

Dadansoddi ar gyfer Polisi



Analysis for Policy



Llywodraeth Cymru
Welsh Government

Ymchwil gymdeithasol

Social research

Number: 09/2014

www.cymru.gov.uk

Data Linking Demonstration Project - Flying Start



**Martin Heaven, Health Information Research Unit, Swansea
University/Data Linkage Fellow, Welsh Government**

**Sarah Lowe, Knowledge and Analytical Services, Welsh
Government**

Views expressed in this report are those of the researcher and not necessarily those
of the Welsh Government

For further information please contact:

Sarah Lowe

Knowledge and Analytical Services

Welsh Government

Cathays Park

Cardiff

CF10 3NQ

Tel: 02920 826229

Email: sarah.lowe@wales.gsi.gov.uk

Welsh Government Social Research, 2014

ISBN 978-1-4734-0907-1

© Crown Copyright 2014

Table of Contents

Glossary of Acronyms	3
1. Introduction	5
2. Making health data available for research	10
3. Methodology	16
4. Address matching and defining populations	25
5. Findings	31
6. Discussion	60
Appendix 1 IGRP membership	64
Appendix 2 IGRP application	65
Appendix 3 Split File Process diagram	72
Appendix 4 List of possible future research projects	75
Appendix 5 Technical Guide	76
Appendix 6 Distribution of RALFs by Geographical Area	83

Glossary of acronyms

'5 in 1'	The '5 in 1' vaccine for Diphtheria, Tetanus, Pertussis, Polio and HIB.
A&E	Accident and Emergency (Data Set)
ALF	Anonymised Linking Field
CHS	Child Health System (Data Set)
EASHR	European Age Standardised Hospitalisation Rate
EASMR	European Age Standardised Mortality Rate
EDDS	Emergency Department Data Set
ESRC	Economic and Social Research Council
FSE	Flying Start Eligible (for the purposes of this report)
HES	Hospital Episode Statistics
HIB	Haemophilus influenzae type B vaccine
HIRU	Health Information Research Unit
IGRP	Information Governance Review Panel
KS1	Key Stage 1 (educational attainment assessment)
LA	Local Authority
LSOA	Lower Super Output Area
MCHC	Multiple Chronic Health Conditions
MMR1	Measles, Mumps and Rubella (1 indicating 1st round of vaccination)
MMR2	Measles, Mumps and Rubella (2 indicating 2nd round of vaccination)
NMD	Next Most Deprived (a comparison group in this project)
NWIS	NHS Wales Information Service
NHSAR	NHS Administrative Register, (now superseded by WDS)
ONS	Office for National Statistics
PAS	Patient Administration System
PEDW	Patient Episode Database for Wales
PHW	Public Health Wales

RALF	Residential Anonymised Linking Field
RoW	Rest of Wales (a comparison group in this project)
SAIL	Secure Anonymised Information Linkage
SQL	Structured Query Language
WDS	Welsh Demographics Service (GP registration history database)
WG	Welsh Government
WIMD	Welsh Index of Multiple Deprivation

1 Introduction

The aim and objectives of the demonstration project

- 1.1 This project is being delivered as part of the Welsh Government Programme to Maximise the Use of Existing Data. It aims to demonstrate the unique contribution data linking can make to the evidence base. The suite of three data linking demonstration projects has examined the anonymised data linkage process from acquiring additional data to carrying out analysis on new data sets created by linking existing administrative data. The projects are intended to stimulate engagement of appropriate WG officials with regard to information governance and practical issues around acquiring, processing and analysing new linked data sets. The projects were delivered by a Knowledge Transfer Research Fellow jointly funded by WG and the Economic and Social Research Council (ESRC) working on a one-year fellowship. The projects were therefore designed to be small in scale and exploratory in nature. These constraints are reflected in their relatively limited scope and in both the practical and analytical decisions made throughout.
- 1.2 Flying Start is the Welsh Government's flagship early years programme and provides a path towards improving the life chances of children in some of our most disadvantages communities. This multi-disciplinary programme provides a 'universal' set of entitlements which all children under the age of 4, and their families within targeted geographical areas can access. The entitlements offer a range of support that will provide a quality early years provision for the child and a range of support for parents to build skills and resilience. The 4 key elements are:
- quality part-time childcare for 2-3 year olds in Flying Start areas;
 - an enhanced Health Visiting service (caseloads are capped at 1:110);
 - access to Parenting Programmes; and
 - support for Early Language Development.
- 1.3 The Flying Start Project aims to demonstrate data linkage in a practical way by establishing whether any impact of the Flying Start programme can be determined from analysing linked longitudinal administrative data sets that already routinely flow into the SAIL databank.

- 1.4 The Project also aims to demonstrate that routinely collected datasets, handled appropriately, can provide the means to retrospectively establish baselines for monitoring the effect of an existing intervention, and to prospectively monitor changes in outcomes over time. The recipient population is oblivious to the assessment process so there are no data collection biases and the cost of ‘keeping in touch’ is avoided altogether.
- 1.5 In order to achieve the aims outlined in 2.3, a number of intermediate objectives had to be met. Firstly, it was necessary to identify a number of outcome variables or ‘indicators’ in existing health and education datasets, so that baseline information about the Flying Start Eligible (FSE), Next Most Deprived (NMD), and Rest of Wales (RoW) populations could be generated (for information about the selection of indicators, see Chapter 3). Changes in the indicators over time could then be identified between the pre- and post-intervention time periods. Eligible children and families can also be followed into the future, providing a low-cost solution to long-term monitoring of the outcomes for which data are – or become – available in SAIL.
- 1.6 The Project documents the step-by-step process of adding potentially identifiable data (addresses) into an anonymised data repository and linking it, without compromising confidentiality, with data already held.
- 1.7 The suite of projects was designed to demonstrate data linkage methods but, in doing so, to use the experimental methods developed to deliver some tentative policy-relevant findings to contribute to the evidence base. In the case of this Project, we hoped to contribute to the evidence base for Flying Start. We would expect to see changes in the indicators for the FSE population if Flying Start is having an effect, however it was by no means certain that it would be possible to do so, given that a) changes over time are complex to study because other unrelated changes can be going on at the same time, and b) changes due to Flying Start interventions would need to reach a certain size before we can be confident they are ‘real’ changes.
- 1.8 It is also possible that other interventions may have influenced the outcomes e.g. Communities First or Families First, but it was not possible within the limited scope

of a demonstration project to identify and control for the effects of other interventions.

- 1.9 The Project is experimental. Due to the limited project scope, the lessons learned about data quality and the challenges associated with the use of newly-linked data sets, as well as the developmental nature of the methods, the reader must note that **the findings must be viewed with caution.**
- 1.10 Given the limited scope of a demonstration project, it was accepted that it would not be possible to address every potential shortcoming of the data and/or analysis. Data is not available on individual receipt of the various component benefits of the Flying Start programme, so the key limitation of the project is that it analyses data based on a child's eligibility for the service, not their or their family's actual receipt of that service¹.
- 1.11 The challenges of working with linked data sets, none of which were designed or collected with research purposes in mind, are already considerable. There are also considerable challenges associated with appropriately and meaningfully analysing data over time (longitudinally). It was therefore accepted from the beginning that the learning about methodology that would come out of these projects would be more important than any substantive research or statistical findings. One key challenge in undertaking the Project has been the selection of suitable outcome indicators e.g. from General Practice data. This is a complex task and there is no precedent to follow in developing an approach. For example, comparing immunisation uptake over time is made more difficult due to the fact that different age groups of children have been targeted over time. For further discussion of indicator selection, see Chapter 3.
- 1.12 Flying Start is a Welsh Government (WG) led initiative that aims to 'make a decisive difference to the life chances of children aged less than four in the areas in which it runs'. It includes the provision of enhanced health visiting, childcare, parenting programmes and basic skills (language and play) programmes, with information

¹ Data on receipt of enhanced services is not held centrally. Health visitor caseload approximately matches the funded places and for 2012-13, virtually all eligible children were offered either a full or reduced place for childcare and take up of these offers ran at 90% across Wales. However, this data is for acceptance of a place regardless of subsequent attendance; no data on subsequent attendance is available.

sharing and referral between all practitioners within the Flying Start delivery team. Family eligibility is determined by address – all families with a child aged less than 4 years living in the most deprived community school catchment areas within each Local Authority (LA) in Wales.

1.13 The programme was initiated in 2006/07 and became operational in 2007/08 when Local Authorities began the process of allocating the required staff and premises to provide the programme components. Survey-based evaluation was designed into the programme with reports available on Welsh Government website². In 2012, the Welsh Government announced a commitment to double the number of children benefiting from the programme from 18,000 to 36,000. This will mean that almost a quarter of children in Wales, under the age of 4, will be receiving services through the programme by 2016. The expansion of the programme is being phased over a number of years to allow for the necessary infrastructure to be put in place, including the recruitment and training of additional health visitors; childcare workers; staff to deliver parenting programmes; and the development of capital infrastructure, such as childcare settings, in new target areas. The Welsh Government has allocated more than £88 million additional grant funding to support the delivery of this commitment up to 2016. In addition, £25 million has been approved for capital developments, including high quality childcare settings in all 22 local authorities, providing facilities in communities where currently none exist. It should be noted, however, that all of the analyses in this report relate to the pre-expansion definitions of Flying Start eligible areas. The Project therefore analysed data for 2004 to 2012, comparing indicators for the pre-Flying Start period of 2004 to 2009 with period of 2010 to 2012 when Flying Start had been rolled out.

1.14 As a number of years will pass before children reach various key stages in education, a key objective of the project is to develop a semi-automated process to provide annual reporting for the future.

1.15 The process, issues, problems and limitations encountered are documented in this Report. Chapter 3 describes the SAIL databank, the anonymisation of the Flying Start Eligible addresses and their use to 'flag' records in other anonymised

² <http://wales.gov.uk/about/aboutresearch/social/latestresearch/EvalFlyStart7-20/?lang=en#>

databases. Chapter 4 describes the Project design, comparing Flying Start Eligible families to other families over time. Chapter 5 reports the Project findings and Chapter 6 discusses the various limitations of the Project. Chapter 7 provides a discussion on what has been learned from the project. The challenges that emerged during the demonstration process will also be included in a Lessons Learned report publication of which will follow.

2 Making administrative data available for research

The development of SAIL

- 2.1 The NHS in Wales generates a large amount of data collected on a routine basis for the purposes of treating patients and making sure the health service runs properly. Almost none of it is collected for research or statistical purposes but it nevertheless has considerable potential to be used for research. The same is true of other public services, such as education, housing and social care. Data collected by the NHS in Wales is held complete with patient identifiers: NHS number, full name, address, postcode, date of birth and gender. Information Governance law therefore prohibits the use of data in this form for research outside the NHS.
- 2.2 To address this problem, the Welsh Government funded the creation and development of the Health Information Research Unit (HIRU) at Swansea University, from 2006. The aim of this unit was to develop a means by which routinely collected health data from many different sources could be utilised in a linked way, but in such a way that each dataset was rendered completely anonymous. The process developed was called Secure Anonymised Information Linkage (SAIL); it led to the collation of a very large collection of anonymised health (and other) datasets stored in Swansea University and made available for research. Each anonymised dataset is stored separately and as far as possible in the form in which it is received from the data provider. The name of 'SAIL' has become synonymous with both the large data repository and the HIRU. 'SAIL' will be used to refer to the databank at Swansea University throughout this Report.
- 2.3 In the data anonymisation process, SAIL use a common procedure for safeguarding respondent privacy during data linking (for a diagram of the process, see Appendix 3). This involves a "split file" process - the separation of all identifying information i.e. name, address, date of birth, NHS number etc. from all other analytical data, whether medical, social, financial, attitudinal etc. in each source to be linked. For each source, this creates two files, the first containing an index plus the identifiable information and the second containing an index plus the analytical data. For each dataset the identifying information is sent to a trusted third party that creates an

anonymous linking field. Once the linking field has been created the identifying information is destroyed leaving only the linking field and the index. The index allows the anonymous linking field to be reattached to the analytical data. The analytical data can then be linked to other anonymised data sets without using any identifying variables.

- 2.4 For SAIL, the NHS Wales Information Service (NWIS) act as the 'trusted third party' organisation, routinely turning lists of personalised details into anonymised codes. NWIS utilise the Welsh Demographic Service (WDS) data as the 'population spine' or 'template' for its anonymisation process. The WDS is a database of everyone registered with a GP in Wales from 1994 to the present day. It includes an anonymised residential address history – an index of numbers, one for each residence in Wales, known as the Residential Anonymised Linking Field (RALF). Individual people who have been registered with a GP in Wales, past and present, are represented in the WDS data as another index of unique numbers, known as the Anonymised Linking Field (ALF). In this way, it is possible to associate RALFs with ALFs, that is: homes with residents.
- 2.5 For the development of the Flying Start demonstration project, WG provided NWIS, (i.e. not SAIL) with the complete list of eligible addresses contained in the pre-expansion Flying Start area definition. NWIS processed these addresses, providing SAIL with the associated RALFs. When NWIS process any address it always generates the same unique number (RALF) representing that address. The SAIL system is able to utilise the RALF for linkage, but as only NWIS hold the key to how the number was generated, it's not possible in SAIL to identify the address it refers to, only the small geographical area, the Lower Super Output Area (LSOA), in which it is located. An LSOA contains on average about 1,600 residents, making the identification of an individual from the LSOA impossible.

Information Governance Issues

- 2.6 SAIL follows the data protection guidance provided by the Data Commissioner's Office, and operates within the Swansea University Data Protection policy which is in line with all the relevant UK laws. The anonymous nature of data held in SAIL is such that it is not governed by the Data Protection Act, and it has been agreed by the National Research Ethics Service that research carried out within SAIL does not require ethical review. However all research carried out within SAIL is still managed through a rigorous control structure to ensure that confidentiality is maintained and potentially disclosive outputs are not produced.
- 2.7 One of the controls in place is a requirement for all proposals involving the analysis of linked data within SAIL to obtain approval from the Information Governance Review Panel (IGRP). The IGRP is a panel of independent specialists in informatics governance and lay people that oversees all research taking place within SAIL. Current membership (June 2013) is listed in Appendix 1. An IGRP application contains an outline of the research rationale for creating the link(s), details of any new datasets that would be accessed, and precisely what variables would be required from the linked datasets. Researchers must indicate in the application that they have considered the handling of sensitive data in the research design. Although the data sets are all held completely anonymously in SAIL, the selection of a very specific sub-group based on age and gender at small area (LSOA) level, looking at a specific condition could return small numbers. Small numbers in a published output could be put together with other local knowledge to establish who the statistic refers to. Researchers are given access to the data at the most detailed level required to complete their analysis, but need to ensure that nothing potentially identifiable is revealed in their reporting. IGRP applications must indicate how the analyst proposes to deal with small numbers (e.g. through grouping and aggregation of cases).
- 2.8 The IGRP application for this Project was agreed without any further clarification questions being asked by the IGRP members (for a copy of the application, see Appendix 2). From an Information Governance perspective, the project is regarded as low risk as no new information is provided about the anonymised individuals that

could compromise the existing confidentiality of the datasets. As noted above, anonymisation of the addresses allows the flagging of anonymous records for the people resident in Flying Start eligible households but data was not available to allow the flagging of specific families who had taken up the benefits of the Flying Start programme. However, gaining IGRP approval still took eight weeks, mainly due to waiting for individual reviewers to respond. Researchers should be aware that gaining IGRP approval can be a time-consuming, iterative process requiring adjustments to their research proposal. Lessons learned during the implementation of this and the other demonstration projects will be presented in the Lessons Learned report, which is to follow.

The SAIL Research Environment

- 2.9 A database “view” is a structured ‘image’ of information stored in the database, including only a subset of the complete dataset. A “view can include data from more than one database, and can be restricted to include specific rows and columns. In this way, the database administrators can very closely control the data with which each researcher is allowed to work. There is, in addition, no way that a researcher can alter the underlying data table providing the “view”.
- 2.10 The “view”, tailored specifically to meet the requirements of the researcher’s project, is loaded into the SAIL databank by the SAIL technical team. The SAIL technical team members provide the hardware and database management support for research and are not data analysts. Separating the data management and research analyst functions prevents the need for technical team members to understand the data and for researchers to access underlying data tables or any intermediate stage data. Access to “views” is controlled and restricted to authorised approved researchers. For the Flying Start project, access was initially restricted to the Author, who is a senior research analyst working in the SAIL databank. Later, access was sought for named WG staff for Quality Assurance purposes.
- 2.11 The database “views” are made available through a secure remote access system, the SAIL Gateway, which can be accessed securely over the internet, using a system where authorised researchers are able to log on to a dedicated computer through a password protected browser. Outputs are ‘locked down’, so that nothing

can be copied and pasted out of the gateway, saved to a port or drive on the remote computer, or sent to a printer.

- 2.12 All analysts who are provided with a SAIL gateway account are given access only after both they and their line manager have signed a detailed agreement outlining the researcher's responsibilities and the agreed usage that can be made of that account. The agreement clearly places the responsibility with the researcher to ensure that no individual could be potentially identifiable from the research outputs. However, in addition, all potential outputs are scrutinised by a SAIL administrator to ensure potentially disclosive information does not leave the secure gateway.
- 2.13 The researcher is required to carry out the analysis within the gateway, in which suitable database, statistical, spreadsheet, word processing, mapping and presentation software are available. The only outputs allowed are summarised or aggregate results. Proposed outputs are processed through the 'request data out' link within the gateway. This is the stage at which outputs are scrutinised by a senior research analyst in SAIL, checking for potential disclosure issues such as small numbers. The 'data out' process does not check that the analysis has been performed correctly and that results are correct, it merely scrutinises outputs for potentially disclosive situations.
- 2.14 It is not possible to put a process in place that would stop researchers taking photographs of a computer screen, for example, or simply writing down results and not following the 'data out' procedure. Given this, the researcher must be trusted to adhere to the terms of the SAIL access agreement. However, when signing the access agreement, researchers and their line managers are agreeing to abide by the statement of procedures in the National Statistics Code of Practice: Protocol on Data Access and Confidentiality, in both letter and spirit, to the maximum extent that they apply. Breaches of these rules would result in penalties and legal action. As part of the creation of the UK Administrative Data Research Network, it is anticipated that researchers accessing databases like SAIL will be required to successfully apply to be part of an official register. Abuse of the privileges of data access would then result in removal from such a register, effectively ending the perpetrator's research career.

