost consequences analysis will be extended from 0-2 to 0-6 years through collection of resource use data from medical and education records (including from the latter, data related to social care usage). These will be summarised against the range of outcomes collected within BB2 without aggregation to allow weighing up changes in the various outcomes reported in BB2 against the changes in costs in a consistent and transparent manner.[29] This will contribute to providing more robust and valid medium-term estimates within the extended period.

ETHICS AND DISSEMINATION

Legal & Ethical considerations

The potential for using routine data in health and social care research has been greatly publicised and study designs utilising these data are encouraged by funders.[30] There are, however, many inherent challenges in working with secondary-use data, in particular for this project the ethical and legal requirements/responsibilities which have fundamentally informed this study design.

Although BB:0-2 linked trial data to HES and ONS data via NHS Digital, the governance requirements around the two applications have differed between the two studies not least because of the difference in consent models. Trial data were provided by NHS Digital and ONS after participant consent to prospective collection and for specified purposes limited to the time-frame of that study. The current follow-on study uses a dissent model under which we are only able to send trial participant identifiers to ICs for matching to outcome data records if there is no objection received from mothers. The study will require all clinical, social and educational data to be held in a data safe haven using encrypted record identifiers and analysis via a securely managed and monitored remote portal. The legal bases for transfer of identifiable data to ICs without explicit consent are as follows; s251 of the 2006 NHS Act 2006 for HES data from NHS Digital, s42(4) of the Statistics and Registration Service Act 2007 through NIHR funding for ONS data via NHS Digital, and 6(1) of Schedule 2 of the 1998 Data Protection Act for NPD data.

Dissemination of findings The Building Blocks: 2-6 Study will generate policy-relevant findings describing the medium-term impact of FNP on measurements of child maltreatment. The findings will also include other policy relevant outcomes from the programme such as health care use, education attainment and changes in social care use over the 6 years of follow up. Such medium-term evaluation remains important as some outcomes for the intervention are expected to arise only after the child's second birthday, including maltreatment. This study will either confirm the largely negative trial findings from BB:0-2 further weakening the justification for FNP Programme continuation or provide a balance to the early measurable outcomes.

In addition to reporting the findings to the funder for this study, the funder for the BB:0-2 trial (DH Policy Research Programme) will also be informed and the FNP National Unit (FNPNU). All local authorities in England will be notified of the results, as (since October 2015) they have responsibility for commissioning public health services for children aged 0-5. Participants will receive a summary of the results and all reports and publications will be made publicly available in full on the Cardiff University website. The research team have previously convened and met twice with a stakeholder group, including relevant policy leads from each country in the UK delivering FNP (England, Scotland, Northern Ireland). We will stage a similar event to present and discuss the implications for practice and policy of the results of this medium-term follow up of participants.

In addition to policy and public outputs, academic outputs will include (i) this protocol paper providing visibility of this medium-term follow up, (ii) a methods paper describing the piloting process of the study (including data quality and success of data matching) and (iii) main study findings. We aim to disseminate in high-quality, peer reviewed journals and present in key conferences.

A particular benefit of this study is understanding of, and learning from, the governance challenges. There is potential to use this method for future trials looking at longer term follow-up. Therefore this study has the potential to add to the understanding of routine data and data linkage methods in future public health and clinical trials and these planned publications will provide a basis for the dissemination of the success of these methods.

Declarations

List of abbreviations

ALF: anonymous linking field;
ALF-E: encrypted ALF;
BB:0-2: Building Blocks trial;
CAPI: computer-assisted personal interview;
CATI: computer-assisted telephone interview;
CIN: Child in Need;

CLA: Child Looked After;
DfE: Department for Education;
DNA: Did not attend;
DoH: Department of Health;
EYFSP: Early Years Foundation Stage Profile;
FNP: Family Nurse Partnership;
FNPNU: Family Nurse Partnership National Unit;
HES: Hospital Episode Statistics;
HRA CAG: Health Research Authority's Confidentiality Advisory Group;
HSCIC: Health and Social Care Information Centre;
IC: information centres;
IG: Information Governance ;
NFP: Nurse Family Partnership;
NPD: National Pupil Database;
ONS: Office for National Statistics;
PCT: Primary Care Trusts;
PRU: pupil referral unit;
QALY: Quality adjusted life years;
s251: Section 251 of the NHS 2006 Act;
SAIL: Secure Anonymised Information Linkage;
SEN: special educational needs;
TTP: trusted third party;
UPN: Unique Pupil Number.

Ethics approval and consent to participate - Ethics approval of the study has been given by the Research Ethics Committee for Wales (14/WA10062) and the transfer and use of identifiable data has been approved by the Health Research Authority [HRA] Confidentiality Advisory Group [CAG] (CAG 10-08(b)/2014).

Consent for publication - Not Applicable

Availability of data and material - Not Applicable

Competing interests - The authors declare that they have no competing interests

Funding - This project was funded by the National Institute for Health Research Public Health Research (NIHR PHR) Programme (project number 11/3002/11). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the NIHR PHR Programme or the Department of Health.

Authors' contributions

MR is the chief investigator of the study. All authors have contributed to and are responsible for the final design of the study. FLW and GM are responsible for study and data management. RCJ is responsible for statistical planning and for data analysis. DF is responsible for the health economics. All authors read and approved the final manuscript.

Acknowledgements

The South East Wales Trials Unit is funded by Health and Care Research Wales. The work was undertaken with the support of The Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement (DECIPHer), a UKCRC Public Health Research Centre of Excellence. Joint funding (MR/KO232331/1) from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, the Welsh Government and the Wellcome Trust, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

References

1 A guide to inter-agency working to safeguard and promote the welfare of children. In: Working together to safeguard children. Department for Children Schools and Families. 2010. http://webarchive.nationalarchives.gov.uk/20130401151715/https://www.education.gov.uk/publications/eorderingdownload/00305-2010dom-en-v3.pdf Accessed Dec 2014.

2 Characteristics of children in need: 2014 to 2015 (Statistics: children in need and child protection) year ending 31 March 2015. In: Department for Education National Statistics. 2015. Department for Education. https://www.gov.uk/government/statistics/characteristics-of-children-in-need-2014-to-2015 Accessed Feb 2016.

3 MacMillan HL, MacMillan JH, Offord DR, Griffith L, et al. Primary prevention of Child Physical Abuse and Neglect: A Critical Review. *J child psychol psyc* 1994;35(5):835-56.

4 Olds DL, Hill PL, O'Brien R, et al. Taking preventive intervention to scale: The nurse-family partnership. *Cogn behav pract* 2003;10(4):278-90.

5 Olds DL, Henderson CR Jr, Chamberlin R, et al. Preventing child abuse and neglect: a randomized trial of nurse home visitation. *Pediatrics* 1986;78: 65–78.

6 Kitzman H, Olds DL, Henderson CR, 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.

7 Olds DL, Robinson J, O'Brien R, et al. Home visiting by paraprofessionals and by nurses: a randomized, controlled trial. *Pediatrics* 2002;110(3):486-96.

8 Olds D, Henderson CRJ, Kitzman H, et al. Effects of Prenatal and Infancy Nurse Home Visitation on Surveillance of Child Maltreatment. *Pediatrics* 1995;95(3).

9 Owen-Jones E, Bekkers MJ, Butler CC, 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.

10 Robling M, Bekkers MJ, Bell K, et al. Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial. *The Lancet* 2016;387(10014):146-55.

11 Olds DL, Eckenrode J, Henderson CR, 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.

12 Zielinski DS, Eckenrode J, Olds DL. Nurse home visitation and the prevention of child maltreatment: Impact on the timing of official reports. *Dev Psychopathol* 2009;21(02):441-53.

13 Data Access Request Service http://www.hscic.gov.uk/dars. Accessed March 2016.

14 *International Statistical Classification of Diseases and Related Health Problems, 10th revision,* Volume 1, Geneva. World Health Organisation 1992.

15 National pupil database: apply for a data extract https://www.gov.uk/guidance/national-pupil-database-apply-for-a-data-extract#data-extracts-available. Accessed April 2016.