3 Methodology

- 3.1 Flying Start areas were defined by WG from school catchments in deprived areas across Wales. A set of control areas was also defined by WG, based on the areas where the 'Next Most Deprived' (NMD) populations were located.
- 3.2 The Flying Start Eligible (FSE) addresses were generated by the WG Geography and Technology Team and sent to NWIS via Secure electronic transfer. As described in Section 3.1, above, NWIS generated the FSE RALFs and provided them to SAIL via a secure electronic data transfer. The NMD areas were provided to SAIL as a list of Lower Super Output Areas (LSOA); i.e. NWIS was not involved in flagging this Group. The NMD RALFs were established by a "LSOA to RALF" look-up table using the WDS as a template. Extracting both Groups of households from the complete list of households in Wales created a third Group – households in the "Rest of Wales" (RoW). RoW is a large Group of RALFs i.e. it relates to a large number of residences, and can be used to help identify any underlying change over time that was not associated with Flying Start.
- 3.3 Linkage of these three Groups of RALFs to the WDS data established which Anonymous Linkage Fields (ALFs) – or people - were associated with the RALFs – or residences – on specific dates. The WDS holds the address history of the registered population in the form of dates they moved into and out of RALFs.
- 3.4 As Flying Start became fully operational across Wales from 1st April 2009, this Project has used data starting from 1st April 2004. This provided a five year period prior to the introduction of Flying Start to establish baseline rates and trends. Some children aged 3-4 years in 2004 would be born in 2000, which is the earliest year from which complete and robust recording of the health data is available.
- 3.5 The project established, for each of the three Groups, which RALFs (residences) contained a child aged less than four years on 1st April 2004. For each residence containing a child of the right age, individuals of all ages were then used to establish three lists of ALFs (people) in the FSE, NMD and RoW Groups on 1st April 2004. These lists of ALFs therefore represent all the 'family' members resident in the RALFs on the 1st April 2004. Due to the anonymisation, we are not able to confirm that the co-resident individuals are in fact a 'family'; we are only able to associate

anonymous people (ALFs) to addresses (RALFs) by virtue of the fact that they were registered as living at the same residence on the same date. So, for example, an unrelated adult 'lodger' might change the assumed household structure from 'lone parent' to 'couple with children'. For the purposes of this document we will refer to the three Groups either as FSE, NMD and RoW or as 'follow-up Groups'.

- 3.6 For the purposes of this Project, the process described above in Section 3.3 was repeated for each year from 2004 to 2012. This provided data for five years of follow-up Groups prior to the official Flying Start start date of 1st April 2009, and four further years of follow-up Groups following the introduction of Flying Start.
- 3.7 The successive annual follow-up Groups generated by this methodology are made up of many of the same people e.g. the 2005 follow-up Groups contain a core set of people who also appear in the 2004 follow-up Groups, but a year older. The 2005 follow-up Groups do not contain people from 2004 if they have moved away or died, but do contain families from RALFs where new births occurred between 2nd April 2004 and 1st April 2005. People who have moved house between FSE, NMD and RoW RALFs will be included in whichever of these three Groups they fall into in each year's follow-up Groups. This method does, to some extent, deal with local migration, but it is based on an annual 'snapshot' of where people lived rather than the actual moving dates, so there is scope for further refinement. However, people who do not move very far (within a distance where it is convenient to remain registered with the same GP) may not think of informing the GP practice of their change of address, so that the data filtering through into SAIL may not be entirely up-to-date for local moves. Also, families receiving the benefits of Flying Start, who move to a new location a short distance away, may continue to experience the benefits, although the new address is officially non-eligible.

Note on Future Reporting

The RoW is made up of families that contained children aged less than 4 years and who lived in relatively less deprived areas of Wales. The fact that they are living in less deprived areas makes them less useful as a 'control' group when attempting to identify the effects of Flying Start. The Next Most Deprived (NMD) Group represents a better 'control' group. However, the expansion of Flying Start from 31st March 2013 will introduce the Flying Start benefits to many of the NMD families. After Flying Start is introduced to an expanded set of areas, many NMD families will become FSE and RoW will become the appropriate 'control' group.

- 3.8 In future years, re-running the analysis will include more recent data about the population of each follow-up Group, e.g. the children aged 3-4 years in the 2004 follow-up Groups will be 16 years old by 2016 and taking their GCSEs, the results of which will flow into the SAIL data set for comparison purposes.

Selection of indicators

- 3.9 The development of a multi-agency approach to Flying Start was considered key to the effectiveness of the programme and to be instrumental in supporting both children and families, enabling early identification, assessment and referral. Local Authorities in Wales were required to use the Flying Start funding to contribute to the delivery of a set of specific outcomes. Flying Start is designed to look at and improve the home environment as part of health visits, which in turn should manifest itself in e.g. reduced accidents in the home. Another strand is to encourage the use of health services where there is a real need as opposed to leaving the child suffer. Taken together, it is theorised that Flying Start may influence a range of health outcomes. In particular, the enhanced health visiting and access to parenting programmes were theorised by the Author to be likely to contribute to improved healthy behaviours, e.g. reduced smoking during pregnancy, and to improvements in the general health of the FSE population across a range of primary and secondary care indicators.

3.10 Potential health-related indicators for monitoring over time were chosen according to:

- Whether they matched a specific Flying Start indicator (see Table 3.1, below).
- Whether they were theorised to align with Flying Start outcomes e.g. 'Flying Start Children to be Healthy and Thriving', even if they did not match a specific Flying Start indicator, or related to information Local Authorities were expected to collect as part of Flying Start performance management arrangements³.
- Whether health events occurred in sufficient numbers (i.e. it would be difficult to demonstrate any significant change over time using events experienced by very small numbers of people). The project has utilised hospital data, which is an all Wales database within SAIL, and GP event data, which is based on a sample of participating GPs and covers around 47% of the population of Wales.

Appendix 5 describes in greater detail the indicators selected for the analysis. The list of Flying Start outcomes and indicators mapped against the Project indicators is shown in Table 3.1, below. The list of Project indicators is shown in Table 3.2, below. The findings are reported in Chapter 6.

³ As listed in guidance issued by Welsh Government to Local Authorities for termly Flying Start returns.

Table 3.1 Flying Start Outcomes and Indicators mapped against Indicators analysed for this Project

Flying Start Children are Healthy and Thriving	
Indicator 1: Percentage of Flying Start children reaching or exceeding their developmental milestones at 2 years .	Although this has been collected since 2012-13 on regular monitoring forms, individual level information was not available for linking when this Project was undertaken.
Indicator 2: Percentage of Flying Start children reaching or exceeding their developmental milestones at 3 years .	Although this has been collected since 2012-13 on regular monitoring forms, individual level information was not available for linking when this Project was undertaken.
Indicator 3: Percentage of children fully immunised at their fourth birthday	As we discuss in greater detail below, further development will be required to finalise this indicator; for this Project, the following are reported: Percentage of children receiving their MMR1 by 13 months of age and by 3.5 years of age Percentage of children receiving their MMR2 by 3.5 years of age Percentage of children receiving their first 5 in 1 vaccination by the age of 3 months
Indicator 4: Percentage of Flying Start children breastfed at stages up to 6 months.	For this Project, percentage of mothers breast feeding 'at birth' and 'at 8 weeks' have been analysed as the closest available indicator. From 2012, the new data definitions for breastfeeding have been introduced, which will lead to the 10 day and 6 week data being improved for future reporting years.
Flying Start Families are Capable and coping	
Indicator 1: Percentage of Flying Start Children presenting at A&E department as having unintentional injuries in the home	Hospital activity for babies aged <1 year: number of Injury and poisoning admissions. For future reporting years, injuries to children aged under 4 years can be reported.
Indicator 2: Percentage of Flying Start children who have an open case with Social Services as Child In	All-Wales data for Social Services is not currently in SAIL but a number of projects to improve linked data for social care are

Need.	underway, so that this indicator can be added for future reporting years.
Indicator 3: Percentage of Flying Start Children on the Child Protection Register	As for previous indicator, this can be added for future reporting years.
Flying Start Children are Reaching Potential	
Indicator 1: Percentage of eligible children taking up Flying Start Childcare and who are registered and started in a Flying Start setting.	Data is not currently available for this indicator.
Indicator 2: Proportion of Flying Start children registered to start Foundation Phase as measured by the number of 3 year old Flying Start children on school rolls.	This indicator relates to Foundation Phase, which replaced KS1 from 2011/12. Data will become available on Foundation Phase in future reporting years
Indicator 3: Percentage of Flying Start children achieving Step 3 or above entry assessment in Foundation Phase.	This indicator relates to Foundation Phase, which replaced KS1 from 2009. Data will become available on Foundation Phase in future reporting years. For this Project, the indicator relates to the previous Key Stage 1 indicator: Percentage of 7 year old children meeting the expectation that they would attain level 2 in each subject at KS1

Table 3.2 Summary of Indicators Reported for this Project

Source Dataset	Indicator name and description*
National Community Child Health Database	Maternal Smoking during pregnancy: Percentage of mothers by category of smoking status (non-smoker, current smoker, ex-smoker, unknown)
	Gestational age: Percentage of pre-term (less than 37 weeks duration) births
	* Teenage mothers: Percentage of teenage mothers (aged <20 years) at birth of first child
	Birth weight: Percentage of low birth weight babies (<2,500 grams)
	* Breast feeding at birth: Percentage of mothers reporting 'at birth' that they planned to breast feed
	* Breast Feeding at 8 weeks: Percentage of mothers recorded as breast feeding their babies at 8 weeks
	* Immunisations: Percentage of children receiving MMR1 by 13 months and 3.5 years of age
	* Immunisations: Percentage of children receiving their MMR2 by 3.5 years of age
	* Immunisations: Percentage of children receiving first 5 in 1 vaccination by 3 months of age
Welsh Demographic Service	Infant mortality rate: The number of babies who died before the age of one year per 1,000 live births**
Patient Episode Database Wales	Hospital activity indicators for infants aged <1 year: The number of: <ul style="list-style-type: none"> ▪ Admissions for All Causes ▪ Admissions for All Emergencies ▪ Infectious disease emergency admissions ▪ Respiratory condition emergency admissions ▪ *Injury and poisoning admissions <p style="text-align: right;">Continued overleaf ...</p> <ul style="list-style-type: none"> ▪ Admissions for 'signs and symptoms (not otherwise classified)'

GP Practice Data	<p>Interactions with primary care for children aged 0-12 years</p> <p>The number of interactions for:</p> <ul style="list-style-type: none"> ▪ Respiratory conditions ▪ Infectious diseases ▪ *Injuries and poisonings in children aged 0-12 years ▪ Preventive procedures⁴
National Pupil Database	<p>Attainment at Key Stage 1:</p> <p>* Percentage of 7 year old children meeting the expectation that they would attain level 2 in each subject at KS1</p>

* items relate to specific Flying Start Indicators or performance management targets.

** based on WDS death record, not verified using ONS death notification.

3.11 For some of the indicators listed in Table 3.1, relatively small changes were found between years. Other indicators display a high level of volatility over time because the numbers of events taking place in any single year are very small. To help in identifying whether there has been a significant change over time and whether it is associated with the introduction of Flying Start, where small changes or small numbers were found, the average percentages are compared for the two sets of grouped years 2004-08 and 2009-12.

3.12 Areas were not randomly assigned to the eligible (FSE) and control (NDM) groups. For the purposes of evaluation, the process of randomising areas (or individuals, in some studies) into either an intervention group or a control group is usually done in order to ensure, to a high degree of confidence, that there are no systematic differences between the groups in terms of both observed and unobserved socio-demographic characteristics; the one difference between the groups would be receipt of the intervention, making it a relatively simple matter to identify any intervention effect. In the absence of randomisation, it was to be expected that both observable and unobservable differences in characteristics would exist between the two groups.

3.13 The key difference we expected to see was that the FSE areas were selected to be the most deprived and the NDM areas the next most deprived community school catchment areas within each LA. At point of writing, no individual-level deprivation

⁴ "Preventative Procedures" are routine examinations that can be carried out on children as they grow and develop e.g. head circumference, breast feeding, sensory and motor control testing, carried out at specified ages.

information and relatively little information about individual socio-demographic characteristics was available in SAIL⁵. Nevertheless, analysis was undertaken to compare the Groups in terms of area deprivation (WIMD tenth) and their age and gender profile as the best available indicator of whether systematic differences existed between the two groups.

- 3.14 Where more detailed analysis of the differences between the two Groups cannot be completed, care must be taken when assessing whether any changes observed represent an effect of the intervention. Due to the limited project scope, the lessons learned about data quality, the developmental nature of the methods and the limitations outlined above, **all of the findings presented in this Report should be viewed with caution.**
- 3.15 Work is continually underway at SAIL to improve the general understanding of the data it receives. With such large datasets it is impossible to entirely quality assure every aspect of a data set prior to its use in research. Even where the data are perfectly clean, the Project has been experimental in nature and strictly circumscribed in scope and, had time permitted, would have benefited from additional testing of the extent to which findings were sensitive to the analytical assumptions made.

⁵ Through the creation of the Wales Administrative Research Centre, it is hoped to gain access to anonymised individual-level benefits and employment information.

4 Address matching and defining populations

Address Matching results

- 4.1 This section describes the results of the address-matching process and the demographic characteristics of the resulting three Groups of anonymous families in Wales containing children aged less than 4 years, using 2009 as an example year.
- 4.2 Choosing the populations from the anonymised addresses was a two stage process, starting with defining the full set of RALFs in the Flying Start Eligible, Next Most Deprived and Rest of Wales areas. From the full sets, the subsets of RALFs where eligible children (aged less than 4 years on target dates) were living were established.
- 4.3 Anonymisation of the FSE addresses by NWIS provided SAIL with 139,137 (99%) matched addresses and 1,215 (1%) non-matched addresses. This means that 99% of the Flying Start eligible addresses provided by WG had been successfully supplied with a RALF to allow linkage to other address-level data in SAIL. However, a further 1.6% (2,000 households) did not have any people living in them who were registered in the WDS data set. This issue remains unexplained and further investigation is underway with NWIS.
- 4.4 The Next Most Deprived (control) Group was defined by WG as a list of LSOAs. All RALFs in these LSOAs were selected from the anonymised WDS address dataset held by SAIL, so did not involve NWIS. The “LSOA to RALF” linkage provided a set of 131,408 RALFs, based on postcodes that were current at 1st April 2009. The RALF to ALF linkage then reduced the list of RALFs to include just those homes where eligible families were residing, based on the presence of at least one child aged less than 4 years on the 1st April of each year of the Project.
- 4.5 Comparison of FSE and NMD RALFs indicated an overlap, with 67 RALFs appearing in both FSE and NMD lists. The overlapping RALFs were from two LSOA areas, which indicates that they were concentrated in two geographical areas. Further investigation of this has not been carried out but is included as a suggestion for further work in Appendix 4. For the purposes of the Project, the 67 overlapping RALFs were left in the FSE list and removed from the NMD list. The total numbers

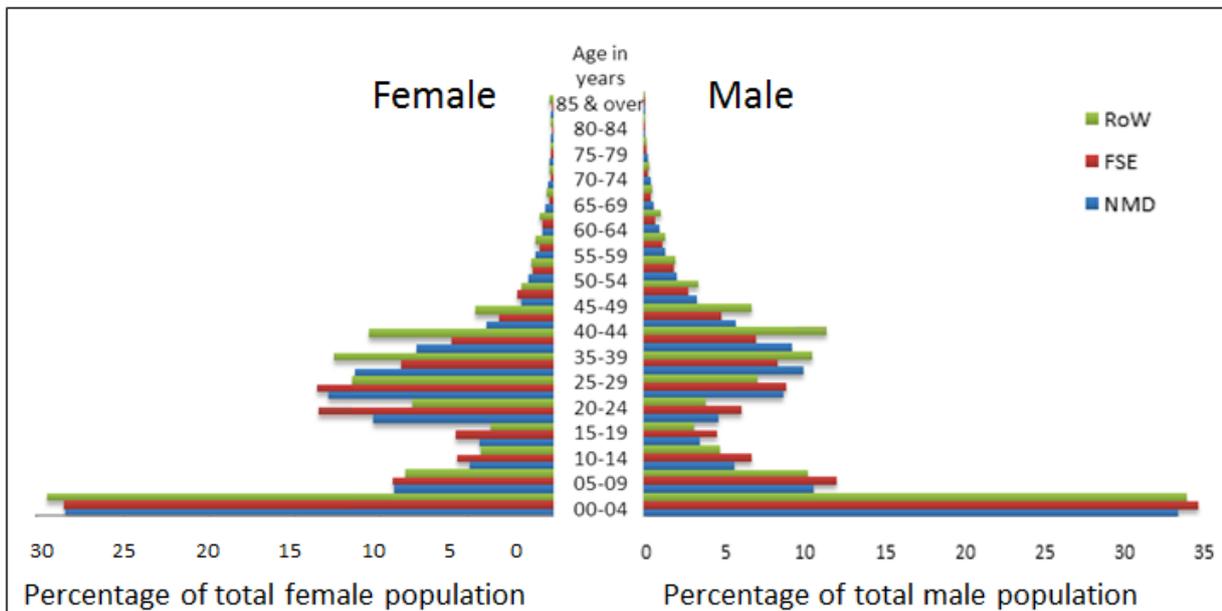
of RALFs from which the populations were selected was 139,039 for FSE, 131,408 for NMD, which left 1,012,967 for the rest of Wales.

- 4.6 As part of the quality assurance for the analysis, the geographical distribution of the RALFs was calculated by counting the number of RALFs per LSOA and aggregating these up to Local Authority level. The resulting distribution is included in Appendix 6. The analysis confirmed that all Local Authority areas in Wales are represented in each of the three populations. Across Wales as a whole, similar numbers of RALFs are included in the selection of NMD as is the case for the FSE.
- 4.7 Subsets of these RALFs were then selected for each year from 2004 onwards. The full population (all ages) living in the selected eligible subsets of RALFs is included in each of the follow-up Groups.

Characteristics of the Project Groups

- 4.8 Population subgroups were defined for the years 2004 through to 2012 for comparison purposes. The 2004-08 subgroups will help to establish the baseline – how things were prior to the introduction of Flying Start. Figure 4.1, below, shows a ‘population pyramid’ for 2009 by age group for the FSE, NMD and RoW Groups. As we would expect, the three populations all have a very young population structure with over 30% of both males and females aged less than 5 years.

Figure 4.1 Population Pyramid for Families with Children Aged Less Than 4 Years by Project Group



4.9 All three population Groups contain large percentages of women aged 20-44 large percentages of men aged 25-44, small percentages of teenagers and relatively small percentages of people aged over 44 years. This is consistent with what we would expect for households containing a child aged less than 4 years. As we would also expect, all three Groups have similar proportions of children aged 0-5 years. However, the FSE Group contains a higher proportion of people aged 5 to 29 years compared with both the NMD and RoWs Groups, and smaller proportion of people aged 30 years and above. The FSE Group is therefore a ‘younger’ population compared with the RoW. This is not surprising, when the RoW, by definition, is made up of the least deprived households that nevertheless contain children aged less than 4 years, since we know that the average age at which women have children tends to increase as socio-economic status increases.

4.10 Overall, there are higher proportions of women than men in all three population Groups. Table 4.1a, below, shows the overall numbers of people assigned to population Groups for the 2009 follow-up Groups with their relative percentages. This shows that there are 10% more females than males aged over 15 years in the NMD and RoW populations, and 15% more females than males over the age of 15 years in the FSE population. This may indicate a higher proportion of lone-parent

females in the FSE Group for 2009. Similar proportions occur in all the annual follow-up Groups from 2004 to 2012. The size of the 2009 follow-up Group shows that the NMD Group has the smallest population, with just over 14,100 fewer people than the FSE Group. Figure 4.1b, below, shows the total population of Wales divided using the Project Group definitions and into the same population groupings. These are the total populations from which families with children aged 0-4 have been drawn.

Table 4.1a Total population of Families with Children Aged Less Than 4 years in 2009, split by 0-14 year olds and those aged 15 years and older

Age: 0-14 years	FSE	NMD	RoW	FSE	NMD	RoW
Girls	15,031	11,229	75,804	48%	49%	49%
Boys	15,994	11,889	80,343	52%	51%	51%
Children	31,025	23,118	156,147	100%	100%	100%
Age: 15 and over						
Women	19,635	15,302	103,980	58%	55%	55%
Men	14,423	12,499	86,210	42%	45%	45%
People aged 15 and over	34,058	27,801	190,190	100%	100%	100%

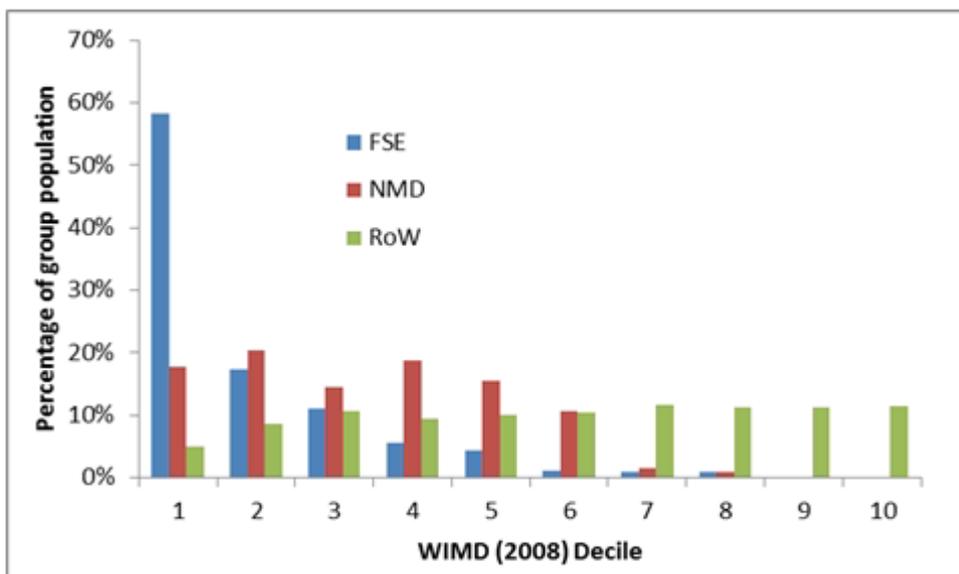
Table 4.1b Total population of Wales in 2009 by Project Group, split by 0-14 year olds and those aged 15 years and older

Age: 0-14 years	FSE	NMD	RoW	FSE	NMD	RoW
Girls	32,278	26,671	194,046	49%	49%	49%
Boys	34,137	28,085	204,613	51%	51%	51%
Children	66,415	54,756	398,659	100%	100%	100%
Age: 15 and over						
Women	130,431	135,940	1,072,061	50%	50%	51%
Men	129,336	135,035	1,050,052	50%	50%	49%
People aged 15 and over	259,767	270,975	2,122,113	100%	100%	100%

4.11 The FSE Group (i.e. households containing a child aged less than 4 years) represents 2.1% of the total population of Wales, the NMD Group represents about 1.6% and the RoW Group 10.9% of the total population of Wales. Households with children aged less than four years therefore make up about 15% of the total population of Wales (see Table 4.1a, above). The FSE families with children aged under 4 represent 20% of the total population of the FSE areas, the NMD group represents around 16% of the total NMD area population, and the RoW group represents 14% of the population making up the remainder of Wales (see Figure 4.1b, above).

4.12 The FSE, NMD and RoW Groups were compared in terms of the Welsh Index of Multiple Deprivation (WIMD). The 2008 version of WIMD was used because this was the closest to the time when the Flying Start programme started in 2009. Although more detailed analysis could be undertaken to compare the Groups in terms of the individual WIMD domains, this was not possible within the limited scope of a demonstration project. The LSOAs of Wales were ranked by overall WIMD score and divided into equal tenths of population⁶. Figure 4.2, below, shows the proportion of the FSE, NMD and RoW Group populations that fall into each of these tenths. The data presented are for 2009, but a similar distribution is seen in the Groups from 2004 to 2012.

Figure 4.2 Total population of Families with Children Aged Less Than 4 years in 2009 by tenth of WIMD Score



4.13 As expected, a large proportion (58%) of the FSE Group is from the most deprived tenth of Wales, and 87% of the FSE population lies within the three most deprived tenths. The NMD population is more evenly distributed across the more deprived tenths, with only 18% in the most deprived tenth and just over half of its population (53%) in the three most deprived tenths. The RoW has only a 5% representation in the most deprived tenth and 24% in the three most deprived tenths. This does

⁶ The alternative way to split the Groups would be by tenths of deprivation score, which, because of the distribution of scores would give a very small first and tenth decile of populations and make interpretation difficult. SAIL holds the tenths as equal sized, 189-190 LSOA groups (i.e. each with a similar population size).

indicate that the NMD are a more deprived group than the RoW, but that it is a significantly less deprived Group than the FSE Group.

- 4.14 Figure 4.2 illustrates that the selection of the FSE Group included some people living in areas classified as less deprived according to WIMD, and that some families living in areas classed as being among the most deprived in Wales are not included in the FSE Group. Area based indicators such as WIMD assign the area characteristics to everyone living in that area, which is an over-simplification of the situation on the ground. In future studies, particularly if we were able to flag the actual families in receipt of Flying Start interventions, could also attempt to include some individual or household indication of deprivation. Such studies would give a much better idea of the true magnitude of the differences between the Groups, and therefore the effectiveness of the interventions.
- 4.15 This chapter has demonstrated the selection of a specific subsection of the population of Wales (families with children aged less than 4 years). Three Project-specific comparator Groups have been established based on anonymised addresses. The Groups were well distributed geographically across Wales, and had very similar population profiles in terms of age and gender. Figure 4.2 illustrates that the NMD, whilst being a more deprived group than the RoW, is somewhat less deprived than the FSE Group. So, whilst the NMD is the best comparator group available, it should be noted that it is not the ideal comparator Group.

5 Findings

5.1 The objectives of the Project were to demonstrate that linked, anonymised, routinely collected data can:

- provide a retrospective baseline against which the impacts of an intervention can be measured;
- be used to identify the impacts of an intervention by:
 - demonstrating differences amongst population subgroups; and
 - demonstrating changes or trends in indicators over time.

This chapter compares the Flying Start Eligible, the Next Most Deprived, and the Rest of Wales population Groups in terms of the indicators listed in Table 3.1, above. Findings relate only to the specific, longitudinal Project Groups (or 'cohorts') created as described in Chapter 3, so are not designed to serve as a report on the health of the population of Wales and will not be comparable to whole-population estimates of e.g. health service use.

5.2 As discussed in Chapter 3, the analysis is longitudinal so we expect to see change over time in each indicator in the absence of any intervention. What we are looking for in terms of identifying an impact of Flying Start, is a, 'excess' change over time in the Flying Start Eligible (FSE) Group over and above any change that is observed to be happening anyway. In order to 'control' for the underlying change over time, we are using the Next Most Deprived (NMD) Group as our 'control' group. Since the NMD Group is, as discussed above in Chapter 4, slightly less deprived than the FSE Group, we expect to see relatively poorer findings for the FSE Group than for the NMD Group before the introduction of Flying Start. Since the majority of Flying Start services were being delivered by 2009, we are looking for an improvement in the indicators for the FSE Group that is greater than any improvement observed in the NMD Group and that begins in or after 2009.

5.3 Interpretation is complicated by two further factors. Firstly, it is possible that the rate of any underlying population improvement may be greater in areas where there is greater scope for improvement, accounting for some of any 'excess' improvement seen in the FSE Group. Secondly, Flying Start may have the effect of supporting parents to make better use of health services e.g. to consult a health professional rather than leaving symptoms to improve on their own. Where some of the

indicators relate to the number of consultations for various conditions, this may mean that an increase in consultations is a positive impact of Flying Start. Where we might expect to see a mixture of increasing consultations due to better use of health services and decreasing consultations due to improvements in health, it may therefore be more difficult to identify a meaningful impact. However, any trends that are identified will nevertheless be meaningful.

- 5.4 The findings relate to two kinds of data. Firstly, to administrative health records relating to the whole population of Wales and secondly to administrative data on GP events, which relates, as discussed in Chapter 3, to about half the population of Wales.
- 5.5 We checked whether any ‘excess’ changes we did observe were big enough for us to be confident they did not happen purely by chance – such changes are referred to as ‘statistically significant’ or ‘significant’⁷. Other findings where a consistent effect over time was observed, but where that effect was not big enough to reach the level of statistical significance, are nevertheless worthy of note so are reported but using the phrase ‘the data suggests’. The analysis compared the average rate for each indicator in the FSE Group for the period before the introduction of Flying Start (2004-2008) with the average rate for the period after Flying Start began (2009-2012), adjusting for any underlying change in population health using the NMD Group.
- 5.6 As discussed in Chapter 3, some indicators suffer from small numbers and for others relatively small changes over time make it difficult to demonstrate a statistically significant difference that can be attributed to Flying Start. In addition, where this Project has used data about eligibility for, rather than receipt of, enhanced Flying Start services, it is likely that any true effect of Flying Start is being diluted by the presence in the analysis of those who were eligible but who did not receive services.

⁷ The significance testing used a 95% confidence interval, so we can be confident that such findings would only occur 1 in 20 times by chance alone.

5.7 Table 5.1, below, summarises the findings. As noted above, due to the experimental nature of the methods, the limited scope of the project, **all findings must be viewed with caution.**

Table 5.1 Summary of Findings

Indicator	Finding
Maternal Smoking in pregnancy	Data was only 25% complete but the proportion of women giving up smoking during pregnancy increased significantly in the FSE Group after Flying Start. However the proportion who reported smoking also increased significantly in the FSE Group, suggesting that these findings may be at least partly explained by improved recording of smoking status. No significant effect of Flying Start can therefore be identified.
Percentage of pre-term births	The percentage of pre-term births: the data indicates a significant decrease in pre-term births in the FSE Group after the introduction of Flying Start.
Percentage of teenage mothers (aged <20 years) at birth of first child	The data indicates a significant decrease in first births to teenage mothers in the FSE Group after Flying Start. Although the decrease appeared to begin in 2008, a continuing trend downward from 2009 suggests some form of Flying Start effect.
Percentage of low birth weight babies (<2,500 grams)	A statistically significant reduction was observed in the number of low birth weight babies born to the FSE Group after 2009.
Breast feeding 'at birth'	For both indicators, the data show a small increase in all Groups so no increase can be attributed to Flying Start. As with smoking, the level of recording of breast feeding also increased over time and may partly explain the rising rates.
Breast Feeding at 8 weeks:	
Infant mortality	No conclusive finding due to small numbers.
Immunisation - % receiving '5 in 1' by the age of 3 months	No observed impact of Flying Start.
Immunisation - % receiving first MMR: <ul style="list-style-type: none"> • before 13 months • before 3.5 years of age 	<ul style="list-style-type: none"> • No observed impact of Flying Start; • No observed impact of Flying Start;
Immunisations – percentage of children receiving their second MMR by 3.5 years of age	No Observed impact of Flying Start.

<p>Hospital activity for babies aged <1 year: number of:</p> <ul style="list-style-type: none"> • Admissions for All Emergencies • Infectious disease emergency admissions • Respiratory condition emergency admissions • Injury and poisoning admissions • Admissions for 'signs and symptoms (not otherwise classified)' 	<ul style="list-style-type: none"> • A small overall decrease was found in all Groups, so when adjusting for the decrease in NMD it was not possible to identify any noteworthy change in the FSE Group. • No clear pattern, but, the figures for 2012 show a reduction in the FSE Group not seen in the other two Groups. Further years of data would be required to establish if this is the beginning of a significant downward trend. • No clear pattern. • The data show a statistically significant decrease in the FSE Group. • Due to small numbers findings are inconclusive.
<p>Primary care interactions for children 0-12 years: number for:</p> <ul style="list-style-type: none"> • Respiratory conditions • Infectious diseases • Injuries and poisonings • Preventive procedures 	<p>The data is likely to be biased, including data for fewer GP practices in more deprived areas. Nevertheless:</p> <ul style="list-style-type: none"> • there was a statistically significant increase in consultations in FSE compared with NMD and RoW; this may partly be explained by parents seeking help for conditions that might previously have gone untreated; • no clear pattern; • the data suggests a slightly greater decrease between 2004-08 and 2009-12 in the FSE Group compared with the NMD Group; • the data suggest that the number of consultations for the FSE Group increased after 2010 to bring them more in line with the numbers seen in the NMD Group.
<p>Percentage of 7 year old children meeting the expectation that they would attain level 2 in each subject at KS1</p>	<p>Data was only available for 2004-08 so it was only possible to make baseline comparisons between the three Groups. Fewer children in the FSE Group attained the expected level at KS1 compared with NMD and RoW.</p>

Indicators from the National Community Child Health Database (NCCHD)

5.8 These indicators relate to data collected from a number of sources during the course of pregnancy, child birth and early years. As such there is some variability in the data quality, in particular, the electronic capture of data for items recorded outside a health service establishment, such as those collected by midwives working in the Community.

Maternal Smoking in Pregnancy

- 5.9 Maternal Smoking is based on a question asked by Community Midwives at the 'initial assessment' of a newly pregnant woman. Around 75% of the records do not have a recorded value (for further discussion, see Appendix 5). Because the data on smoking in pregnancy was only complete for around 25% of cases, we have no way of knowing whether, for example, smoking status was better recorded if women were smokers or ex-smokers so that all the 'missing' cases are more likely to be non-smokers. The analysis for this Section is based on all births (i.e. including those cases for whom the data is missing); this is because by selecting only those cases for whom smoking status was recorded, the graph would be telling us more about the differences in 'missingness' than about smoking status. However, the key point for this indicator is that, when this Project was undertaken, the data quality was not sufficient to allow us to be confident in making conclusions about any real differences.
- 5.10 Figures 5.1 to 5.3, below, do show differences between the three Groups in the percentage of births where the mother was recorded as having been a smoker and a non-smoker. The percentage of FSE mothers who reported smoking during pregnancy was consistently twice that for the RoW, and the percentage in the NMD Group was consistently 2% to 3% higher than in the rest of Wales (see Figure 5.1). On average, there were 9% fewer non-smokers in the FSE Group than in the RoW Group and 4% fewer non-smokers in the NMD Group compared with the RoW Group (see Figure 5.2). This is as we would expect based on the relative deprivation of these population Groups and what is known about smoking in Wales.
- 5.11 The percentage of births where the mother reported giving up smoking during the pregnancy shows quite similar rates over time in the three Groups (see Figure 5.3). The percentage in the FSE Group was lower than in the other two Groups in 2004, but by 2008 the rate was the same in the FSE Group as in the Rest of Wales. The rate was higher in the FSE Group than either of the other two Groups for all years since 2009, i.e. had increased significantly in the FSE Group after the introduction of Flying Start. However, since there is an increase over time in all three smoking status categories, i.e. in smoking *and* non-smoking *and* giving up smoking, some of

this effect is likely to be explained by improved recording, so no significant effect of Flying Start can therefore be identified.

Figure 5.1 Percentage of babies whose mother reported smoking during pregnancy

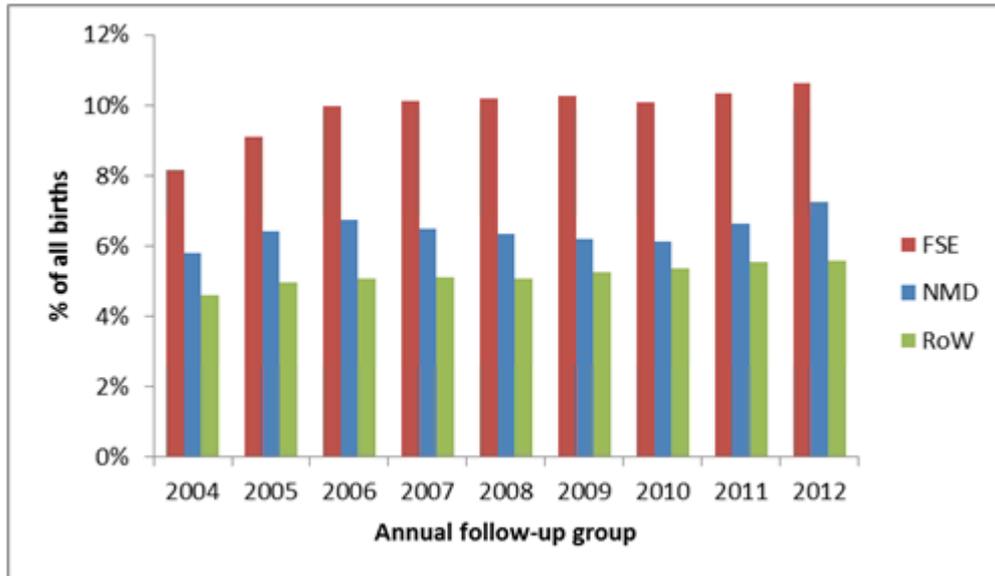


Figure 5.2 Percentage of babies whose mother reported they were a non-smoker

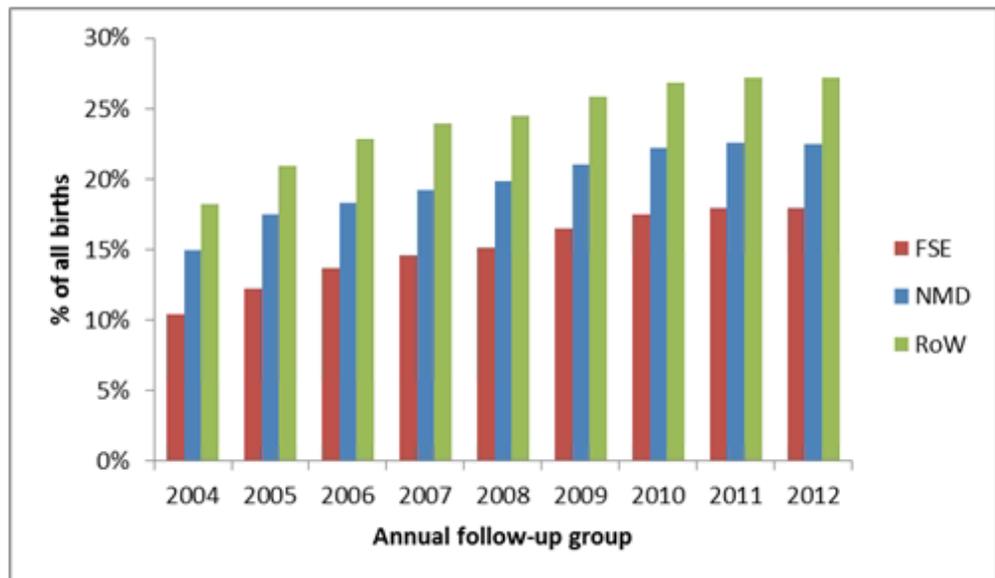
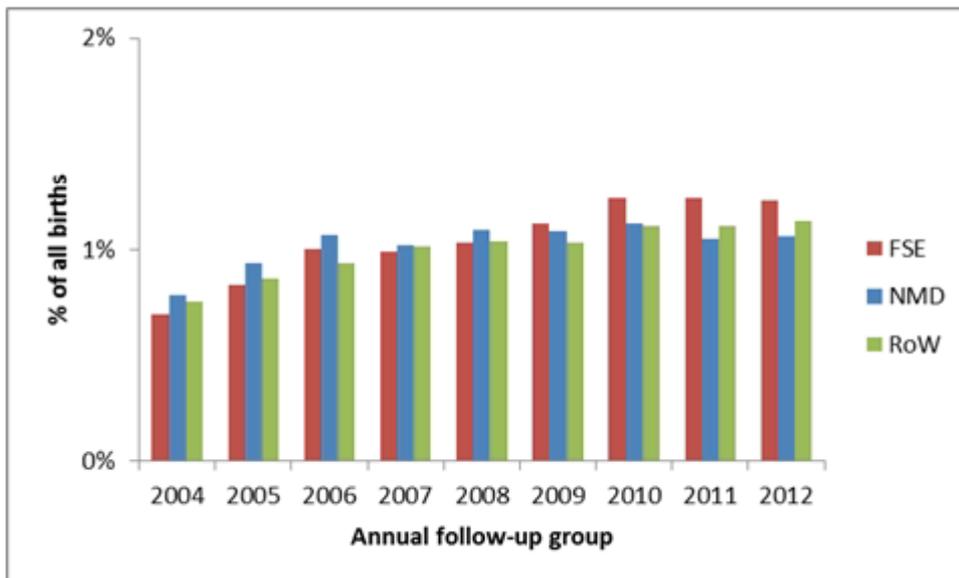