16 Children in need census 2015 to 2016: guide https://www.gov.uk/government/publications/children-in-need-census-2015-to-2016-guide Accessed April 2016.

17 Children looked after return 2015 to 2016: guide. https://www.gov.uk/government/publications/children-looked-after-return-2015-to-2016-guide. Accessed April 2016.

18 Shaw IG, Holland S. Doing qualitative research in social work. Sage; 2014.

19 Information Governance Toolkit https://www.igt.hscic.gov.uk/. Accessed March 2016.

20 Ford DV, Jones KH, Verplancke JP, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;9(1):157.

21 Lyons RA, Jones KH, John G, et al. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis* 2009;9(1):1.

22 Jones KH, Ford DV, Jones C, et al. A case study of the secure anonymous information linkage (SAIL) gateway: A privacy-protecting remote access system for health-related research and evaluation. *J Biomed Inform* 2014;50:196-204.

23 Meyn SP, Tweedie RL. Markov chains and stochastic stability. Springer Science & Business Media; 2012.

24 Chaffin M, Bard D. Impact of intervention surveillance bias on analyses of child welfare report outcomes. *Child maltreatment* 2006;11(4):301-12.

25 MacMillan HL, Wathen CN, Barlow J, et al. Interventions to prevent child maltreatment and associated impairment. *The Lancet* 2009;373:250-66.

513 26 Drummond M, Weatherly H, Claxton K, et al. Assessing the challenges of applying
514 standard methods of economic evaluation to public health interventions. Public Health
515 Research Consortium; 2006. http://phrc.lshtm.ac.uk/papers/PHRC_D1-05_Final_Report.pdf

516 27 National Institute for Health and Care Excellence (2012) Methods for the development of
517 NICE public health guidance (Third edition)
518 https://www.nice.org.uk/article/pmg4/chapter/1%20introduction. Accessed October 2016.

519 28 Kelly MP, McDaid D, Ludbrook A and Powell J (2003) Economic appraisal of public
520 health interventions.
521 http://www.cawt.com/Site/11/Documents/Publications/Population%20Health/Economics%20
522 of%20Health%20Improvement/Economic_appraisal_of_public_health_interventions.pdf.
523 Accessed October 2016.

524 29 National Institute for Health and Care Excellence. Developing NICE guidelines: the
525 manual [PMG20]. October 2014 Updated April 2016 Accessed at
526 https://www.nice.org.uk/process/pmg20/chapter/incorporating-economic-
527 evaluation#approaches-to-bespoke-economic-evaluation

528 30 Methods research to support use of observational data in clinical decision making.
529 http://www.mrc.ac.uk/funding/how-we-fund-research/highlight-notices/methods-research-
530 observational-data/. Accessed March 2016.

531

532

(Title) Figure 1. Follow up and datasets over the six years

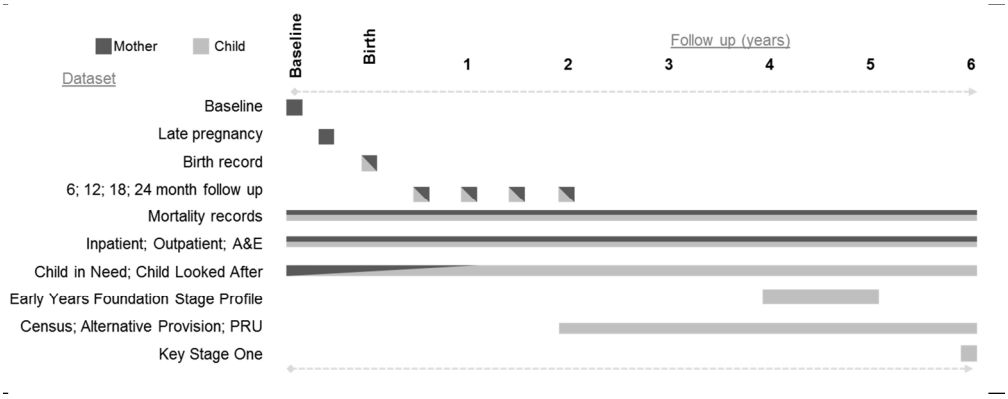
(Legend) A&E Accident and Emergency; PRU Pupil Referral Unit

(Title) Figure 2. Governance and Information centre requirements prior to application approval.

(Legend) s251 Section 251 of the NHS 2006 Act; ONS Office for National Statistics; NPD National Pupil Database; IG Information Governance

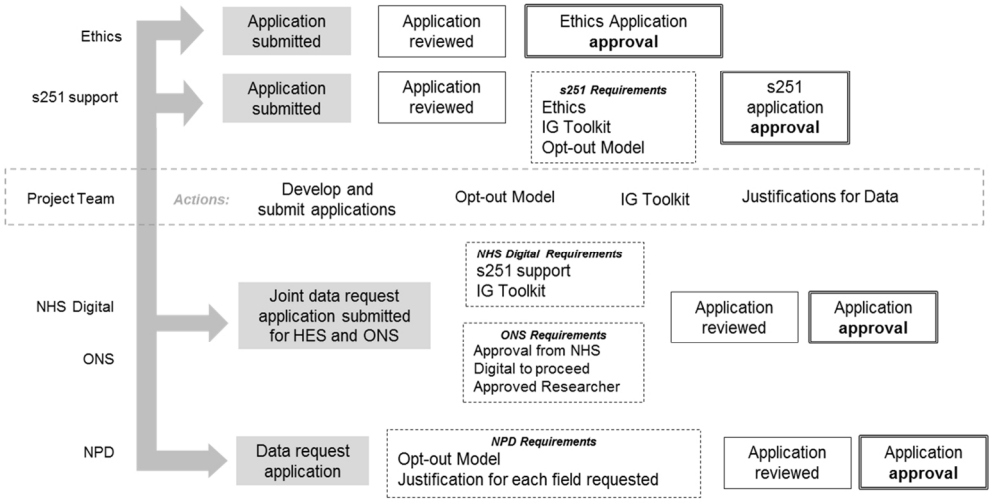
Figure 3. Data Flow

(Legend) [1]Participant identifiable information securely transferred for linkage; [2]De-identification and Standardisation applied (e.g. date of birth to week of birth); [3]Information centres confirm matching of participant identifiers; [4]Hosted on SAIL secure platform. ALF-Anonymised Linking Field; BB:0-2 – The Building Blocks trial; DfE – Department for Education; DPO – Data Providing Organisation; HES – Hospital Episode Statistics; ONS – Office for National Statistics; SAIL – Secure Anonymised Information Linkage.



(Title) Figure 1. Follow up and datasets over the six years
(Legend) A&E Accident and Emergency; PRU Pupil Referral Unit

252x98mm (150 x 150 DPI)



(Title) Figure 2. Governance and Information centre requirements prior to application approval.
(Legend) s251 Section 251 of the NHS 2006 Act; ONS Office for National Statistics; NPD National Pupil Database; IG Information Governance

236x127mm (150 x 150 DPI)