Figure 5.3 Percentage of Babies whose Mother Reported Giving Up Smoking During Pregnancy by Project Group

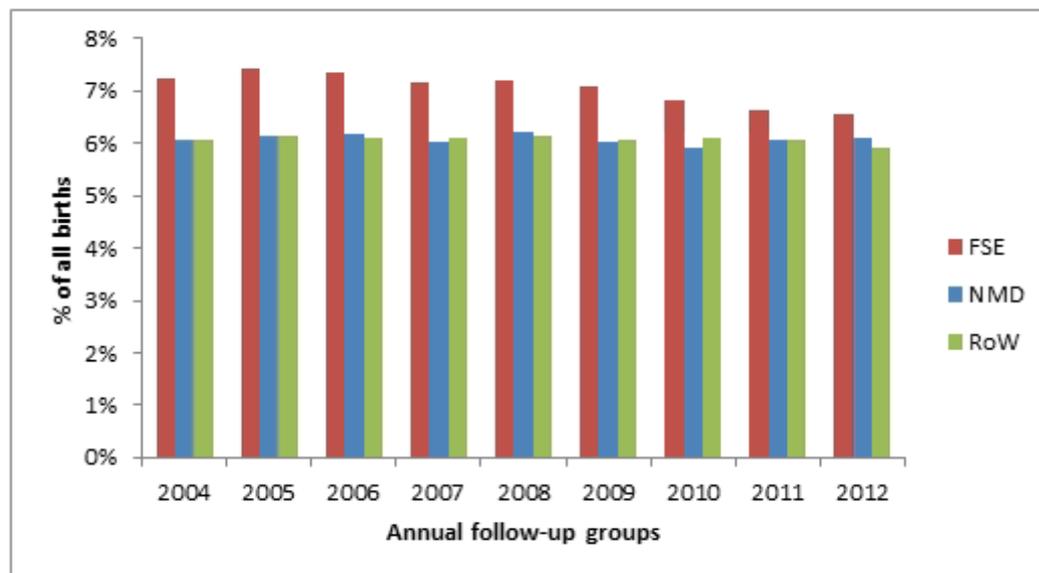


Pre-term Births

5.12 The annual percentage of births occurring at a gestational age of less than 37 weeks for the three Groups over the period 2004 to 2012 is shown in Figure 5.4, below⁸. On average over the whole period around 1 percentage point more pre-term babies were born in the FSE Group than in the NMD and RoW Groups. Comparing the average percentages for the years 2004-08 with the years 2009-12, there is a larger reduction (0.5 percentage point) for the FSE Group compared with both the NMD and the RoW Groups (both show a 0.1 percentage point reduction). Using the reduction in the NMD Group to allow for underlying change in the population, the data indicates a significant decrease in pre-term births in the FSE Group from 2009 onwards.

⁸ These percentages are based on average figures of 227 pre-term births per annum in NMD, 317 in FSE and 1520 in the RoW.

Figure 5.4 Percentage of Babies Born at a Gestational Age of <37 Weeks by Project Group



Teenage Mothers

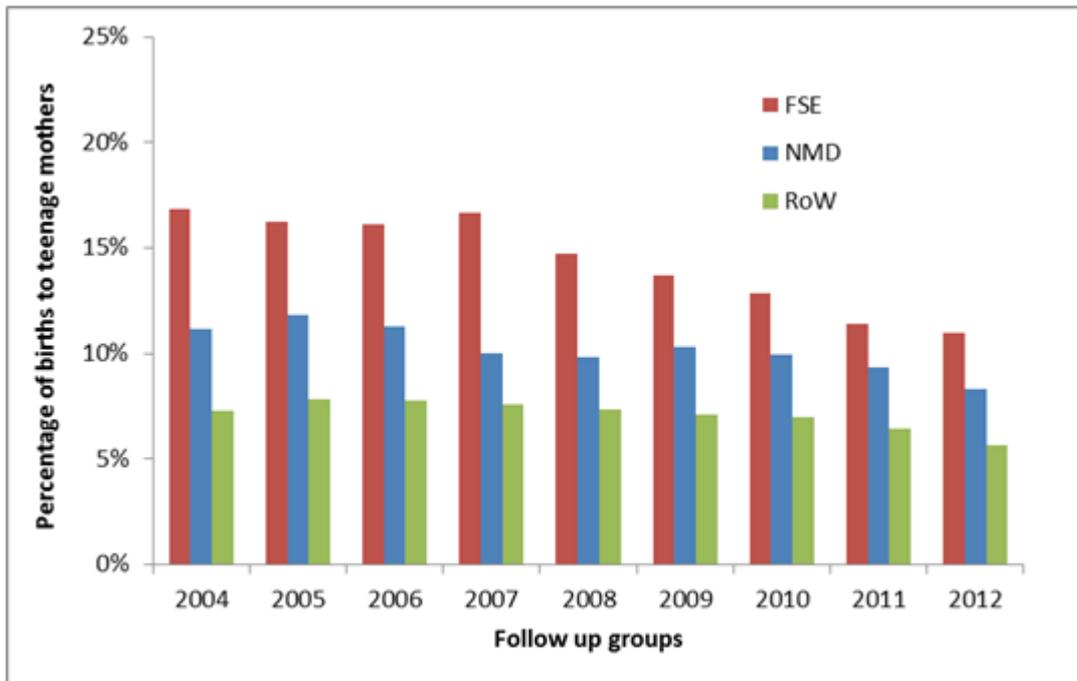
5.13 The percentage of teenage mothers (aged less than 20 years) is shown in Figure 5.5, below. The graph shows that for 2004, the percentage of births to teenage mothers in the FSE Group was more than double that for the RoW Group, with the NMD Group lying midway between the two. The percentage has steadily reduced in all three Groups since 2009. The drop in births to teenage mothers when comparing the average for 2004-08 and for 2009-12 was 5.3% for the FSE Group, 1.7% for the NMD Group and 1.1% in the RoW Group.

5.14 There are two reasons for these changes. The number of births per annum was rising by about 2% per year (3% in FSE and NMD), and the number of teenage mothers was falling in Wales (by about 0.1% per year for NMD and RoW and 0.4% for FSE). Nevertheless, the data indicates a significant decrease in first births to teenage mothers in the FSE Group after the introduction of Flying Start. Although the decrease appeared to begin in 2008, a continuing trend downward from 2009 suggests there may be some form of Flying Start effect. However, it should be noted that teenage conceptions have also been falling in poor communities in England where there was no Flying Start⁹, so further analysis, possibly using further years of

⁹ <http://www.ons.gov.uk/ons/rel/vsob1/conception-statistics--england-and-wales/2011/index.html>

data, would be required to examine the extent to which any Flying Start effect is being seen.

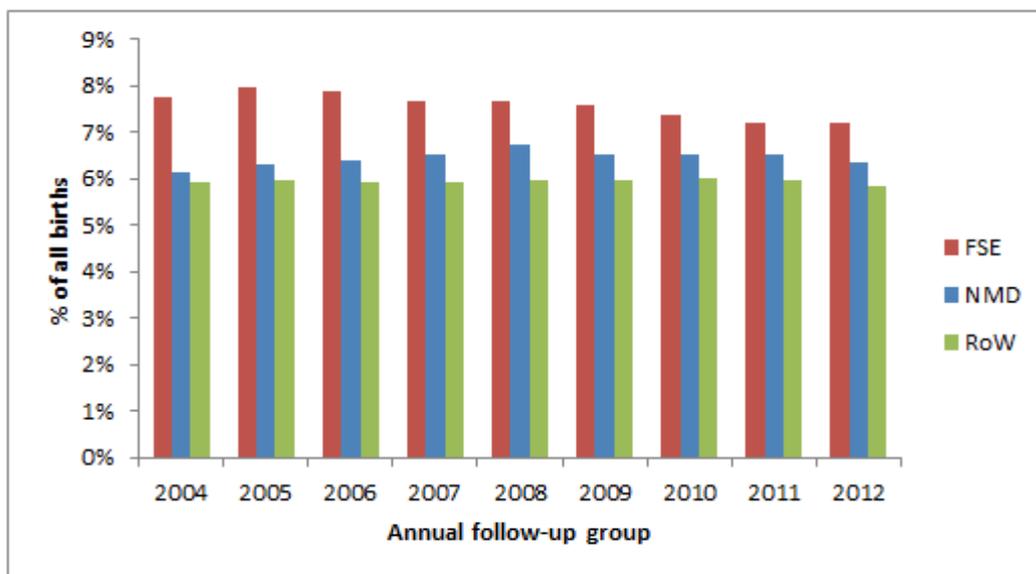
Figure 5.5 Percentage of Births to Teenage Mothers by Project Group



Low Birth Weight

5.15 The percentage of low birth weight babies born for each Group over time is shown in Figure 5.6, below. In the years 2004 to 2012 the Group with the highest percentage of low birth weight babies is consistently FSE, followed by NMD and then the RoW. It should also be noted that around 4-8% of birth weight records are not complete (see Appendix 5 for further information); if Flying Start is driving up recording completeness in the FSE Group, and most births are of normal weight, the improved recording might produce a significant drop in low birth weight babies entirely explained by improved recording. Including further years into this analysis in the future will establish if there is a true drop in low birth weight babies.

Figure 5.6 Percentage of Babies with Low Birth Weight (<2500 grams) by Project Group



5.16 As noted above, we are analysing relatively small percentage changes so in order to help in identifying a trend, the average percentage has been calculated for the years 2004-08 and 2009-12. For the FSE Group, the average percentage for the period 2009-12 was 0.4 percentage points lower than for the period 2004-08. A corresponding reduction is not seen in the NMD and RoW Groups. Allowing for the underlying population change seen in the NMD Group, a statistically significant reduction was observed in the number of low birth weight babies born to the FSE Group after 2009.

Breast Feeding 'at Birth'

5.17 The percentage of babies whose mothers reported breast feeding 'at birth' was lower in the FSE Group than the NMD Group and both were lower than in the RoW Group across the whole Project period. In all three Groups there was a small increase in the average breast feeding rate in the period 2009-12 compared with 2004-08 (see Table 5.2, below) However, there were also increases in the percentages of births where the mother was NOT breast feeding - these rates have increased by 2-3% for all the Groups between the two periods. The reduction in 'not known' or missing values in the data suggests that the changes observed may be entirely explained by better data recording in the later years rather than by any real change in breastfeeding rates.

Table 5.2 Percentage Point Changes in Breast Feeding at Birth Between 2004-08 and 2009-12 by Project Group

	FSE			NMD			RoW		
	No	Yes	Not known	No	Yes	Not known	No	Yes	Not known
	%	%	%	%	%	%	%	%	%
Average rate 2004-08	50	32	18	41	42	17	34	46	20
Average rate 2009-12	52	35	13	43	47	9	37	52	11
Percentage point difference	2	3	-5	3	5	-8	3	6	-9

Breast Feeding: 'at 8 weeks'

5.18 Breast feeding at 8 weeks is shown in Table 5.3, below. For all three Groups, the average percentage of women breastfeeding before Flying Start (2004-08) was slightly lower than the average following the introduction of Flying Start. The proportion of missing data is also consistently higher than for 'breast feeding at birth' and the proportion of missing data has decreased for the same two time periods. So, as with breastfeeding at birth, we can't be sure whether the change is real or due to better recording and will certainly be partly explained by better recording rather than real changes in breastfeeding practices.

Table 5.3 Percentage Point Changes in Breast Feeding at 8 Weeks between 2004-08 and 2009-12 by Project Group

	FSE			NMD			RoW		
	No	Yes	Not known	No	Yes	Not known	No	Yes	Not known
Average rate 2004-08	45	14	41	45	19	35	37	21	42
Average rate 2009-12	51	15	34	51	22	27	44	23	33
Percentage point difference	5	2	-7	6	2	-9	7	2	-8

Infant Mortality

5.19 Based on both a birth and a date of death having been recorded in the WDS dataset (i.e. this has not been verified using the official ONS notification), the number of deaths in babies aged less than 1 year per 1000 live births was calculated. This

calculation is commonly carried out in public health using ONS births data but WDS has been used because it is the data set that has been chosen to measure the population for the purposes of this Project.

5.20 As the number of infant deaths is (thankfully) small (averaging 9 children per year in Wales for the period 2004-12), a comparison has been made between figures for the 5-year period of 2004-08 and the 4-year period of 2009-12. Deaths were higher in the FSE Group than in the other two Groups. Although there was a small reduction of 0.14 deaths per 1,000 live births in the FSE Group in the period after the introduction of Flying Start, the absolute numbers are very small and this issue is complicated by the issue of late death reporting¹⁰, so it is not possible to demonstrate any effect of Flying Start on Infant Mortality.

Immunisation

5.21 Vaccines are administered on the basis of a rolling programme of different vaccine courses over time. Which course of vaccine a child received will depend on the timing of their vaccination and which programmes were being rolled out at the time (e.g. some children will have received their 'first' MMR vaccination from the course of treatment designated for 'second' MMR treatment). For each child, the 'course' of vaccine e.g. MMR, is recorded against the immunisation date, rather than the specific jab e.g. MMR1 or MMR2. This introduces uncertainty into the calculation of immunisation rates by age, as described below. However, the evidence suggests that uptake of immunisations is lower in more deprived areas of Wales¹¹.

5.22 For this Project, the ages by which vaccine uptake has been analysed were chosen as within the first year of life as a national target, and as 3.5 years as an approximation for pre-school school status, on the basis that Flying Start objectives are about enhancing 'preparedness for school'. For future reporting years, immunisation uptake can be reported by single year of age and for all Flying Start Eligible children (i.e. all children ages under 4 years).

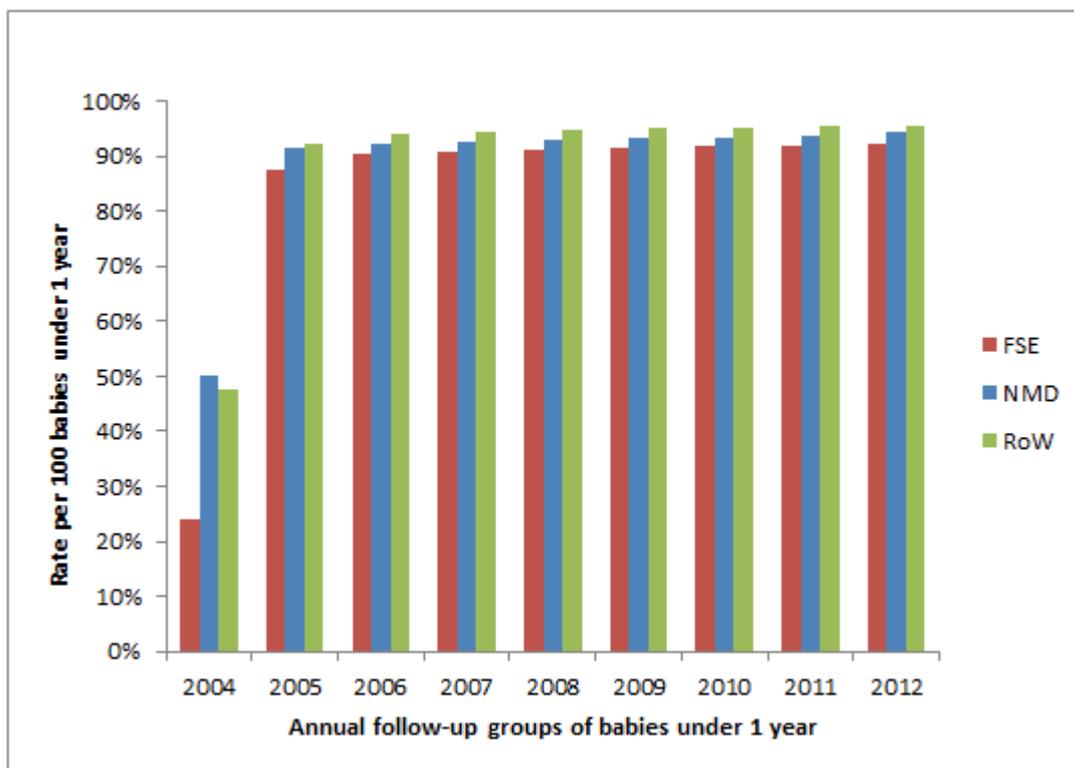
¹⁰ Deaths involving coroner's inquest are not reported in ONS mortality records until after the coroner's verdict. This can delay death registrations by up to two years.

¹¹ <http://www.wales.nhs.uk/sites3/page.cfm?orgid=457&pid=54144>; see Annual Reports, 2012-13 Figure 8

Diphtheria, Tetanus, Pertussis, Polio and H.I.B. ('5 in 1') Vaccine

5.23 The '5 in 1' vaccine protects babies from Diphtheria, Tetanus, Pertussis, Polio and Haemophilus influenzae type B (HIB). It is one of the first vaccinations that babies receive. It is administered in three doses, at 2, 3, and 4 months. We have included babies in this analysis according to the earliest recorded date for a vaccination, and calculated their age on the date of immunisation, for the '5 in 1' course that was introduced in 2004. Figure 5.7, below, shows the proportion of babies in each Group that had received this vaccination before the age of 3 months (or 90 days). The rates are very high (e.g. for 2012, 94%, 92% and 96% respectively in the NMD, FSE and RoW Groups). In 2004, the rates are low because the '5 in 1' was introduced part way through the year, so some babies were treated using the previous vaccination. The graph shows a consistent difference of around 4% between FSE and the RoW from 2005 onwards. There is no observed impact of Flying Start on this indicator.

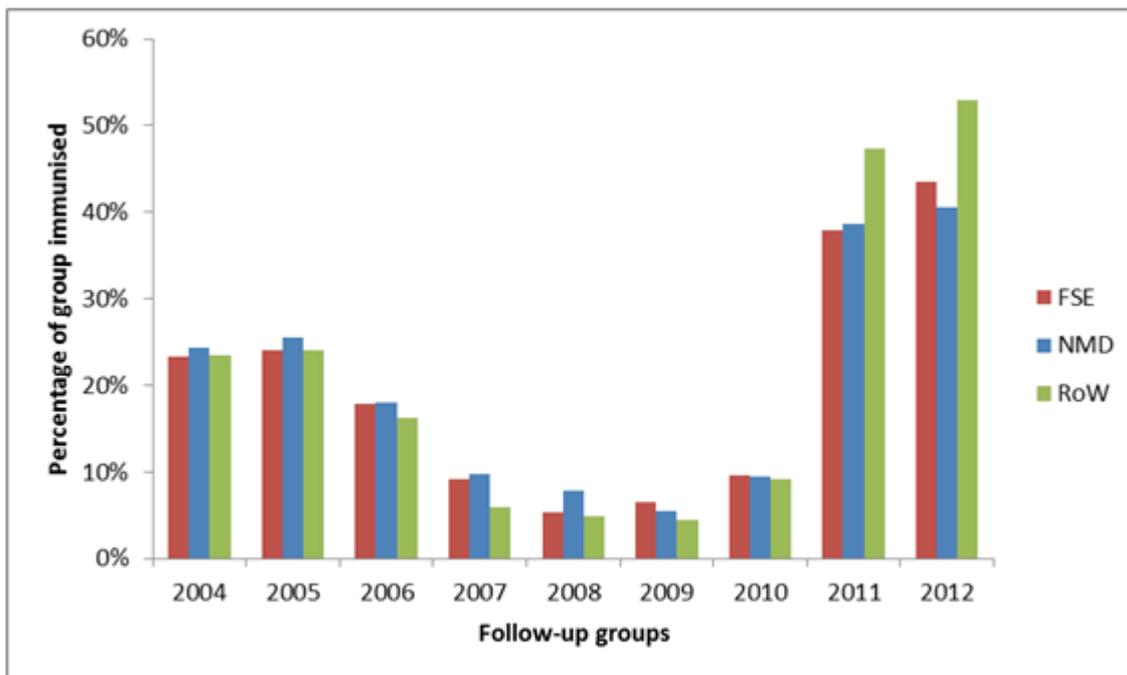
Figure 5.7 Percentage of Babies Receiving their First '5 in 1' Vaccine within 3 Months of Birth by Project Group



Measles Mumps and Rubella (MMR) 1

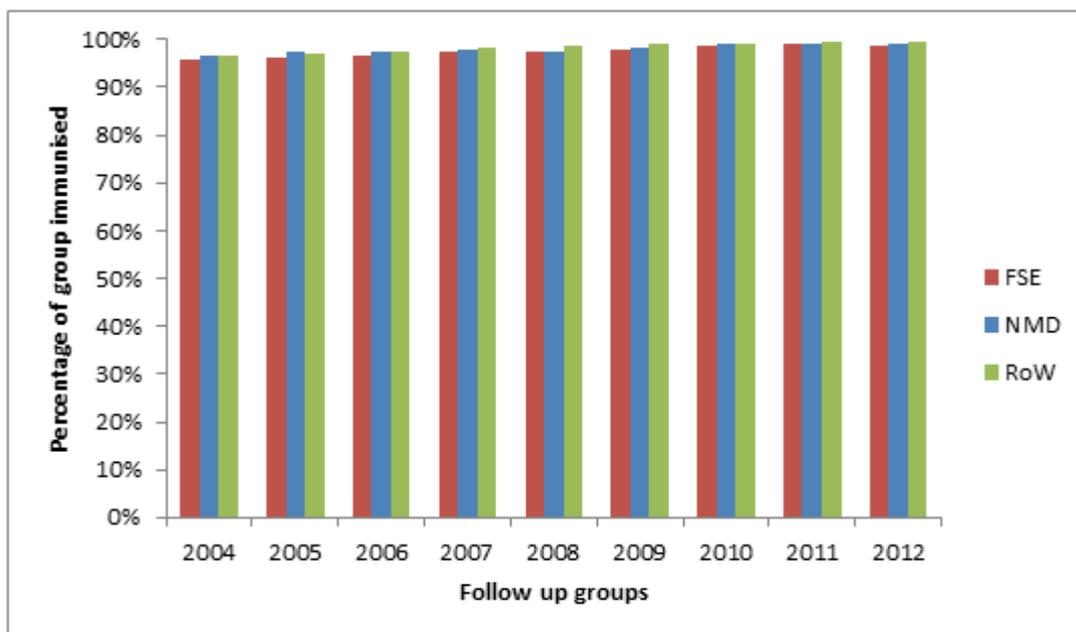
- 5.24 The Measles Mumps and Rubella vaccine is administered in two stages, the first dose (MMR1) being ideally administered before the age of 13 months, and the second dose (MMR2) ideally being administered before the child goes to school, and certainly before the age of 5 years. Two main courses of vaccine have been identified in the data for each of these two doses, but a further ten codes were found, relating to measles, mumps and rubella immunisations singularly or in paired combinations. It is also apparent, as noted above, that the first vaccine some children received was recorded as 'MMR2', so there are some confusing elements to the data. The analysis looks at the uptake of MMR1 and currently ignores the smaller numbers of children recorded as receiving other vaccination courses.
- 5.25 Figure 5.8, below, shows the percentage of babies receiving MMR1 by the age of 13 months. There has been considerable change in the percentage of children receiving this immunisation by age 13 months, dropping from over 20% to around 5% in the middle of the decade and rising sharply since 2010. This may reflect public distrust of the vaccination but could also be due to changes in the delivery of this vaccination over time such that children started to receive it at an earlier age. There certainly appears to have been a change in 2011 and 2012 when achievement of the 13-month target improved significantly in all three Groups. In the early years (2004-05) the FSE Group had a lower uptake than the NMD Groups, but from 2009 onwards they have higher or very similar uptake than the NMD Groups. In particular, in 2012 there is distinctly better uptake in the FSE Group than in the NMD Group. However, it is not possible to establish whether this finding is either noteworthy or statistically significant without more years of data and carrying out a more sophisticated analysis including all possible courses of vaccine (proposed additional work to take additional vaccines into account is outlined in Appendix 4).

Figure 5.8 Percentage of Babies Receiving their First MMR Vaccination before 13 Months by Project Group



5.26 The percentage of children receiving MMR1 by the age of 3.5 years (i.e. pre-school) is shown in Figure 5.9, below. Coverage was well over 90% and from 2010 onwards there was no real difference in uptake between the three Groups. In the earlier years, there was a slightly lower uptake in the FSE Groups so this does appear to represent an improvement since Flying Start, but again, the differences are quite small and the analysis would require further refinements before any statistical significance could be placed on this change.

Figure 5.9 Percentage of Babies Receiving their First MMR Vaccination before 3.5 Years by Project Group

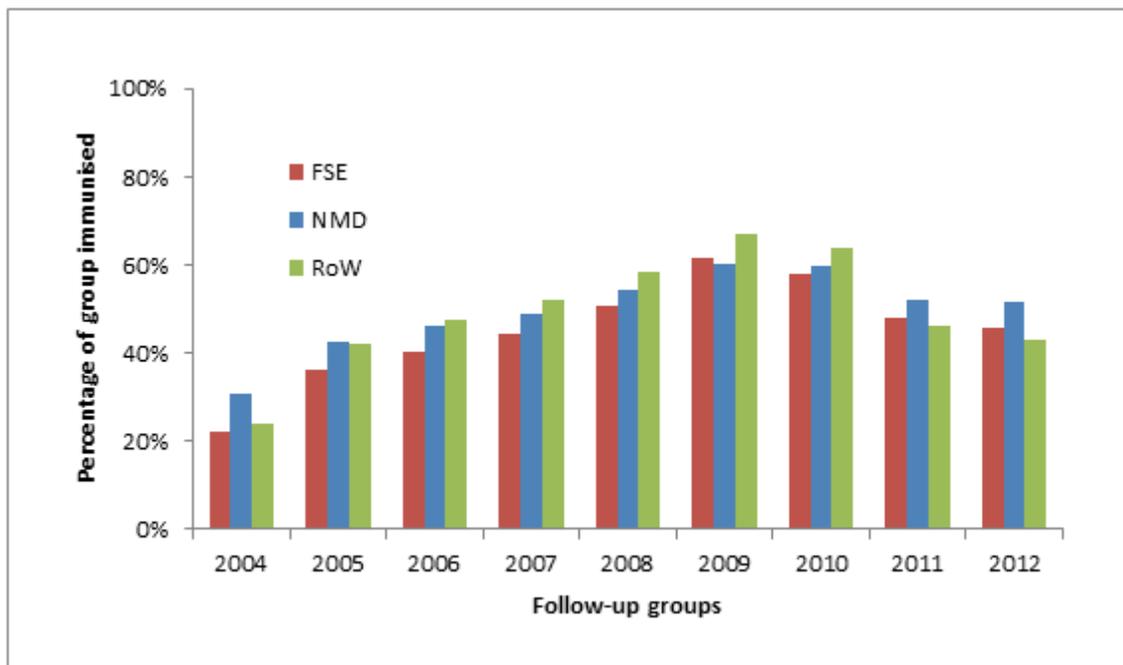


Measles Mumps and Rubella (MMR) 2

5.27 A similar process to that adopted for MMR1 was followed in analysing MMR2. This vaccination should occur before the child starts school. Taking the age of 3.5 years as the mid-point of 3 year olds, the percentage of children receiving MMR2 was calculated. The results are shown in Figure 5.10 below. In all three Groups, there was a rising trend in uptake until 2009 and then a reducing trend until 2012. It should be noted that there were two other courses of immunisation called ‘additional MMR’ being administered at this time and it may be that some people were immunised using these courses instead. Further analysis would be required to determine this and could not be completed within the limited scope of this Project.

5.28 However, using data that is limited to one of several possible immunisations, the data suggest that uptake in the FSE Group lagged slightly behind that in the other Groups until 2009 but that the FSE Group had a higher uptake in 2009 compared with the NMD Group. In 2011 and 2012 the uptake in all three Groups declined but the decline was smaller in both the FSE and NMD Groups. Since the NMD Group showed the smallest decline of the three Groups, it is not possible to identify a positive effect associated with of Flying Start.

Figure 5.10 Percentage of Children Receiving their Second MMR Vaccination by 3.5 Years by Project Group



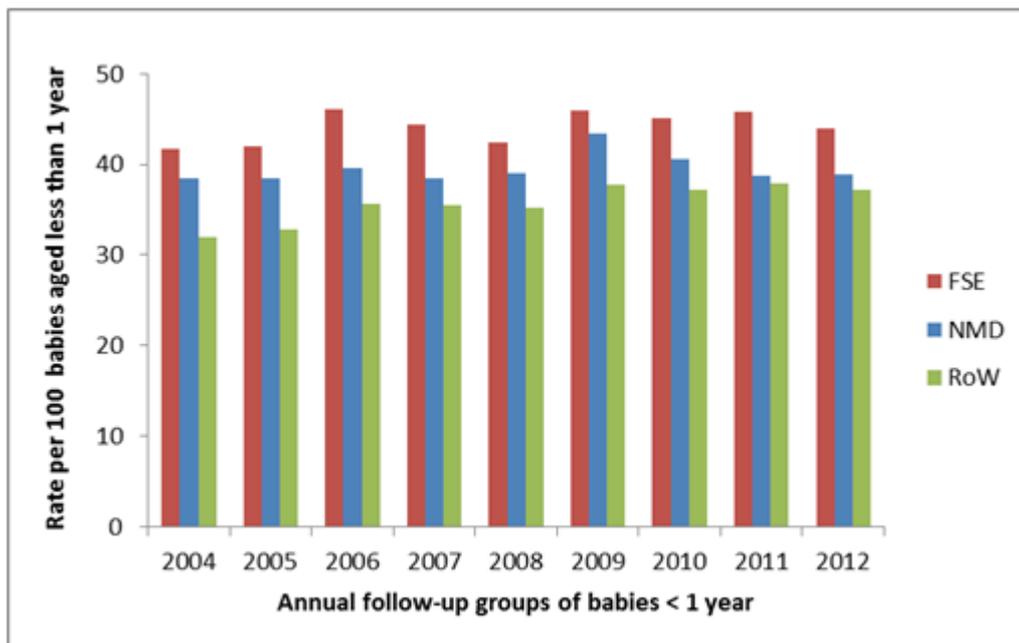
5.29 Between the ages of 3.5 and 5 years, the uptake in all three Groups was similar for each year, with over 90% of children being immunised with MMR2 (table not shown).

Hospital Activity

5.30 In Wales, there are around 16,000 hospital admissions every year of babies aged less than 1 year. The overall admission rate (i.e. for any cause and for all Wales) per 100 babies aged less than 1 year is provided in Appendix 5. The data show a step change after 2007 which appears to be due to changes in the recording of 'other' admissions, i.e. not emergency or elective spells of care, but maternity and transfers between hospitals.

5.31 For the purposes of this report, the analysis has concentrated on emergency admissions in babies aged less than 1 year, which reflect the acute events that Flying Start interventions would hope to reduce through the provision of preventative support. Every year, there are around 11,000 *emergency* admissions of babies aged less than 1 year in Wales. The emergency admission rates for the Project Groups over time are shown in Figure 5.11, below.

Figure 5.11 All Emergency Admissions of Babies Aged Less than 1 Year by Project Group

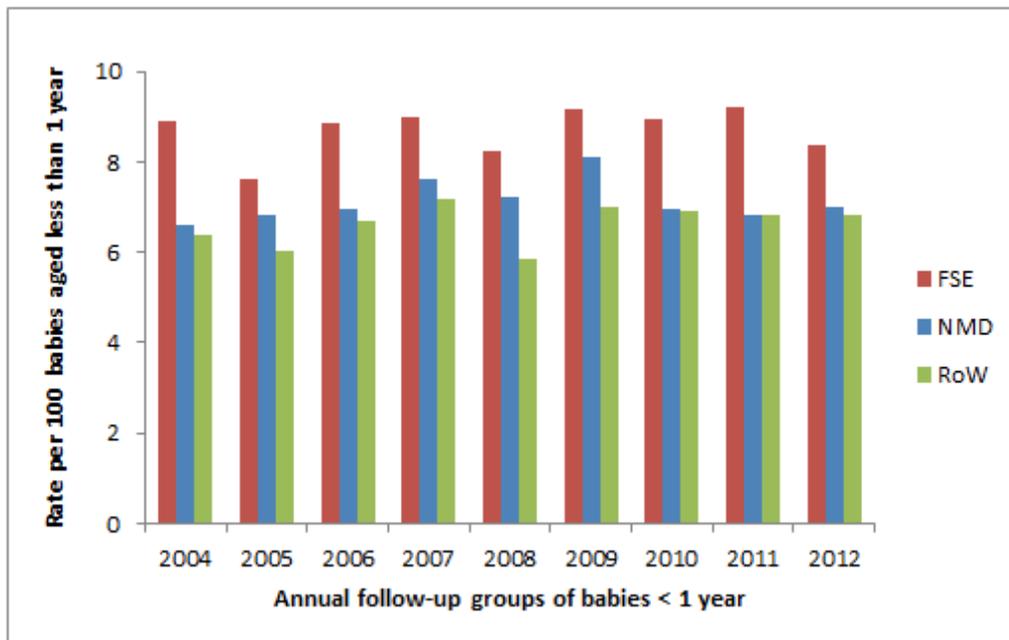


5.32 A slight reduction (2 percentage points) in emergency admissions for all causes was found in the FSE Group between 2009 and 2012. However, there was a greater reduction in the NMD Group (4.5 percentage points) over the same time period, so it was not possible to attribute any effect to Flying Start. It is possible that a greater reduction in the FSE Group might have been offset by a positive, increased use of services. It is also possible that where health issues arise e.g. out of hours, the NMD Group may have been more likely to consult a GP, while the FSE Group may have attended A&E. Further monitoring as time goes on may help us to understand these kinds of changes.

Emergency Admissions for Infectious Diseases

5.33 Infectious diseases in babies aged less than 1 year result in around 2,100 hospital emergency admissions annually in Wales (emergency admissions with a principal diagnosis A00-B99). This category is mainly made up of unspecified viral infections (54%) and unspecified intestinal infections (27%). The rest of the admissions are for small more specific diagnoses, the largest being Rotavirus enteritis (3%), 'other gastroenteritis' (2%) and Varicella (2%). The changing admission rates over time for the three Groups are shown in Figure 5.12, below.

Figure 5.12 Emergency admissions for Infectious Diseases in Babies Aged Less Than 1 year, by Project Group

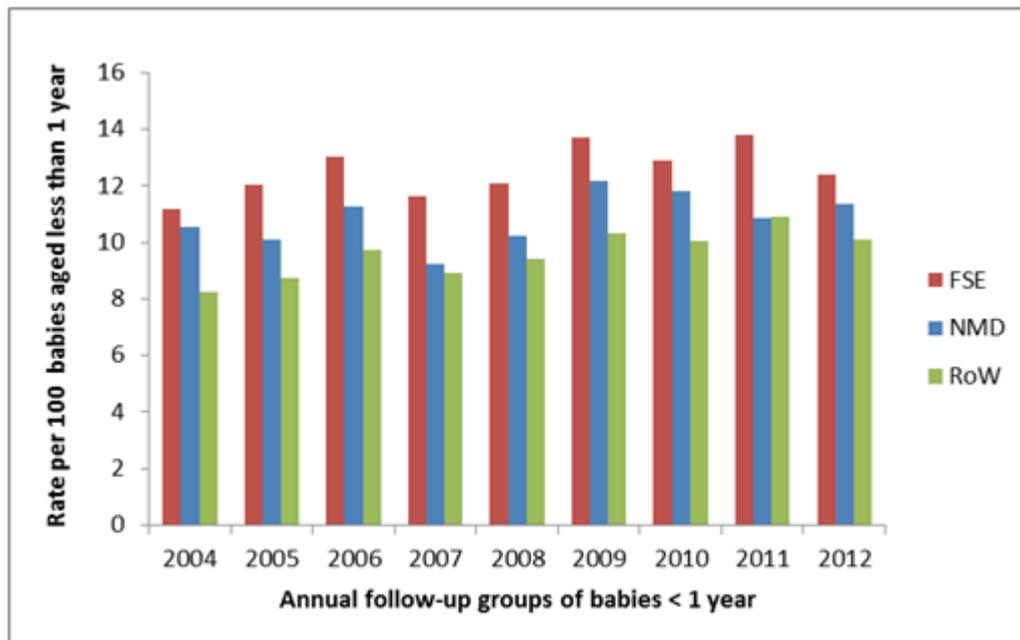


5.34 The admissions rate is higher among the FSE Group compared with the NMD Group; both are higher compared with the RoW until 2009, when the rate for the NMD and RoW became similar. The graph indicates a fairly constant trend over this time period for all three Groups with some annual fluctuation. Although no clear impact of Flying Start is suggested by the data, the figures for 2012 show a reduction in the FSE Group not seen in the other two Groups. Further years of data would be required to establish if this is the beginning of a significant downward trend.