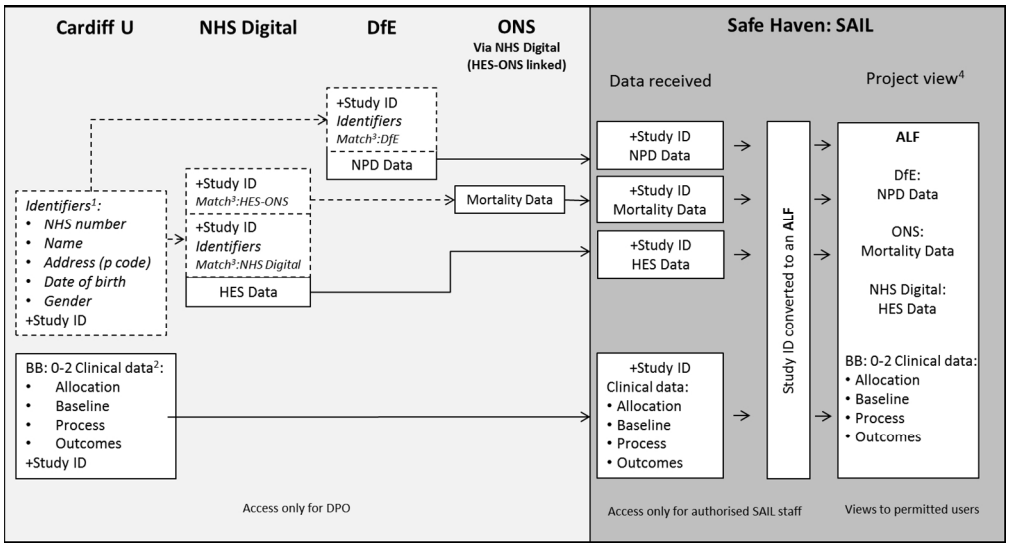


Figure 3. Data Flow

(Legend) 1Participant identifiable information securely transferred for linkage; 2De-identification and Standardisation applied (e.g. date of birth to week of birth); 3Information centres confirm matching of participant identifiers; 4Hosted on SAIL secure platform. ALF- Anonymised Linking Field; BB:0-2 – The Building Blocks trial; DfE – Department for Education; DPO – Data Providing Organisation; HES – Hospital Episode Statistics; ONS – Office for National Statistics; SAIL – Secure Anonymised Information Linkage.

266x143mm (150 x 150 DPI)

BMJ Open

Assessing the medium-term impact of a home-visiting programme on child maltreatment in England: protocol for a routine data linkage study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-015728.R1
Article Type:	Protocol
Date Submitted by the Author:	30-Mar-2017
Complete List of Authors:	Lugg-Widger, Fiona; Cardiff University, Centre for Trials Research Cannings John, Rebeca; Cardiff University, Centre for Trials Research Channon, Sue; Cardiff University, Centre for Trials Research Fitzsimmons, Deborah; Swansea University, Swansea Centre for Health Economics Hood, Kerenza; Cardiff University, Centre for Trials Research Jones, Kerina; Swansea University, College of Medicine Kemp, Alison; Cardiff University School of Medicine, Division of Population Medicine Kenkre, Joyce; University of South Wales, 5. Faculty of Life Sciences and Education Longo, Mirella; Cardiff University School of Medicine, Marie Curie Palliative Care Research Centre McEwan, Kirsten; University of Derby, Department of Psychology Moody, Gwenllian; Cardiff University, Centre for Trials Research Owen-Jones, Eleri; Cardiff University, Centre for Trials Research Sanders, Julia; Cardiff University, School of Healthcare Sciences Segrott, Jeremy; Cardiff University, Centre for Trials Research; Cardiff University, DECIPHer Centre, Cardiff School of Social Sciences Robling, Michael; Cardiff University, Centre for Trials Research
Primary Subject Heading:	Public health
Secondary Subject Heading:	Health informatics, Paediatrics, Research methods
Keywords:	Child maltreatment, Family Nurse Partnership, Early years prevention, Medical Record Linkage, Follow-Up Studies

Assessing the medium-term impact of a home-visiting programme on child maltreatment in England: protocol for a routine data linkage study

Corresponding author

Fiona Lugg-Widger[1] LuggFV@cardiff.ac.uk Telephone: 02920687519

Address: 4th Floor Neuadd Meirionnydd, Heath Park, Cardiff. CF14 4YS.

Rebecca Cannings-John[1] CanningsRL@cardiff.ac.uk
Sue Channon[1] ChannonS2@cardiff.ac.uk
Deborah Fitzsimmons[2] D.Fitzsimmons@swansea.ac.uk
Kerenza Hood[1] HoodK1@cardiff.ac.uk
Kerina Jones[3] k.h.jones@swansea.ac.uk
Alison Kemp[4] KempAM@cf.ac.uk
Joyce Kenkre[5] joyce.kenkre@southwales.ac.uk
Mirella Longo[6] LongoM1@cardiff.ac.uk
Kirsten McEwan[7] k.mcewan@derby.ac.uk
Gwenllian Moody[1] MoodyG@cardiff.ac.uk
Eleri Owen-Jones[1] Owen-JonesCE@cardiff.ac.uk
Julia Sanders[8] SandersJ3@cardiff.ac.uk
Jeremy Segrott[1,9] SegrottJ@cardiff.ac.uk
Michael Robling[1] RoblingMR@cardiff.ac.uk

Author Affiliations

1. South East Wales Trials Unit, Centre for Trials Research, Cardiff University, Cardiff, UK.
2. Swansea Centre for Health Economics, Swansea University, Swansea, UK
3. College of Medicine, Swansea University, Swansea, UK
4. Division of Population Medicine, School of Medicine, Cardiff University, Cardiff, UK.
5. Faculty of Life Sciences and Education, University of South Wales, Pontypridd, UK
6. Marie Curie Palliative Care Research Centre, Cardiff University, Cardiff, UK
7. Department of Psychology, University of Derby, UK
8. School of Healthcare Sciences, Cardiff University, Cardiff, UK.
9. DECIPHer Centre, Cardiff School of Social Sciences, Cardiff University, Cardiff, UK.

Word Count: 4853

Keywords: Child maltreatment, Family Nurse Partnership, Early years prevention, Medical Record Linkage, Follow-Up Studies

ABSTRACT

Introduction: Child maltreatment involves acts of omission (neglect) or commission (abuse) often by caregivers that results in potential or actual harm to a child. The Building Blocks trial [ISRCTN 23019866] assessed the short-term impact of an intensive programme of antenatal and postnatal visiting by specially trained nurses to support young pregnant women in England. The Building Blocks: 2-6 study will assess the medium-term impacts of the programme for mothers and children (n=1562), through the linkage of routinely collected data to the trial data, with a particular emphasis on the programme's impact upon preventing child maltreatment.

Methods and analysis: We have developed bespoke model of data linkage whereby outcome data for the trial cohort will be retrieved by linked anonymous data abstraction from NHS Digital, Office for National Statistics and the Department for Education's National Pupil Database. Participants will be given reasonable opportunity to opt-out of this study prior to data transfer. The information centres will match participants to the information held in their databases using standard identifiers, and send extracts to a third party safe haven. The study will have 80% power to detect a 4% difference (4% vs 8%) for the binary primary outcome of Child in Need status (from birth to key stage one) at a two-sided 5% alpha level by following up 602 children in each trial arm. Analysis will be by intention to treat using logistic multilevel modelling. A cost and consequences analysis will extend the time-frame of the economic analysis from the original trial.

Ethics and Dissemination: The study protocol has been approved by NHS Wales Research Ethics Committee and the Health Research Authority's Confidentiality Advisory Group. Methods of innovative study design and findings will be disseminated through peer-review journals and conferences, results will be of interest to clinical and policy stakeholders in the UK.

Strengths and limitations of this study

- This study aims to provide much needed evidence about the medium-term benefits of the Family Nurse Partnership programme in England. This study has the capacity to either confirm the current perspective on the value of the intervention or demonstrate clinically meaningful benefits to children in vulnerable young families.
- There are distinct benefits associated with using routine data including a reduction in cost and participant burden over prospectively collected data, and relative completeness and therefore minimisation of bias over self-report, particularly for such sensitive outcomes.
- The establishment of a regulatory secure research database for this cohort of trial participants also offers the prospect of further data being added over the longer term and of broadening the scope of the dataset to other outcome domains relevant to this intervention, such as criminal justice and welfare benefits.
- The extent of this benefit will be balanced by our ability to adequately access the data from information centres in a timely fashion, the quality of matching conducted as well as the quality of the data ultimately retrieved.

INTRODUCTION

Maltreatment

Child maltreatment involves acts of omission (neglect) or commission (abuse) often by caregivers that inflict harm, or fail to act to prevent harm to a child.[1] Abuse may be physical, emotional or sexual. Neglect represents persistent failure to meet basic physical or psychological needs, often resulting in serious impairment of the child's health and/or development.[1] Neglect may involve failing to: protect a child from physical and emotional harm or danger; provide adequate supervision; or ensure access to appropriate medical care. In the year ending 31st March 2015 in England there were 635,600 referrals to children's social care services, 403,400 children starting an episode of need (an overall rate

of 348.0 per 10,000) and 62,200 children became subject of a child protection plan.[2] Of children who became subject of a child protection plan, the most common initial category of abuse was neglect (43.2%) followed by emotional abuse (33.7%).