Hospital Admissions for Respiratory System Diseases

5.35 Every year, there are on average just over 3,000 emergency admissions for respiratory system diseases (ICD 10 codes J00-J99) in babies aged less than 1 year in Wales (see Figure 5.13, below).

Figure 5.13 Emergency Admissions for Respiratory Conditions in Babies Aged Less Than 1 year by Project Group

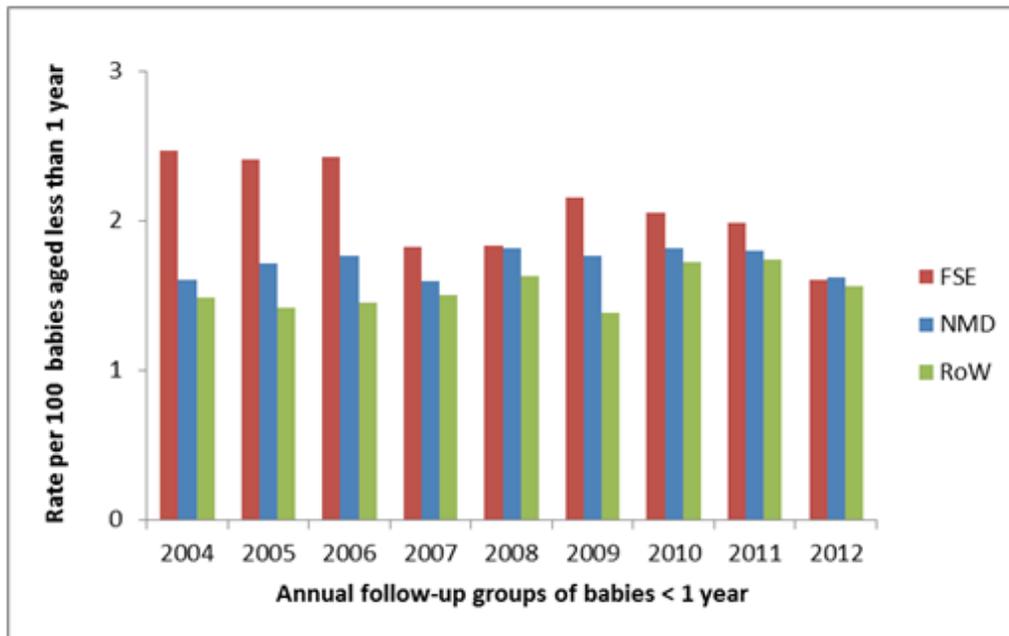


5.36 Annual rates of admission for respiratory conditions are higher across the whole period of 2004 to 2012 for the FSE Group compared with the NMD Group; both are higher compared with the 'Rest of Wales' (RoW). The rates of admission are very variable over time, with, for example, admissions for both the FSE Group and the 'rest of Wales' rising between 2010 and 2011 then falling in 2012, while annual rates for the NMD Group fall from 2010 to 2011, then rise in 2012. Where these rates are so variable, it is difficult to identify any trend that might be worthy of note.

Injury and Poisoning

5.37 Emergency admissions for babies aged less than 1 year include an average of around 500 'Injury and poisoning' (ICD10 S00-T99) admissions per year in Wales.

Figure 5.14 Emergency Admissions for Injuries and Poisoning in Babies Aged Less Than 1 year, by Project Group 2004-12



5.38 The main cause of injury and poisoning admissions was head injury (e.g. 55% in 2012). The numbers are relatively small so, as can be seen in Figure 5.14, above, the annual rate is a relatively volatile measure. The rate of admissions for 'Injury and poisoning' seems to have been decreasing to some degree across the whole Project Period in the FSE Group and there is a decrease in the annual rate of admissions from 2009-12 for all three Groups. Nevertheless, the decrease in the FSE Group between 2009 and 2012, even taking into account the underlying population decrease seen in the NMD Group, was found to be statistically significant

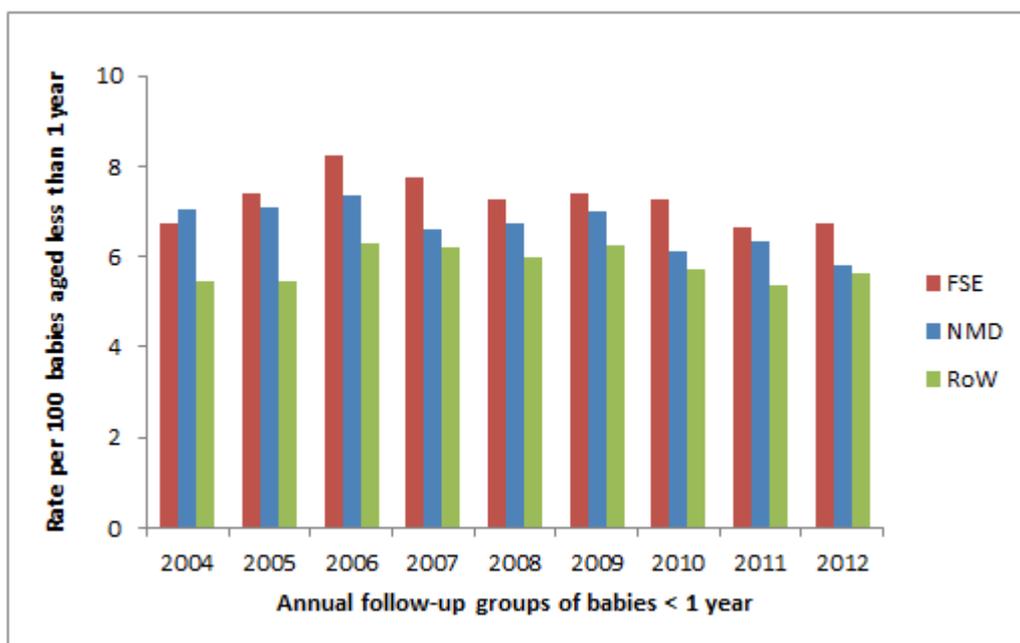
5.39 The analysis for this Project was limited by the fact that if children up to age 4 years had been included, only children born in 2008 or before would have been old enough (by 2012) to be included. For future reporting years, further work will be required to refine this indicator and to complete the analysis for children aged under 4 years; when sufficient data becomes available in future years, this would provide larger numbers and therefore more robust analysis.

Hospital Admissions: Symptoms, Signs and Abnormal Clinical Findings'

5.40 'Symptoms, signs and abnormal clinical findings' is a classification used to record findings that have not been attributed to a specifically diagnosed cause (see Figure 5.15, below). Emergency admissions in babies aged less than 1 year include

around 1,850 such admissions per year in Wales. Generally, there were more admissions of this nature in the FSE Group, and these admissions decreased from 2006 onwards in both the FSE and NMD Groups compared with the 'Rest of Wales', so that no noteworthy trend could be identified over time.

Figure 5.15 Emergency Admissions for 'Symptoms and Signs' in Babies Aged Less Than 1 Year by Project Group



Primary Care

5.41 A great deal of the NHS support for young families is provided through Primary Care. Around 47% of the population is now included in the 'GP Event' data that participating GP practices in Wales contribute into SAIL and the proportion of the population covered is increasing as time goes on¹². Around 50% of the NMD and RoW Groups are present in this data, but only 40% of the FSE Group. A possible explanation for this is the difficulty in recruiting GPs into deprived practices leading to less time for practices to get involved in 'added value' projects like SAIL. If GPs from more deprived areas are less likely to provide data, we might expect it to be the data for the most deprived areas within the FSE areas that is missing. This would mean that indicators developed using GP data would underestimate both the size of any gap between FSE and NMD and any effect of Flying Start.

¹² NISCHR has recently invested further funding for GP engagement posts with the aim of significantly increasing the number of GP practices contributing to the SAIL databank.

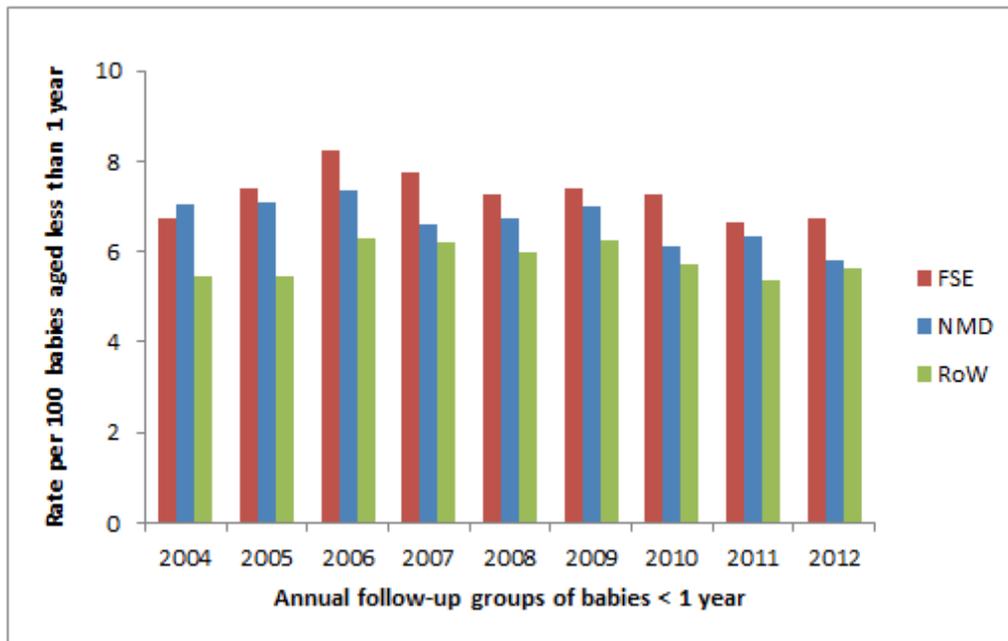
- 5.42 The GP Event data is held as a long list of recorded events, diagnoses, signs, symptoms, treatments, test results, prescriptions, observations and administrative procedures. This makes it complex to establish indicators that properly reflect levels of activity. Some experimental demonstration indicators are presented here. Through experimenting with the diagnostic codes recorded, respiratory diseases, infectious diseases and 'injuries and poisonings' have been chosen as the conditions that a) would be most likely to require a face-to-face consultation with a primary care health professional, and b) generated sufficient primary care records to allow robust analysis for children in the Project Groups. Also included is a comparative analysis of 'Preventative Procedures', another potential indicator of activity that was chosen due to its likely relevance to the objectives of Flying Start.
- 5.43 The next steps in developing the GP-based indicators will include trying to understand the trends seen below, and to look more closely at which specific health events are recorded in sufficient numbers to allow robust comparisons between Project Groups and over time.

Respiratory Conditions

- 5.44 'Respiratory conditions' is the group of diagnostic codes that appears most frequently in the GP data for children aged less than 13 years. The number of consultations for respiratory diseases per 100 babies aged less than 1 year for the three Groups is shown in Figure 5.16, below. The figures include 246 different respiratory diagnosis codes. Some children have multiple consultations for the same or different respiratory diagnostic codes which result in the high rates shown.
- 5.45 The FSE Group has a consistently higher consultation rate than the other two Groups, and comparing 2009-12 with 2004-2008, the FSE Group shows a significant increase in respiratory consultations. In 2012, the rates for both NMD and RoW decrease compared with previous years, whereas the FSE rate does not. It is not clear whether we should interpret this finding as positive or negative, since the introduction of Flying Start may have influenced parents to make better use of health services e.g. leading to more parents consulting a healthcare professional about their child's respiratory condition instead of leaving it to get better on its own. This could lead to a sustained level of recording for FSE at a time when the number

of consultations for respiratory conditions was falling for the rest of the population, as illustrated by the rate in the other two Groups. Further research would be required to establish whether better use of health services or increased respiratory disease in the FSE Group is the explanation for this trend.

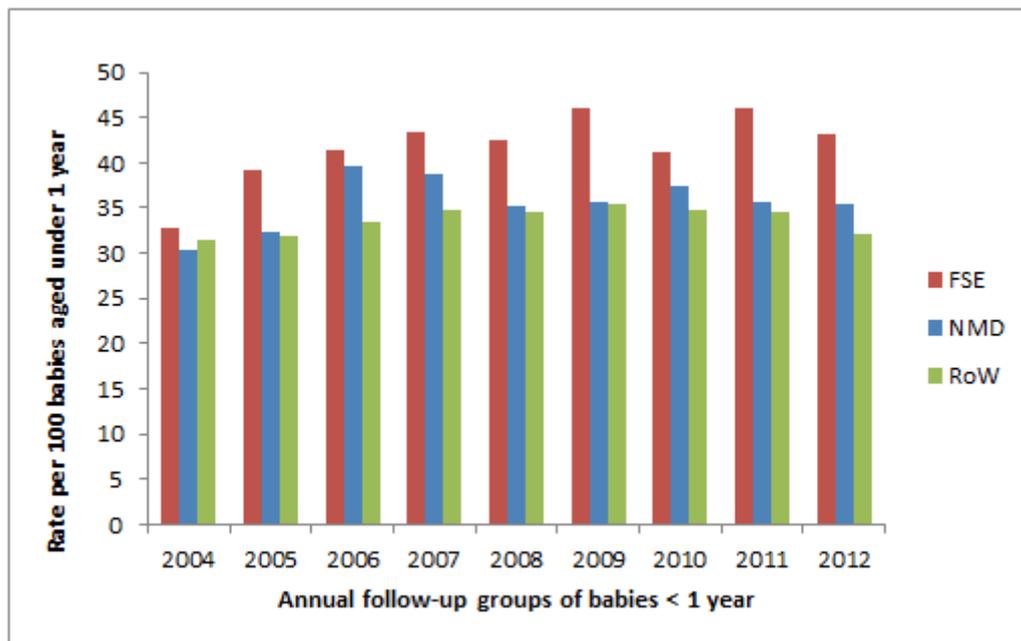
Figure 5.16 Primary Care Consultations for Respiratory Conditions in Babies Aged Less Than 1 year by Project Group.



Primary Care Consultations for Infectious Diseases

5.46 This indicator counts the number of primary care consultations for infectious diseases per 100 babies aged less than 1 year. The rates for the FSE Group were generally higher than for the other two Groups across the Project Period (see Figure 5.17, below). No statistically significant effect of Flying Start is discernible after implementation in 2009.

Figure 5.17 Primary Care Infectious Disease Consultations for Babies Aged Less Than 1 year

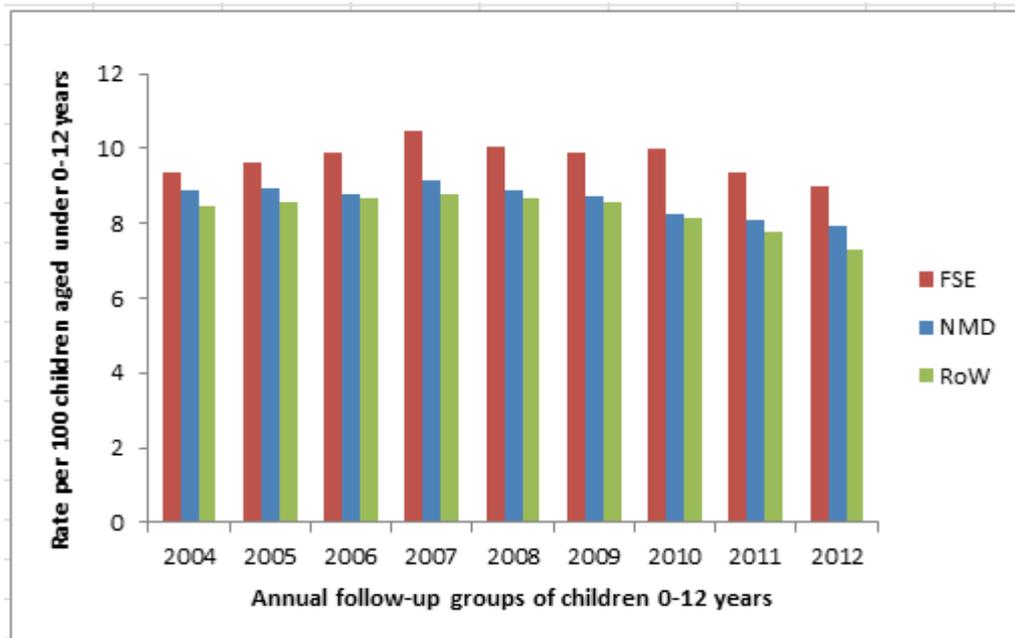


Primary Care Consultations for Injury and Poisoning

5.47 This indicator counts the number of consultations for 'Injury and Poisoning' codes per 100 children aged 0-12 years. It is likely to capture less serious injury and poisoning events than the hospital admissions indicator. The greater age range has been used to provide larger, more robust numbers. However, it should be noted that children over the age of 4 years from 2009 onwards, will be included in the analysis but could only have been directly influenced by Flying Start where a different Flying Start eligible child was in the same household, so any Flying Start effect will be diluted. This indicator requires further experimentation in the choice of age range, but will become more useful as time progresses and more years of data become available.

5.48 The data shows a reduction in this kind of consultation from 2007 onwards for all three Groups (see Figure 5.18, below). The FSE Group had consistently higher rates of 'injury and poisoning' consultations than the other two Groups. The data suggests a slightly greater decrease between 2004-08 and 2009-12 in the FSE Group compared with the NMD Group.

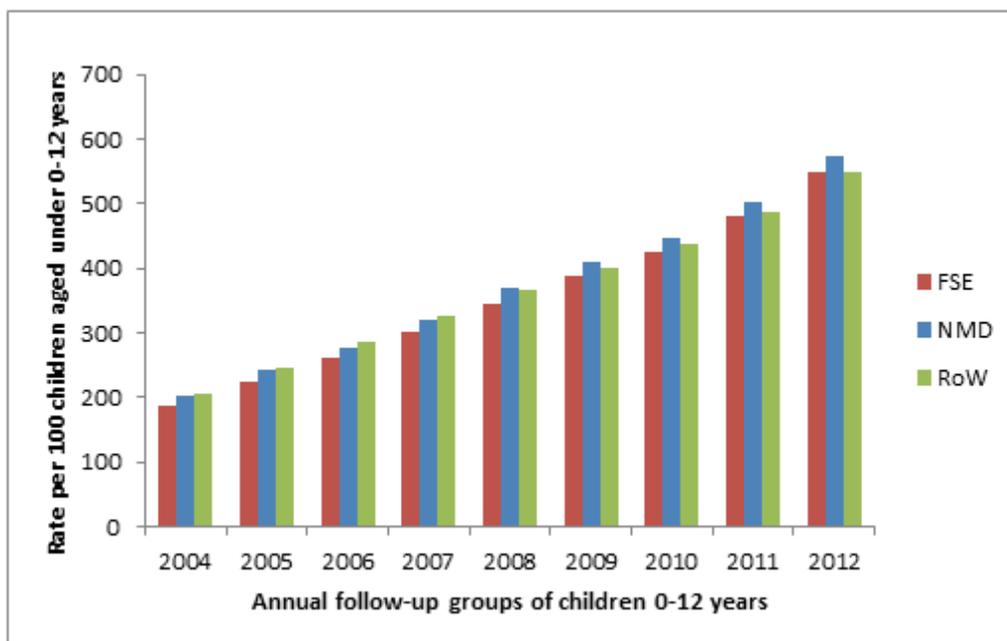
Figure 5.18 Primary Care Consultations for Injury and Poisoning for Children Aged 0-12 years by Project Group (rate per 100 children)



Primary Care ‘Preventative Procedures’

5.49 Figure 5.19, below, shows the rate of Primary Care “Preventative Procedures: Child Health Care” per child aged 0-12 years for each of the three Groups. The FSE Group has a lower rate than the RoW each year up until 2010, but after this the FSE has a more similar rate to the RoW. The NMD Group has a higher rate than the RoW from 2008 onwards. We are observing small relative changes between the Groups but the data suggest that the number of consultations for the FSE Group increased after 2010 to bring them more in line with the numbers seen in the RoW Group. Further investigation would be required to establish whether the generally increasing trend over time seen in the graph is due to a ‘real’ change in primary care activity or to greater use of this diagnostic code.

Figure 5.19 Primary Care Consultations for ‘Preventative Procedures’ for Children Aged 0-12 Years by Project Group (Rate per 100 children)



Educational attainment

5.50 At the time of writing, a major refresh of education data was expected into SAIL which will be incorporated into reporting for 2014. The findings for this Project are therefore based on data for 2004 to 2008. This only allows us to examine the baseline differences between the Groups in the years leading up to the implementation of Flying Start. For the purposes of this Project, the scope of the analysis for educational attainment is therefore limited to KS1.

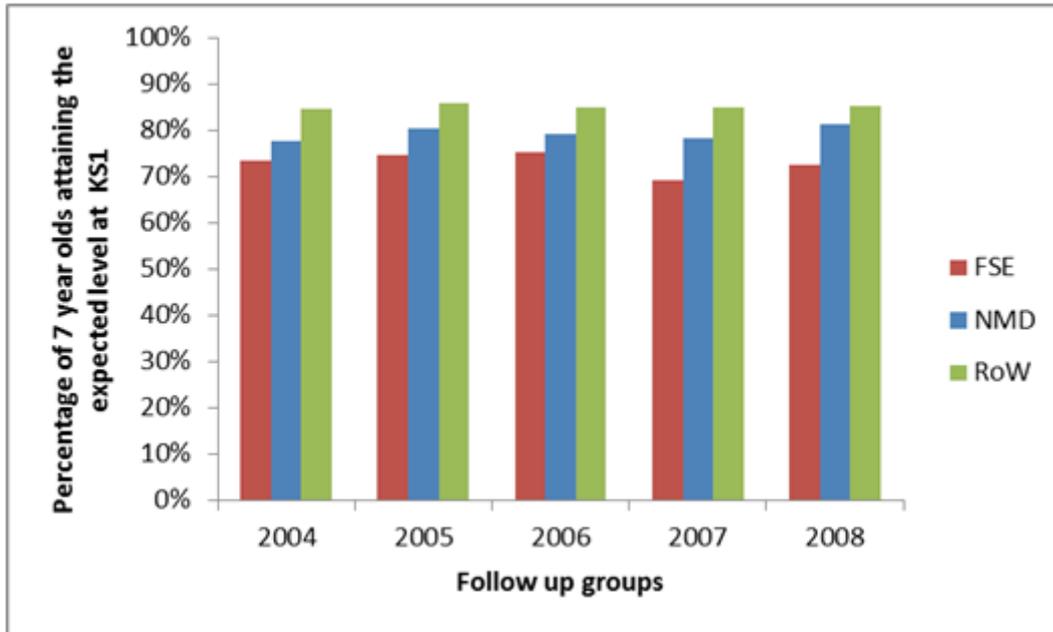
5.51 As noted above, a key strength associated with using linked administrative data is the ability to follow the flagged FSE children over time to identify longer-term outcomes so the reporting tool developed as part of this Project has the potential to track pupils through to school leaving age and eventually report on long-term effects of Flying Start.

Key Stage 1

5.52 The level of attainment at Key Stage 1 for the three Groups in years 2004 to 2008 is shown below in Figure 5.20, below. The graph clearly illustrates that a much smaller percentage of the FSE Group were attaining the expected level at KS1 compared with the NMD and RoW Groups. The relatively low attainment in the more deprived

Groups is as expected, based on what is known about the relationship between relative deprivation and educational outcomes, and reflects the motivation for creating programmes such as Flying Start.

Figure 5.20 Percentage of Children attaining the expected level at KS1 by Project Group



Methodology and future reporting

5.53 The Project has demonstrated that indicators can be sourced from multiple data sets and population Groups compared over time to show both ‘between-Group differences’ and changing trends over time.

5.54 Overall, the analyses shown above demonstrate relatively consistent relationships between the three Groups with, as might have been expected, the FSE Group tending to have the worst outcomes, followed by the NMD Group, with the RoW having the best outcomes. Even if the indicators we chose were the best possible ones, we might not expect to see an improvement in every one of them. In addition, even if there was an improvement in the real world, we wouldn’t necessarily detect it in the administrative data, bearing in mind the issues about recording and data quality outlined above. Other interventions or period effects may also be acting to offset any effect of Flying Start.

- 5.55 Further work is required in consultation with various health and educational specialists to finalise a set of outcome indicators that allow us to measure change against the original objectives of Flying Start. The decision must also be based on sufficient numbers being available for analysis and the probability of the indicator continuing to be recorded consistently into the future.
- 5.56 As discussed in Chapter 3 and at the beginning of this Chapter, **all findings must be viewed with caution**. Nevertheless, the Project does suggest some positive impacts of Flying Start on the health of children in Wales.

6 Discussion

- 6.1 Dividing households containing children aged under 4 years into the most deprived (FSE), Next Most Deprived (NMD), and the Rest of Wales (RoW) produced three Groups with surprisingly consistent differences across many of the indicators, with FSE having more problems than the NMD Group and NMD, in turn, performing worse than the more affluent (by default) RoW Group.
- 6.2 As noted above, the Project focused on eligibility for, rather than receipt of, Flying Start services. In terms of targeting the *areas* where individuals were most in need, Flying Start does seem to have worked well. Prior to the implementation of Flying Start, the Flying Start Eligible (FSE) Group had higher rates of infant mortality, lower and later uptake of vaccinations, less breast feeding at birth and at 8 weeks, fewer non-smoking mothers, more low-birth weight babies, more pre-term births, more births to teenage mothers, lower uptake of primary care preventative procedures and more hospital admissions and primary care consultations about injuries and poisonings.
- 6.3 If, in future, data on the receipt of Flying Start services becomes available, the analysis can be completed with greater precision and without the 'diluting' effect on findings of eligible individuals who have not received services.
- 6.4 As noted in Chapter 1, it is possible that other interventions may have influenced the outcomes; future studies could attempt to identify such interventions and to flag recipients, allowing the effects to be disentangled.
- 6.5 Although only small changes have been detected and not in all of the chosen indicators, the Project does suggest some narrowing in health inequalities between the FSE and NMD Groups in the years following the introduction of Flying Start. Although not all of the findings reach the level of statistical significance, the consistent way in which the three Groups tend to come 'first, second and third' in terms of most indicators (with FSE having the worst outcomes and RoW the best) and the way the gap between FSE and NMD narrows on a number of indicators, taken together, represent a somewhat more conclusive picture.

- 6.6 After the 2009 introduction of Flying Start, the FSE Group does show significant signs of improvement relative to the NMD Group in terms of reducing pre-term births, births to teenage mothers and low birth weight babies as well as in hospital admissions and primary care consultations for 'injury and poisoning'. The data also suggests improvements in the number of preventative procedures and the number of Injury and Poisonings recorded in primary care.
- 6.7 Having said that, there are some indicators where changes are apparent in the NMD and RoW Groups that are not reflected in the FSE Group e.g. primary care consultations for respiratory conditions; it is not clear whether this should be interpreted as a positive finding (i.e. more FSE children being taken to see the doctor instead of left to get over a cough), or a negative finding (i.e. indicating worse respiratory condition rates in these children).
- 6.8 Looking beyond Flying Start to the contribution the analysis of linked administrative data can make to the evidence base, the Project has demonstrated that data linking can provide significant added value. In particular, that linked administrative data can be used to:
- establish retrospective baselines;
 - create 'control' groups for comparison purposes; and
 - anonymously flag individuals eligible to receive services (and, when data become available, individuals receiving services) to identify changes in service use resulting from the policy intervention.

Reporting in future years

- 6.9 There is still some work to do to choose the final indicators for which reporting in future years should be provided. As time goes on, data for additional years will become available, allowing us to examine not just families newly joining the three population Groups but also the progress of those Groups over a greater follow-up period. For example, to establish comparable immunisation data will require further analysis to include alternative immunisations to MMR. Suitably robust indicators for GP data will also need further development.

- 6.10 WG and other organisations have plans to anonymise data sets for additional topics into SAIL in the coming years, which will provide the ability to monitor a wider range of outcomes. The national recording of heights and weights in children and the data from the stop smoking service are two examples.
- 6.11 Coupled with the developments described above, further analyses could already be done, if additional resources were made available. A list of possible additional analysis is provided at Appendix 4.
- 6.12 A technical appendix is included for the Project, describing the variables used, limitations of the Project design, and other issues dealt with to perform this analysis (see Appendix 5).
- 6.13 The SQL coding for the project has been documented for future use¹³. This documentation describes what has been done so far, and how and where code should be adjusted to change certain reporting parameters. For further information, please use the contact information at the beginning of the Report.

Key Lessons Learned

- 6.14 As noted above, lessons learned during the implementation of this and the other demonstration projects will be presented in a Lessons Learned report, which is to follow. However, the key lessons are as follows:
- 6.15 This kind of analysis takes a lot longer than you think! When this Project was proposed it was envisaged that it would be a simple matter to establish populations based on addresses provided and to provide some trend information from existing data sources about a set of indicators. The reality is that using linked administrative data and undertaking longitudinal analysis both add massive complexity to any analytical task. More work will be needed for WG and SAIL to jointly decide which indicators and from which data sets would provide the best, most comparable outcome indicators over time.

¹³ Within the SAIL Gateway working environment, analysts post completed pieces of code to a 'Wiki'-style bulletin board, where a library of concepts is being developed. When reusable code is developed (e.g. to select cases with a specific condition from complex primary care codes), this is made available to other researchers for peer review, and to eventually develop standard methodologies for data manipulation and case selection. These will be shared across International data linkage forums to help develop encourage global standardisation and can be made available on request to approved researchers undertaking similar projects

6.16 Dedicated time for programming: the interrogation of large, complex data sets complete with duplication and erroneous data requires the use of database querying tools. This means that the analysis of each indicator requires the construction of specific computer code. It is far from ideal to be 'picking up' and 'putting down' such a demanding and complex task between other jobs. Although it is possible to annotate the work, it's difficult to do so in a timely manner to a level where it can easily be picked up at a later date. Providing dedicated time periods for the completion of the programming steps would be far more efficient.

These lessons will be kept in mind by WG and SAIL when developing further data linking projects and will be fed into the development of the UK Administrative Data Research Network and, in particular, into the work of the Wales Administrative Data Research Centre.

Appendix 1 Membership of the IGRP

Membership of the Information Governance Review Panel as at 10th June 2013. The IGRP provides independent advice on Information Governance and reviews all proposals to use SAIL data to ensure that they are appropriate and in the public interest.

The current panel comprises:

Organisation	Name
British Medical Association	Dr Tony Calland
National Research Ethics Service	Corrine Scott
Public Health Wales	Dr Judith Greenacre
NHS Wales Informatics Service	Martin Murphy Darren Lloyd
SAIL Consumer Panel	Dr Neil McKenzie Dot Williams

Appendix 2 IGRP Application Form

Centre for Health Information Research and Evaluation (CHIRAL)

College of Medicine

Swansea University

Health Information Research Unit (HIRU)

HIRU Enquiry form

Template review chronology

Version no.	Effective date	Reason for change
1.0	29/11/07	N/A
2.0	1/5/08	Establishment of CRS necessitating changes to content and layout
3.0	14/10/09	Recommendations of IGRP
3.1	05/04/11	Annual review

HIRU Enquiry Form

The following form has been designed to collect the information needed from individuals and organisations interested in collaborating with HIRU on work involving the SAIL databank. The information you provide will facilitate consideration of your enquiry. Please complete sections A & B and provide additional documents as requested.

SECTION A

1a. Contact details of project lead:

Name:

Job title:

Organisation:

Address:

Tel:

Fax:

Email:

1b. The project lead will be the only person accessing the data:

Yes [] No []

1c. Please provide contact details of the person(s) who will be accessing the data (apart from the project lead):

Name:

Job title:

Organisation:

Address:

Tel:

Fax:

Email:

2. Does your proposed work with HIRU constitute?

Part of a larger project?

If yes, please complete all questions

The entire project?

If yes, please complete all questions except 3a, 5a and 7a

3a. Full title of the *main* project:

3b. Full title of the (part of the) project involving HIRU (if different):

4a. Who is commissioning the project (if relevant)?

4b. Why is the project being done?

5a. Aim of the *main* project, including anticipated outcomes:

5b. Aim of the (part of the) project involving HIRU, including anticipated outcomes (if different):

Please include a copy of the protocol/plan for the proposed work with HIRU, including the contact details of any co-applicants when you return your completed form.

6. Lay summary of the project involving HIRU: (approximately 150 words)

7. Please list the relevant permissions you have obtained or that are being sought:

Obtained

Being sought

Not required

<i>Research ethics</i>	[<input type="checkbox"/>]	[<input type="checkbox"/>]	[<input checked="" type="checkbox"/>]
<i>Independent peer review</i>	[<input type="checkbox"/>]	[<input type="checkbox"/>]	[<input checked="" type="checkbox"/>]
<i>Permission from data-holding organisation to use their datasets</i>	[<input type="checkbox"/>]	[<input type="checkbox"/>]	[<input checked="" type="checkbox"/>]

Please state the name of the organisation/committee that is being applied to, or that has given approval, as applicable:

Research ethics:

Peer review:

Data organisation permission:

If you have ticked 'not required' please specify the reasons.

Please note that it is the responsibility of the project lead to ensure that the relevant permissions are obtained.

8a. At what stage is the main project?

Protocol/plan being developed	[<input checked="" type="checkbox"/>]
Protocol/plan in place but project not commenced	[<input type="checkbox"/>]
Project underway	[<input type="checkbox"/>]

If underway, what was the start date of the main project (dd/mm/yy)?

8b. Please indicate a prospective start date for the (part of the) project involving HIRU:

(dd/mm/yy)

8c. Over what period do you anticipate you will require the assistance of HIRU?

Start and end dates in dd/mm/yy: [] to []

9a. What data do you require for the proposed work with HIRU?

Please list:

The datasets you require information from

The types of variable you need

The datasets that will need to be linked

9b. Will you also be providing other datasets to be incorporated into the SAIL databank?

Yes [] No []

If yes, please specify:

9c. Please provide an outline of your analysis plan including the anticipated outcomes

9d. Are the results/methods developed likely to have other potential applications?

Yes [] No []

If yes, please specify:

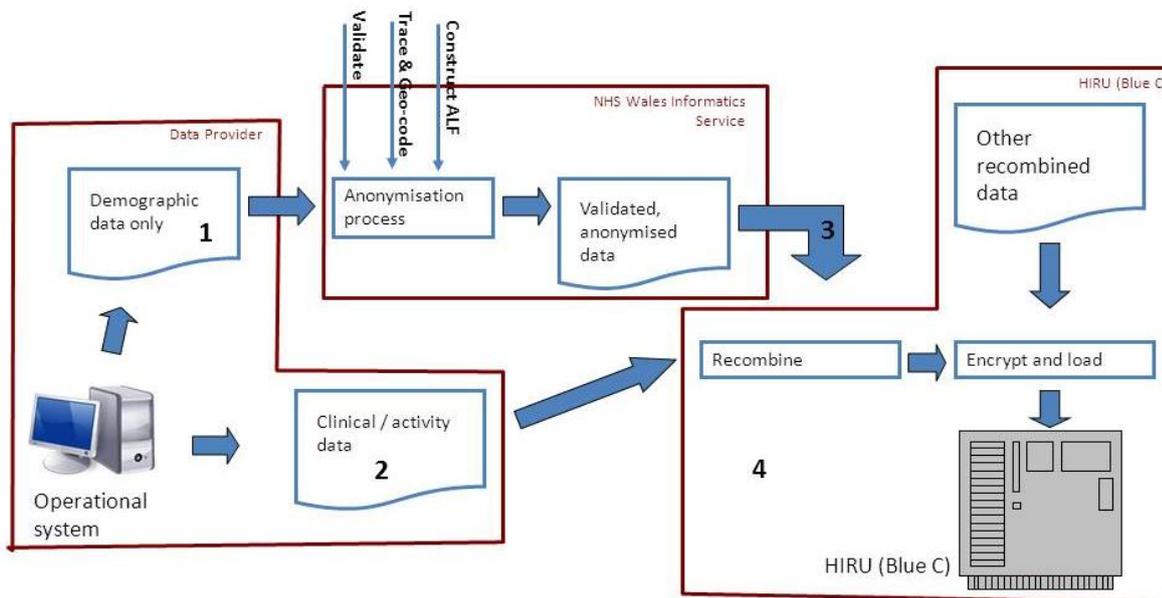
10a. Please indicate your plans for publishing the results of your project, e.g. target journal or intended recipients of report:

10b. What are the potentially sensitive issues that need to be taken into account when publicising the findings of the project?

Please outline the issues and your proposed solutions:

Appendix 3 The SAIL Split File Process

Figure 1: Overview of file processing into SAIL



Example Tables containing fictitious information

File 1, Demographic data sent to NWIS in Cardiff, contains no clinical data.

Index	First Name	Surname	Address line 1	Address line2	...Postcode	Gender	NHS number if known
...							
334	Mary	Smith	The Grange	Grangetown	CF31 5TP	2	1234123456
335	Fred	Bloggs	15 High Street	Aberdabbie	SA56 6GH	1	3453435975
336	Burt	Reynolds	6 Tee Road	Brecon	SA29 8PQ	1
...							...