In the UK, preventing maltreatment is an important focus of Government concern. The Children Act 1989 specifies agencies' responsibilities to cooperate in the interests of vulnerable children, for Children in Need (section 17) and children suffering or likely to suffer from significant harm (section 47). A child in need is defined as a child who is unlikely to achieve or maintain a reasonable level of health or development, or whose health and development is likely to be significantly or further impaired, without the provision of services; or is a child who is disabled. Local authority provisions may include supervision of activities, financial help, provision of family accommodation, respite or home help in addition to advice and guidance from social workers.

Family Nurse Partnership home-visiting programme

There has been increasing emphasis upon the primary prevention of child maltreatment, including interventions directed at general populations and those targeting high-risk groups.[3] One such intervention is the Family Nurse Partnership (FNP) programme (developed in the US as the Nurse Family Partnership or NFP) – a home-visiting approach with three overarching goals: to improve birth outcomes, optimise child health and development - including reducing maltreatment - and promote economic self-sufficiency of mothers.[4]

In three US trials (in Elmira, Memphis and Denver),[5-7] the NFP has demonstrated improvements in prenatal health behaviours and birth outcomes, sensitive child care, maternal life course (e.g. greater workforce participation) and child and adolescent functioning. It has also shown positive effects in relation to reductions in rates of child injuries, abuse and neglect. In the first US trial in 1977, a sub-group analysis of poor unmarried teens (54 families) revealed that by age 2 there was verified abuse / neglect in 19% of control children compared to 4% in the group in receipt of NFP in both pregnancy

and infancy (treatment difference of 0.15, 95% confidence interval of -0.01 to 0.31) and 56% relative reduction in emergency department encounters for injuries and ingestions during the second year of life.[5] Amongst the sub-group of children (56 families) with a state-verified report of maltreatment by age 4, the NFP group of children exhibited fewer risks for harm than the control group (e.g. fewer attendances with injuries / ingestions, safer home environment) at follow-up points between 25 and 50 months of life.[8] This was considered to be due to the earlier and more comprehensive detection of maltreatment by nurse-visited families.

The NFP programme was adapted for implementation as the Family Nurse Partnership and was introduced in England in 2007. Our Building Blocks trial (ISRCTN23019866) was the first trial of FNP in England and evaluated short-term outcomes to age 2 – the duration of the FNP programme.[9] The trial reported no difference for four primary outcomes: biomarker-calibrated self-reported tobacco use by the mother at late pregnancy, birth weight of the baby, the proportion of women with a second pregnancy within 24 months post-partum, and emergency attendances and hospital admissions for the child within 24 months post-partum.[10] We observed some differences for secondary child development outcomes including the rate of safeguarding events reported in primary care records. While the current evidence does not support continuation of the programme in England, previous evaluations have demonstrated benefit over the longer-term (e.g. up to 15 years of age).[11] For maltreatment outcomes this benefit has been increasingly evident after age 4 years,[12] therefore, the current study will establish whether FNP has moderated maltreatment outcomes over a medium-term period of follow-up (i.e. to the point where the child is aged six years old).

METHODS AND ANALYSIS

Research objective

The Building Blocks: 2-6 Study will use data linkage of routinely collected national datasets to assess the medium-term impact of the FNP intervention upon child maltreatment outcomes and key indicators of neglect.

Study design

This is a data linkage study, which will generate a linked anonymised database hosted by an independent Trusted Third Party (TTP). Participant mothers and children from Building Blocks: 0-2 (BB:0-2) will be followed up for a further four years using routine data only. Data from various routine public sector sources will be retrieved and linked to the trial data to enable children and mothers to be followed until the child reaches key stage one (the two years of schooling when pupils are aged between 5 and 7). The study formally started in February 2014 and will report to the funder in May 2018. Participants were recruited to the trial between June 2009, and July 2010 and the six year follow up ends (i.e. the last child will have turned six) in March 2017. A summary of the data sources is provided in Table 1 and the time period for each dataset are shown in Figure 1. Study outcomes are summarised in Table 2.

Table 1. Summary of data sources

BB: 0-2	BB: 2-6	Provided by	Dataset		Time period*	Eligibility / Coverage		Mother	Child	Indicative / key data items
✓		Trial participants maternal self-report	Baseline		2009-2013	Trial participants		Yes	No	Socioeconomic; Maternal health & well-being; Health behaviour; pregnancy complications, Neonatal outcomes; Feeding & development.
✓			Late Pregnancy						Yes	
✓			6 month	Post-birth						
✓			12 month							
✓			18 month							
✓			24 month							
✓		Maternity records	Maternal outcomes		2009-2010	UK		Yes	Yes	Maternal health & well-being; Neonatal outcomes
✓		GP records	GP consultations		2009-2013	UK		Yes	Yes	Immunisations; safeguarding
✓		PCTs	Immunisation		2009-2013	England		No	Yes	Immunisations
✓	✓	DoH	Abortions		2009-2013	England and Wales All abortions performed in the NHS or an approved independent sector		Yes	No	Abortions
✓	✓	ONS	Mortality records		2009 - 2017	UK		Yes	Yes	Mortality data
✓	✓	NHS Digital / HES	Inpatient;		2009 - 2017	Any NHS hospital in England		Yes	Yes	Injuries and ingestions; subsequent pregnancies;
✓	✓		Outpatient;							
✓	✓		A&E							
	✓	Dept. for Education / NPD	CIN;		2009 - 2017	< 18 years Registered with social services in England		Yes	Yes	Child in need status and child looked after status
	✓		CLA							
	✓		EYFSP		2013-2017	Public Schools in England	4 yrs	No	Yes	Indicators of maltreatment; educational development and attainment; eligible for free school meals;
	✓		Census				2-19 yrs			
	✓		Alt Provision				2-19 yrs			
	✓		PRU				2-19 yrs			
	✓		Key stage One		2016-2017		5-7 yrs	No	Yes	

*Trial started 2009; 2 year follow up ended 2013; 6 year follow up ends in 2017. PCTs – Primary Care Trusts ONS- Office for National Statistics; HES-Hospital Episode statistics; NPD-national pupil database; CIN-child in need; CLA-child looked after; PRU-pupil referral unit; EYFSP-Early Years Foundation Stage Profile DoH – Department of Health

Table 2. Study Outcomes

Domains	Outcomes	HES	ONS	NPD
Primary: Child in Need status recorded at any time during the follow-up period.	CIN status as of 31 March each year			✓
Secondary: (i) Objective measures of maltreatment	Child Protection registration			✓
	Details of a child protection plan			✓
	CIN categorisation			✓
	CIN duration			✓
	Looked after status			✓
	CLA period of care			✓
	Legal status of CLA			✓
	Cause of death		✓	
(ii) Associated measures of maltreatment	DNA appointments	✓		
	Injuries and ingestions	✓		
(iii) Intermediate FNP programme outcomes	Subsequent pregnancies	✓		
(iv) Costs	Health and Social Care resource use	✓		✓
(v) Child health, developmental and educational outcomes	Special Educational Needs			✓
	Disability	✓		✓
	Day care attendance			✓
	Early Years assessment			✓
	School attendance			✓
	Key stage one attainment			✓

CIN – Child in Need; CLA – Child Looked After; DNA – Did not attend; FNP – Family Nurse Partnership; HES – Hospital Episode Statistics; NPD – National Pupil Database; ONS – Office for National Statistics

Data providers and datasets

The BB:0-2 Trial Data

Data collected for the initial trial will be used in the present study.[9-10] A baseline home assessment was conducted upon trial entry using Computer-Assisted Personal Interview (CAPI). Follow-up was by computer-assisted telephone interview (CATI) at 34-36 weeks gestation and 6, 12 and 18 months postnatal. A final home-based CAPI was conducted at 2 years after birth. Several routinely collected datasets were accessed and data obtained from the following sources: maternity records (medical and obstetric history items, antenatal

attendances and maternal and neonatal outcomes), primary care notes for each mother and child dyad (consultations, immunisations, pregnancies, safeguarding), abortions data from the Department of Health Abortions Statistics Team, and immunisation data via COVER (Cover of Vaccination Evaluated Rapidly) contacts.

NHS Digital

The Hospital Episode Statistics (HES) datasets hold records on over 125 million hospital admissions, outpatient and accident and emergency episodes each year. Data can be requested from NHS Digital (formerly known as the Health and Social Care Information Centre), the executive non-departmental public body established under the Health and Social Care Act 2012.[13] All available records belonging to cohort members (mothers and children) will be obtained from study entry of the mother, which occurred between June 2009 and July 2010 until the date the child turns six. The data requested include diagnoses, procedures, length of episode and external causes of injuries coded according to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems [ICD-10] codes.[14]

NHS Digital has responsibility for collecting these data from across the health and social care system to allow NHS hospitals to be paid for the care they deliver. At the end of the financial year (March) a final dataset is collated. This dataset is cleaned and validated before being available for research at the end of the year (December).

Office for National Statistics (ONS)

The Office for National Statistics (ONS) collects information on cause of death from civil registration records. Mortality data can be accessed through NHS Digital. For registered deaths, the underlying cause of death is derived from the sequence of conditions leading directly to the death and is recorded on the death certificate. Deaths are subsequently coded in line with the ICD-10.

Department for Education (DfE)

The Department for Education (DfE) holds information on pupils throughout the different phases of education. Records are sourced from publicly funded schools, local authorities and awarding bodies and held in the National Pupil Database (NPD). Datasets are available on various aspects of education such as school census data, absence data, and school attainment.[15] All available records for the children in the cohort will be obtained from the various datasets held. Data coverage will vary depending on the dataset in question. For example, the School Census returns data on maintained schools (funding and oversight is through the local authority) which represents the majority of schools, Academies (funding and oversight is from the Department for Education), City Technology Colleges, maintained and non-maintained special schools and hospital special schools. Schools that are entirely privately funded and home education are not included in the data, this represents 7% of English students.[16]

In the UK education is mandatory from the first school term after their 5th birthday. Prior to this, some children may not have received formal early years provision and therefore may not appear in the datasets. A survey conducted in 2014-15 commissioned by Department for Education reported that 25 per cent of children aged 0-4, were not in receipt of any early years' provision. Older preschool children (aged 3 to 4) however, were far more likely to receive early years provision (92%) than younger pre-school children (aged 0 to 2) (61%).[17] We would therefore expect similar coverage rates for this study.

The data requested includes the number of hours attended, early educational development, eligibility for free school meals, and special educational needs (SEN) provision type. Datasets are collated throughout the year and are available at set time points annually.

Social Care Data

Social care data from local authorities is available through the NPD via two datasets, Child in Need (CIN) and Child Looked After (CLA). The CIN census captures individual level

information on children referred to and assessed by children's social care services within each 12month period.[18] CLA is collected in the SSDA903 return – an annual statutory data collection for all local authorities.[19] Any child in the cohort who is in one of these datasets will be identified. Mothers who were <18 years at the time of participation in the BB:0-2 trial will also be identified in these datasets. There will not be the coverage issues as seen in the education data returns and importantly, the primary outcome will be sourced from these social care datasets.

Study participants: Inclusion and exclusion criteria

Eligible participants are those mothers recruited to the BB:0-2 trial and their first child (or twins, if relevant) and who were not mandatorily withdrawn from the study, or electively withdrew including their consent for use of their data. Women were recruited as nulliparous women aged 19 or under, living in one of 18 local authority FNP catchment areas; recruited by 24^{+6} weeks gestation, have conversational level of English and were able to consent to research.[10]

Children in medium-term foster placements or adopted within the six year study period can be linked up to the date of adoption. Maternal or child death will be captured as an outcome.

Recruitment / Dissent

Participants previously consented to enter into the BB:0-2 trial and provide self-report and access to their routine records for the period up to two years postpartum. In order to obtain an unbiased estimate of the medium-term effect of FNP on objective and associated maltreatment outcomes we have received section 251 (s251) support of the 2006 NHS Act approval from the Health Research Authority's Confidentiality Advisory Group (HRA CAG) to pass identifiable participant data legally held by Cardiff University to the information centres (IC) to link to routine data. This is without obtaining further consent from participants, instead using an opt-out/dissent model.

Justification of approach

Consent for longer-term follow-up (i.e. beyond 24 months postpartum) was originally proposed in the BB:0-2 trial. However, upon ethical review it was considered that greater specificity about exact outcomes than could be provided at recruitment was required. Additionally, providing meaningful consent for much longer follow-up was also challenging, particularly on behalf of yet to be born children.

Developing the opt-out approach was necessary due to (i) the child protection focus of the study and the consequent sensitivity and impracticality in asking directly for consent, (ii) the mobility and relative difficulty in ongoing direct access to these participants (iii) the consequent introduction of non-ascertainment bias on sample representativeness – resulting in a non-random sample, and (iv) the likely cost and logistical requirements of securing even modest levels of additional consent.

Methods of notifying participants

We discussed the issue of dissent and fair processing with the HRA CAG and have subsequently attempted to contact all mothers recruited to the original BB:0-2 trial to inform them that medium-term follow-up using anonymised records will be undertaken.

Details of participants' residential addresses were updated using their most recent address registered with their GP. Where available, mobile number and email addresses collected for the trial were used to send SMS and emails to participants. All three modes of contact were used over a two-day period and participants were provided with a two-month window in which to contact the project team to discuss the project and opt-out if they wished. A website was also available with the same information which directed participants to contact the project team if they wished.

Development of opt-out letter

A group of care-experienced young people (CASCADE Voices)[20] advised on the layout, wording and tone of a letter to be sent to all participants. A key consideration was to communicate the focus of this follow-on study in a sensitive manner. The final letter was

approved by both an NHS Research Ethics Committee and CAG committee as part of overall governance approval for the study. The letter contained both information on the trial, the follow-on study and a flowchart for what to do if women wished to discuss the project and/or opt out.

Process to manage dissent

Women notifying the study team of their dissent will be recorded as "opted out", removed from all project datasets for this follow-up work and identifiable datasets to be sent to ICs. They will not be included in any of the datasets or analyses for this follow-on study.

Governance and compliance

Following Ethical approval (14/WA10062) and s251 support (CAG 10-08(b)/2014), data request applications were submitted to DfE, NHS Digital and ONS.

In order to satisfy the requirements of the s251 support and NHS Digital contract, the Information Governance (IG) Toolkit self-assessment[21] (commissioned by the Department for Health for NHS Digital to develop and maintain) was required. This organisation-level assessment provides reassurance of satisfactory information governance within the host trials unit. Both the s251 support and IG Toolkit are assessed and renewed on an annual basis. The opt-out model was also required to satisfy s251 support as well as the DfE assessment of compliance with principle one of the Data Protection Act 1998. Governance and IC requirements prior to application approval are shown in Figure 2.

Data matching

Maternal and child identifiers will be sent to both NHS Digital and DfE for matching with their databases. Each IC holds differing identifiers including a unique identifier for each individual (NHS Number; Unique Pupil Number UPN).

Matching with HES data will be by exact matching on NHS Number; Date of Birth; Postcode and Gender. This was conducted for BB:0-2 and achieved a high match rate where 99.6% of mothers and babies' records were matched fully (i.e. matched on all identifiers provided)

or partially (i.e. matched on a reduced, but acceptable number of identifiers provided). This will be repeated for this study. NHS Digital will then exact match with ONS using NHS number in order to obtain mortality data.