File 2, sent directly to SAIL in Swansea, containing no identifiable data

Index	Diagnosis 1	Diagnosis 2	Operation 1	Operation 2	Date of admission	Date of Discharge	Method of admission
...							
334	Q223	S342	A123	B342	13/12/2009	19/12/2009	21
335	S987	T754	P321	V450	31/12/2009	16/01/2010	13
336	R342	T34x	10/09/2008	14/09/2008	12
...							...

File 3, created from File 1 by NWIS and sent to SAIL, is linked to File 2 using the Index field

Index	Anonymised Linking Field	Index	Diagnosis 1	Diagnosis 2	Operation 1	Operation 2	Date of admission	Date of Discharge	Method of admission
...		...							
334	1345234198	334	Q223	S342	A123	B342	13/12/2009	19/12/2009	21
335	2908070687	335	S987	T754	P321	V450	31/12/2009	16/01/2010	13
336	3210456342	336	R342	T34x	10/09/2008	14/09/2008	12
...	

Description

The data supplier number the individual records in their data with an index. They then split the data into two files, file 1, which has the index and the identifiable details, and file 2 which has the index and the detail data, but no identifiable data. File 1 is sent to NWIS who process the names and addresses and produce file 3 which has the original index and the anonymised linking field, which is a number unique to each person in Wales.

The supplier sends file 2 to Swansea University, and NWIS supply file 3 to Swansea. They are linked together using the original index, and the index is then discarded, leaving file 4 (below) with the detail data linked to the Anonymised Linking field, ready to be linked to other datasets that have been processed in the same way.

File 4 Resulting file in SAIL, with index removed, can be linked to other data from different sources processed in the same way

Anonymised linking field	Diagnosis 1	Diagnosis 2	Operation 1	Operation 2	Date of admission	Date of Discharge	Method of admission
...							
1345234198	Q223	S342	A123	B342	13/12/2009	19/12/2009	21
2908070687	S987	T754	P321	V450	31/12/2009	16/01/2010	13
3210456342	R342	T34x	10/09/2008	14/09/2008	12
...							...

Appendix 4 – List of possible future research projects

Proposed refinements to existing work

Immunisation

- The recording of immunisations is on a rolling programme of different vaccines, such that over time not all participants are provided with their immunisations from the same round of immunisation cover. For the purposes of this project the majority of vaccinations have been captured but there are other small groups of immunisations that have not. It would be possible to build a much more sophisticated selection tool that would capture cover on a case by case basis and may provide a more sensitive comparative measure.
- The addition of all immunisations programmes administered to children is perfectly possible with the current data available, but would benefit from the selection tool described above being created

Geographical variation

- The Local Authority has been captured in the data sets allowing some kind of variation by geography to be explored, and the combination of unique identifiers for people and households means that approximated family composition could be investigated for most of the households. Accident and emergency and outpatient data sets could also be linked.

Appendix 5 Technical guide

Limitations of the Project

- Flying Start is an area-based intervention, so this Project has not harnessed the full power of data linkage at the individual level. If, in future, data were to be collected about individual receipt of Flying Start services, it would be possible to track exactly which benefits families were exposed to and over what time periods. It would then be possible to examine not only whether Flying Start overall was having a positive impact on health and education, but also which of the Flying Start interventions was having the most impact.
- In reality, only some families living in Flying Start areas will have used the enhanced services offered by the Programme, and the presence in the analysis of those who didn't receive services will dilute the effects we are trying to identify; this means that any effect we are able to observe is likely to be underestimated. It is possible that other interventions may have influenced the outcomes; future studies could attempt to identify such interventions and to flag recipients, allowing the effects to be disentangled.
- It is also possible that the actual identification of Flying Start areas on the ground, by those delivering the services, is not easy. If the delivery of Flying Start benefits does not perfectly match the areas designated, there may be Flying Start beneficiaries included in the NMD or RoW Groups. If this were the case, data based on the official area definitions would be likely to underestimate any effects the programme might have.
- The anonymisation process has to remove dates of birth because it is an identifiable field. Within SAIL, the dates of birth are altered to the date of the Monday of the week of birth. This means that the selection in SAIL of children aged less than 4 years on the 1st April suffers from a slight inaccuracy, as a baby might be six days older than the week of birth states, depending on what day of the week 1st April fell in each year. This discrepancy is thought to even itself out over the years of the Project.

- The three population Groups were defined on a single date (1st April each year). So, a family may have moved into a Flying Start address one day before the 1st April or may move out one day after. As noted above, the bulk of children aged less than 4 years, progress into the next year's cohort and the set of households chosen on each 1st April expands and contracts to include those households where a new child is born or moves-in or moves-out. There is, however, some movement within years that the current methodology does not account for. A more detailed study would be required to establish the effects of these movements. For the purposes of this Project, we have assumed that similar migratory patterns have occurred between all the Groups and that the overall effects would therefore cancel each other out.
- Some house moves took families from FSE to NMD or RoW Groups. The Project methodology deals with this to some extent by re-allocating that individual to the other population Group (if the eligible child remains under 4 years) but, due to the methodology used, families may have contributed data to the wrong population Group for almost a year. Analysis demonstrated that there were very few moves between FSE and NMD Groups, so it is unlikely such moves are affecting the outcomes¹⁴. It is also possible that word of mouth led to improvements in areas that border with Flying Start areas or, as mentioned above, other interventions e.g. Communities First or Families First, could be influencing outcomes; further research would be required to test this theory.
- Using administrative data rather than longitudinal survey data, has the disadvantage that the data are not collected for research purposes. Often changes are made to data collections to reflect changes in procedure and policy. For example, as noted above, the immunisation data changes over time as different courses of vaccine are introduced. This makes the identification of a Flying Start effect on immunisations problematic. We now know that child assessment at Key Stage 1 has been

¹⁴ Over the period 2004 to 2012 there were a total of 417,543 children aged under 4 years included in the Project, 43,952 (11%) of whom moved between the Groups. The vast majority (40,377 [92%]) of these moves were either into or out of the RoW. In a typical year (2008) the number of children aged 0-4 years included in the Project was 131,657. Of these, 218 (0.17%) children had moved from the FSE into the NMD Group by 2009, and 184 (0.14%) children had moved from the NMD to the FSE Group by 2009. These lower numbers in the latter two groups may be explained in part by the fact that the two areas are generally closer together geographically and so people may move house much more regularly than they move GP (which is how moves are picked up in SAIL).

superseded and the new assessment is not comparable with KS1. The effect of this may be to hide any effect that Flying Start has on educational attainment. For this Project, every effort has been made to address such issues where possible and to document them. As time progresses, we will incorporate any improvements into the monitoring tool, and will feed back lessons learned to the data linkage research community, especially the Administrative Data Research Centre and Administrative Data Research Network.

Outcome indicators

Definitional and data completeness issues for the indicators listed in Table 4.1 are described in this Section.

Indicators from the NCCHD

Maternal smoking during pregnancy:

- This is based on a question asked of women by the Community Midwife at booking – this is usually when the woman is around 8 to 10 weeks pregnant.
- Smoking status is coded as shown in Table A.3, below.

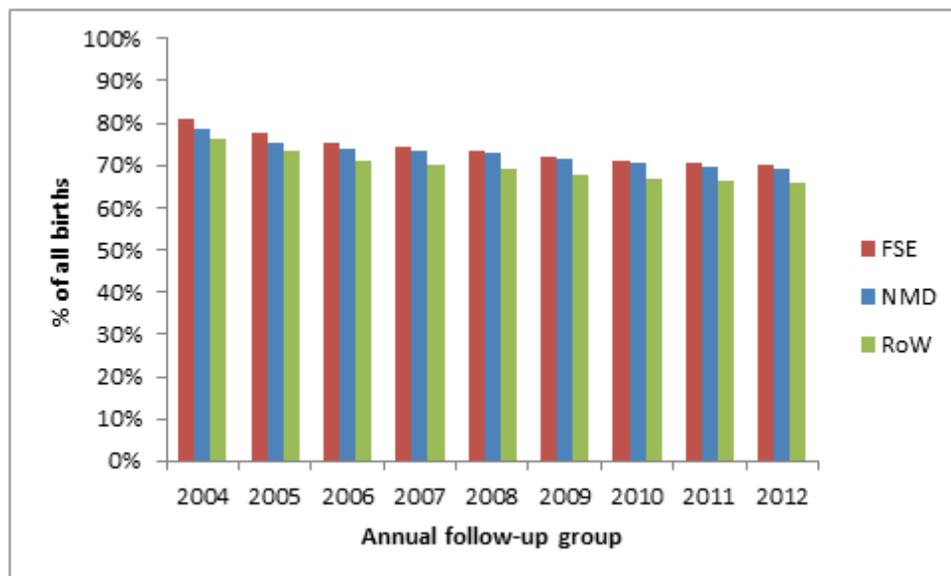
Table A.3: Coding of smoking status in NCCHD

Maternal Smoking codes	Meaning
0	Non smoker
1	Gave up in pregnancy
2	0-9 a day
3	10 or more a day
9	Not known

- Around 75% of the records do not have a recorded value. The percentage of missing values is fairly consistent across the three population Groups and the percentage of missing data is steadily reducing over time (see Figure A.4, below). Although it would be possible within SAIL to test whether some of the characteristics of the group for which this field has been completed differ compared with the group for which it was

left blank e.g. age and gender, in order to assess whether bias was evident, this was not possible within the limited scope of the demonstration project.

Figure A.4: Percentage of missing data on maternal smoking status



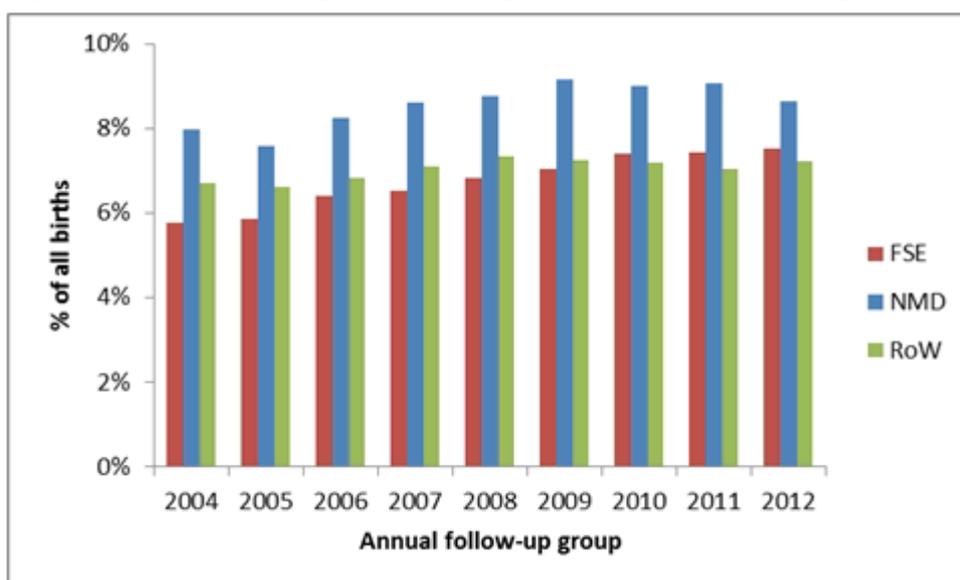
Gestational age

- A gestational age of less than 37 weeks is considered 'pre-term'. The proportion of pre-term births in each Group has been calculated for all follow-up Groups.
- This is calculated at the time of the birth from the estimated date of conception based upon the details collected during the course of the pregnancy. Dates of conception are not specifically collected.
- The data is recorded in weeks.
- Around 11% of records have no recorded gestational age, but the level of missing data decreased slightly in the later years.

Low Birth Weight

Figure A.5, below, shows the percentage of each Project Group where birth weight is missing. The graph shows that the FSE group had less non-recording of birth weight than the Rest of Wales in the years before 2009 but more non-recording from 2009 onwards. Further investigation would be required to try to establish what possible mechanism might result in this pattern of non-recording e.g. where recording was worse in the NMD Group. Nationally, around 0.1% of birth weights for live births are missing¹⁵, so further investigation is also needed to examine why the data in SAIL includes greater proportions of missing data.

Figure A.5: Percentage of Missing Values for Birth Weight by Project Group



Breast feeding: 'at birth'

- This is a question normally established in hospital directly after the delivery, and recorded as 'yes' or 'no'. There is a substantial amount of missing data. The relative proportions of 'yes', 'no' and 'not known' have been compared.

Breast feeding: 'at 8 weeks'

- This is recorded by health visitors during home visits or during a specific clinic attendance planned at around 8 weeks. There is a large proportion of missing data in

¹⁵ See Table 6 and Annex 1 of the NCCHD annual births statistics release: <http://wales.gov.uk/statistics-and-research/births-national-community-child-health-database/?lang=en>.

this field. The data is recorded as 'yes', 'no' or 'missing' and has been analysed in a similar way to breast feeding at birth.

Infant mortality

- According to the mortality data in WDS there has been an average of 9 deaths per year in babies aged less than one year in Wales, over the period 2004-12. As the numbers are so small, the only comparison possible was to look at deaths during the periods before and after Flying Start was rolled out. The number of deaths per 1000 live births has therefore been calculated for 2004-08 and 2009-12.

Immunisation

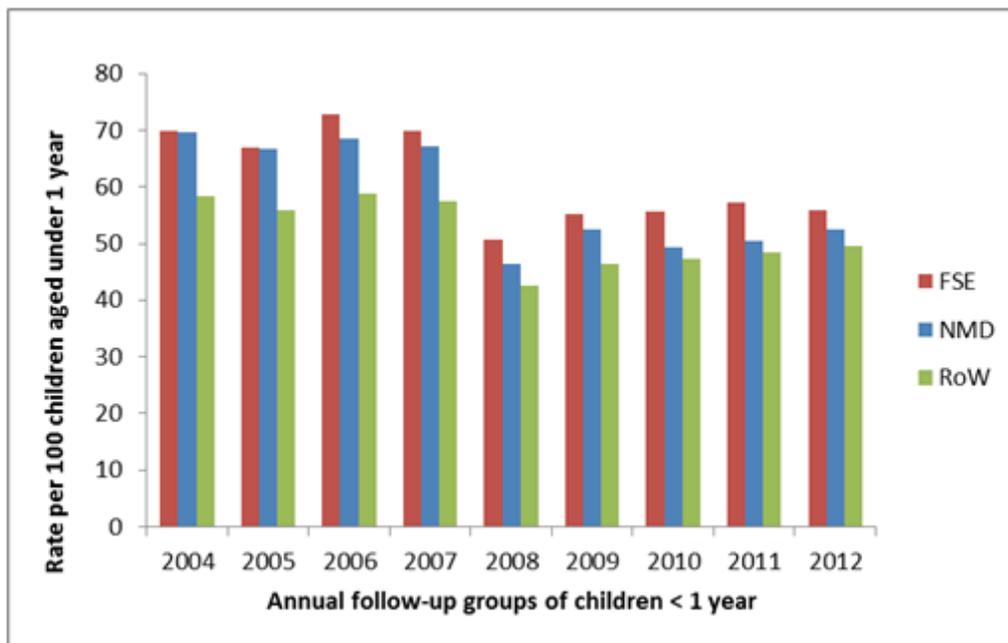
- There are a number of childhood vaccinations that could have been analysed as part of this Project, but early examination of the immunisation data revealed some complexity in the recording of each child's immunisation history and in particular which 'vaccination round' they belonged to for the provision of which dose of cover. For example some children who appear to be late receiving their first MMR vaccination, but have received the second MMR vaccination within the timeframe for the first vaccination. So they were treated on time by being sent along to a clinic dispensing the second round of the vaccine. The percentage of immunisations that occurred in relevant time periods have been explored for MMR and the '5 in 1' immunisation. In future, additional vaccinations can be reported on once a suitable comparative indicator is finalised.
- Measles Mumps and Rubella (MMR) vaccine is administered in two courses, MMR1 which should be given before the baby reaches 13 months, and MMR2 which should be given before the child starts school and before their fifth birthday. Data is recorded by batch of vaccine which is referred to as a 'course' on the NCCHD. Each course of vaccine has a start date, and runs until the vaccine is used up, which will take a number of years. The time periods for some courses of vaccine overlap, so that although course (coded as '25') of MMR started being administered first, course '26' started before course '25' finished. Children from the follow-up Groups may have been assigned to '25' or '26' for the first MMR or second MMR and, unless they have a record on both courses, it is not clear from the data exactly which intervention was intended, (first or second MMR cover). To allow meaningful comparisons to be made,

the proportion of children immunised on either course '25' or course '26' by 13 months, 3.5 years (pre-school), and 5 years has been utilised.

Hospital activity

- For the purposes of this Report, this has been calculated using a count of spells of hospital care in babies aged less than 1 year. Most of the analyses are presented as rates per 100 babies aged less than 1 year. To increase the number of admissions compared, admissions for injuries have been calculated per 100 children aged 0-12 years.
- In order to provide an overview of the full admissions picture, rates of hospital admission from 'All causes' for babies less than 1 year for successive follow-up Groups are compared (Figure A.6).
- The overall admission rates show consistently higher admission rates for the FSE Group than for the NMD and RoW Groups over time.

Figure A.6: Admissions to hospital (all causes) by Project Group



Appendix 6 Distribution of RALFs by Geographical Area (Local Authority)

Local Authority	FSE		NMD		RoW		All Ralfs
	Ralfs included	% of LA population included	Ralfs included	% of LA population included	Ralfs included	% of LA population included	
Blaenau Gwent	5,244	16%	5,582	19%	28,203	65%	39,029
Bridgend	5,017	9%	7,439	13%	71,889	79%	84,345
Caerphilly	9,381	12%	8,748	12%	76,828	75%	94,957
Cardiff	16,521	10%	15,403	12%	197,356	77%	229,280
Carmarthenshire	7,009	8%	9,411	12%	116,714	80%	133,134
Ceredigion	3,348	9%	2,177	6%	92,897	85%	98,422
Conwy	3,635	7%	7,306	15%	78,385	78%	89,326
Denbighshire	4,277	11%	5,970	14%	63,943	75%	74,190
Flintshire	5,164	8%	1,870	3%	85,592	89%	92,626
Gwynedd	3,292	6%	2,808	5%	115,718	88%	121,818
Isle of Anglesey	2,007	7%	1,665	6%	51,750	87%	55,422
Merthyr Tydfil	3,284	14%	4,518	18%	29,036	68%	36,838
Monmouthshire	2,284	6%	1,259	3%	50,231	91%	53,774
Neath Port Talbot	9,238	15%	7,773	13%	64,757	72%	81,768
Newport	9,407	17%	3,221	6%	67,300	78%	79,928
Pembrokeshire	5,731	11%	7,974	15%	82,878	75%	96,583
Powys	3,578	6%	2,553	4%	107,687	90%	113,818
Rhondda Cynon Taf	13,433	13%	14,506	14%	113,888	73%	141,827
Swansea	9,802	9%	12,489	12%	137,988	79%	160,279
The Vale of Glamorgan	5,302	10%	6,153	12%	66,915	79%	78,370
Torfaen	5,452	14%	542	2%	44,326	84%	50,320
Wrexham	4,435	9%	2,041	4%	88,845	88%	95,321
Totals	136,841		131,408		1,833,126		2,101,375