As the NPD does not include NHS numbers, initially exact matching on first name and surname, date of birth and postcode (of both mother and child for social care data; all other datasets just child) will be undertaken. Further matching required will be by fuzzy matching of first name. The CIN and CLA datasets do not contain names or postcodes therefore the matching will be in two phases: i) Participants will be matched with the NPD, the UPN added to all participants and ii) this will be used to identify individuals in the CIN and CLA datasets.

Data matching at DfE and NHS Digital/ONS are independent therefore match rates at the participant level are expected to vary (some may match to NPD but not HES). Educational records should be available for all children in the trial cohort whereas health and social care derived data will only exist where the child has received a relevant episode of care. Participants will be compared using trial baseline data to check for any bias between those who are matched and not-matched for those datasets where they would all be expected to be present (e.g. school census for all children).

The Pseudonymised Dataset

A unique study ID will be attached to each participant's record prior to data transfer to ICs. Once ICs have matched records to their database, only the unique study ID is retained. Data extracts from both ICs plus data files from the trial (following a process of de-identification and standardisation in Cardiff to reduce risk of later unintentional participant level identification) will all be securely transferred to a data safe haven,[22] the Secure Anonymised Information Linkage (SAIL) databank, for linking and storage. The data flow is shown in Figure 3.

A SAIL data analyst will re-assign the study ID with a new anonymous linking field [ALF] and store the corresponding ID in a separate encrypted password protected file.[23]

Participants will not be identifiable to the study team, or to the SAIL analyst, but incoming datasets can be linked at the individual level using the ALF. The study team will have controlled remote access to these data thus ensuring the security of the pseudonymised database.[24] All data cleaning and analysis will be carried out via the remote portal by the study data manager and statistician.

Data from NHS Digital and NPD will be requested at two time points. The first data extract will confirm the data flow model, assess data quality and the suitability of data for answering key study analyses. The second data request will be made once all children in the study have reached key stage one (April 2017) and on which the study findings will be reported on (in 2018).

Control of data

Cardiff University control under contract the identifiable trial data that are being transferred to the ICs and to the safe haven. Data held by NHS Digital, ONS and DfE, for which they are the controllers, are de-identified and then sent to SAIL to be linked and held (including the de-identified trial data) in a secure anonymised standalone database for use by nominated study team members. SAIL will control the safe haven environment, and will process the pseudonymised data for secure use by study team. Cardiff University will control the purposes to which the data are put in answering research questions as per the study protocol. Once linked in the data safe haven, the ability to submit queries to each IC about individual records will be more limited than if identifiable data were returned to the research team in Cardiff. Data cleaning will remain possible however as will generic queries about data provided in batch. The quality of matching conducted by NPD and NHS Digital/ONS will be a key factor in the success of the study.

Analysis

Power Calculation

Primary outcome (CIN status at any point between birth and six years): For CIN status, available UK data on rates are not specific to the age-range of interest, but the rate in the general population aged 5-9 years is 4.6% (for local authorities comprising study sites in BB:0-2). The rate of CIN status would be expected to be greater in the specific study sample, and therefore we have assumed a rate of 8%. We hypothesise that FNP would reduce the detection of CIN in the first six years and thus assumed a difference of 4% as being important. To detect a difference of 4% (FNP: 4% vs Usual Care: 8%) would require 602 children in each arm (1204 in total) using 80% power and a two-sided 5% alpha level.

BB:0-2 recruited 1645 women, with 1562 available for follow-up (i.e. excluding those subject to a mandatory withdrawal). Follow-up through medical records (assuming 10% loss in tracking & linkage) would result in 1405 participants, thus securing enough data to test the primary outcome.

Main analysis

Analyses will be conducted on an intention-to-treat basis and due emphasis placed on confidence intervals for the between-arm comparisons. Descriptive statistics of demographic and baseline measures will be used to ascertain any marked imbalance between the trial arms. The primary comparative analysis on CIN status at any point between birth and six years will use logistic multilevel modelling to investigate differences between the groups, and odds ratios alongside 95% confidence intervals (CIs) will be reported. Multilevel modelling will allow for clustering of effect within a site and family nurse. Modelling the impact of key subgroups (deprivation, looked after status of mother, adaptive functioning, Not in Education, Employment, or Training (NEET) status and age) and different intervention elements (e.g. gestational age at programme entry, dosage) on outcome will be undertaken by extending

the primary models and testing for interaction effects. The role of potential moderators of programme effect (e.g. domestic violence) will also be explored.

Secondary outcomes will assess group differences in objective and associated measures of maltreatment, intermediate FNP programme outcomes as well as child health, development and educational outcomes (as detailed in table 2). The majority of these are binary outcomes (presence/absence of a status, meeting the Key stage one standard or not)) and will be analysed using a multilevel logistic regression model. The distribution of potential continuous outcomes such as Early Year assessment scores will be assessed before analysing using linear regression. Count data such as the number of attendances for injuries and ingestions will be analysed using a Poisson or negative binomial multilevel regression modelling. A detailed statistical plan will be written and signed off prior to any analysis.

A state transition model using Markov chains will be used to assess the probabilities of moving from one stage marker (states) to another.[25] The transition probabilities (the probability of the various state-changes) in our model will be derived from our data and compared between groups.

Bias in the followed-up BB:2-6 sample will be quantified by examining group differences (participants and non-participants) in baseline variables such as age, deprivation, gestational age, and education. Surveillance bias in detection of maltreatment during the child's infancy and toddlerhood can be assessed by examining subsequent reporting.[26] The duration between birth and the date of first referral to CSC will be calculated and group differences examined using Cox regression analysis to calculate hazard ratios for referral, together with 95% CIs. Surveillance bias is most likely to occur during the intervention phase, although improved handover to other services at 2 years may lead to higher identification in the following year. Severity of the referral will also be compared between the two groups (an approach used in US trials of NFP to explore surveillance bias).

Health economics

The economic evaluation will consider costs and consequences of the FNP over the full follow-up period (BB:0-2 & BB:2-6). The current BB:0-2 study reported 1) a within trial cost utility analysis assessing NHS costs against quality adjusted life years (QALY) from the perspective of the mother, and 2) a within trial cost consequences analysis relating all costs (including those to the social care, education and criminal justice sectors as well as health) against the full range of effects.[12] Cost and consequences framework is deemed the most appropriate economic evaluation framework for public health interventions[27] and preferred by NICE[28] because it enables capture of equity consideration as well as intersectoral costs and consequences[29] yet applications are still limited.[27]

The absence of additional data on Health Related Quality of Life within the BB2 study means that it will not be possible to estimate QALYs beyond 24 months postpartum and hence extend the within trial cost utility analysis. However, the within trial cost consequences analysis will be extended from 0-2 to 0-6 years through collection of resource use data from medical and education records (including from the latter, data related to social care usage). Costs will be summarised against the range of outcomes collected within BB2 without aggregation to allow weighing up changes in the various outcomes reported in BB2 against the changes in costs in a consistent and transparent manner.[30] This will contribute to providing more robust and valid medium-term estimates within the extended period.

ETHICS AND DISSEMINATION

Legal & Ethical considerations

The potential for using routine data in health and social care research has been greatly publicised and study designs utilising these data are encouraged by funders.[31] There are, however, many inherent challenges in working with secondary-use data, in particular for this project the ethical and legal requirements/responsibilities which have fundamentally informed this study design.

Although BB:0-2 linked trial data to HES and ONS data via NHS Digital, the governance requirements around the two applications have differed between the two studies not least because of the difference in consent models. Trial data were provided by NHS Digital and ONS after participant consent to prospective collection and for specified purposes limited to the time-frame of that study. The current follow-on study uses a dissent model under which we are only able to send trial participant identifiers to ICs for matching to outcome data records if there is no objection received from mothers. This is especially important, as following an opportunity to object to being included in the current study, those women who withdrew from the original Building Blocks will be retained. The study will require all clinical, social and educational data to be held in a data safe haven using encrypted record identifiers and analysis via a securely managed and monitored remote portal. The legal bases for transfer of identifiable data to ICs without explicit consent are as follows; s251 of the 2006 NHS Act 2006 for HES data from NHS Digital, s42(4) of the Statistics and Registration Service Act 2007 through NIHR funding for ONS data via NHS Digital, and 6(1) of Schedule 2 of the 1998 Data Protection Act for NPD data.

Dissemination of findings The Building Blocks: 2-6 Study will generate policy-relevant findings describing the medium-term impact of FNP on measurements of child maltreatment. The findings will also include other policy relevant outcomes from the programme such as health care use, education attainment and changes in social care use over the 6 years of follow up. Such medium-term evaluation remains important as some outcomes for the intervention are expected to arise only after the child's second birthday, including maltreatment. This study will either confirm the largely negative trial findings from BB:0-2 further weakening the justification for FNP Programme continuation or provide a balance to the early measurable outcomes.

In addition to reporting the findings to the funder for this study, the funder for the BB:0-2 trial (DH Policy Research Programme) will also be informed and the FNP National Unit (FNPNU). All local authorities in England will be notified of the results, as (since October 2015) they

have responsibility for commissioning public health services for children aged 0-5. Participants will receive a summary of the results and all reports and publications will be made publicly available in full on the Cardiff University website. The research team have previously convened and met twice with a stakeholder group, including relevant policy leads from each country in the UK delivering FNP (England, Scotland, Northern Ireland). We will stage a similar event to present and discuss the implications for practice and policy of the results of this medium-term follow up of participants.

In addition to policy and public outputs, academic outputs will include (i) this protocol paper providing visibility of this medium-term follow up, (ii) a methods paper describing the piloting process of the study (including data quality and success of data matching) and (iii) main study findings. We aim to disseminate in high-quality, peer reviewed journals and present in key conferences.

A particular benefit of this study is understanding of, and learning from, the governance challenges. There is potential to use this method for future trials looking at longer term follow-up. Therefore this study has the potential to add to the understanding of routine data and data linkage methods in future public health and clinical trials and these planned publications will provide a basis for the dissemination of the success of these methods.

Finally, publishing protocol papers in medical journals were an important innovation for trials. They convey a number of benefits including transparency about what was intended by researchers and therefore comparison to what was actually reported. While protocols are more commonly published for trials, we consider that the protections afforded are similar for other study types. This may include inhibiting ‘data dredging’ and post-hoc revisions to original study plans. In our study, which links a trial cohort to routine data we consider that this is especially important, particularly because of the broad range of outcomes that are potentially impacted by this complex home visiting intervention.

Declarations

List of abbreviations

ALF: anonymous linking field;
ALF-E: encrypted ALF;
BB:0-2: Building Blocks trial;
CAPI: computer-assisted personal interview;
CATI: computer-assisted telephone interview;
CIN: Child in Need;
CLA: Child Looked After;
DfE: Department for Education;
DNA: Did not attend;
DoH: Department of Health;
EYFSP: Early Years Foundation Stage Profile;
FNP: Family Nurse Partnership;
FNPNU: Family Nurse Partnership National Unit;
HES: Hospital Episode Statistics;
HRA CAG: Health Research Authority's Confidentiality Advisory Group;
HSCIC: Health and Social Care Information Centre;
IC: information centres;
IG: Information Governance ;
NEET: Not in Education, Employment, or Training
NFP: Nurse Family Partnership;
NPD: National Pupil Database;
ONS: Office for National Statistics;
PCT: Primary Care Trusts;
PRU: pupil referral unit;
QALY: Quality adjusted life years;
s251: Section 251 of the NHS 2006 Act;
SAIL: Secure Anonymised Information Linkage;
SEN: special educational needs;
TTP: trusted third party;
UPN: Unique Pupil Number.

Ethics approval and consent to participate - Ethics approval of the study has been given by the Research Ethics Committee for Wales (14/WA10062) and the transfer and use of identifiable data has been approved by the Health Research Authority [HRA] Confidentiality Advisory Group [CAG] (CAG 10-08(b)/2014).

Consent for publication - Not Applicable

Availability of data and material - Not Applicable

Competing interests - The authors declare that they have no competing interests

Funding - This project was funded by the National Institute for Health Research Public Health Research (NIHR PHR) Programme (project number 11/3002/11). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the NIHR PHR Programme or the Department of Health.

Authors' contributions

MR is the chief investigator of the study. All authors have contributed to and are responsible for the final design of the study. FLW and GM are responsible for study and data management. RCJ is responsible for statistical planning and for data analysis. DF is responsible for the health economics. All authors read and approved the final manuscript.

Acknowledgements

The South East Wales Trials Unit is funded by Health and Care Research Wales. The work was undertaken with the support of The Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement (DECIPHer), a UKCRC Public Health Research Centre of Excellence. Joint funding (MR/KO232331/1) from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, the Welsh Government and the Wellcome Trust, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

References

1 A guide to inter-agency working to safeguard and promote the welfare of children. In: Working together to safeguard children. Department for Children Schools and Families. 2010. http://webarchive.nationalarchives.gov.uk/20130401151715/https://www.education.gov.uk/publications/eorderingdownload/00305-2010dom-en-v3.pdf Accessed Dec 2014.

2 Characteristics of children in need: 2014 to 2015 (Statistics: children in need and child protection) year ending 31 March 2015. In: Department for Education National Statistics. 2015. Department for Education. https://www.gov.uk/government/statistics/characteristics-of-children-in-need-2014-to-2015 Accessed Feb 2016.

3 MacMillan HL, MacMillan JH, Offord DR, Griffith L, et al. Primary prevention of Child Physical Abuse and Neglect: A Critical Review. *J child psychol psyc* 1994;35(5):835-56.

4 Olds DL, Hill PL, O'Brien R, et al. Taking preventive intervention to scale: The nurse-family partnership. *Cogn behav pract* 2003;10(4):278-90.

5 Olds DL, Henderson CR Jr, Chamberlin R, et al. Preventing child abuse and neglect: a randomized trial of nurse home visitation. *Pediatrics* 1986;78: 65–78.

6 Kitzman H, Olds DL, Henderson CR, 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.

7 Olds DL, Robinson J, O'Brien R, et al. Home visiting by paraprofessionals and by nurses: a randomized, controlled trial. *Pediatrics* 2002;110(3):486-96.

8 Olds D, Henderson CRJ, Kitzman H, et al. Effects of Prenatal and Infancy Nurse Home Visitation on Surveillance of Child Maltreatment. *Pediatrics* 1995;95(3).

9 Owen-Jones E, Bekkers MJ, Butler CC, 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.

10 Robling M, Bekkers MJ, Bell K, et al. Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial. *The Lancet* 2016;387(10014):146-55.

11 Olds DL, Eckenrode J, Henderson CR, 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.

12 Zielinski DS, Eckenrode J, Olds DL. Nurse home visitation and the prevention of child maltreatment: Impact on the timing of official reports. *Dev Psychopathol* 2009;21(02):441-53.

13 Data Access Request Service http://www.hscic.gov.uk/dars. Accessed March 2016.

14 *International Statistical Classification of Diseases and Related Health Problems, 10th revision,* Volume 1, Geneva. World Health Organisation 1992.

15 National pupil database: apply for a data extract https://www.gov.uk/guidance/national-pupil-database-apply-for-a-data-extract#data-extracts-available. Accessed April 2016.

16 Gorard S, Smith E, Benadusi L, et al. Equity in Education: an international comparison of pupil perspectives. Springer 2010.

17 Huskinson T, Hobden S, Oliver D et al. Department for Education: Childcare and early years survey of parents 2014 to 2015. 2016

18 Children in need census 2015 to 2016: guide https://www.gov.uk/government/publications/children-in-need-census-2015-to-2016-guide Accessed April 2016.

19 Children looked after return 2015 to 2016: guide. https://www.gov.uk/government/publications/children-looked-after-return-2015-to-2016-guide. Accessed April 2016.

20 Shaw IG, Holland S. Doing qualitative research in social work. Sage; 2014.

21 Information Governance Toolkit https://www.igt.hscic.gov.uk/. Accessed March 2016.

22 Ford DV, Jones KH, Verplancke JP, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;9(1):157.

23 Lyons RA, Jones KH, John G, et al. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis* 2009;9(1):1.

24 Jones KH, Ford DV, Jones C, et al. A case study of the secure anonymous information linkage (SAIL) gateway: A privacy-protecting remote access system for health-related research and evaluation. *J Biomed Inform* 2014;50:196-204.

25 Meyn SP, Tweedie RL. Markov chains and stochastic stability. Springer Science & Business Media; 2012.

26 Chaffin M, Bard D. Impact of intervention surveillance bias on analyses of child welfare report outcomes. *Child maltreatment* 2006;11(4):301-12.

27 Drummond M, Weatherly H, Claxton K, et al. Assessing the challenges of applying standard methods of economic evaluation to public health interventions. Public Health Research Consortium; 2006. http://phrc.lshtm.ac.uk/papers/PHRC_D1-05_Final_Report.pdf

28 National Institute for Health and Care Excellence (2012) Methods for the development of NICE public health guidance (Third edition) https://www.nice.org.uk/article/pmg4/chapter/1%20introduction. Accessed October 2016.

29 Kelly MP, McDaid D, Ludbrook A and Powell J (2003) Economic appraisal of public health interventions. http://www.cawt.com/Site/11/Documents/Publications/Population%20Health/Economics%20of%20Health%20Improvement/Economic_appraisal_of_public_health_interventions.pdf. Accessed October 2016.

30 National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [PMG20]. October 2014 Updated April 2016 Accessed at https://www.nice.org.uk/process/pmg20/chapter/incorporating-economic-evaluation#approaches-to-bespoke-economic-evaluation

31 Methods research to support use of observational data in clinical decision making. http://www.mrc.ac.uk/funding/how-we-fund-research/highlight-notices/methods-research-observational-data/. Accessed March 2016.

(Title) Figure 1. Follow up and datasets over the six years

(Legend) A&E Accident and Emergency; PRU Pupil Referral Unit

(Title) Figure 2. Governance and Information centre requirements prior to application approval.

(Legend) s251 Section 251 of the NHS 2006 Act; ONS Office for National Statistics; NPD National Pupil Database; IG Information Governance

Figure 3. Data Flow

(Legend) [1]Participant identifiable information securely transferred for linkage; [2]De-identification and Standardisation applied (e.g. date of birth to week of birth); [3]Information centres confirm matching of participant identifiers; [4]Hosted on SAIL secure platform. ALF-Anonymised Linking Field; BB:0-2 – The Building Blocks trial; DfE – Department for Education; DPO – Data Providing Organisation; HES – Hospital Episode Statistics; ONS – Office for National Statistics; SAIL – Secure Anonymised Information Linkage.

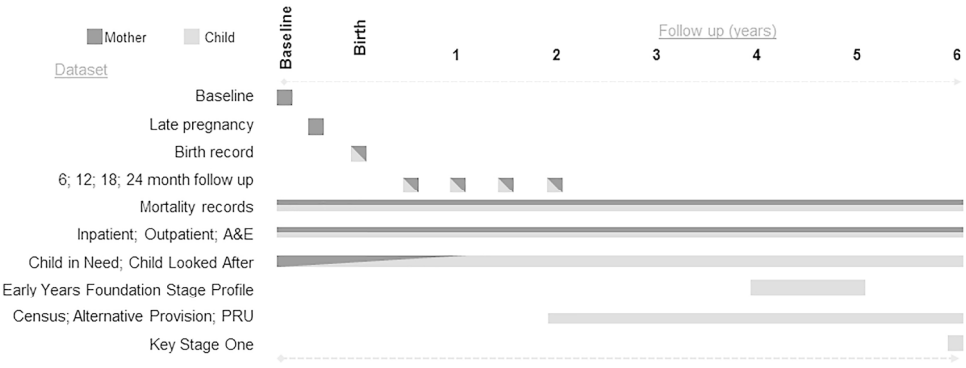


Figure 1. Follow up and datasets over the six years.
A&E - Accident and Emergency; PRU - Pupil Referral Unit

254x190mm (300 x 300 DPI)

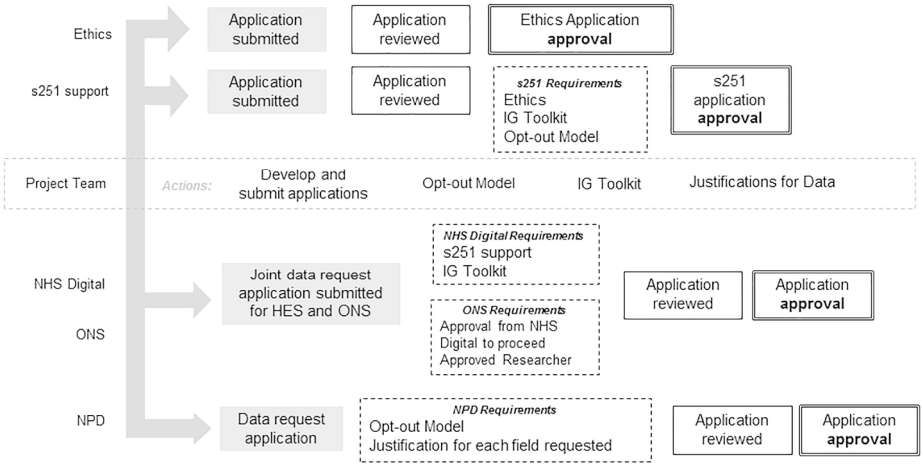


Figure 2. Governance and Information centre requirements prior to application approval.
s251 - Section 251 of the NHS 2006 Act; ONS - Office for National Statistics; NPD - National Pupil Database; IG - Information Governance

254x190mm (300 x 300 DPI)

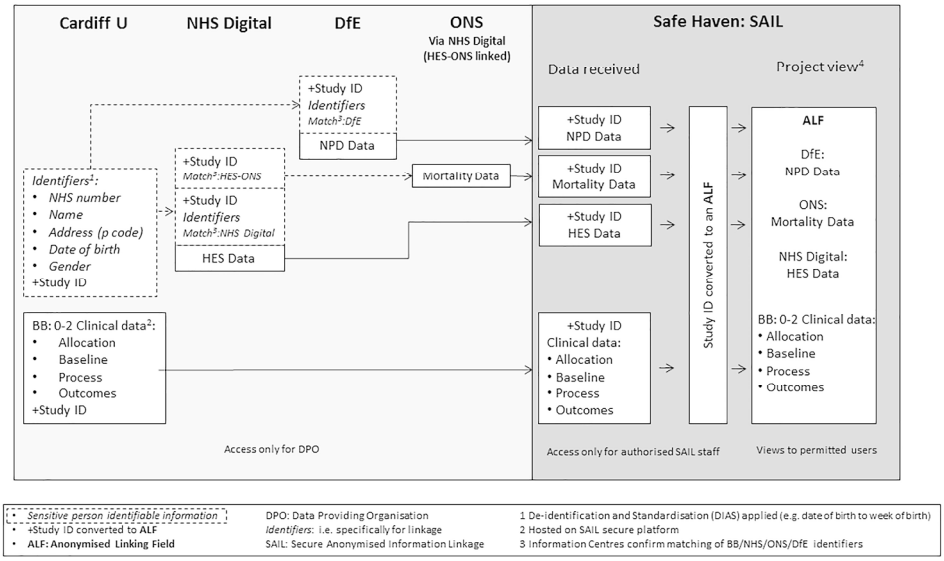


Figure 3. Data Flow.
1 - Participant identifiable information securely transferred for linkage; 2 - De-identification and Standardisation applied (e.g. date of birth to week of birth); 3 - Information centres confirm matching of participant identifiers; 4 -Hosted on SAIL secure platform. ALF- Anonymised Linking Field; BB:0-2 – The Building Blocks trial; DfE – Department for Education; DPO – Data Providing Organisation; HES – Hospital Episode Statistics; ONS – Office for National Statistics; SAIL – Secure Anonymised Information Linkage.

299x190mm (300 x 300 DPI)